Advances in Applied Mathematics

Advanced Engineering Mathematics

A Second Course

Dean G. Duffy

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Advanced Engineering Mathematics

Advances in Applied Mathematics

Series Editors:
Daniel Zwillinger, H. T. Banks

Advanced Mathematical Modeling with Technology
William P. Fox, Robert E. Burks

Introduction to Quantum Control and Dynamics
Domenico D'Alessandro

Handbook of Radar Signal Analysis
Bassem R. Mahafza, Scott C. Winton, Atef Z. Elsherbeni

Separation of Variables and Exact Solutions to Nonlinear PDEs
Andrei D. Polyanin, Alexei I. Zhurov

Boundary Value Problems on Time Scales, Volume I
Svetlin Georgiev, Khaled Zennir

Boundary Value Problems on Time Scales, Volume II
Svetlin Georgiev, Khaled Zennir

Observability and Mathematics

Fluid Mechanics, Solutions of Navier-Stokes Equations, and Modeling
Boris Khots

Handbook of Differential Equations, 4th Edition
Daniel Zwillinger, Vladimir Dobrushkin

Experimental Statistics and Data Analysis for Mechanical and Aerospace Engineers
James Middleton

Advanced Engineering Mathematics with MATLAB
Dean G. Duffy

Handbook of Fractional Calculus for Engineering and Science
Harendra Singh, H. M. Srivastava, Juan J Nieto

Advanced Engineering Mathematics
A Second Course
Dean G. Duffy

https://www.routledge.com/Advances-in-Applied-Mathematics/book-series/CRCADVAPPMTH?pd=published,forthcoming&pg=1&pp=12&so=pub&view=list

Advances in Applied Mathematics

Advanced Engineering Mathematics

A Second Course

Dean G. Duffy

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

First edition published 2022
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

ISBN: 9781032133423 (hbk)
ISBN: 9781032223452 (pbk)
ISBN: 9781003272205 (ebk)

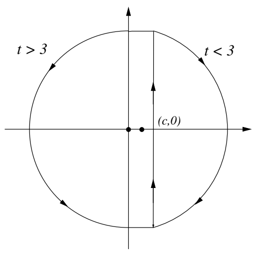
DOI: 10.1201/9781003272205

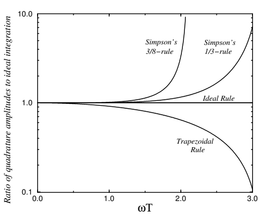
Publisher's note: This book has been prepared from camera-ready copy provided by the authors

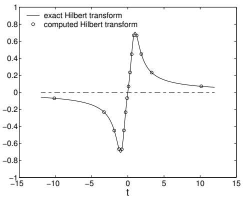
Dedicated to the Brigade of Midshipmen

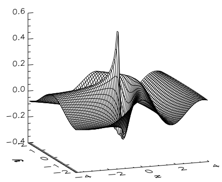
and the Corps of Cadets

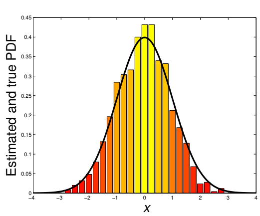
Contents

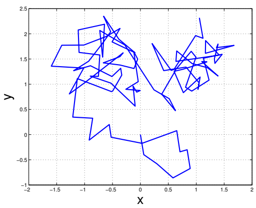

t > 3
t < 3
(c,0)

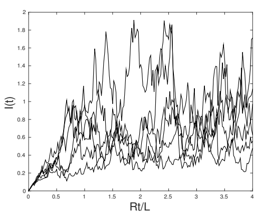

Ratio of quadrature amplitudes to ideal integration
10.0
1.0
0.1
0.0
1.0
2.0
3.0
ωT
Simpson's 3/8−rule
Simpson's 1/3−rule
Ideal Rule
Trapezoidal Rule


exact Hilbert transform
computed Hilbert transform
t




Estimated and true PDF
x


y
x


I(t)
Rt/L

Acknowledgments

I would like to thank the many midshipmen and cadets who have taken engineering mathematics from me. They have been willing or unwilling guinea pigs in testing out many of the ideas and problems in this book. Special thanks go to Prof. William S. Price of the University of Western Sydney (Australia) for his suggestions concerning a chapter on Green's functions. Most of the plots and calculations were done using MATLAB®.

Author

Dean G. Duffy received his bachelor of science in geophysics from Case Institute of Technology (Cleveland, Ohio) and his doctorate of science in meteorology from the Massachusetts Institute of Technology (Cambridge, Massachusetts). He served in the United States Air Force from September 1975 to December 1979 as a numerical weather prediction officer. After his military service, he began a twenty-five-year (1980 to 2005) association with NASA at the Goddard Space Flight Center (Greenbelt, Maryland) where he focused on numerical weather prediction, oceanic wave modeling, and dynamical meteorology. He also wrote papers in the areas of Laplace transforms, antenna theory, railroad tracks, and heat conduction. In addition to his NASA duties, he taught engineering mathematics, differential equations, and calculus at the United States Naval Academy (Annapolis, Maryland) and the United States Military Academy (West Point, New York). Drawing from his teaching experience, he has written several books on transform methods, engineering mathematics, Green's functions, and mixed-boundary-value problems.

Introduction

For the last twenty-five years I have written a series of engineering mathematics books. When it came to revising my fourth edition, I realized that something radical must be done. The encyclopedic engineering mathematics tome is dead, killed by the growth of the Internet and students' unwillingness to buy these books. When I surveyed the current variety of engineering mathematics courses, I realized that I needed to write two books. The companion to the present volume (*Advanced Engineering Mathematics with MATLAB*) would only focus on those topics that are currently taught in most advanced engineering mathematics courses. When one had studied that volume, he/she could feel confident that he/she had a solid knowledge of those mathematical techniques used in current engineering and scientific courses.

This volume (*Advanced Engineering Mathematics: A Second Course*) is my attempt to look into the future of advanced engineering mathematics courses. Some of the material, such as complex variables and probability, is currently taught to engineers, although not usually in courses entitled advanced engineering mathematics. I have included these topics because they are required for transform methods and random processes.

One trend that I see is that entering freshmen are increasingly likely to have had calculus in high school. This means that they will probably place out of the traditional differential and integral calculus courses in their freshman year, allowing them to take multivariable calculus and differential equations during their freshman year and taking advanced engineering mathematics courses during their sophomore year. Therefore, the question arises as to nature of these courses. The answer appears to be that the current traditional engineering mathematics course will occur during the fall semester and some other mathematics course will occur during the spring semester. The present volume is designed to meet this need, as well as stand as the advanced engineering mathematics text on its own. For those past formal education, this book provides the professional with powerful mathematical techniques.

The first five chapters are aimed at the systems, communications and electrical engineering crowd: those involved in the digital revolution. First, that portion of complex variable theory is presented so that the reader will feel prepared in dealing with transform methods. For example, Chapter 2 shows how complex variables can be used to invert particularly complicated Fourier and Laplace transforms. This chapter also illustrates how transform methods can solve the heat, wave and Laplace's equations.

In Chapters 3 and 4 we study two transforms, the z- and Hilbert transforms, that are currently important in the digital revolution. Chapter 3 introduces the z-transform by first giving its definition and then developing some of its general properties. We also illustrate how to compute the inverse by long division, partial fractions, and contour integration. Finally, we use z-transforms to solve difference equations, especially with respect to the stability of the system.

The Hilbert transform is important in the explosion of interest in communications. The Hilbert transform is introduced in Section 4.1 and its properties are explored in Section 4.2. Two important applications of Hilbert transforms are introduced in Sections 4.3 and 4.4, namely the concept of analytic signals and the Kramers-Kronig relationship.

To round out this area we present Green's function in Chapter 5. Green's function gives the response of a system to impulse forcing without the clouding effects of a particular forcing function or initial conditions. Each successive section deals with ordinary, wave, heat and Helmholtz's equations. The solution to general problems follows from the superposition integral.

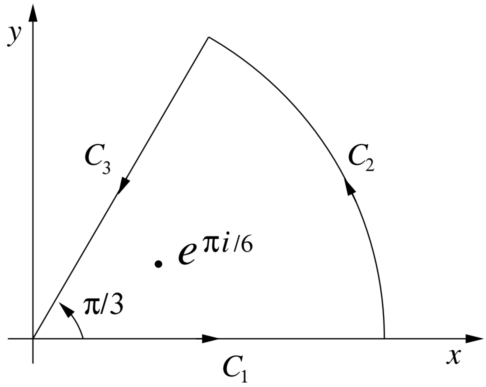
The book concludes by turning to the future. It is now recognized that random processes are useful in describing many physical systems. We begin by introducing the fundamental concepts behind probability in Chapter 6 and random processes in Chapter 7. Chapter 6 introduces the student to the concepts of probability distributions, mean, and variance because these topics appear so frequently in random processes. Chapter 7 explores common random processes such as Poisson processes and birth and death.

A unique aspect of this book appears in Chapter 8, which is devoted to stochastic calculus. We start by exploring deterministic differential equations with a stochastic forcing. Next, the important stochastic process of Brownian motion is developed in depth. Using this Brownian motion, we introduce the concept of (Itô) stochastic integration, Itô's lemma, and stochastic differential equations. The chapter concludes with various numerical methods to integrate stochastic differential equations.

MATLAB is still employed to reinforce the concepts that are taught. Of course, this book still continues my principle of including a wealth of examples from the scientific and engineering literature. Worked solutions to all of the problems are given at the end.

List of Definitions

Function	Definition
$\delta(t-a)$	$=\begin{cases}\infty, & t=a,\\ 0, & t\neq a,\end{cases} \qquad \int_{-\infty}^{\infty}\delta(t-a)\,dt=1$
$\operatorname{erf}(x)$	$=\dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^x e^{-y^2}\,dy$
$\Gamma(x)$	gamma function
$H(t-a)$	$=\begin{cases}1, & t>a,\\ 0, & t<a.\end{cases}$
$\Im(z)$	imaginary part of the complex variable z
$I_n(x)$	modified Bessel function of the first kind and order n
$J_n(x)$	Bessel function of the first kind and order n
$K_n(x)$	modified Bessel function of the second kind and order n
$P_n(x)$	Legendre polynomial of order n
$\Re(z)$	real part of the complex variable z
$\operatorname{sgn}(t-a)$	$=\begin{cases}-1, & t<a,\\ 1, & t>a.\end{cases}$
$Y_n(x)$	Bessel function of the second kind and order n



Chapter 1
Complex Variables

The theory of complex variables was originally developed by mathematicians as an aid in understanding functions. Functions of a complex variable enjoy many powerful properties that their real counterparts do not. That is *not* why we will study them. For us they provide the keys for the complete mastery of transform methods and differential equations.

In this chapter all of our work points to one objective: integration on the complex plane by the method of residues. For this reason we minimize discussions of limits and continuity, which play such an important role in conventional complex variables, in favor of the computational aspects. We begin by introducing some simple facts about complex variables. Then we progress to differential and integral calculus on the complex plane.

1.1 COMPLEX NUMBERS

A *complex number* is any number of the form $a+bi$, where a and b are real and $i=\sqrt{-1}$. We denote any member of a *set* of complex numbers by the *complex variable* $z = x + iy$. The real part of z, usually denoted by $\Re(z)$, is x while the imaginary part of z, $\Im(z)$, is y. The *complex conjugate*, $\overline{z}$ or z^*, of the complex number $a + bi$ is $a - bi$.

Complex numbers obey the fundamental rules of algebra. Thus, two complex numbers $a + bi$ and $c + di$ are equal if and only if $a = c$ and $b = d$. Just as real numbers have the fundamental operations of addition, subtraction, multiplication, and division, so too do complex numbers. These operations are defined:

Addition

$$(a + bi) + (c + di) = (a + c) + (b + d)i \qquad \textbf{(1.1.1)}$$

Subtraction

$$(a + bi) - (c + di) = (a - c) + (b - d)i \qquad \textbf{(1.1.2)}$$

Multiplication

$$(a+bi)(c+di) = ac + bci + adi + i^2bd = (ac-bd) + (ad+bc)i \tag{1.1.3}$$

Division

$$\frac{a+bi}{c+di} = \frac{a+bi}{c+di}\,\frac{c-di}{c-di} = \frac{ac-adi+bci-bdi^2}{c^2+d^2} = \frac{ac+bd+(bc-ad)i}{c^2+d^2}. \tag{1.1.4}$$

The *absolute value* or *modulus* of a complex number $a+bi$, written $|a+bi|$, equals $\sqrt{a^2+b^2}$. Additional properties include:

$$|z_1z_2z_3\cdots z_n| = |z_1||z_2||z_3|\cdots|z_n| \tag{1.1.5}$$

$$|z_1/z_2| = |z_1|/|z_2| \quad \text{if} \quad z_2 \neq 0 \tag{1.1.6}$$

$$|z_1+z_2+z_3+\cdots+z_n| \leq |z_1|+|z_2|+|z_3|+\cdots+|z_n| \tag{1.1.7}$$

and

$$|z_1+z_2| \geq |z_1|-|z_2|. \tag{1.1.8}$$

The use of inequalities with complex variables has meaning only when they involve absolute values.

It is often useful to plot the complex number $x+iy$ as a point (x,y) in the xy-plane, now called the *complex plane.* Figure 1.1.1 illustrates this representation.

This geometrical interpretation of a complex number suggests an alternative method of expressing a complex number: the polar form. From the polar representation of x and y,

$$x = r\cos(\theta) \quad \text{and} \quad y = r\sin(\theta), \tag{1.1.9}$$

where $r=\sqrt{x^2+y^2}$ is the *modulus*, *amplitude*, or *absolute value* of z and θ is the *argument* or *phase*, we have that

$$z = x+iy = r[\cos(\theta)+i\sin(\theta)]. \tag{1.1.10}$$

However, from the Taylor expansion of the exponential in the real case,

$$e^{i\theta} = \sum_{k=0}^{\infty}\frac{(\theta i)^k}{k!}. \tag{1.1.11}$$

Expanding Equation 1.1.11,

$$e^{i\theta} = 1-\frac{\theta^2}{2!}+\frac{\theta^4}{4!}-\frac{\theta^6}{6!}+\cdots+i\left(\theta-\frac{\theta^3}{3!}+\frac{\theta^5}{5!}-\frac{\theta^7}{7!}+\cdots\right) \tag{1.1.12}$$

$$= \cos(\theta)+i\sin(\theta). \tag{1.1.13}$$

Equation 1.1.13 is *Euler's formula.* Consequently, we may express Equation 1.1.10 as

$$z = re^{i\theta}, \tag{1.1.14}$$

which is the *polar form* of a complex number. Furthermore, because

$$z^n = r^ne^{in\theta} \tag{1.1.15}$$

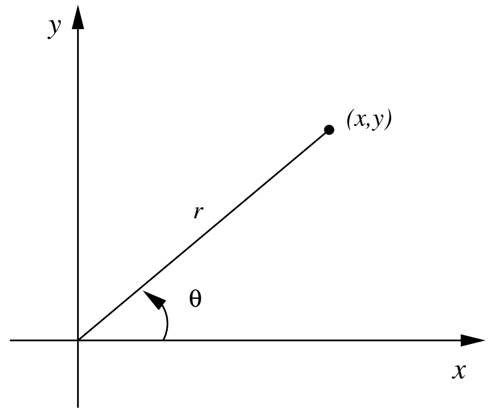


Figure 1.1.1: The complex plane.

by the law of exponents,

$$z^n = r^n[\cos(n\theta) + i\sin(n\theta)]. \tag{1.1.16}$$

Equation 1.1.16 is *De Moivre's theorem.*

• **Example 1.1.1**

Let us simplify the following complex number:

$$\frac{3-2i}{-1+i} = \frac{3-2i}{-1+i} \times \frac{-1-i}{-1-i} = \frac{-3-3i+2i+2i^2}{1+1} = \frac{-5-i}{2} = -\frac{5}{2} - \frac{i}{2}. \tag{1.1.17}$$

□

• **Example 1.1.2**

Let us reexpress the complex number $-\sqrt{6} - i\sqrt{2}$ in polar form. From Equation 1.1.9 $r = \sqrt{6+2}$ and $\theta = \tan^{-1}(b/a) = \tan^{-1}(1/\sqrt{3}) = \pi/6$ or $7\pi/6$. Because $-\sqrt{6} - i\sqrt{2}$ lies in the third quadrant of the complex plane, $\theta = 7\pi/6$ and

$$-\sqrt{6} - i\sqrt{2} = 2\sqrt{2}e^{7\pi i/6}. \tag{1.1.18}$$

Note that Equation 1.1.18 is not a unique representation because $\pm 2n\pi$ may be added to $7\pi/6$ and we still have the same complex number since

$$e^{i(\theta \pm 2n\pi)} = \cos(\theta \pm 2n\pi) + i\sin(\theta \pm 2n\pi) = \cos(\theta) + i\sin(\theta) = e^{i\theta}. \tag{1.1.19}$$

For uniqueness we often choose $n = 0$ and define this choice as the *principal branch.* Other branches correspond to different values of n. □

• **Example 1.1.3**

Find the curve described by the equation $|z - z_0| = a$.

From the definition of the absolute value,

$$\sqrt{(x-x_0)^2 + (y-y_0)^2} = a \tag{1.1.20}$$

or

$$(x - x_0)^2 + (y - y_0)^2 = a^2. \tag{1.1.21}$$

Equation 1.1.21, and hence $|z - z_0| = a$, describes a circle of radius a with its center located at (x_0, y_0). Later on, we shall use equations such as this to describe curves in the complex plane. □

• Example 1.1.4

As an example in manipulating complex numbers, let us show that

$$\left|\frac{a + bi}{b + ai}\right| = 1. \tag{1.1.22}$$

We begin by simplifying

$$\frac{a + bi}{b + ai} = \frac{a + bi}{b + ai} \times \frac{b - ai}{b - ai} = \frac{2ab}{a^2 + b^2} + \frac{b^2 - a^2}{a^2 + b^2} i. \tag{1.1.23}$$

Therefore,

$$\left|\frac{a + bi}{b + ai}\right| = \sqrt{\frac{4a^2b^2}{(a^2 + b^2)^2} + \frac{b^4 - 2a^2b^2 + a^4}{(a^2 + b^2)^2}} = \sqrt{\frac{a^4 + 2a^2b^2 + b^4}{(a^2 + b^2)^2}} = 1. \tag{1.1.24}$$

MATLAB can also be used to solve this problem. Typing the commands

```
>> syms a b real
>> abs((a+b*i)/(b+a*i))
```

yields

```
ans =
1
```

Note that you must declare a and b real in order to get the final result.

Problems

Simplify the following complex numbers. Represent the solution in the Cartesian form $a+bi$. Check your answers using MATLAB.

1. $\dfrac{5i}{2 + i}$ 2. $\dfrac{5 + 5i}{3 - 4i} + \dfrac{20}{4 + 3i}$ 3. $\dfrac{1 + 2i}{3 - 4i} + \dfrac{2 - i}{5i}$

4. $(1 - i)^4$ 5. $i(1 - i\sqrt{3})(\sqrt{3} + i)$ 6. $\dfrac{(7 + i)(1 - 5i)}{(4 - i)(6 + i)}$

Represent the following complex numbers in polar form:

7. $-i$ 8. -4 9. $2 + 2\sqrt{3}\,i$

10. $-5 + 5i$ 11. $2 - 2i$ 12. $-1 + \sqrt{3}\,i$

13. By the law of exponents, $e^{i(\alpha+\beta)} = e^{i\alpha}e^{i\beta}$. Use Euler's formula to obtain expressions for $\cos(\alpha + \beta)$ and $\sin(\alpha + \beta)$ in terms of sines and cosines of α and β.

14. Use De Moivre's theorem with $r = 1$ to express $\cos(4\theta)$ and $\sin(4\theta)$ in terms of $\cos(\theta)$ and $\sin(\theta)$.

15. Using the property that $\sum_{n=0}^{N} q^n = (1-q^{N+1})/(1-q)$ and the geometric series $\sum_{n=0}^{N} e^{int}$, obtain the following sums of trigonometric functions:

$$\sum_{n=0}^{N} \cos(nt) = \cos\left(\frac{Nt}{2}\right)\frac{\sin[(N+1)t/2]}{\sin(t/2)} \quad \text{and} \quad \sum_{n=1}^{N} \sin(nt) = \sin\left(\frac{Nt}{2}\right)\frac{\sin[(N+1)t/2]}{\sin(t/2)}.$$

These results are often called *Lagrange's trigonometric identities.*

16. (a) Using the property that $\sum_{n=0}^{\infty} q^n = 1/(1-q)$, if $|q| < 1$, and the geometric series $\sum_{n=0}^{\infty} \epsilon^n e^{int}$, $|\epsilon| < 1$, show that

$$\sum_{n=0}^{\infty} \epsilon^n \cos(nt) = \frac{1-\epsilon\cos(t)}{1+\epsilon^2-2\epsilon\cos(t)} \quad \text{and} \quad \sum_{n=1}^{\infty} \epsilon^n \sin(nt) = \frac{\epsilon\sin(t)}{1+\epsilon^2-2\epsilon\cos(t)}.$$

(b) Let $\epsilon = e^{-a}$, where $a > 0$. Show that

$$2\sum_{n=1}^{\infty} e^{-na}\sin(nt) = \frac{\sin(t)}{\cosh(a)-\cos(t)}.$$

1.2 FINDING ROOTS

The concept of finding roots of a number, which is rather straightforward in the case of real numbers, becomes more difficult in the case of complex numbers. By finding the *roots* of a complex number, we wish to find all of the solutions w of the equation $w^n = z$, where n is a positive integer for a given z.

We begin by writing z in the polar form:

$$z = re^{i\varphi}, \tag{1.2.1}$$

while we write

$$w = Re^{i\Phi} \tag{1.2.2}$$

for the unknown. Consequently,

$$w^n = R^n e^{in\Phi} = re^{i\varphi} = z. \tag{1.2.3}$$

We satisfy Equation 1.2.3 if

$$R^n = r \quad \text{and} \quad n\Phi = \varphi + 2k\pi, \quad k = 0, \pm1, \pm2, \ldots, \tag{1.2.4}$$

because the addition of any multiple of 2π to the argument is also a solution. Thus, $R = r^{1/n}$, where R is the uniquely determined real positive root, and

$$\Phi_k = \frac{\varphi}{n} + \frac{2\pi k}{n}, \qquad k = 0, \pm1, \pm2, \ldots. \tag{1.2.5}$$

Because $w_k = w_{k\pm n}$, it is sufficient to take $k = 0, 1, 2, \ldots, n-1$. Therefore, there are exactly n solutions:

$$w_k = Re^{\Phi_k i} = r^{1/n} \exp\left[i\left(\frac{\varphi}{n} + \frac{2\pi k}{n}\right)\right] \tag{1.2.6}$$

with $k = 0, 1, 2, \ldots, n-1$. They are the n roots of z. Geometrically we can locate these solutions w_k on a circle, centered at the point $(0, 0)$, with radius R and separated from each other by $2\pi/n$ radians. These roots also form the vertices of a regular polygon of n sides inscribed inside a circle of radius R. (See Example 1.2.1.)

In summary, the method for finding the n roots of a complex number z_0 is as follows. First, write z_0 in its polar form: $z_0 = re^{i\varphi}$. Then multiply the polar form by $e^{2i\pi k}$. Using the law of exponents, take the $1/n$ power of both sides of the equation. Finally, using Euler's formula, evaluate the roots for $k = 0, 1, \ldots, n-1$.

• Example 1.2.1

Let us find all of the values of z for which $z^5 = -32$ and locate these values on the complex plane.

Because

$$-32 = 32e^{\pi i} = 2^5 e^{\pi i}, \tag{1.2.7}$$

$$z_k = 2\exp\left(\frac{\pi i}{5} + \frac{2\pi i k}{5}\right), \qquad k = 0, 1, 2, 3, 4, \tag{1.2.8}$$

or

$$z_0 = 2\exp\left(\frac{\pi i}{5}\right) = 2\left[\cos\left(\frac{\pi}{5}\right) + i\sin\left(\frac{\pi}{5}\right)\right], \tag{1.2.9}$$

$$z_1 = 2\exp\left(\frac{3\pi i}{5}\right) = 2\left[\cos\left(\frac{3\pi}{5}\right) + i\sin\left(\frac{3\pi}{5}\right)\right], \tag{1.2.10}$$

$$z_2 = 2e^{\pi i} = -2, \tag{1.2.11}$$

$$z_3 = 2\exp\left(\frac{7\pi i}{5}\right) = 2\left[\cos\left(\frac{7\pi}{5}\right) + i\sin\left(\frac{7\pi}{5}\right)\right] \tag{1.2.12}$$

and

$$z_4 = 2\exp\left(\frac{9\pi i}{5}\right) = 2\left[\cos\left(\frac{9\pi}{5}\right) + i\sin\left(\frac{9\pi}{5}\right)\right]. \tag{1.2.13}$$

Figure 1.2.1 shows the location of these roots in the complex plane. □

• Example 1.2.2

Let us find the cube roots of $-1 + i$ and locate them graphically.

Because $-1 + i = \sqrt{2}\,\exp(3\pi i/4)$,

$$z_k = 2^{1/6}\exp\left(\frac{\pi i}{4} + \frac{2i\pi k}{3}\right), \qquad k = 0, 1, 2, \tag{1.2.14}$$

or

$$z_0 = 2^{1/6}\exp\left(\frac{\pi i}{4}\right) = 2^{1/6}\left[\cos\left(\frac{\pi}{4}\right) + i\sin\left(\frac{\pi}{4}\right)\right], \tag{1.2.15}$$

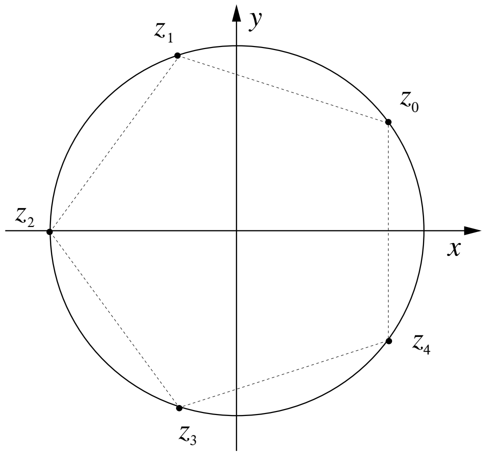


Figure 1.2.1: The zeros of $z^5 = -32$.

$$z_1 = 2^{1/6} \exp\left(\frac{11\pi i}{12}\right) = 2^{1/6}\left[\cos\left(\frac{11\pi}{12}\right) + i\sin\left(\frac{11\pi}{12}\right)\right], \tag{1.2.16}$$

and

$$z_2 = 2^{1/6} \exp\left(\frac{19\pi i}{12}\right) = 2^{1/6}\left[\cos\left(\frac{19\pi}{12}\right) + i\sin\left(\frac{19\pi}{12}\right)\right]. \tag{1.2.17}$$

Figure 1.2.2 gives the location of these zeros on the complex plane. □

• **Example 1.2.3**

The routine `solve` in MATLAB can also be used to compute the roots of complex numbers. For example, let us find all of the roots of $z^4 = -a^4$.

The MATLAB commands are as follows:

```
>> syms a z
>> solve(z^4+a^4)
```

This yields the solution

```
ans=
[  (1/2*2^(1/2)+1/2*i*2^(1/2))*a]
[ (-1/2*2^(1/2)+1/2*i*2^(1/2))*a]
[  (1/2*2^(1/2)-1/2*i*2^(1/2))*a]
[ (-1/2*2^(1/2)-1/2*i*2^(1/2))*a]
```

Problems

Extract all of the possible roots of the following complex numbers. Verify your answer using MATLAB.

1. $8^{1/6}$ 2. $(-1)^{1/3}$ 3. $(-i)^{1/3}$ 4. $(-27i)^{1/6}$

5. Find algebraic expressions for the square roots of $a - bi$, where $a > 0$ and $b > 0$.

6. Find all of the roots for the algebraic equation $z^4 - 3iz^2 - 2 = 0$. Then check your answer using `solve` in MATLAB.

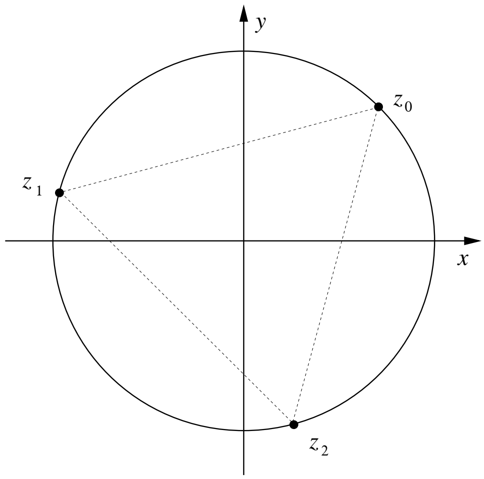


Figure 1.2.2: The zeros of $z^3 = -1 + i$.

7. Find all of the roots for the algebraic equation $z^4 + 6iz^2 + 16 = 0$. Then check your answer using `solve` in MATLAB.

1.3 THE DERIVATIVE IN THE COMPLEX PLANE: THE CAUCHY-RIEMANN EQUATIONS

In the previous two sections, we introduced complex arithmetic. We are now ready for the concept of function as it applies to complex variables.

We already defined the complex variable $z = x+iy$, where x and y are variable. We now introduce another complex variable $w = u + iv$ so that for each value of z there corresponds a value of $w = f(z)$. From all of the possible complex functions that we might invent, we focus on those functions where for each z there is one, and only one, value of w. These functions are *single-valued*. They differ from functions such as the square root, logarithm, and inverse sine and cosine, where there are multiple answers for each z. These *multivalued functions* do arise in various problems. However, they are beyond the scope of this book and we shall always assume that we are dealing with single-valued functions.

A popular method for representing a complex function involves drawing some closed domain in the z-plane and then showing the corresponding domain in the w-plane. This procedure is called *mapping* and the z-plane illustrates the *domain* of the function while the w-plane illustrates its *image* or *range*. Figure 1.3.1 shows the z-plane and w-plane for $w = z^2$; a pie-shaped wedge in the z-plane maps into a semicircle on the w-plane.

• Example 1.3.1

Given the complex function $w = e^{-z^2}$, let us find the corresponding $u(x, y)$ and $v(x, y)$. From Euler's formula,

$$w = e^{-z^2} = e^{-(x+iy)^2} = e^{y^2-x^2}e^{-2ixy} = e^{y^2-x^2}[\cos(2xy) - i\sin(2xy)]. \tag{1.3.1}$$

Therefore, by inspection,

$$u(x, y) = e^{y^2-x^2}\cos(2xy), \qquad \text{and} \qquad v(x, y) = -e^{y^2-x^2}\sin(2xy). \tag{1.3.2}$$

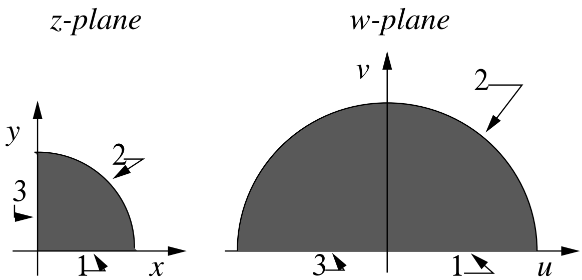


Figure 1.3.1: The complex function $w = z^2$.

Note that there is no i in the expression for $v(x, y)$. The function $w = f(z)$ is single-valued because for each distinct value of z, there is a unique value of $u(x, y)$ and $v(x, y)$. □

• Example 1.3.2

As counterpoint, let us show that $w = \sqrt{z}$ is a multivalued function.

We begin by writing $z = re^{i\theta+2\pi ik}$, where $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$. Then,

$$w_k = \sqrt{r}e^{i\theta/2+\pi ik}, \qquad k = 0, 1, \tag{1.3.3}$$

or

$$w_0 = \sqrt{r}\,[\cos(\theta/2) + i\sin(\theta/2)] \qquad \text{and} \qquad w_1 = -w_0. \tag{1.3.4}$$

Therefore,

$$u_0(x, y) = \sqrt{r}\cos(\theta/2), \qquad v_0(x, y) = \sqrt{r}\sin(\theta/2), \tag{1.3.5}$$

and

$$u_1(x, y) = -\sqrt{r}\cos(\theta/2), \qquad v_1(x, y) = -\sqrt{r}\sin(\theta/2). \tag{1.3.6}$$

Each solution w_0 or w_1 is a *branch* of the multivalued function $\sqrt{z}$. We can make $\sqrt{z}$ single-valued by restricting ourselves to a single branch, say w_0. In that case, the $\Re(w) > 0$ if we restrict $-\pi < \theta < \pi$. Although this is not the only choice that we could have made, it is a popular one. For example, most digital computers use this definition in their complex square root function. The point here is our ability to make a multivalued function single-valued by defining a particular branch. □

Although the requirement that a complex function be single-valued is important, it is still too general and would cover all functions of two real variables. To have a useful theory, we must introduce additional constraints. Because an important property associated with most functions is the ability to take their derivative, let us examine the derivative in the complex plane.

Following the definition of a derivative for a single real variable, the derivative of a complex function $w = f(z)$ is defined as

$$\frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}. \tag{1.3.7}$$

A function of a complex variable that has a derivative at every point within a region of the complex plane is said to be *analytic* (or *regular* or *holomorphic*) over that region. If the function is analytic everywhere in the complex plane, it is *entire*.

Because the derivative is defined as a limit and limits are well behaved with respect to elementary algebraic operations, the following operations carry over from elementary calculus:

$$\frac{d}{dz}\left[cf(z)\right] = cf'(z), \qquad c \text{ a constant} \tag{1.3.8}$$

$$\frac{d}{dz}\left[f(z) \pm g(z)\right] = f'(z) \pm g'(z) \tag{1.3.9}$$

$$\frac{d}{dz}\left[f(z)g(z)\right] = f'(z)g(z) + f(z)g'(z) \tag{1.3.10}$$

$$\frac{d}{dz}\left[\frac{f(z)}{g(z)}\right] = \frac{g(z)f'(z) - g'(z)f(z)}{g^2(z)} \tag{1.3.11}$$

$$\frac{d}{dz}\left\{f[g(z)]\right\} = f'[g(z)]g'(z), \qquad \text{the chain rule.} \tag{1.3.12}$$

Another important property that carries over from real variables is l'Hôpital's rule: Let $f(z)$ and $g(z)$ be analytic at z_0, where $f(z)$ has a zero[1] of order m and $g(z)$ has a zero of order n. Then, if $m > n$,

$$\lim_{z \to z_0} \frac{f(z)}{g(z)} = 0; \tag{1.3.13}$$

if $m = n$,

$$\lim_{z \to z_0} \frac{f(z)}{g(z)} = \frac{f^{(m)}(z_0)}{g^{(m)}(z_0)}; \tag{1.3.14}$$

and if $m < n$,

$$\lim_{z \to z_0} \frac{f(z)}{g(z)} = \infty. \tag{1.3.15}$$

• Example 1.3.3

Let us evaluate $\lim_{z \to i}(z^{10} + 1)/(z^6 + 1)$. From l'Hôpital's rule,

$$\lim_{z \to i} \frac{z^{10} + 1}{z^6 + 1} = \lim_{z \to i} \frac{10z^9}{6z^5} = \frac{5}{3} \lim_{z \to i} z^4 = \frac{5}{3}. \tag{1.3.16}$$

□

So far, we introduced the derivative and some of its properties. But how do we actually know whether a function is analytic or how do we compute its derivative? At this point we must develop some relationships involving the known quantities $u(x, y)$ and $v(x, y)$.

We begin by returning to the definition of the derivative. Because $\Delta z = \Delta x + i\Delta y$, there is an infinite number of different ways of approaching the limit $\Delta z \to 0$. Uniqueness of that limit requires that Equation 1.3.7 must be independent of the manner in which Δz approaches zero. A simple example is to take Δz in the x-direction so that $\Delta z = \Delta x$; another is to take Δz in the y-direction so that $\Delta z = i\Delta y$. These examples yield

$$\frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \lim_{\Delta x \to 0} \frac{\Delta u + i\Delta v}{\Delta x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} \tag{1.3.17}$$

---

[1] An analytic function $f(z)$ has a zero of order m at z_0 if and only if $f(z_0) = f'(z_0) = \cdots = f^{(m-1)}(z_0) = 0$ and $f^{(m)}(z_0) \neq 0$.

Although educated as an engineer, Augustin-Louis Cauchy (1789–1857) would become a mathematician's mathematician, publishing 789 papers and 7 books in the fields of pure and applied mathematics. His greatest writings established the discipline of mathematical analysis as he refined the notions of limit, continuity, function, and convergence. It was this work on analysis that led him to develop complex function theory via the concept of residues. (Portrait courtesy of the Archives de l'Académie des sciences, Paris.)

and

$$\frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \lim_{\Delta y \to 0} \frac{\Delta u + i\Delta v}{i\Delta y} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}. \tag{1.3.18}$$

In both cases we are approaching zero from the positive side. For the limit to be unique and independent of path, Equation 1.3.17 must equal Equation 1.3.18, or

$$\boxed{\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.} \tag{1.3.19}$$

These equations that u and v must both satisfy are the *Cauchy-Riemann* equations. They are necessary but not sufficient to ensure that a function is differentiable. The following example illustrates this.

• **Example 1.3.4**

Consider the complex function

$$w = \begin{cases} z^5/|z|^4, & z \neq 0 \\ 0, & z = 0. \end{cases} \tag{1.3.20}$$

The derivative at $z = 0$ is given by

$$\frac{dw}{dz} = \lim_{\Delta z \to 0} \frac{(\Delta z)^5/|\Delta z|^4 - 0}{\Delta z} = \lim_{\Delta z \to 0} \frac{(\Delta z)^4}{|\Delta z|^4}, \tag{1.3.21}$$

Despite his short life, (Georg Friedrich) Bernhard Riemann's (1826–1866) mathematical work contained many imaginative and profound concepts. It was in his doctoral thesis on complex function theory (1851) that he introduced the Cauchy-Riemann differential equations. Riemann's later work dealt with the definition of the integral and the foundations of geometry and non-Euclidean (elliptic) geometry. (Portrait courtesy of Photo AKG, London, with permission.)

provided that this limit exists. However, this limit does not exist because, in general, the numerator depends upon the path used to approach zero. For example, if $\Delta z = re^{\pi i/4}$ with $r \to 0$, $dw/dz = -1$. On the other hand, if $\Delta z = re^{\pi i/2}$ with $r \to 0$, $dw/dz = 1$.

Are the Cauchy-Riemann equations satisfied in this case? To check this, we first compute

$$u_x(0,0) = \lim_{\Delta x \to 0} \left(\frac{\Delta x}{|\Delta x|} \right)^4 = 1, \qquad v_y(0,0) = \lim_{\Delta y \to 0} \left(\frac{i\Delta y}{|\Delta y|} \right)^4 = 1, \tag{1.3.22}$$

$$u_y(0,0) = \lim_{\Delta y \to 0} \Re\left[\frac{(i\Delta y)^5}{\Delta y |\Delta y|^4} \right] = 0, \quad \text{and} \quad v_x(0,0) = \lim_{\Delta x \to 0} \Im\left[\left(\frac{\Delta x}{|\Delta x|} \right)^4 \right] = 0. \tag{1.3.23}$$

Hence, the Cauchy-Riemann equations are satisfied at the origin. Thus, even though the derivative is not uniquely defined, Equation 1.3.21 happens to have the same value for paths taken along the coordinate axes so that the Cauchy-Riemann equations are satisfied. □

In summary, if a function is differentiable at a point, the Cauchy-Riemann equations hold. Similarly, if the Cauchy-Riemann equations are not satisfied at a point, then the function is not differentiable at that point. This is one of the important uses of the Cauchy-Riemann equations: the location of nonanalytic points. Isolated nonanalytic points of an otherwise analytic function are called *isolated singularities*. Functions that contain isolated singularities are called *meromorphic*.

The Cauchy-Riemann condition can be modified so that it is sufficient for the derivative to exist. Let us require that u_x, u_y, v_x, and v_y be continuous in some region surrounding a point z_0 and satisfy the Cauchy-Riemann equations there. Then

$$\begin{aligned}
f(z) - f(z_0) &= [u(z) - u(z_0)] + i[v(z) - v(z_0)] && \textbf{(1.3.24)}\\
&= [u_x(z_0)(x - x_0) + u_y(z_0)(y - y_0) + \epsilon_1(x - x_0) + \epsilon_2(y - y_0)]\\
&+ i[v_x(z_0)(x - x_0) + v_y(z_0)(y - y_0) + \epsilon_3(x - x_0) + \epsilon_4(y - y_0)] && \textbf{(1.3.25)}\\
&= [u_x(z_0) + iv_x(z_0)](z - z_0) + (\epsilon_1 + i\epsilon_3)(x - x_0)\\
&+ (\epsilon_2 + i\epsilon_4)(y - y_0), && \textbf{(1.3.26)}
\end{aligned}$$

where we used the Cauchy-Riemann equations and $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4 \to 0$ as $\Delta x, \Delta y \to 0$. Hence,

$$f'(z_0) = \lim_{\Delta z \to 0} \frac{f(z) - f(z_0)}{\Delta z} = u_x(z_0) + iv_x(z_0), \qquad \textbf{(1.3.27)}$$

because $|\Delta x| \leq |\Delta z|$ and $|\Delta y| \leq |\Delta z|$. Using Equation 1.3.27 and the Cauchy-Riemann equations, we can obtain the derivative from any of the following formulas:

$$\boxed{\frac{dw}{dz} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y},} \qquad \textbf{(1.3.28)}$$

and

$$\boxed{\frac{dw}{dz} = \frac{\partial v}{\partial y} + i\frac{\partial v}{\partial x} = \frac{\partial u}{\partial x} - i\frac{\partial u}{\partial y}.} \qquad \textbf{(1.3.29)}$$

Furthermore, $f'(z_0)$ is continuous because the partial derivatives are.

• **Example 1.3.5**

Let us show that $\sin(z)$ is an entire function.

$$\begin{aligned}
w &= \sin(z) && \textbf{(1.3.30)}\\
u + iv &= \sin(x + iy) = \sin(x)\cos(iy) + \cos(x)\sin(iy) && \textbf{(1.3.31)}\\
&= \sin(x)\cosh(y) + i\cos(x)\sinh(y), && \textbf{(1.3.32)}
\end{aligned}$$

because

$$\cos(iy) = \tfrac{1}{2}\left[e^{i(iy)} + e^{-i(iy)}\right] = \tfrac{1}{2}\left[e^{y} + e^{-y}\right] = \cosh(y), \qquad \textbf{(1.3.33)}$$

and

$$\sin(iy) = \tfrac{1}{2i}\left[e^{i(iy)} - e^{-i(iy)}\right] = -\tfrac{1}{2i}\left[e^{y} - e^{-y}\right] = i\sinh(y), \qquad \textbf{(1.3.34)}$$

so that

$$u(x,y) = \sin(x)\cosh(y), \quad \text{and} \quad v(x,y) = \cos(x)\sinh(y). \qquad \textbf{(1.3.35)}$$

Differentiating both $u(x,y)$ and $v(x,y)$ with respect to x and y, we have that

$$\frac{\partial u}{\partial x} = \cos(x)\cosh(y), \qquad \frac{\partial u}{\partial y} = \sin(x)\sinh(y), \qquad \textbf{(1.3.36)}$$

$$\frac{\partial v}{\partial x} = -\sin(x)\sinh(y), \qquad \frac{\partial v}{\partial y} = \cos(x)\cosh(y), \tag{1.3.37}$$

and $u(x,y)$ and $v(x,y)$ satisfy the Cauchy-Riemann equations for all values of x and y. Furthermore, u_x, u_y, v_x, and v_y are continuous for all x and y. Therefore, the function $w = \sin(z)$ is an entire function. □

• Example 1.3.6

Consider the function $w = 1/z$. Then

$$w = u + iv = \frac{1}{x+iy} = \frac{x}{x^2+y^2} - \frac{iy}{x^2+y^2}. \tag{1.3.38}$$

Therefore,

$$u(x,y) = \frac{x}{x^2+y^2}, \qquad \text{and} \qquad v(x,y) = -\frac{y}{x^2+y^2}. \tag{1.3.39}$$

Now

$$\frac{\partial u}{\partial x} = \frac{(x^2+y^2)-2x^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2}, \tag{1.3.40}$$

$$\frac{\partial v}{\partial y} = -\frac{(x^2+y^2)-2y^2}{(x^2+y^2)^2} = \frac{y^2-x^2}{(x^2+y^2)^2} = \frac{\partial u}{\partial x}, \tag{1.3.41}$$

$$\frac{\partial v}{\partial x} = -\frac{0-2xy}{(x^2+y^2)^2} = \frac{2xy}{(x^2+y^2)^2}, \tag{1.3.42}$$

and

$$\frac{\partial u}{\partial y} = \frac{0-2xy}{(x^2+y^2)^2} = -\frac{2xy}{(x^2+y^2)^2} = -\frac{\partial v}{\partial x}. \tag{1.3.43}$$

The function is analytic at all points except the origin because the function itself ceases to exist when both x and y are zero and the modulus of w becomes infinite. □

• Example 1.3.7

Let us find the derivative of $\sin(z)$.

Using Equation 1.3.28 and Equation 1.3.32,

$$\frac{d}{dz}\left[\sin(z)\right] = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = \cos(x)\cosh(y) - i\sin(x)\sinh(y) = \cos(x+iy) = \cos(z). \tag{1.3.44}$$

Similarly,

$$\frac{d}{dz}\left(\frac{1}{z}\right) = \frac{y^2-x^2}{(x^2+y^2)^2} + \frac{2ixy}{(x^2+y^2)^2} = -\frac{1}{(x+iy)^2} = -\frac{1}{z^2}. \tag{1.3.45}$$

□

The results in the above examples are identical to those for z real. As we showed earlier, the fundamental rules of elementary calculus apply to complex differentiation. Consequently, it is usually simpler to apply those rules to find the derivative rather than breaking $f(z)$ down into its real and imaginary parts, applying either Equation 1.3.28 or Equation 1.3.29, and then putting everything back together.

An additional property of analytic functions follows by cross differentiating the Cauchy-Riemann equations, or

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y} = -\frac{\partial^2 u}{\partial y^2}, \quad \text{or} \quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \tag{1.3.46}$$

and

$$\frac{\partial^2 v}{\partial x^2} = -\frac{\partial^2 u}{\partial x \partial y} = -\frac{\partial^2 v}{\partial y^2}, \quad \text{or} \quad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0. \tag{1.3.47}$$

Any function that has continuous partial derivatives of second order and satisfies Laplace's equation, Equation 1.3.46 or Equation 1.3.47, is called a *harmonic function*. Because both $u(x, y)$ and $v(x, y)$ satisfy Laplace's equation if $f(z) = u + iv$ is analytic, $u(x, y)$ and $v(x, y)$ are called *conjugate harmonic functions*.

• Example 1.3.8

Given that $u(x, y) = e^{-x}[x \sin(y) - y \cos(y)]$, let us show that u is harmonic and find a conjugate harmonic function $v(x, y)$ such that $f(z) = u + iv$ is analytic.

Because

$$\frac{\partial^2 u}{\partial x^2} = -2e^{-x} \sin(y) + xe^{-x} \sin(y) - ye^{-x} \cos(y), \tag{1.3.48}$$

and

$$\frac{\partial^2 u}{\partial y^2} = -xe^{-x} \sin(y) + 2e^{-x} \sin(y) + ye^{-x} \cos(y), \tag{1.3.49}$$

it follows that $u_{xx} + u_{yy} = 0$. Therefore, $u(x, y)$ is harmonic. From the Cauchy-Riemann equations,

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x} = e^{-x} \sin(y) - xe^{-x} \sin(y) + ye^{-x} \cos(y), \tag{1.3.50}$$

and

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} = e^{-x} \cos(y) - xe^{-x} \cos(y) - ye^{-x} \sin(y). \tag{1.3.51}$$

Integrating Equation 1.3.50 with respect to y,

$$v(x, y) = ye^{-x} \sin(y) + xe^{-x} \cos(y) + g(x). \tag{1.3.52}$$

Using Equation 1.3.51,

$$\begin{aligned} v_x &= -ye^{-x} \sin(y) - xe^{-x} \cos(y) + e^{-x} \cos(y) + g'(x) \\ &= e^{-x} \cos(y) - xe^{-x} \cos(y) - ye^{-x} \sin(x). \end{aligned} \qquad (1.3.53$$

Therefore, $g'(x) = 0$ or $g(x) = \text{constant}$. Consequently,

$$v(x, y) = e^{-x}[y \sin(y) + x \cos(y)] + \text{constant}. \tag{1.3.54}$$

Hence, for our real harmonic function $u(x, y)$, there are infinitely many harmonic conjugates $v(x, y)$, which differ from each other by an additive constant.

Problems

Show that the following functions are entire:

1. $f(z) = iz + 2$ 2. $f(z) = e^{-z}$ 3. $f(z) = z^3$ 4. $f(z) = \cosh(z)$

Find the derivative of the following functions:

5. $f(z) = (1 + z^2)^{3/2}$ 6. $f(z) = (z + 2z^{1/2})^{1/3}$ 7. $f(z) = (1 + 4i)z^2 - 3z - 2$

8. $f(z) = (2z - i)/(z + 2i)$ 9. $f(z) = (iz - 1)^{-3}$ 10. $f(z) = z/(z^3 + 1)$

Evaluate the following limits:

11. $\displaystyle\lim_{z\to i} \frac{z^2 - 2iz - 1}{z^4 + 2z^2 + 1}$ 12. $\displaystyle\lim_{z\to 0} \frac{z - \sin(z)}{z^3}$ 13. $\displaystyle\lim_{z\to n} \frac{z - n}{\sin(\pi z)}$

Here, n is an integer.

14. Show that the function $f(z) = z^*$ is nowhere differentiable.

For each of the following $u(x, y)$, show that it is harmonic and then find a corresponding $v(x, y)$ such that $f(z) = u + iv$ is analytic.

15. $u(x, y) = x^2 - y^2$ 16. $u(x, y) = x^4 - 6x^2y^2 + y^4 + x$

17. $u(x, y) = x\cos(x)e^{-y} - y\sin(x)e^{-y}$ 18. $u(x, y) = (x^2 - y^2)\cos(y)e^x - 2xy\sin(y)e^x$

1.4 LINE INTEGRALS

So far, we discussed complex numbers, complex functions, and complex differentiation. We are now ready for integration.

Just as we have integrals involving real variables, we can define an integral that involves complex variables. Because the z-plane is two-dimensional, there is clearly greater freedom in what we mean by a complex integral. For example, we might ask whether the integral of some function between points A and B depends upon the curve along which we integrate. (In general it does.) Consequently, an important ingredient in any complex integration is the *contour* that we follow during the integration.

The result of a line integral is a complex number or expression. Unlike its counterpart in real variables, there is no physical interpretation for this quantity, such as area under a curve. Generally, integration in the complex plane is an intermediate process with a physically realizable quantity occurring only after we take its real or imaginary part. For example, in potential fluid flow, the lift and drag are found by taking the real and imaginary parts of a complex integral, respectively.

How do we compute $\int_C f(z)\,dz$? Let us deal with the definition; we illustrate the actual method by examples.

A popular method for evaluating complex line integrals consists of breaking everything up into real and imaginary parts. This reduces the integral to line integrals of real-valued functions, which we know how to handle. Thus, we write $f(z) = u(x, y) + iv(x, y)$ as usual, and because $z = x + iy$, formally $dz = dx + i\,dy$. Therefore,

$$\int_C f(z)\,dz = \int_C [u(x, y) + iv(x, y)][dx + i\,dy] \tag{1.4.1}$$

$$= \int_C u(x, y)\,dx - v(x, y)\,dy + i\int_C v(x, y)\,dx + u(x, y)\,dy. \tag{1.4.2}$$

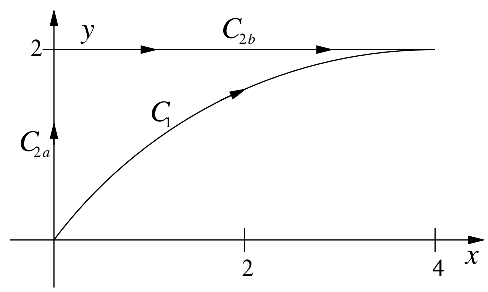


Figure 1.4.1: Contour used in Example 1.4.1.

The exact method used to evaluate Equation 1.4.2 depends upon the exact path specified.

From the definition of the line integral, we have the following self-evident properties:

$$\int_C f(z)\,dz = -\int_{C'} f(z)\,dz, \tag{1.4.3}$$

where C' is the contour C taken in the opposite direction of C and

$$\int_{C_1+C_2} f(z)\,dz = \int_{C_1} f(z)\,dz + \int_{C_2} f(z)\,dz. \tag{1.4.4}$$

• **Example 1.4.1**

Let us evaluate $\int_C z^* dz$ from $z = 0$ to $z = 4+2i$ along two different contours. The first consists of the parametric equation $z = t^2 + it$. The second consists of two "dog legs": the first leg runs along the imaginary axis from $z = 0$ to $z = 2i$ and then along a line parallel to the x-axis from $z = 2i$ to $z = 4 + 2i$. See Figure 1.4.1.

For the first case, the points $z = 0$ and $z = 4+2i$ on C_1 correspond to $t = 0$ and $t = 2$, respectively. Then the line integral equals

$$\int_{C_1} z^* dz = \int_0^2 (t^2 + it)^*\, d(t^2 + it) = \int_0^2 (2t^3 - it^2 + t)\,dt = 10 - \tfrac{8i}{3}. \tag{1.4.5}$$

The line integral for the second contour C_2 equals

$$\int_{C_2} z^* dz = \int_{C_{2a}} z^* dz + \int_{C_{2b}} z^* dz, \tag{1.4.6}$$

where C_{2a} denotes the integration from $z = 0$ to $z = 2i$ while C_{2b} denotes the integration from $z = 2i$ to $z = 4 + 2i$. For the first integral,

$$\int_{C_{2a}} z^* dz = \int_{C_{2a}} (x - iy)(dx + i\,dy) = \int_0^2 y\,dy = 2, \tag{1.4.7}$$

because $x = 0$ and $dx = 0$ along C_{2a}. On the other hand, along C_{2b}, $y = 2$ and $dy = 0$ so that

$$\int_{C_{2b}} z^* dz = \int_{C_{2b}} (x - iy)(dx + i\,dy) = \int_0^4 x\,dx + i\int_0^4 -2\,dx = 8 - 8i. \tag{1.4.8}$$

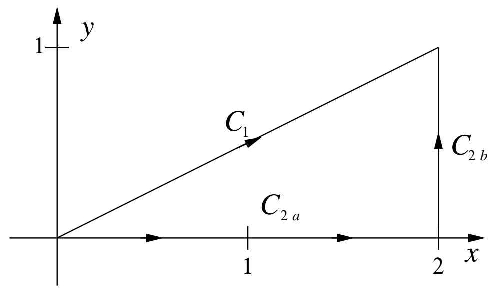


Figure 1.4.2: Contour used in Example 1.4.2.

Thus the value of the entire C_2 contour integral equals the sum of the two parts, or $10 - 8i$.

The point here is that integration along two different paths has given us different results even though we integrated from $z = 0$ to $z = 4 + 2i$ both times. This result foreshadows a general result that is extremely important. Because the integrand contains nonanalytic points along and inside the region enclosed by our two curves, as shown by the Cauchy-Riemann equations, the results depend upon the path taken. Since complex integrations often involve integrands that have nonanalytic points, many line integrations depend upon the contour taken. □

• Example 1.4.2

Let us integrate the *entire* function $f(z) = z^2$ along the two paths from $z = 0$ to $z = 2 + i$ shown in Figure 1.4.2. For the first integration, $x = 2y$, while along the second path we have two straight paths: $z = 0$ to $z = 2$ and $z = 2$ to $z = 2 + i$.

For the first contour integration,

$$\int_{C_1} z^2 dz = \int_0^1 (2y + iy)^2 (2\,dy + i\,dy) = \int_0^1 (3y^2 + 4y^2 i)(2\,dy + i\,dy) \tag{1.4.9}$$

$$= \int_0^1 6y^2\,dy + 8y^2 i\,dy + 3y^2 i\,dy - 4y^2\,dy = \int_0^1 2y^2\,dy + 11y^2 i\,dy \tag{1.4.10}$$

$$= \tfrac{2}{3} y^3 \big|_0^1 + \tfrac{11}{3} i y^3 \big|_0^1 = \tfrac{2}{3} + \tfrac{11i}{3}. \tag{1.4.11}$$

For our second integration,

$$\int_{C_2} z^2\,dz = \int_{C_{2a}} z^2\,dz + \int_{C_{2b}} z^2\,dz. \tag{1.4.12}$$

Along C_{2a} we find that $y = dy = 0$ so that

$$\int_{C_{2a}} z^2\,dz = \int_0^2 x^2\,dx = \tfrac{1}{3} x^3 \big|_0^2 = \tfrac{8}{3}, \tag{1.4.13}$$

and

$$\int_{C_{2b}} z^2\,dz = \int_0^1 (2 + iy)^2 i\,dy = i \left(4y + 2iy^2 - \frac{y^3}{3} \right) \Bigg|_0^1 = 4i - 2 - \tfrac{i}{3}, \tag{1.4.14}$$

because $x = 2$ and $dx = 0$. Consequently,

$$\int_{C_2} z^2\,dz = \frac{2}{3} + \frac{11i}{3}. \tag{1.4.15}$$

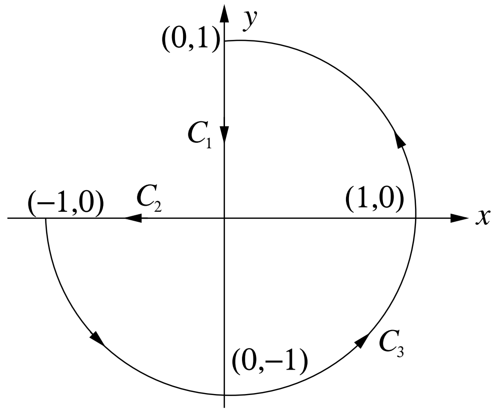


Figure 1.4.3: Contour used in Example 1.4.3.

In this problem we obtained the same results from two different contours of integration. Exploring other contours, we would find that the results are always the same; the integration is path independent. But what makes these results path independent while the integration in Example 1.4.1 was not? Perhaps it is the fact that the integrand is analytic everywhere on the complex plane and there are no nonanalytic points. We will explore this later. □

Finally, an important class of line integrals involves *closed contours*. We denote this special subclass of line integrals by placing a circle on the integral sign: $\oint$. Consider now the following examples:

• Example 1.4.3

Let us integrate $f(z) = z$ around the closed contour shown in Figure 1.4.3.
From Figure 1.4.3,

$$\oint_C z\,dz = \int_{C_1} z\,dz + \int_{C_2} z\,dz + \int_{C_3} z\,dz. \tag{1.4.16}$$

Now

$$\int_{C_1} z\,dz = \int_1^0 iy\,(i\,dy) = -\int_1^0 y\,dy = -\left.\frac{y^2}{2}\right|_1^0 = \frac{1}{2}, \tag{1.4.17}$$

$$\int_{C_2} z\,dz = \int_0^{-1} x\,dx = \left.\frac{x^2}{2}\right|_0^{-1} = \frac{1}{2}, \tag{1.4.18}$$

and

$$\int_{C_3} z\,dz = \int_{-\pi}^{\pi/2} e^{\theta i} i e^{\theta i} d\theta = \left.\frac{e^{2\theta i}}{2}\right|_{-\pi}^{\pi/2} = -1, \tag{1.4.19}$$

where we used $z = e^{\theta i}$ around the portion of the unit circle. Therefore, the closed line integral equals zero. □

• Example 1.4.4

Let us integrate $f(z) = 1/(z - a)$ around any circle centered on $z = a$. The Cauchy-Riemann equations show that $f(z)$ is a meromorphic function. It is analytic everywhere except at the isolated singularity $z = a$.

If we introduce polar coordinates by letting $z-a=re^{\theta i}$ and $dz=ire^{\theta i}d\theta$,

$$\oint_C \frac{dz}{z-a} = \int_0^{2\pi} \frac{ire^{\theta i}}{re^{\theta i}}\,d\theta = i\int_0^{2\pi} d\theta = 2\pi i. \qquad \textbf{(1.4.20)}$$

Note that the integrand becomes undefined at $z=a$. Furthermore, the answer is independent of the size of the circle. Our example suggests that when we have a closed contour integration, it is the behavior of the function within the contour rather than the exact shape of the closed contour that is of importance. We will return to this point in later sections.

Problems

1. Evaluate $\oint_C (z^*)^2\,dz$ around the circle $|z|=1$ taken in the counterclockwise direction.

2. Evaluate $\oint_C |z|^2\,dz$ around the square with vertices at (0,0), (1,0), (1,1), and (0,1) taken in the counterclockwise direction.

3. Evaluate $\int_C |z|\,dz$ along the right half of the circle $|z|=1$ from $z=-i$ to $z=i$.

4. Evaluate $\int_C e^z\,dz$ along the line $y=x$ from $(-1,-1)$ to $(1,1)$.

5. Evaluate $\int_C (z^*)^2\,dz$ along the line $y=x^2$ from $(0,0)$ to $(1,1)$.

6. Evaluate $\int_C z^{-1/2}\,dz$, where C is (a) the upper semicircle $|z|=1$ and (b) the lower semicircle $|z|=1$. If $z=re^{\theta i}$, restrict $-\pi<\theta<\pi$. Take both contours in the counterclockwise direction.

1.5 THE CAUCHY-GOURSAT THEOREM

In the previous section we showed how to evaluate line integrations by brute-force reduction to real-valued integrals. In general, this direct approach is quite difficult and we would like to apply some of the deeper properties of complex analysis to work smarter. In the remaining portions of this chapter, we introduce several theorems that will do just that.

If we scan over the examples worked in the previous section, we see considerable differences when the function was analytic inside and on the contour and when it was not. We may formalize this anecdotal evidence into the following theorem:

Cauchy-Goursat theorem:[2] *Let $f(z)$ be analytic in a domain D and let C be a simple Jordan curve*[3] *inside D so that $f(z)$ is analytic on and inside of C. Then $\oint_C f(z)\,dz=0$.*

Proof: Let C denote the contour around which we will integrate $w=f(z)$. We divide the region within C into a series of infinitesimal rectangles. See Figure 1.5.1. The integration

---

[2] Goursat, E., 1900: Sur la définition générale des fonctions analytiques, d'après Cauchy. *Trans. Am. Math. Soc.*, **1**, 14–16.

[3] A Jordan curve is a simply closed curve. It looks like a closed loop that does not cross itself. See Figure 1.5.2.

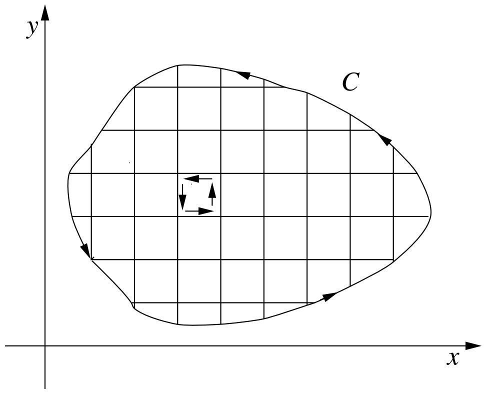


Figure 1.5.1: Diagram used in proving the Cauchy-Goursat theorem.

around each rectangle equals the product of the average value of w on each side and its length,

$$\begin{aligned}\left[w + \frac{\partial w}{\partial x}\frac{dx}{2}\right] dx + \left[w + \frac{\partial w}{\partial x}dx + \frac{\partial w}{\partial (iy)}\frac{d(iy)}{2}\right] d(iy) \\ + \left[w + \frac{\partial w}{\partial x}\frac{dx}{2} + \frac{\partial w}{\partial (iy)} d(iy)\right] (-dx) + \left[w + \frac{\partial w}{\partial (iy)}\frac{d(iy)}{2}\right] d(-iy) \\ = \left(\frac{\partial w}{\partial x} - \frac{\partial w}{i\partial y}\right)(i\,dx\,dy). \qquad (\mathbf{1.5.1})\end{aligned}$$

Substituting $w = u + iv$ into Equation 1.5.1,

$$\frac{\partial w}{\partial x} - \frac{\partial w}{i\,\partial y} = \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + i\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right). \qquad (\mathbf{1.5.2})$$

Because the function is analytic, the right side of Equation 1.5.1 and Equation 1.5.2 equals zero. Thus, the integration around each of these rectangles also equals zero.

We note next that in integrating around adjoining rectangles, we transverse each side in opposite directions, the net result being equivalent to integrating around the outer curve C. We therefore arrive at the result $\oint_C f(z)\,dz = 0$, where $f(z)$ is analytic within and on the closed contour. □

The Cauchy-Goursat theorem has several useful implications. Suppose that we have a domain where $f(z)$ is analytic. Within this domain, let us evaluate a line integral from point A to B along two different contours C_1 and C_2. Then, the integral around the closed contour formed by integrating along C_1 and then back along C_2, only in the opposite direction, is

$$\oint_C f(z)\,dz = \int_{C_1} f(z)\,dz - \int_{C_2} f(z)\,dz = 0 \qquad (\mathbf{1.5.3})$$

or

$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz. \qquad (\mathbf{1.5.4})$$

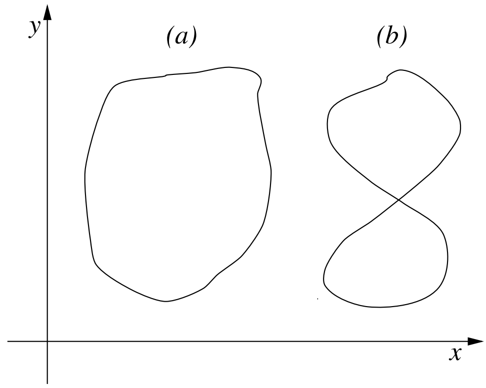


Figure 1.5.2: Examples of a (a) simply closed curve and (b) not simply closed curve.

Because C_1 and C_2 are completely arbitrary, we have the result that if, in a domain, $f(z)$ is analytic, the integral between any two points within the domain is *path independent.*

One obvious advantage of path independence is the ability to choose the contour so that the computations are made easier. This obvious choice immediately leads to the following principle:

The principle of deformation of contours: *The value of a line integral of an analytic function around any simple closed contour remains unchanged if we deform the contour in such a manner that we do not pass over a nonanalytic point.* □

• Example 1.5.1

Let us integrate $f(z) = z^{-1}$ around the closed contour C in the counterclockwise direction. This contour consists of a square, centered on the origin, with vertices at $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$.

The direct integration of $\oint_C z^{-1} dz$ around the original contour is very cumbersome. However, because the integrand is analytic everywhere except at the origin, we may deform the origin contour into a circle of radius r, centered on the origin. Then, $z = re^{\theta i}$ and $dz = rie^{\theta i} d\theta$ so that

$$\oint_C \frac{dz}{z} = \int_0^{2\pi} \frac{rie^{\theta i}}{re^{\theta i}}\, d\theta = i \int_0^{2\pi} d\theta = 2\pi i. \tag{1.5.5}$$

The point here is that no matter how bizarre the contour is, as long as it encircles the origin and is a simply closed contour, we can deform it into a circle and we get the same answer for the contour integral. This suggests that it is not the shape of the closed contour that makes the difference but whether we enclose any singularities (points where $f(z)$ becomes undefined) that matters. We shall return to this idea many times in the next few sections.□

Finally, suppose that we have a function $f(z)$ such that $f(z)$ is analytic in some domain. Furthermore, let us introduce the analytic function $F(z)$ such that $f(z) = F'(z)$. We would like to evaluate $\int_a^b f(z)\, dz$ in terms of $F(z)$.

We begin by noting that we can represent F, f as $F(z) = U + iV$ and $f(z) = u + iv$. From Example 1.3.28 we have that $u = U_x$ and $v = V_x$. Therefore,

$$\int_a^b f(z)\, dz = \int_a^b (u + iv)(dx + i\, dy) = \int_a^b U_x\, dx - V_x\, dy + i \int_a^b V_x\, dx + U_x\, dy \tag{1.5.6}$$

$$= \int_a^b U_x\,dx + U_y\,dy + i\int_a^b V_x\,dx + V_y\,dy = \int_a^b dU + i\int_a^b dV = F(b) - F(a) \tag{1.5.7}$$

or

$$\int_a^b f(z)\,dz = F(b) - F(a). \tag{1.5.8}$$

Equation 1.5.8 is the complex variable form of the fundamental theorem of calculus. Thus, if we can find the antiderivative of a function $f(z)$ that is analytic within a specific region, we can evaluate the integral by evaluating the antiderivative at the endpoints for any curves within that region.

• Example 1.5.2

Let us evaluate $\int_0^{\pi i} z\,\sin(z^2)\,dz$.

The integrand $f(z) = z\,\sin(z^2)$ is an entire function and its antiderivative equals $-\frac{1}{2}\cos(z^2)$. Therefore,

$$\int_0^{\pi i} z\,\sin(z^2)\,dz = -\tfrac{1}{2}\cos(z^2)\Big|_0^{\pi i} = \tfrac{1}{2}[\cos(0) - \cos(-\pi^2)] = \tfrac{1}{2}[1 - \cos(\pi^2)]. \tag{1.5.9}$$

Problems

For the following integrals, show that they are path independent and determine the value of the integral:

1. $\displaystyle\int_{1-\pi i}^{2+3\pi i} e^{-2z}\,dz$
2. $\displaystyle\int_0^{2\pi} [e^z - \cos(z)]\,dz$
3. $\displaystyle\int_0^{\pi} \sin^2(z)\,dz$
4. $\displaystyle\int_{-i}^{2i} (z+1)\,dz$
5. $\displaystyle\int_1^{2+2i} (z^2 - z + 8)\,dz$
6. $\displaystyle\int_1^{2i} [(1-i)z^2 + 2iz - 4]\,dz$
7. $\displaystyle\int_0^{i} z^2\cos(z^3)\,dz$
8. $\displaystyle\int_i^{1+i} ze^{-z^2}\,dz$

1.6 CAUCHY'S INTEGRAL FORMULA

In the previous section, our examples suggested that the presence of a singularity within a contour really determines the value of a closed contour integral. Continuing with this idea, let us consider a class of closed contour integrals that explicitly contains a single singularity within the contour, namely $\oint_C g(z)\,dz$, where $g(z) = f(z)/(z - z_0)$, and $f(z)$ is analytic within and on the contour C. We closed the contour in the *positive sense* where the enclosed area lies to your left as you move along the contour.

We begin by examining a closed contour integral where the closed contour consists of the C_1, C_2, C_3, and C_4 as shown in Figure 1.6.1. The gap or cut between C_2 and C_4 is very small. Because $g(z)$ is analytic within and on the closed integral, we have that

$$\int_{C_1} \frac{f(z)}{z - z_0}\,dz + \int_{C_2} \frac{f(z)}{z - z_0}\,dz + \int_{C_3} \frac{f(z)}{z - z_0}\,dz + \int_{C_4} \frac{f(z)}{z - z_0}\,dz = 0. \tag{1.6.1}$$

It can be shown that the contribution to the integral from the path C_2 going into the singularity cancels the contribution from the path C_4 going away from the singularity as the gap between them vanishes. Because $f(z)$ is analytic at z_0, we can approximate its

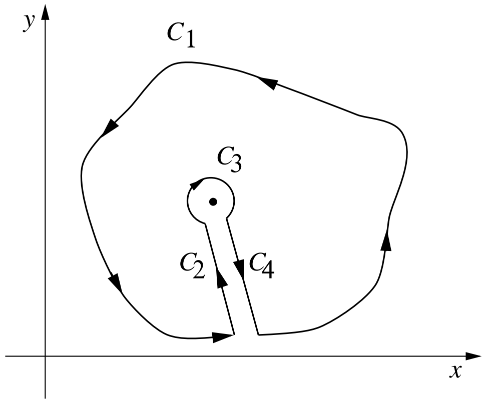


Figure 1.6.1: Diagram used to prove Cauchy's integral formula.

value on C_3 by $f(z) = f(z_0) + \delta(z)$, where δ is a small quantity. Substituting into Equation 1.6.1,

$$\oint_{C_1} \frac{f(z)}{z - z_0}\, dz = -f(z_0) \int_{C_3} \frac{1}{z - z_0}\, dz - \int_{C_3} \frac{\delta(z)}{z - z_0}\, dz. \qquad \mathbf{(1.6.2)}$$

Consequently, as the gap between C_2 and C_4 vanishes, the contour C_1 becomes the closed contour C so that Equation 1.6.2 may be written

$$\oint_C \frac{f(z)}{z - z_0}\, dz = 2\pi i f(z_0) + i \int_0^{2\pi} \delta\, d\theta, \qquad \mathbf{(1.6.3)}$$

where we set $z - z_0 = \epsilon e^{\theta i}$ and $dz = i\epsilon e^{\theta i} d\theta$.

Let M denote the value of the integral on the right side of Equation 1.6.3 and Δ equal the greatest value of the modulus of δ along the circle. Then

$$|M| < \int_0^{2\pi} |\delta|\, d\theta \le \int_0^{2\pi} \Delta\, d\theta = 2\pi\Delta. \qquad \mathbf{(1.6.4)}$$

As the radius of the circle diminishes to zero, Δ also diminishes to zero. Therefore, $|M|$, which is positive, becomes less than any finite quantity, however small, and M itself equals zero. Thus, we have that

$$\boxed{f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0}\, dz.} \qquad \mathbf{(1.6.5)}$$

This equation is *Cauchy's integral formula.* By taking n derivatives of Equation 1.6.5, we can extend Cauchy's integral formula[4] to

$$\boxed{f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^{n+1}}\, dz} \qquad \mathbf{(1.6.6)}$$

[4] See Carrier, G. F., M. Krook, and C. E. Pearson, 1966: *Functions of a Complex Variable: Theory and Technique.* McGraw-Hill, pp. 39–40 for the proof.

for $n = 1, 2, 3, \ldots$. For computing integrals, it is convenient to rewrite Equation 1.6.6 as

$$\oint_C \frac{f(z)}{(z-z_0)^{n+1}}\,dz = \frac{2\pi i}{n!} f^{(n)}(z_0). \tag{1.6.7}$$

• **Example 1.6.1**

Let us find the value of the integral

$$\oint_C \frac{\cos(\pi z)}{(z-1)(z-2)}\,dz, \tag{1.6.8}$$

where C is the circle $|z| = 5$. Using partial fractions,

$$\frac{1}{(z-1)(z-2)} = \frac{1}{z-2} - \frac{1}{z-1}, \tag{1.6.9}$$

and

$$\oint_C \frac{\cos(\pi z)}{(z-1)(z-2)}\,dz = \oint_C \frac{\cos(\pi z)}{z-2}\,dz - \oint_C \frac{\cos(\pi z)}{z-1}dz. \tag{1.6.10}$$

By Cauchy's integral formula with $z_0 = 2$ and $z_0 = 1$,

$$\oint_C \frac{\cos(\pi z)}{z-2}\,dz = 2\pi i\ \cos(2\pi) = 2\pi i, \tag{1.6.11}$$

and

$$\oint_C \frac{\cos(\pi z)}{z-1}\,dz = 2\pi i \cos(\pi) = -2\pi i, \tag{1.6.12}$$

because $z_0 = 1$ and $z_0 = 2$ lie inside C and $\cos(\pi z)$ is analytic there. Thus the required integral has the value

$$\oint_C \frac{\cos(\pi z)}{(z-1)(z-2)}\,dz = 4\pi i. \tag{1.6.13}$$

□

• **Example 1.6.2**

Let us use Cauchy's integral formula to evaluate

$$I = \oint_{|z|=2} \frac{e^z}{(z-1)^2(z-3)}\,dz. \tag{1.6.14}$$

We need to convert Equation 1.6.14 into the form Equation 1.6.7. To do this, we rewrite Equation 1.6.14 as

$$\oint_{|z|=2} \frac{e^z}{(z-1)^2(z-3)}\,dz = \oint_{|z|=2} \frac{e^z/(z-3)}{(z-1)^2}\,dz. \tag{1.6.15}$$

Therefore, $f(z) = e^z/(z-3)$, $n = 1$, and $z_0 = 1$. The function $f(z)$ is analytic within the closed contour because the point $z = 3$ lies outside of the contour. Applying Cauchy's integral formula,

$$\oint_{|z|=2} \frac{e^z}{(z-1)^2(z-3)}\,dz = \frac{2\pi i}{1!}\,\frac{d}{dz}\left(\frac{e^z}{z-3}\right)\bigg|_{z=1} = 2\pi i\left[\frac{e^z}{z-3} - \frac{e^z}{(z-3)^2}\right]\bigg|_{z=1} = -\frac{3\pi i e}{2}. \tag{1.6.16}$$

Project: Computing Derivatives of Any Order of a Complex or Real Function

The most common technique for computing a derivative is finite differencing. Recently Mahajerin and Burgess[5] showed how Cauchy's integral formula can be used to compute the derivatives of any order of a complex or real function via numerical quadrature. In this project you will derive the algorithm, write code implementing it, and finally test it.

Step 1: Consider the complex function $f(z) = u + iv$, which is analytic inside the closed circular contour C of radius R centered at z_0. Using Cauchy's integral formula, show that

$$f^{(n)}(z_0) = \frac{n!}{2\pi R^n}\int_0^{2\pi} [u(x,y) + iv(x,y)][\cos(n\theta) - i\sin(n\theta)]\,d\theta,$$

where $x = x_0 + R\cos(\theta)$, and $y = y_0 + R\sin(\theta)$.

Step 2: Using five-point Gaussian quadrature, write code to implement the results from Step 1.

Step 3: Test out this scheme by finding the first, sixth, and eleventh derivative of $f(x) = 8x/(x^2+4)$ for $x = 2$. The exact answers are 0, 2.8125, and 1218.164, respectively. What is the maximum value of R? How does the accuracy vary with the number of subdivisions used in the numerical integration? Is the algorithm sensitive to the value of R and the number of subdivisions? For a fixed number of subdivisions, is there an optimal R?

Problems

Use Cauchy's integral formula to evaluate the following integrals. Assume all of the contours are in the positive sense.

1. $\displaystyle\oint_{|z|=1} \frac{\sin^6(z)}{z-\pi/6}\,dz$

2. $\displaystyle\oint_{|z|=1} \frac{\sin^6(z)}{(z-\pi/6)^3}\,dz$

3. $\displaystyle\oint_{|z|=1} \frac{1}{z(z^2+4)}\,dz$

4. $\displaystyle\oint_{|z|=1} \frac{\tan(z)}{z}\,dz$

5. $\displaystyle\oint_{|z-1|=1/2} \frac{1}{(z-1)(z-2)}\,dz$

6. $\displaystyle\oint_{|z|=5} \frac{\exp(z^2)}{z^3}\,dz$

7. $\displaystyle\oint_{|z-1|=1} \frac{z^2+1}{z^2-1}\,dz$

8. $\displaystyle\oint_{|z|=2} \frac{z^2}{(z-1)^4}\,dz$

9. $\displaystyle\oint_{|z|=2} \frac{z^3}{(z+i)^3}\,dz$

10. $\displaystyle\oint_{|z|=1} \frac{\cos(z)}{z^{2n+1}}\,dz$

11. $\displaystyle\oint_{|z|=1} \frac{z^2+3z-1}{z(z^2-3)}\,dz$

12. $\displaystyle\oint_{|z|=3} \frac{ie^z}{(z-2+i)^4}\,dz$

---

[5] Mahajerin, E., and G. Burgess, 1993: An algorithm for computing derivatives of any order of a complex or real function. *Computers & Struct.*, **49**, 385–387.

1.7 TAYLOR AND LAURENT EXPANSIONS AND SINGULARITIES

In the previous section we showed what a crucial role singularities play in complex integration. Before we can find the most general way of computing a closed complex integral, our understanding of singularities must deepen. For this, we employ power series.

One reason why power series are so important is their ability to provide locally a general representation of a function even when its arguments are complex. For example, when we were introduced to trigonometric functions in high school, it was in the context of a right triangle and a real angle. However, when the argument becomes complex, this geometrical description disappears and power series provide a formalism for defining the trigonometric functions, regardless of the nature of the argument.

Let us begin our analysis by considering the complex function $f(z)$, which is analytic everywhere on the boundary, and the interior of a circle whose center is at $z = z_0$. Then, if z denotes any point within the circle, we have from Cauchy's integral formula that

$$f(z) = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{\zeta - z}\, d\zeta = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{\zeta - z_0} \left[\frac{1}{1 - (z - z_0)/(\zeta - z_0)} \right] d\zeta, \tag{1.7.1}$$

where C denotes the closed contour. Expanding the bracketed term as a geometric series, we find that

$$f(z) = \frac{1}{2\pi i} \left[\oint_C \frac{f(\zeta)}{\zeta - z_0}\, d\zeta + (z - z_0) \oint_C \frac{f(\zeta)}{(\zeta - z_0)^2}\, d\zeta + \cdots + (z - z_0)^n \oint_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\, d\zeta + \cdots \right]. \tag{1.7.2}$$

Applying Cauchy's integral formula to each integral in Equation 1.7.2, we finally obtain

$$f(z) = f(z_0) + \frac{(z - z_0)}{1!} f'(z_0) + \cdots + \frac{(z - z_0)^n}{n!} f^{(n)}(z_0) + \cdots \tag{1.7.3}$$

or the familiar formula for a Taylor expansion. Consequently, *we can expand any analytic function into a Taylor series.* Interestingly, the radius of convergence[6] of this series may be shown to be the distance between z_0 and the nearest nonanalytic point of $f(z)$.

• Example 1.7.1

Let us find the expansion of $f(z) = \sin(z)$ about the point $z_0 = 0$.

Because $f(z)$ is an entire function, we can construct a Taylor expansion anywhere on the complex plane. For $z_0 = 0$,

$$f(z) = f(0) + \tfrac{1}{1!} f'(0) z + \tfrac{1}{2!} f''(0) z^2 + \tfrac{1}{3!} f'''(0) z^3 + \cdots. \tag{1.7.4}$$

Because $f(0) = 0$, $f'(0) = 1$, $f''(0) = 0$, $f'''(0) = -1$ and so forth,

$$f(z) = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \cdots. \tag{1.7.5}$$

Because $\sin(z)$ is an entire function, the radius of convergence is $|z - 0| < \infty$, i.e., all z. □

---

[6] A positive number h such that the series diverges for $|z - z_0| > h$ but converges absolutely for $|z - z_0| < h$.

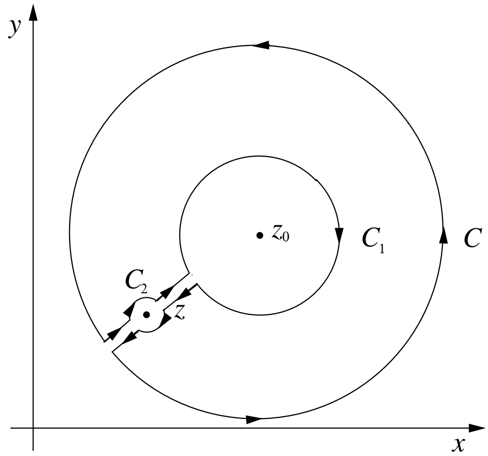


Figure 1.7.1: Contour used in deriving the Laurent expansion.

• **Example 1.7.2**

Let us find the expansion of $f(z) = 1/(1-z)$ about the point $z_0 = 0$.
From the formula for a Taylor expansion,

$$f(z) = f(0) + \tfrac{1}{1!}f'(0)z + \tfrac{1}{2!}f''(0)z^2 + \tfrac{1}{3!}f'''(0)z^3 + \cdots. \quad \textbf{(1.7.6)}$$

Because $f^{(n)}(0) = n!$, we find that

$$f(z) = 1 + z + z^2 + z^3 + z^4 + \cdots = \frac{1}{1-z}. \quad \textbf{(1.7.7)}$$

Equation 1.7.7 is the familiar result for a geometric series. Because the only nonanalytic point is at $z = 1$, the radius of convergence is $|z-0| < 1$, the unit circle centered at $z = 0$.□

Consider now the situation where we draw two concentric circles about some arbitrary point z_0; we denote the outer circle by C while we denote the inner circle by C_1. See Figure 1.7.1. Let us assume that $f(z)$ is analytic inside the annulus between the two circles. Outside of this area, the function may or may not be analytic. Within the annulus we pick a point z and construct a small circle around it, denoting the circle by C_2. As the gap or *cut* in the annulus becomes infinitesimally small, the line integrals that connect the circle C_2 to C_1 and C sum to zero, leaving

$$\oint_C \frac{f(\zeta)}{\zeta - z}\,d\zeta = \oint_{C_1} \frac{f(\zeta)}{\zeta - z}\,d\zeta + \oint_{C_2} \frac{f(\zeta)}{\zeta - z}\,d\zeta. \quad \textbf{(1.7.8)}$$

Because $f(\zeta)$ is analytic everywhere within C_2,

$$2\pi i f(z) = \oint_{C_2} \frac{f(\zeta)}{\zeta - z}\,d\zeta. \quad \textbf{(1.7.9)}$$

Using the relationship:

$$\oint_{C_1} \frac{f(\zeta)}{\zeta - z}\,d\zeta = -\oint_{C_1} \frac{f(\zeta)}{z - \zeta}\,d\zeta, \quad \textbf{(1.7.10)}$$

Equation 1.7.8 becomes

$$f(z) = \frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{\zeta - z}\, d\zeta + \frac{1}{2\pi i} \oint_{C_1} \frac{f(\zeta)}{z - \zeta}\, d\zeta. \tag{1.7.11}$$

Now,

$$\frac{1}{\zeta - z} = \frac{1}{\zeta - z_0 - z + z_0} = \frac{1}{\zeta - z_0} \frac{1}{1 - (z - z_0)/(\zeta - z_0)} \tag{1.7.12}$$

$$= \frac{1}{\zeta - z_0}\left[1 + \left(\frac{z - z_0}{\zeta - z_0}\right) + \left(\frac{z - z_0}{\zeta - z_0}\right)^2 + \cdots + \left(\frac{z - z_0}{\zeta - z_0}\right)^n + \cdots\right], \tag{1.7.13}$$

where $|z - z_0|/|\zeta - z_0| < 1$ and

$$\frac{1}{z - \zeta} = \frac{1}{z - z_0 - \zeta + z_0} = \frac{1}{z - z_0} \frac{1}{1 - (\zeta - z_0)/(z - z_0)} \tag{1.7.14}$$

$$= \frac{1}{z - z_0}\left[1 + \left(\frac{\zeta - z_0}{z - z_0}\right) + \left(\frac{\zeta - z_0}{z - z_0}\right)^2 + \cdots + \left(\frac{\zeta - z_0}{z - z_0}\right)^n + \cdots\right], \tag{1.7.15}$$

where $|\zeta - z_0|/|z - z_0| < 1$. Upon substituting these expressions into Equation 1.7.11,

$$\begin{aligned} f(z) = &\left[\frac{1}{2\pi i} \oint_C \frac{f(\zeta)}{\zeta - z_0}\, d\zeta + \frac{z - z_0}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z_0)^2}\, d\zeta + \cdots \right. \\ &\left. + \frac{(z - z_0)^n}{2\pi i} \oint_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\, d\zeta + \cdots\right] \\ &+ \left[\frac{1}{z - z_0} \frac{1}{2\pi i} \oint_{C_1} f(\zeta)\, d\zeta + \frac{1}{(z - z_0)^2} \frac{1}{2\pi i} \oint_{C_1} f(\zeta)(\zeta - z_0)\, d\zeta + \cdots \right. \\ &\left. + \frac{1}{(z - z_0)^n} \frac{1}{2\pi i} \oint_{C_1} f(\zeta)(\zeta - z_0)^{n-1}\, d\zeta + \cdots\right] \end{aligned} \tag{1.7.16}$$

or

$$f(z) = \frac{a_1}{z - z_0} + \frac{a_2}{(z - z_0)^2} + \cdots + \frac{a_n}{(z - z_0)^n} + \cdots + b_0 + b_1(z - z_0) + \cdots + b_n(z - z_0)^n + \cdots. \tag{1.7.17}$$

Equation 1.7.17 is a *Laurent expansion*.[7] If $f(z)$ is analytic at z_0, then $a_1 = a_2 = \cdots = a_n = \cdots = 0$ and the Laurent expansion reduces to a Taylor expansion. If z_0 is a singularity of $f(z)$, then the Laurent expansion includes both positive and *negative* powers. The coefficient of the $(z - z_0)^{-1}$ term, a_1, is the *residue*, for reasons that will appear in the next section.

Unlike the Taylor series, a Laurent series provides no straightforward method for obtaining the coefficients. For the remaining portions of this section we illustrate their construction. These techniques include replacing a function by its appropriate power series, the use of geometric series to expand the denominator, and the use of algebraic tricks to assist in applying the first two methods.

[7] Laurent, M., 1843: Extension du théorème de M. Cauchy relatif à la convergence du développement d'une fonction suivant les puissances ascendantes de la variable x. *C. R. l'Acad. Sci.*, **17**, 938–942.

• Example 1.7.3

Laurent expansions provide a formalism for the classification of singularities of a function. *Isolated singularities* fall into three types; they are as follows:

• *Essential Singularity*: Consider the function $f(z) = \cos(1/z)$. Using the expansion for cosine,

$$\cos\left(\frac{1}{z}\right) = 1 - \frac{1}{2!z^2} + \frac{1}{4!z^4} - \frac{1}{6!z^6} + \cdots \qquad (\mathbf{1.7.18})$$

for $0 < |z| < \infty$. Note that this series never truncates in the inverse powers of z. Essential singularities have Laurent expansions, which have an infinite number of inverse powers of $z - z_0$. The value of the residue for this essential singularity at $z = 0$ is zero.

• *Removable Singularity*: Consider the function $f(z) = \sin(z)/z$. This function has a singularity at $z = 0$. Upon applying the expansion for sine,

$$\frac{\sin(z)}{z} = \frac{1}{z}\left(z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \frac{z^9}{9!} - \ldots\right) = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \frac{z^6}{7!} + \frac{z^8}{9!} - \ldots \qquad (\mathbf{1.7.19})$$

for all z, if the division is permissible. We made $f(z)$ analytic by defining it by Equation 1.7.19 and, in the process, removed the singularity. The residue for a removable singularity always equals zero.

• *Pole of order n*: Consider the function

$$f(z) = \frac{1}{(z-1)^3(z+1)}. \qquad (\mathbf{1.7.20})$$

This function has two singularities: one at $z = 1$ and the other at $z = -1$. We shall only consider the case $z = 1$. After a little algebra,

$$f(z) = \frac{1}{(z-1)^3}\,\frac{1}{2+(z-1)} = \frac{1}{2}\,\frac{1}{(z-1)^3}\,\frac{1}{1+(z-1)/2} \qquad (\mathbf{1.7.21})$$

$$= \frac{1}{2}\,\frac{1}{(z-1)^3}\left[1 - \frac{z-1}{2} + \frac{(z-1)^2}{4} - \frac{(z-1)^3}{8} + \cdots\right] \qquad (\mathbf{1.7.22})$$

$$= \frac{1}{2(z-1)^3} - \frac{1}{4(z-1)^2} + \frac{1}{8(z-1)} - \frac{1}{16} + \cdots \qquad (\mathbf{1.7.23})$$

for $0 < |z - 1| < 2$. Because the largest inverse (negative) power is three, the singularity at $z = 1$ is a third-order pole; the value of the residue is 1/8. Generally, we refer to a first-order pole as a *simple* pole. □

• Example 1.7.4

Let us find the Laurent expansion for

$$f(z) = \frac{z}{(z-1)(z-3)} \qquad (\mathbf{1.7.24})$$

about the point $z = 1$.

We begin by rewriting $f(z)$ as

$$f(z) = \frac{1+(z-1)}{(z-1)[-2+(z-1)]} = -\frac{1}{2}\,\frac{1+(z-1)}{(z-1)[1-\frac{1}{2}(z-1)]} \tag{1.7.25}$$

$$= -\frac{1}{2}\,\frac{1+(z-1)}{(z-1)}[1+\tfrac{1}{2}(z-1)+\tfrac{1}{4}(z-1)^2+\cdots] \tag{1.7.26}$$

$$= -\frac{1}{2}\,\frac{1}{z-1} - \frac{3}{4} - \frac{3}{8}(z-1) - \frac{3}{16}(z-1)^2 - \cdots \tag{1.7.27}$$

provided $0 < |z-1| < 2$. Therefore we have a simple pole at $z = 1$ and the value of the residue is $-1/2$. A similar procedure would yield the Laurent expansion about $z = 3$. □

• **Example 1.7.5**

Let us find the Laurent expansion for

$$f(z) = \frac{z^n + z^{-n}}{z^2 - 2z\cosh(\alpha) + 1}, \quad \alpha > 0, \quad n \geq 0, \tag{1.7.28}$$

about the point $z = 0$.

We begin by rewriting $f(z)$ as

$$f(z) = \frac{z^n + z^{-n}}{(z-e^{\alpha})(z-e^{-\alpha})} = \frac{1}{2\sinh(\alpha)}\left(\frac{z^n+z^{-n}}{z-e^{\alpha}} - \frac{z^n+z^{-n}}{z-e^{-\alpha}}\right). \tag{1.7.29}$$

Because

$$\frac{1}{z-e^{\alpha}} = -\frac{e^{-\alpha}}{1-ze^{-\alpha}} = -e^{-\alpha}\left(1+ze^{-\alpha}+z^2e^{-2\alpha}+\cdots\right) \tag{1.7.30}$$

if $|z| < e^{\alpha}$ and

$$\frac{1}{z-e^{-\alpha}} = -\frac{e^{\alpha}}{1-ze^{\alpha}} = -e^{\alpha}\left(1+ze^{\alpha}+z^2e^{2\alpha}+\cdots\right) \tag{1.7.31}$$

if $|z| < e^{-\alpha}$,

$$f(z) = \frac{e^{\alpha}}{2\sinh(\alpha)}\left(z^n + z^{n+1}e^{\alpha} + z^{n+2}e^{2\alpha} + \cdots + z^{-n} + z^{1-n}e^{\alpha} + z^{2-n}e^{2\alpha} + \cdots\right) \tag{1.7.32}$$
$$- \frac{e^{-\alpha}}{2\sinh(\alpha)}\left(z^n + z^{n+1}e^{-\alpha} + z^{n+2}e^{-2\alpha} + \cdots + z^{-n} + z^{1-n}e^{-\alpha} + z^{2-n}e^{-2\alpha} + \cdots\right),$$

if $|z| < e^{-\alpha}$. Clearly we have an nth-order pole at $z = 0$. The residue, the coefficient of all of the z^{-1} terms in Equation 1.7.32, is found directly and equals

$$\operatorname{Res}[f(z); 0] = \frac{\sinh(n\alpha)}{\sinh(\alpha)}. \tag{1.7.33}$$

□

For complicated complex functions, it is very difficult to determine the nature of the singularities by finding the complete Laurent expansion, and we must try another method. We shall call it “a poor man’s Laurent expansion.” The idea behind this method is the fact that we generally need only the first few terms of the Laurent expansion to discover its nature. Consequently, we compute these terms through the application of power series where we retain only the leading terms. Consider the following example.

• Example 1.7.6

Let us discover the nature of the singularity at $z = 0$ of the function

$$f(z) = \frac{e^{tz}}{z \sinh(az)}, \tag{1.7.34}$$

where a and t are real.

We begin by replacing the exponential and hyperbolic sine by their Taylor expansion about $z = 0$. Then

$$f(z) = \frac{1 + tz + t^2z^2/2 + \cdots}{z(az + a^3z^3/6 + \cdots)}. \tag{1.7.35}$$

Factoring out az in the denominator,

$$f(z) = \frac{1 + tz + t^2z^2/2 + \cdots}{az^2(1 + a^2z^2/6 + \cdots)}. \tag{1.7.36}$$

Within the parentheses, all of the terms except the leading one are small. Therefore, by long division, we formally have that

$$f(z) = \frac{1}{az^2}(1 + tz + t^2z^2/2 + \cdots)(1 - a^2z^2/6 + \cdots) \tag{1.7.37}$$

$$= \frac{1}{az^2}(1 + tz + t^2z^2/2 - a^2z^2/6 + \cdots) = \frac{1}{az^2} + \frac{t}{az} + \frac{3t^2 - a^2}{6a} + \cdots. \tag{1.7.38}$$

Thus, we have a second-order pole at $z = 0$ and the residue equals t/a.

Problems

1. Find the Taylor expansion of $f(z) = (1 - z)^{-2}$ about the point $z = 0$.

2. Find the Taylor expansion of $f(z) = (z - 1)e^z$ about the point $z = 1$. (Hint: Don't find the expansion by taking derivatives.)

By constructing a Laurent expansion, describe the type of singularity and give the residue at z_0 for each of the following functions:

3. $f(z) = z^{10}e^{-1/z}; \quad z_0 = 0$

4. $f(z) = z^{-3}\sin^2(z); \quad z_0 = 0$

5. $f(z) = \dfrac{\cosh(z) - 1}{z^2}; \quad z_0 = 0$

6. $f(z) = \dfrac{z}{(z + 2)^2}; \quad z_0 = -2$

7. $f(z) = \dfrac{e^z + 1}{e^{-z} - 1}; \quad z_0 = 0$

8. $f(z) = \dfrac{e^{iz}}{z^2 + b^2}; \quad z_0 = bi$

9. $f(z) = \dfrac{1}{z(z - 2)}; \quad z_0 = 2$

10. $f(z) = \dfrac{\exp(z^2)}{z^4}; \quad z_0 = 0$

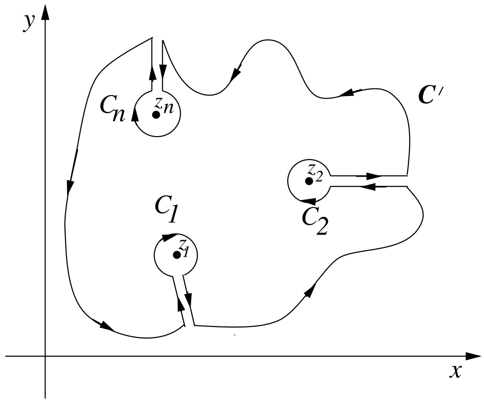


Figure 1.8.1: Contour used in deriving the residue theorem.

1.8 THEORY OF RESIDUES

Having shown that around any singularity we may construct a Laurent expansion, we now use this result in the integration of closed complex integrals. Consider a closed contour in which the function $f(z)$ has a number of isolated singularities. As we did in the case of Cauchy's integral formula, we introduce a new contour C' that excludes all of the singularities because they are isolated. See Figure 1.8.1. Therefore,

$$\oint_C f(z)\,dz - \oint_{C_1} f(z)\,dz - \cdots - \oint_{C_n} f(z)\,dz = \oint_{C'} f(z)\,dz = 0. \qquad \mathbf{(1.8.1)}$$

Consider now the mth integral, where $1 \le m \le n$. Constructing a Laurent expansion for the function $f(z)$ at the isolated singularity $z = z_m$, this integral equals

$$\oint_{C_m} f(z)\,dz = \sum_{k=1}^{\infty} a_k \oint_{C_m} \frac{1}{(z-z_m)^k}\,dz + \sum_{k=0}^{\infty} b_k \oint_{C_m} (z-z_m)^k\,dz. \qquad \mathbf{(1.8.2)}$$

Because $(z - z_m)^k$ is an entire function if $k \ge 0$, the integrals equal zero for each term in the second summation. We use Cauchy's integral formula to evaluate the remaining terms. The analytic function in the numerator is 1. Because $d^{k-1}(1)/dz^{k-1} = 0$ if $k > 1$, all of the terms vanish except for $k = 1$. In that case, the integral equals $2\pi i a_1$, where a_1 is the value of the residue for that particular singularity. Applying this approach to each of the singularities, we obtain the following:

Cauchy's residue theorem:[8] *If $f(z)$ is analytic inside and on a closed contour C (taken in the positive sense) except at points $z_1, z_2, \ldots, z_n$ where $f(z)$ has singularities, then*

$$\boxed{\oint_C f(z)\,dz = 2\pi i \sum_{j=1}^{n} \operatorname{Res}[f(z); z_j],} \qquad \mathbf{(1.8.3)}$$

[8] See Mitrinović, D. S., and J. D. Kečkić, 1984: *The Cauchy Method of Residues: Theory and Applications*. D. Reidel Publishing, 361 pp. Section 10.3 gives the historical development of the residue theorem.

where $\text{Res}[f(z); z_j]$ *denotes the residue of the* j*th isolated singularity of* $f(z)$ *located at* $z = z_j$*.* □

• Example 1.8.1

Let us compute $\oint_{|z|=2} z^2/(z+1)\, dz$ by the residue theorem, assuming that we take the contour in the positive sense.

Because the contour is a circle of radius 2, centered on the origin, the singularity at $z = -1$ lies within the contour. If the singularity were not inside the contour, then the integrand would have been analytic inside and on the contour C. In this case, the answer would then be zero by the Cauchy-Goursat theorem.

Returning to the original problem, we construct the Laurent expansion for the integrand around the point $z = 1$ by noting that

$$\frac{z^2}{z+1} = \frac{[(z+1)-1]^2}{z+1} = \frac{1}{z+1} - 2 + (z+1). \tag{1.8.4}$$

The singularity at $z = -1$ is a simple pole and by inspection, the value of the residue equals 1. Therefore,

$$\oint_{|z|=2} \frac{z^2}{z+1}\, dz = 2\pi i. \tag{1.8.5}$$

□

As it presently stands, it would appear that we must always construct a Laurent expansion for each singularity if we wish to use the residue theorem. This becomes increasingly difficult as the structure of the integrand becomes more complicated. In the following paragraphs we show several techniques that avoid this problem in practice.

We begin by noting that many functions which we will encounter consist of the ratio of two *polynomials*, i.e., rational functions: $f(z) = g(z)/h(z)$. Generally, we can write $h(z)$ as $(z-z_1)^{m_1}(z-z_2)^{m_2}\cdots$. Here we assumed that we divided out any common factors between $g(z)$ and $h(z)$ so that $g(z)$ does not vanish at $z_1, z_2, \ldots$. Clearly $z_1, z_2, \ldots$, are singularities of $f(z)$. Further analysis shows that the nature of the singularities are a pole of order m_1 at $z = z_1$, a pole of order m_2 at $z = z_2$, and so forth.

Having found the nature and location of the singularity, we compute the residue as follows. Suppose that we have a pole of order n. Then we know that its Laurent expansion is

$$f(z) = \frac{a_n}{(z-z_0)^n} + \frac{a_{n-1}}{(z-z_0)^{n-1}} + \cdots + b_0 + b_1(z-z_0) + \cdots. \tag{1.8.6}$$

Multiplying both sides of Equation 1.8.6 by $(z-z_0)^n$,

$$F(z) = (z-z_0)^n f(z) = a_n + a_{n-1}(z-z_0) + \cdots + b_0(z-z_0)^n + b_1(z-z_0)^{n+1} + \cdots. \tag{1.8.7}$$

Because $F(z)$ is analytic at $z = z_0$, it has the Taylor expansion

$$F(z) = F(z_0) + F'(z_0)(z-z_0) + \cdots + \frac{F^{(n-1)}(z_0)}{(n-1)!}(z-z_0)^{n-1} + \cdots. \tag{1.8.8}$$

Matching powers of $z - z_0$ in Equation 1.8.7 and Equation 1.8.8, the residue equals

$$\text{Res}[f(z); z_0] = a_1 = \frac{F^{(n-1)}(z_0)}{(n-1)!}. \tag{1.8.9}$$

Substituting in $F(z) = (z - z_0)^n f(z)$, we can compute the residue of a pole of order n by

$$\boxed{\operatorname{Res}[f(z); z_j] = \frac{1}{(n-1)!} \lim_{z \to z_j} \frac{d^{n-1}}{dz^{n-1}} \left[(z - z_j)^n f(z)\right].} \tag{1.8.10}$$

For a simple pole, Equation 1.8.10 simplifies to

$$\boxed{\operatorname{Res}[f(z); z_j] = \lim_{z \to z_j} (z - z_j) f(z).} \tag{1.8.11}$$

Quite often, $f(z) = p(z)/q(z)$. From l'Hôpital's rule, it follows that Equation 1.8.11 becomes

$$\boxed{\operatorname{Res}[f(z); z_j] = \frac{p(z_j)}{q'(z_j)}.} \tag{1.8.12}$$

Recall that these formulas work only for finite-order poles. For an essential singularity we must compute the residue from its Laurent expansion; however, essential singularities are very rare in applications.

• **Example 1.8.2**

Let us evaluate

$$\oint_C \frac{e^{iz}}{z^2 + a^2}\, dz, \tag{1.8.13}$$

where C is any contour that includes both poles at $z = \pm ai$ and is in the positive sense.

From Cauchy's residue theorem,

$$\oint_C \frac{e^{iz}}{z^2 + a^2}\, dz = 2\pi i \left[\operatorname{Res}\left(\frac{e^{iz}}{z^2 + a^2}; ai\right) + \operatorname{Res}\left(\frac{e^{iz}}{z^2 + a^2}; -ai\right)\right]. \tag{1.8.14}$$

The singularities at $z = \pm ai$ are simple poles. The corresponding residues are

$$\operatorname{Res}\left(\frac{e^{iz}}{z^2 + a^2}; ai\right) = \lim_{z \to ai} (z - ai) \frac{e^{iz}}{(z - ai)(z + ai)} = \frac{e^{-a}}{2ia} \tag{1.8.15}$$

and

$$\operatorname{Res}\left(\frac{e^{iz}}{z^2 + a^2}; -ai\right) = \lim_{z \to -ai} (z + ai) \frac{e^{iz}}{(z - ai)(z + ai)} = -\frac{e^{a}}{2ia}. \tag{1.8.16}$$

Consequently,

$$\oint_C \frac{e^{iz}}{z^2 + a^2}\, dz = -\frac{2\pi}{2a}\left(e^a - e^{-a}\right) = -\frac{2\pi}{a}\sinh(a). \tag{1.8.17}$$

□

• **Example 1.8.3**

Let us evaluate

$$\frac{1}{2\pi i}\oint_C \frac{e^{tz}}{z^2(z^2+2z+2)}\,dz, \tag{1.8.18}$$

where C includes all of the singularities and is in the positive sense.

The integrand has a second-order pole at $z=0$ and two simple poles at $z=-1\pm i$, which are the roots of $z^2+2z+2=0$. Therefore, the residue at $z=0$ is

$$\operatorname{Res}\left[\frac{e^{tz}}{z^2(z^2+2z+2)};0\right] = \lim_{z\to 0}\frac{1}{1!}\frac{d}{dz}\left\{(z-0)^2\left[\frac{e^{tz}}{z^2(z^2+2z+2)}\right]\right\} \tag{1.8.19}$$

$$= \lim_{z\to 0}\left[\frac{te^{tz}}{z^2+2z+2} - \frac{(2z+2)e^{tz}}{(z^2+2z+2)^2}\right] = \frac{t-1}{2}. \tag{1.8.20}$$

The residue at $z=-1+i$ is

$$\operatorname{Res}\left[\frac{e^{tz}}{z^2(z^2+2z+2)};-1+i\right] = \lim_{z\to -1+i}[z-(-1+i)]\frac{e^{tz}}{z^2(z^2+2z+2)} \tag{1.8.21}$$

$$= \left(\lim_{z\to -1+i}\frac{e^{tz}}{z^2}\right)\left(\lim_{z\to -1+i}\frac{z+1-i}{z^2+2z+2}\right) \tag{1.8.22}$$

$$= \frac{\exp[(-1+i)t]}{2i(-1+i)^2} = \frac{\exp[(-1+i)t]}{4}. \tag{1.8.23}$$

Similarly, the residue at $z=-1-i$ is

$$\operatorname{Res}\left[\frac{e^{tz}}{z^2(z^2+2z+2)};-1-i\right] = \lim_{z\to -1-i}[z-(-1-i)]\frac{e^{tz}}{z^2(z^2+2z+2)} \tag{1.8.24}$$

$$= \left(\lim_{z\to -1-i}\frac{e^{tz}}{z^2}\right)\left(\lim_{z\to -1-i}\frac{z+1+i}{z^2+2z+2}\right) \tag{1.8.25}$$

$$= \frac{\exp[(-1-i)t]}{(-2i)(-1-i)^2} = \frac{\exp[(-1-i)t]}{4}. \tag{1.8.26}$$

Then by the residue theorem,

$$\frac{1}{2\pi i}\oint_C \frac{e^{tz}}{z^2(z^2+2z+2)}\,dz = \operatorname{Res}\left[\frac{e^{tz}}{z^2(z^2+2z+2)};0\right] + \operatorname{Res}\left[\frac{e^{tz}}{z^2(z^2+2z+2)};-1+i\right]$$

$$+ \operatorname{Res}\left[\frac{e^{tz}}{z^2(z^2+2z+2)};-1-i\right] \tag{1.8.27}$$

$$= \frac{t-1}{2} + \frac{\exp[(-1+i)t]}{4} + \frac{\exp[(-1-i)t]}{4} \tag{1.8.28}$$

$$= \tfrac{1}{2}\left[t-1+e^{-t}\cos(t)\right]. \tag{1.8.29}$$

Problems

Assuming that all of the following closed contours are in the positive sense, use the residue theorem to evaluate the following integrals:

1. $\displaystyle\oint_{|z|=1} \frac{z+1}{z^4-2z^3}\,dz$
2. $\displaystyle\oint_{|z|=1} \frac{(z+4)^3}{z^4+5z^3+6z^2}\,dz$
3. $\displaystyle\oint_{|z|=1} \frac{1}{1-e^z}\,dz$
4. $\displaystyle\oint_{|z|=2} \frac{z^2-4}{(z-1)^4}\,dz$
5. $\displaystyle\oint_{|z|=2} \frac{z^3}{z^4-1}\,dz$
6. $\displaystyle\oint_{|z|=1} z^n e^{2/z}\,dz, \quad n>0$
7. $\displaystyle\oint_{|z|=1} e^{1/z}\cos(1/z)\,dz$
8. $\displaystyle\oint_{|z|=2} \frac{2+4\cos(\pi z)}{z(z-1)^2}\,dz$
9. $\displaystyle\oint_{|z-1|=\frac{1}{2}} \frac{z+1}{z-1}\frac{dz}{\sin(\pi z)}$

Hint for Problem 9: $\sin(\pi z) = -\sin[\pi(z-1)]$ and $z+1 = (z-1)+2$.

1.9 EVALUATION OF REAL DEFINITE INTEGRALS

One of the important applications of the theory of residues consists of the evaluation of certain types of real definite integrals. Similar techniques apply when the integrand contains a sine or cosine.

• Example 1.9.1

Let us evaluate the integral

$$\int_0^\infty \frac{dx}{x^2+1} = \frac{1}{2}\int_{-\infty}^\infty \frac{dx}{x^2+1}. \tag{1.9.1}$$

This integration occurs along the real axis. In terms of complex variables, we can rewrite Equation 1.9.1 as

$$\int_0^\infty \frac{dx}{x^2+1} = \frac{1}{2}\int_{C_1} \frac{dz}{z^2+1}, \tag{1.9.2}$$

where the contour C_1 is the line $\Im(z)=0$. However, the use of the residue theorem requires an integration along a closed contour. Let us choose the one pictured in Figure 1.9.1. Then

$$\oint_C \frac{dz}{z^2+1} = \int_{C_1} \frac{dz}{z^2+1} + \int_{C_2} \frac{dz}{z^2+1}, \tag{1.9.3}$$

where C denotes the complete closed contour and C_2 denotes the integration path along a semicircle at infinity. Clearly we want the second integral on the right side of Equation 1.9.3 to vanish; otherwise, our choice of the contour C_2 is poor. Because $z = Re^{\theta i}$ and $dz = iRe^{\theta i}\,d\theta$,

$$\left|\int_{C_2} \frac{dz}{z^2+1}\right| = \left|\int_0^\pi \frac{iR\exp(\theta i)}{1+R^2\exp(2\theta i)}\,d\theta\right| \le \int_0^\pi \frac{R}{R^2-1}\,d\theta, \tag{1.9.4}$$

which tends to zero as $R\to\infty$. On the other hand, the residue theorem gives

$$\oint_C \frac{dz}{z^2+1} = 2\pi i\,\mathrm{Res}\left(\frac{1}{z^2+1};i\right) = 2\pi i \lim_{z\to i}\frac{z-i}{z^2+1} = 2\pi i \times \frac{1}{2i} = \pi. \tag{1.9.5}$$

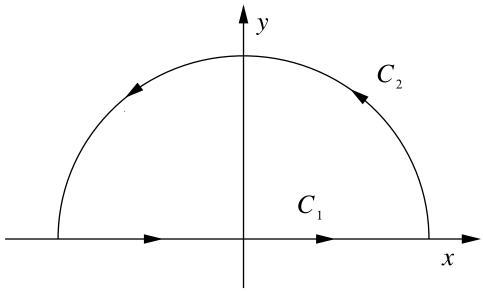


Figure 1.9.1: Contour used in evaluating the integral, Equation 1.9.1.

Therefore,

$$\int_0^\infty \frac{dx}{x^2+1} = \frac{\pi}{2}. \tag{1.9.6}$$

Note that we only evaluated the residue in the upper half-plane because it is the only one inside the contour. □

This example illustrates the basic concepts of evaluating definite integrals by the residue theorem. We introduce a closed contour that includes the real axis and an additional contour. We must then evaluate the integral along this additional contour as well as the closed contour integral. If we properly choose our closed contour, this additional integral vanishes. For certain classes of general integrals, we shall now show that this additional contour is a circular arc at infinity.

Theorem: *If, on a circular arc C_R with a radius R and center at the origin, $zf(z) \to 0$ uniformly with $|z| \in C_R$ and as $R \to \infty$, then*

$$\lim_{R\to\infty} \int_{C_R} f(z)\, dz = 0. \tag{1.9.7}$$

The proof is as follows: If $|zf(z)| \le M_R$, then $|f(z)| \le M_R/R$. Because the length of C_R is αR, where α is the subtended angle,

$$\left| \int_{C_R} f(z)\, dz \right| \le \frac{M_R}{R}\, \alpha R = \alpha M_R \to 0, \tag{1.9.8}$$

because $M_R \to 0$ as $R \to \infty$. □

• Example 1.9.2

A simple illustration of this theorem is the integral

$$\int_{-\infty}^\infty \frac{dx}{x^2+x+1} = \int_{C_1} \frac{dz}{z^2+z+1}. \tag{1.9.9}$$

A quick check shows that $z/(z^2+z+1)$ tends to zero uniformly as $R \to \infty$. Therefore, if we use the contour pictured in Figure 1.9.1,

$$\int_{-\infty}^\infty \frac{dx}{x^2+x+1} = \oint_C \frac{dz}{z^2+z+1} = 2\pi i \operatorname{Res}\left(\frac{1}{z^2+z+1}; -\tfrac{1}{2} + \tfrac{\sqrt{3}}{2} i \right) \tag{1.9.10}$$

$$= 2\pi i \lim_{z\to -\frac{1}{2}+\frac{\sqrt{3}}{2}i} \left(\frac{1}{2z+1} \right) = \frac{2\pi}{\sqrt{3}}. \tag{1.9.11}$$

□

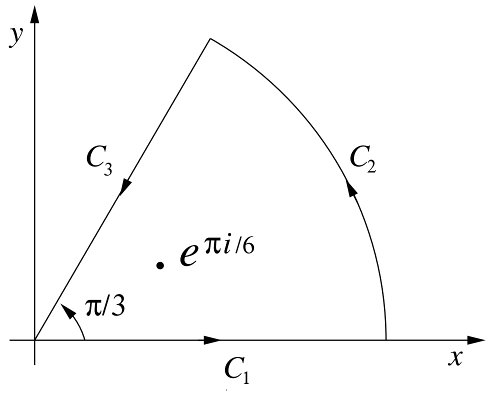


Figure 1.9.2: Contour used in evaluating the integral, Equation 1.9.13.

• **Example 1.9.3**

Let us evaluate

$$\int_0^\infty \frac{dx}{x^6+1}. \tag{1.9.12}$$

In place of an infinite semicircle in the upper half-plane, consider the following integral

$$\oint_C \frac{dz}{z^6+1}, \tag{1.9.13}$$

where we show the closed contour in Figure 1.9.2. We chose this contour for two reasons. First, we only have to evaluate one residue rather than the three enclosed in a traditional upper half-plane contour. Second, the contour integral along C_3 simplifies to a particularly simple and useful form.

Because the only enclosed singularity lies at $z = e^{\pi i/6}$,

$$\oint_C \frac{dz}{z^6+1} = 2\pi i \text{ Res}\left(\frac{1}{z^6+1}; e^{\pi i/6}\right) = 2\pi i \lim_{z\to e^{\pi i/6}} \frac{z-e^{\pi i/6}}{z^6+1} \tag{1.9.14}$$

$$= 2\pi i \lim_{z\to e^{\pi i/6}} \frac{1}{6z^5} = -\frac{\pi i}{3} e^{\pi i/6}. \tag{1.9.15}$$

Let us now evaluate Equation 1.9.12 along each of the legs of the contour:

$$\int_{C_1} \frac{dz}{z^6+1} = \int_0^\infty \frac{dx}{x^6+1}, \tag{1.9.16}$$

$$\int_{C_2} \frac{dz}{z^6+1} = 0, \tag{1.9.17}$$

because of Equation 1.9.7 and

$$\int_{C_3} \frac{dz}{z^6+1} = \int_\infty^0 \frac{e^{\pi i/3}\, dr}{r^6+1} = -e^{\pi i/3} \int_0^\infty \frac{dx}{x^6+1}, \tag{1.9.18}$$

since $z = re^{\pi i/3}$.

Substituting into Equation 1.9.15,

$$\left(1-e^{\pi i/3}\right)\int_0^\infty \frac{dx}{x^6+1} = -\frac{\pi i}{3}e^{\pi i/6} \tag{1.9.19}$$

or

$$\int_0^\infty \frac{dx}{x^6+1} = \frac{\pi i}{6}\,\frac{2ie^{\pi i/6}}{e^{\pi i/6}\left(e^{\pi i/6}-e^{-\pi i/6}\right)} = \frac{\pi}{6\sin(\pi/6)} = \frac{\pi}{3}. \tag{1.9.20}$$

□

• **Example 1.9.4**

Rectangular closed contours are best for the evaluation of integrals that involve hyperbolic sines and cosines. To illustrate[9] this, let us evaluate the integral

$$2\int_0^\infty \frac{\sin(ax)\,\sinh(x)}{[b+\cosh(x)]^2}\,dx = \int_{-\infty}^\infty \frac{\sin(ax)\,\sinh(x)}{[b+\cosh(x)]^2}\,dx = \Im\left[\int_{-\infty}^\infty \frac{\sinh(x)e^{iax}}{[b+\cosh(x)]^2}\,dx\right], \tag{1.9.21}$$

where $a>0$ and $b>1$.

We begin by determining the value of

$$\oint_C \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz$$

about the closed contour shown in Figure 1.9.3. Writing this contour integral in terms of the four line segments that constitute the closed contour, we have

$$\begin{aligned}\oint_C \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz &= \int_{C_1} \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz + \int_{C_2} \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz \\ &\quad + \int_{C_3} \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz + \int_{C_4} \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz.\end{aligned} \tag{1.9.22}$$

Because the integrand behaves as e^{-R} as $R\to\infty$, the integrals along C_2 and C_4 vanish. On the other hand,

$$\int_{C_1} \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz = \int_{-\infty}^\infty \frac{\sinh(x)e^{iax}}{[b+\cosh(x)]^2}\,dx, \tag{1.9.23}$$

and

$$\int_{C_3} \frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}\,dz = -e^{-2\pi a}\int_{-\infty}^\infty \frac{\sinh(x)e^{iax}}{[b+\cosh(x)]^2}\,dx, \tag{1.9.24}$$

[9] This is a slight variation on a problem solved by Spyrou, K. J., B. Cotton, and B. Gurd, 2002: Analytical expressions of capsize boundary for a ship with roll bias in beam waves. *J. Ship Res.*, **46**, 167–174.

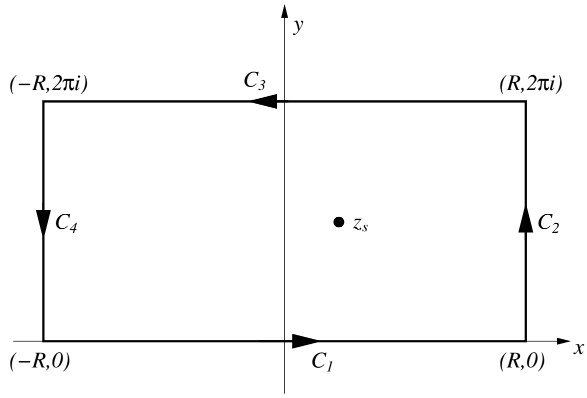


Figure 1.9.3: Rectangular closed contour used to obtain Equation 1.9.31.

because $\cosh(x+2\pi i) = \cosh(x)$ and $\sinh(x+2\pi i) = \sinh(x)$.

Within the closed contour C, we have a single singularity where $b + \cosh(z_s) = 0$ or $e^{z_s} = -b - \sqrt{b^2-1}$ or $z_s = \ln(b + \sqrt{b^2-1}\,) + \pi i$. To discover the nature of this singularity, we expand $b + \cosh(z)$ in a Taylor expansion and find that

$$b + \cosh(z) = \sinh(z_s)(z - z_s) + \tfrac{1}{2}\cosh(z_s)(z - z_s)^2 + \cdots. \tag{1.9.25}$$

Therefore, we have a second-order pole at $z = z_s$. Therefore, the value of the residue there is

$$\operatorname{Res}\left[\frac{\sinh(z)e^{iaz}}{[b+\cosh(z)]^2}; z_s\right] = \lim_{z\to z_s} \frac{d}{dz}\left[\frac{\sinh(z)e^{iaz}}{\sinh^2(z_s) + \sinh(z_s)\cosh(z_s)(z-z_s) + \cdots}\right] \tag{1.9.26}$$

$$= \frac{ia\,e^{-\pi a}}{\sinh(z_s)} \exp[ia\cosh^{-1}(b)]. \tag{1.9.27}$$

Therefore,

$$\int_{-\infty}^{\infty} \frac{\sinh(x)e^{iax}}{[b+\cosh(x)]^2}\,dx = -\frac{2\pi a \exp[-\pi a + ai\cosh^{-1}(b)]}{(1-e^{-2\pi a})\sinh(z_s)} = \frac{\pi a \exp[ai\cosh^{-1}(b)]}{\sqrt{b^2-1}\,\sinh(\pi a)}, \tag{1.9.28}$$

because

$$\sinh(z_s) = \frac{1}{2}\left[-b - \sqrt{b^2-1} + \frac{1}{b+\sqrt{b^2-1}}\right] = -\sqrt{b^2-1}. \tag{1.9.29}$$

Substituting Equation 1.9.28 into Equation 1.9.21 yields

$$\int_0^{\infty} \frac{\sin(ax)\,\sinh(x)}{[b+\cosh(x)]^2}\,dx = \frac{\pi a \sin[a\cosh^{-1}(b)]}{2\sqrt{b^2-1}\,\sinh(\pi a)}. \tag{1.9.30}$$

□

• **Example 1.9.5**

The method of residues is also useful in the evaluation of definite integrals of the form $\int_0^{2\pi} F[\sin(\theta), \cos(\theta)]\, d\theta$, where F is a quotient of polynomials in $\sin(\theta)$ and $\cos(\theta)$. For example, let us evaluate the integral[10]

$$I = \int_0^{2\pi} \frac{\cos^3(\theta)}{\cos^2(\theta) - a^2}\, d\theta, \qquad a > 1. \tag{1.9.31}$$

We begin by introducing the complex variable $z = e^{i\theta}$. This substitution yields the closed contour integral

$$I = \frac{1}{2i} \oint_C \frac{(z^2+1)^3}{(z^2+1)^2 - 4a^2z^2} \frac{dz}{z^2}, \tag{1.9.32}$$

where C is a circle of radius 1 taken in the positive sense. The integrand of Equation 1.9.32 has five singularities: a second-order pole at $z_5 = 0$ and simple poles located at

$$z_1 = -a - \sqrt{a^2-1}, \qquad z_2 = -a + \sqrt{a^2-1}, \tag{1.9.33}$$

$$z_3 = a - \sqrt{a^2-1}, \quad \text{and} \quad z_4 = a + \sqrt{a^2-1}. \tag{1.9.34}$$

Only the singularities z_2, z_3, and z_5 lie within C. Consequently, the value of I equals $2\pi i$ times the sum of the residues at these three singularities. The residues equal

$$\operatorname{Res}\left\{ \frac{(z^2+1)^3}{z^2[(z^2+1)^2 - 4a^2z^2]}; -a + \sqrt{a^2-1} \right\}$$

$$= \lim_{z\to -a+\sqrt{a^2-1}} \frac{(z^2+1)^3}{z^2} \lim_{z\to -a+\sqrt{a^2-1}} \frac{z + a - \sqrt{a^2-1}}{(z^2+1)^2 - 4a^2z^2} \tag{1.9.35}$$

$$= \lim_{z\to -a+\sqrt{a^2-1}} \frac{(z^2+1)^3}{4z^3(z^2+1-2a^2)} \tag{1.9.36}$$

$$= -\frac{a^2(a - \sqrt{a^2-1}\,)^3}{(2a^2 - 1 - 2a\sqrt{a^2-1}\,)(a^2 - 1 - a\sqrt{a^2-1}\,)}, \tag{1.9.37}$$

$$\operatorname{Res}\left\{ \frac{(z^2+1)^3}{z^2[(z^2+1)^2 - 4a^2z^2]}; a - \sqrt{a^2-1} \right\}$$

$$= \lim_{z\to a-\sqrt{a^2-1}} \frac{(z^2+1)^3}{z^2} \lim_{z\to a-\sqrt{a^2-1}} \frac{z - a + \sqrt{a^2-1}}{(z^2+1)^2 - 4a^2z^2} \tag{1.9.38}$$

$$= \lim_{z\to a-\sqrt{a^2-1}} \frac{(z^2+1)^3}{4z^3(z^2+1-2a^2)} \tag{1.9.39}$$

$$= \frac{a^2(a - \sqrt{a^2-1}\,)^3}{(2a^2 - 1 - 2a\sqrt{a^2-1}\,)(a^2 - 1 - a\sqrt{a^2-1}\,)}, \tag{1.9.40}$$

---

[10] Simplified version of an integral presented by Jiang, Q. F., and R. B. Smith, 2000: V-waves, bow shocks, and wakes in supercritical hydrostatic flow. *J. Fluid Mech.*, **406**, 27–53.

and

$$\begin{aligned}
&\operatorname{Res}\left\{\frac{(z^2+1)^3}{z^2[(z^2+1)^2-4a^2z^2]};0\right\} \\
&\quad = \lim_{z\to 0}\frac{d}{dz}\left[\frac{(z^2+1)^3}{(z^2+1)^2-4a^2z^2}\right] && \mathbf{(1.9.41)}\\
&\quad = \lim_{z\to 0}\frac{6z[(z^2+1)^4-4a^2z^2(z^2+1)^2]-4z(z^2+1)^3(z^2+1-2a^2)}{[(z^2+1)^2-4a^2z^2]^2} && \mathbf{(1.9.42)}\\
&\quad = 0. && \mathbf{(1.9.43)}
\end{aligned}$$

Summing the residues, we obtain 0. Therefore,

$$\int_0^{2\pi}\frac{\cos^3(\theta)}{\cos^2(\theta)-a^2}\,d\theta = 0, \qquad a > 1. \qquad \mathbf{(1.9.44)}$$

Problems

Use the residue theorem to verify the following integrals:

1. $\displaystyle\int_0^\infty \frac{dx}{x^4+1} = \frac{\pi\sqrt{2}}{4}$

2. $\displaystyle\int_{-\infty}^\infty \frac{dx}{(x^2+4x+5)^2} = \frac{\pi}{2}$

3. $\displaystyle\int_{-\infty}^\infty \frac{x\,dx}{(x^2+1)(x^2+2x+2)} = -\frac{\pi}{5}$

4. $\displaystyle\int_0^\infty \frac{x^2}{x^6+1}\,dx = \frac{\pi}{6}$

5. $\displaystyle\int_0^\infty \frac{dx}{(x^2+1)^2} = \frac{\pi}{4}$

6. $\displaystyle\int_0^\infty \frac{dx}{(x^2+1)(x^2+4)^2} = \frac{5\pi}{288}$

7.
$$\int_{-\infty}^\infty \frac{x^2\;dx}{(x^2+a^2)(x^2+b^2)^2} = \frac{\pi}{2b(a+b)^2}, \qquad a, b > 0$$

8.
$$\int_0^\infty \frac{t^2}{(t^2+1)[t^2(a/h+1)+(a/h-1)]}\,dt = \frac{\pi}{4}\left[1-\sqrt{\frac{a-h}{a+h}}\right], \quad a > h$$

9. Show that
$$\int_0^{\pi/2}\frac{d\theta}{a+\sin^2(\theta)} = \frac{\pi}{2\sqrt{a+a^2}}, \qquad a > 0.$$

Step 1: Convert the real integral into a closed contour integration:

$$\int_0^{\pi/2}\frac{d\theta}{a+\sin^2(\theta)} = i\oint_{|z|=1}\frac{z}{(z^2-1)^2-4az^2}\,dz,$$

where $z = e^{\theta i}$.

Step 2: Show that the integrand has four poles: $z = \pm\sqrt{a}\pm\sqrt{1+a}$. Only two are located inside the contour: $z_1 = -\sqrt{a}+\sqrt{1+a}$ and $z_2 = \sqrt{a}-\sqrt{1+a}$.

Step 3: Show that the corresponding residues are

$$\operatorname{Res}\left[\frac{z}{(z^2-1)^2-4az^2}; z_1\right] = \operatorname{Res}\left[\frac{z}{(z^2-1)^2-4az^2}; z_2\right] = -\frac{1}{8\sqrt{a+a^2}}.$$

Step 4: Obtain the final result by applying the residue theorem and the results from Step 1 through Step 3.

10. Show that

$$\int_0^{\pi/2} \frac{d\theta}{a^2\cos^2(\theta)+b^2\sin^2(\theta)} = \frac{\pi}{2ab}, \qquad b \ge a > 0.$$

Step 1: Convert the real integral into a closed contour integration:

$$\int_0^{\pi/2} \frac{d\theta}{a^2\cos^2(\theta)+b^2\sin^2(\theta)} = -i\oint_{|z|=1} \frac{z}{a^2(z^2+1)^2-b^2(z^2-1)^2}\,dz,$$

where $z = e^{\theta i}$.

Step 2: Show that the integrand has four simple poles located at $z_+^2 = (b+a)/(b-a)$, and $z_-^2 = (b-a)/(b+a)$. Only two are located inside the contour: $z_-^{(1)} = \sqrt{(b-a)/(b+a)}$, and $z_-^{(2)} = -\sqrt{(b-a)/(b+a)}$.

Step 3: Show that the corresponding residues are

$$\operatorname{Res}\left[\frac{z}{a^2(z^2+1)^2-b^2(z^2-1)^2}; z_-^{(1)}\right] = \operatorname{Res}\left[\frac{z}{a^2(z^2+1)^2-b^2(z^2-1)^2}; z_-^{(2)}\right] = \frac{1}{8ab}.$$

Step 4: Obtain the final result by employing the residue theorem and the results from Step 1 through Step 3.

11. Show that

$$\int_0^{\pi} \frac{\sin^2(\theta)}{a+b\cos(\theta)}\,d\theta = \frac{\pi}{b^2}\left(a-\sqrt{a^2-b^2}\right), \qquad a > b > 0.$$

Step 1: Convert the real integral into a closed contour integration:

$$\int_0^{\pi} \frac{\sin^2(\theta)}{a+b\cos(\theta)}\,d\theta = \frac{i}{4}\oint_{|z|=1} \frac{(z^2-1)^2}{[b(z^2+1)+2az]z^2}\,dz,$$

where $z = e^{\theta i}$.

Step 2: Show that the integrand has a second-order pole at $z = 0$ and simple poles at $z_{1,2} = \left(-a \pm \sqrt{a^2-b^2}\right)/b$. Only the poles located at $z = 0$ and $z_1 = \left(-a+\sqrt{a^2-b^2}\right)/b$ lie within the closed contour.

Step 3: Show that the corresponding residues are

$$\operatorname{Res}\left[\frac{(z^2-1)^2}{[b(z^2+1)+2az]z^2}; 0\right] = -\frac{2a}{b^2}, \quad \text{and} \quad \operatorname{Res}\left[\frac{(z^2-1)^2}{[b(z^2+1)+2az]z^2}; z_1\right] = \frac{2\sqrt{a^2-b^2}}{b^2}.$$

Step 4: Obtain the final results by employing the residue theorem and Step 1 through Step 3.

12. Show that

$$\int_0^{2\pi} \frac{e^{in\theta}}{1+2r\cos(\theta)+r^2}\,d\theta = 2\pi\frac{(-r)^n}{1-r^2}, \qquad 1>|r|, \quad n=0,1,2,\ldots.$$

Step 1: Convert the real integral into a closed contour integration:

$$\int_0^{2\pi} \frac{e^{in\theta}}{1+2r\cos(\theta)+r^2}\,d\theta = -i\oint_{|z|=1}\frac{z^n}{r(z^2+1)+(1+r^2)z}\,dz,$$

where $z=e^{\theta i}$.

Step 2: Show that the integrand has two simple poles: $z_+=-r$, and $z_-=-1/r$. Why is the z_+ pole the only one inside the contour?

Step 3: Show that the corresponding residue is

$$\text{Res}\left[\frac{z^n}{r(z^2+1)+(1+r^2)z};z_+\right] = \frac{(-r)^n}{1-r^2}.$$

Step 4: Obtain the final result by using the residue theorem and Step 1 through Step 3.

13. Show that

$$\int_0^{2\pi}\sin^{2n}(\theta)\,d\theta = \frac{2\pi(2n)!}{(2^n n!)^2}.$$

Step 1: Convert the real integral into a closed contour integration:

$$\int_0^{2\pi}\sin^{2n}(\theta)\,d\theta = \frac{-i}{(-1)^n 2^{2n}}\oint_{|z|=1}\frac{(z^2-1)^{2n}}{z^{2n+1}}\,dz,$$

where $z=e^{\theta i}$.

Step 2: Show that the integrand has a pole of order $2n+1$ at $z=0$.

Step 3: Because

$$(z^2-1)^{2n} = z^{4n}-2nz^{4n-1}+\cdots+\frac{(2n)!(-1)^n}{n!n!}z^{2n}+\cdots,$$

show that

$$\text{Res}\left[\frac{(z^2-1)^{2n}}{z^{2n+1}};0\right] = \frac{(2n)!(-1)^n}{n!n!}.$$

Step 4: Obtain the final result by using the residue theorem and Step 1 through Step 3.

14. Show that

$$\int_{-\pi}^{\pi}\frac{\cos(n\theta)}{\cos(\theta)+\alpha}\,d\theta = 2\pi\frac{\left(-\alpha+\sqrt{\alpha^2-1}\,\right)^n}{\sqrt{\alpha^2-1}}, \quad \alpha>1, \quad n\geq 0.$$

Step 1: Convert the real integral into a closed contour integration:

$$\int_{-\pi}^{\pi} \frac{\cos(n\theta)}{\cos(\theta)+\alpha}\, d\theta = \frac{1}{i}\oint_{|z|=1} \frac{z^n + z^{-n}}{z^2+2\alpha z+1}\, dz,$$

where $z = e^{\theta i}$.

Step 2: Assume that $n \neq 0$. Show that the integrand has an n-order pole at $z = 0$ and simple poles at $z_{1,2} = -\alpha \pm \sqrt{\alpha^2 - 1}$. Why is $z_1 = -\alpha + \sqrt{\alpha^2-1}$ the only simple pole that lies inside the contour?

Step 3: Because

$$\frac{z^n + z^{-n}}{z^2+2\alpha z+1} = \frac{1}{2\sqrt{\alpha^2-1}}\left(\frac{z^n+z^{-n}}{z-z_1} - \frac{z^n+z^{-n}}{z-z_2}\right),$$

show that

$$\operatorname{Res}\left(\frac{z^n+z^{-n}}{z-z_1}; z_1\right) = \frac{z_1^{2n}+1}{z_1^n}.$$

Step 4: Because

$$\frac{z^n+z^{-n}}{z-z_1} = -\left(z^n+z^{-n}\right)\frac{1}{z_1}\left[1+\left(\frac{z}{z_1}\right)+\left(\frac{z}{z_1}\right)^2+\left(\frac{z}{z_1}\right)^3+\cdots\right],$$

show that

$$\operatorname{Res}\left(\frac{z^n+z^{-n}}{z-z_1}; 0\right) = -\frac{1}{z_1^n}, \qquad \text{and} \qquad \operatorname{Res}\left(\frac{z^n+z^{-n}}{z-z_2}; 0\right) = -\frac{1}{z_2^n} = -z_1^n.$$

Step 5: Use the residue theorem plus Steps 1 through 4 to obtain the final result when $n \neq 0$.

Step 6: Redo the problem when $n = 0$. In this case we only have the pole at $z = z_1$.

15. Show that

$$\int_0^{\pi} \frac{\cos(n\theta)}{\cosh(\alpha)-\cos(\theta)}\, d\theta = \frac{\pi}{\sinh(\alpha)} e^{-n\alpha}, \quad \alpha \neq 0, \quad n \geq 0.$$

Step 1: Convert the real integral into a closed contour integration:

$$\int_0^{\pi} \frac{\cos(n\theta)}{\cosh(\alpha)-\cos(\theta)}\, d\theta = -\frac{1}{i}\oint_{|z|=1} \frac{z^n+z^{-n}}{z^2-2z\cosh(\alpha)+1}\, dz,$$

where $z = e^{\theta i}$.

Step 2: Show that the integrand has an n-order pole at $z = 0$ and simple poles at $z_{1,2} = e^{\alpha}, e^{-\alpha}$. Because α can be taken as positive without loss of generality, then only the poles located at $z = 0$ and $z = e^{-\alpha}$ lie within the closed contour.

Step 3: Because

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

show that the corresponding residues are

$$\operatorname{Res}\left[\frac{z^n + z^{-n}}{z^2 - 2z\cosh(\alpha) + 1}; 0\right] = \frac{\sinh(n\alpha)}{\sinh(\alpha)},$$

from Example 1.7.5 and

$$\operatorname{Res}\left[\frac{z^n + z^{-n}}{z^2 - 2z\cosh(\alpha) + 1}; e^{-\alpha}\right] = -\frac{\cosh(n\alpha)}{\sinh(\alpha)}.$$

Step 4: Use the residue theorem plus Steps 1 through 3 to finish the problem.

16. Show that

$$\int_0^\infty \frac{x^2}{(1 - x^2)^2 + a^2x^2}\, dx = \frac{\pi}{2|a|},$$

where a is real and not equal to zero.

Step 1: Show that

$$\int_0^\infty \frac{x^2}{(1 - x^2)^2 + a^2x^2}\, dx = \tfrac{1}{2}\oint_C \frac{z^2}{(1 - z^2)^2 + a^2z^2}\, dz,$$

where C denotes a semicircle of infinite radius in the upper half of the complex plane. Along the real axis, the contour slightly above $y = 0$ when $x < 0$ and slightly below $y = 0$ when $x > 0$.

Step 2: Show that the poles of the integrand are simple and equal

$$z_n = \begin{cases} \pm\frac{1}{2}\left(\pm\sqrt{4 - a^2} + |a|i\right), & \text{if} \quad 0 < |a| < 2, \\ \pm\frac{i}{2}\left(|a| \pm \sqrt{a^2 - 4}\,\right), & \text{if} \quad 2 < |a|. \end{cases}$$

If $|a| = 2$, we have second-order poles at $z_n = \pm i$.

Step 3: Show that the residues for the poles in the upper half plane are

$$\operatorname{Res}\left[f(z); \tfrac{1}{2}\left(\pm\sqrt{4 - a^2} + |a|i\right)\right] = \pm\frac{(\pm\sqrt{4 - a^2}\, + |a|i)/2}{2|a|i\sqrt{4 - a^2}},$$

$$\operatorname{Res}\left[f(z); \tfrac{i}{2}\left(|a| \pm \sqrt{a^2 - 4}\,\right)\right] = \mp\frac{i(|a| \pm \sqrt{a^2 - 4}\,)/2}{2|a|\sqrt{a^2 - 4}},$$

and

$$\operatorname{Res}[f(z); i\,] = -\frac{i}{4}.$$

Step 4: Show that when you sum the residues for the cases for $0 < |a| < 2$ and $2 < |a|$, you obtain $-i/(2|a|)$

Step 5: Redo the calculation when $|a| = 2$.

17. Evaluating the closed contour integral

$$\oint_C \frac{e^{iaz}}{\cosh^2(bz)}\,dz,$$

around the *rectangular* contour with vertices at $(\infty, 0)$, $(-\infty, 0)$, $(\infty, \pi/b)$, and $(-\infty, \pi/b)$, show that

$$\int_0^\infty \frac{\cos(ax)}{\cosh^2(bx)}\,dx = \frac{\pi a}{2b^2 \sinh[a\pi/(2b)]}, \qquad a, b > 0.$$

Step 1: Show that

$$\oint_C \frac{e^{iaz}}{\cosh^2(bz)}\,dz = \int_{C_1} \frac{e^{iaz}}{\cosh^2(bz)}\,dz + \int_{C_2} \frac{e^{iaz}}{\cosh^2(bz)}\,dz + \int_{C_3} \frac{e^{iaz}}{\cosh^2(bz)}\,dz + \int_{C_4} \frac{e^{iaz}}{\cosh^2(bz)}\,dz,$$

where C_1, C_2, C_3 and C_4 are the contours along the bottom, right side, top, and left side of the rectangle.

Step 2: Show that the integrals along C_2 and C_4 vanish. Why?

Step 3: Show that

$$\int_{C_1} \frac{e^{iaz}}{\cosh^2(bz)}\,dz = \int_{-\infty}^\infty \frac{e^{iax}}{\cosh^2(bx)}\,dx,$$

and

$$\int_{C_3} \frac{e^{iaz}}{\cosh^2(bz)}\,dz = -e^{-\pi a/b} \int_{-\infty}^\infty \frac{e^{iax}}{\cosh^2(bx)}\,dx.$$

Step 4: Setting $z_s = \pi i/(2b)$, show that the Laurent expansion for $e^{iaz}/\cosh^2(bz)$ at z_s is

$$\frac{e^{iaz}}{\cosh^2(bz)} = -\frac{e^{iaz_s}}{b^2(z - z_s)^2} - \frac{ia\,e^{iaz_s}}{b^2(z - z_s)} + \cdots.$$

Hence, we have a second-order pole there.

Step 5: Show that

$$(1 - e^{-\pi a/b}) \int_{-\infty}^\infty \frac{e^{iax}}{\cosh^2(bx)}\,dx = \frac{2\pi a e^{-\pi a/(2b)}}{b^2}.$$

Step 6: Simplify Step 5 to obtain the desired result.

18. Using the closed contour integral

$$\oint_C \frac{z}{\cosh(z)\cosh(z + a)}\,dz,$$

where C is a *rectangular* contour with vertices at $(\infty, 0)$, $(-\infty, 0)$, (∞, π), and $(-\infty, \pi)$, show[11] that

$$\int_0^\infty \frac{dx}{\cosh(x)\cosh(x+a)} = \begin{cases} 2a/\sinh(a), & \text{if} \quad a \neq 0, \\ 2, & \text{if} \quad a = 2. \end{cases}$$

Step 1: If $a \neq 0$, show that

$$\oint_C f(z)\,dz = 2\pi i \operatorname{Res}[f(z); z_1] + 2\pi i \operatorname{Res}[f(z); z_2],$$

where z_1 and z_2 are simple poles with $z_1 = \pi i/2$ and $z_2 = -a + \pi i/2$, respectively.

Step 2: Show that in this case,

$$\operatorname{Res}[f(z); z_1] = -\frac{\pi i}{2\sinh(a)}, \qquad \text{and} \qquad \operatorname{Res}[f(z); z_2] = \left(\frac{\pi i}{2} - a\right)\frac{1}{\sinh(a)}.$$

Step 3: If $a = 0$, show that we a second-order pole located at $z_1 = \pi i/2$ within the closed contour with

$$\begin{aligned} \operatorname{Res}[f(z); z_1] &= \lim_{z \to \pi i/2} \frac{d}{dz}\left[\frac{z(z-\pi i/2)^2}{\cosh^2(z)}\right] = -\lim_{\eta \to 0} \frac{d}{d\eta}\left[\frac{\eta^2(\eta + \pi i/2)}{\sinh^2(\eta)}\right] \\ &= -\lim_{\eta \to 0} \frac{d}{d\eta}\left[\frac{\eta + \pi i/2}{(1 + \eta^2/6 + \cdots)^2}\right] \\ &= -\lim_{\eta \to 0}\left[\frac{(1 + \eta^2/6 + \cdots) - (\eta + \pi i/2)(3\eta + \cdots)}{(1 + \eta^2/6 + \cdots)^4}\right] = -1. \end{aligned}$$

Step 4: Denoting the contour along and parallel to the y-axis at $x = \infty$ as C_2 and the contour along and parallel to the y-axis at $x = -\infty$ as C_4, show that

$$\int_{C_2} f(z)\,dz \to 0 \quad \text{as} \quad x \to \infty, \qquad \text{and} \qquad \int_{C_4} f(z)\,dz \to 0 \quad \text{as} \quad x \to -\infty.$$

Step 5: Along the real axis, call it C_1, show that

$$\int_{C_1} f(z)\,dz = \int_{-\infty}^{\infty} \frac{x}{\cosh(x)\cosh(x+a)}\,dx,$$

while along the contour C_3 (which runs parallel to the real axis but π units above it),

$$\int_{C_3} f(z)\,dz = -\int_{-\infty}^{\infty} \frac{x + \pi i}{\cosh(x)\cosh(x+a)}\,dx.$$

Step 6: If $a \neq 0$, show that

$$-\pi i \int_{-\infty}^{\infty} \frac{dx}{\cosh(x)\cosh(x+a)} = -\frac{2\pi a i}{\sinh(a)},$$

---

[11] See Yan, J. R., X. H. Yan, J. Q. You, and J. X. Zhong, 1993: On the interaction between two nonpropagating hydrodynamic solitons. *Phys. Fluids A*, **5**, 1651–1656.

while for $a = 0$, show that

$$-\pi i \int_{-\infty}^{\infty} \frac{dx}{\cosh^2(x)} = -2\pi i.$$

19. During an electromagnetic calculation, Strutt[12] needed to prove that

$$\pi \frac{\sinh(\sigma x)}{\cosh(\sigma\pi)} = 2\sigma \sum_{n=0}^{\infty} \frac{\cos\left[\left(n+\frac{1}{2}\right)(x-\pi)\right]}{\sigma^2 + \left(n+\frac{1}{2}\right)^2}, \qquad |x| \le \pi.$$

Verify his proof by doing the following:

Step 1: Using the residue theorem, show that

$$\frac{1}{2\pi i} \oint_{C_N} \pi \frac{\sinh(xz)}{\cosh(\pi z)} \frac{dz}{z-\sigma} = \pi \frac{\sinh(\sigma x)}{\cosh(\sigma\pi)} - \sum_{n=-N-1}^{N} \frac{(-1)^n \sin\left[\left(n+\frac{1}{2}\right)x\right]}{\sigma - i\left(n+\frac{1}{2}\right)},$$

where C_N is a circular contour that includes the poles $z = \sigma$ and $z_n = \pm i\left(n+\frac{1}{2}\right)$, $n = 0, 1, 2, \ldots, N$.

Step 2: Show that in the limit of $N \to \infty$, the contour integral vanishes. Hint: Examine the behavior of $z \sinh(xz)/[(z-\sigma)\cosh(\pi z)]$ as $|z| \to \infty$. Use Equation 1.9.7 where C_R is the circular contour.

Step 3: Break the infinite series in Step 1 into two parts and simplify.

You would obtain the same series by computing the Fourier series of $\sinh(\sigma x)/\cosh(\sigma\pi)$ and using direct integration.

1.10 CAUCHY'S PRINCIPAL VALUE INTEGRAL

The conventional definition of the integral of a function $f(x)$ of the real variable x over a finite interval $a \le x \le b$ assumes that $f(x)$ has a definite finite value at each point within the interval. We shall now extend this definition to cover cases when $f(x)$ is infinite at a finite number of points within the interval.

Consider the case when there is only one point c at which $f(x)$ becomes infinite. If c is not an endpoint of the interval, we take two small positive numbers ϵ and η and examine the expression

$$\int_a^{c-\epsilon} f(x)\,dx + \int_{c+\eta}^{b} f(x)\,dx. \tag{1.10.1}$$

If Equation 1.10.1 exists and tends to a unique limit as ϵ and η tend to zero independently, we say that the improper integral of $f(x)$ over the interval exists, its value being defined by

$$\int_a^b f(x)\,dx = \lim_{\epsilon\to 0} \int_a^{c-\epsilon} f(x)\,dx + \lim_{\eta\to 0} \int_{c+\eta}^{b} f(x)\,dx. \tag{1.10.2}$$

[12] Strutt, M. J. O., 1934: Berechnung des hochfrequenten Feldes einer Kreiszylinderspule in einer konzentrischen leitenden Schirmhülle mit ebenen Deckeln. *Hochfrequenztechn. Elecktroak.*, **43**, 121–123.

If, however, the expression does not tend to a limit as ϵ and η tend to zero independently, it may still happen that

$$\lim_{\epsilon\to 0}\left\{\int_a^{c-\epsilon} f(x)\,dx + \int_{c+\epsilon}^{b} f(x)\,dx\right\} \tag{1.10.3}$$

exists. When this is the case, we call this limit the *Cauchy principal value* of the improper integral and denote it by

$$PV\int_a^b f(x)\,dx. \tag{1.10.4}$$

Finally, if $f(x)$ becomes infinite at an endpoint, say a, of the range of integration, we say that $f(x)$ is integrable over $a \leq x \leq b$ if

$$\lim_{\epsilon\to 0^+}\int_{a+\epsilon}^{b} f(x)\,dx \tag{1.10.5}$$

exists.

• Example 1.10.1

Consider the integral $\int_{-1}^{2} dx/x$. This integral does not exist in the ordinary sense because of the strong singularity at the origin. However, the integral would exist if

$$\lim_{\epsilon\to 0}\int_{-1}^{\epsilon}\frac{dx}{x} + \lim_{\delta\to 0}\int_{\delta}^{2}\frac{dx}{x} \tag{1.10.6}$$

existed and had a unique value as ϵ and δ independently approach zero. Because this limit equals

$$\lim_{\epsilon,\delta\to 0}\left[\ln(\epsilon) + \ln(2) - \ln(\delta)\right] = \lim_{\epsilon,\delta\to 0}\left[\ln(2) - \ln(\delta/\epsilon)\right], \tag{1.10.7}$$

our integral would have the value of $\ln(2)$ if $\delta = \epsilon$. This particular limit is the Cauchy principal value of the improper integral, which we express as

$$PV\int_{-1}^{2}\frac{dx}{x} = \ln(2). \tag{1.10.8}$$

□

We can extend these ideas to complex integrals used to determine the value or principal value of an improper integral by Cauchy's residue theorem when the integrand has a singularity on the contour of integration. We avoid this difficulty by deleting from the area within the contour, that portion which also lies within a small circle $|z - c| = \epsilon$, and then integrating around the boundary of the remaining region. This process is called *indenting* the contour.

The integral around the indented contour is calculated by the theorem of residues and then the radius of each indentation is made to tend to zero. This process gives the Cauchy principal value of the improper integral. The details of this method are shown in the following examples.

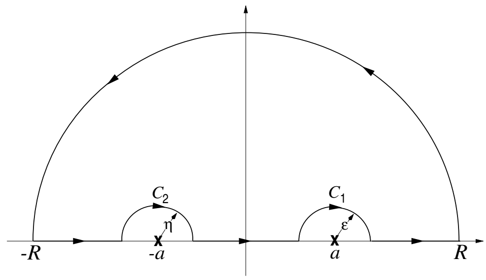


Figure 1.10.1: Contour C used in Example 1.10.2.

• Example 1.10.2

Let us show that

$$PV \int_{-\infty}^{\infty} \frac{\cos(x)}{a^2 - x^2}\, dx = \frac{\pi \sin(a)}{a}, \qquad a > 0. \tag{1.10.9}$$

Consider the integral

$$\oint_C \frac{e^{iz}}{a^2 - z^2}\, dz, \tag{1.10.10}$$

where the closed contour C consists of the real axis from $-R$ to R and a semicircle in the upper half of the z-plane where this segment is its diameter. See Figure 1.10.1. Because the integrand has poles at $z = \pm a$, which lie on this contour, we modify C by making an indentation of radius ϵ at a and another of radius η at $-a$. The integrand is now analytic within and on C and Equation 1.10.10 equals zero by the Cauchy-Goursat theorem.

Evaluating each part of the integral, Equation 1.10.10, we have that

$$\begin{aligned} &\int_0^{\pi} \frac{e^{iR\cos(\theta) - R\sin(\theta)}}{a^2 - R^2 e^{2\theta i}} iRe^{\theta i}\, d\theta + \int_{C_1} \frac{e^{iz}}{a^2 - z^2}\, dz + \int_{C_2} \frac{e^{iz}}{a^2 - z^2}\, dz \\ &\quad + \int_{-R}^{-a-\eta} \frac{e^{ix}}{a^2 - x^2}\, dx + \int_{-a+\eta}^{a-\epsilon} \frac{e^{ix}}{a^2 - x^2}\, dx + \int_{a-\epsilon}^{R} \frac{e^{ix}}{a^2 - x^2}\, dx = 0, \end{aligned} \tag{1.10.11}$$

where C_1 and C_2 denote the integrals around the indentations at a and $-a$, respectively. The modulus of the first term on the left side of Equation 1.10.11 is less than $\pi R/(R^2 - a^2)$, so this term tends to zero as $R \to \infty$. To evaluate C_1, we observe that $z = a + \epsilon e^{\theta i}$ along C_1, where θ decreases from π to 0. Hence,

$$\int_{C_1} \frac{e^{iz}}{a^2 - z^2}\, dz = \lim_{\epsilon \to 0} \int_{\pi}^{0} \exp\left(ia + i\epsilon e^{\theta i}\right) \frac{\epsilon i e^{\theta i}}{-2a\epsilon e^{\theta i} - \epsilon^2 e^{2\theta i}}\, d\theta \tag{1.10.12}$$

$$= \lim_{\epsilon \to 0} \int_0^{\pi} \exp\left(ia + i\epsilon e^{\theta i}\right) \frac{i}{2a + \epsilon e^{\theta i}}\, d\theta = \frac{\pi i e^{ia}}{2a}. \tag{1.10.13}$$

Similarly,

$$\int_{C_2} \frac{e^{iz}}{a^2 - z^2}\, dz = -\frac{\pi i e^{-ia}}{2a}, \tag{1.10.14}$$

as η tends to zero.

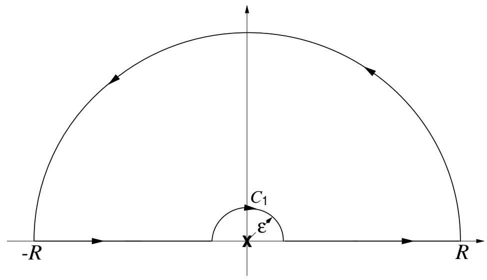


Figure 1.10.2: Contour C used in Example 1.10.3.

Upon letting $R \to \infty$, $\epsilon \to 0$, and $\eta \to 0$, we find that

$$PV \int_{-\infty}^{\infty} \frac{e^{ix}}{a^2 - x^2}\, dx = -\frac{\pi i}{2a}\left(e^{ia} - e^{-ia}\right) = \frac{\pi \sin(a)}{a}. \tag{1.10.15}$$

Finally, equating the real and imaginary parts, we obtain

$$PV \int_{-\infty}^{\infty} \frac{\cos(x)}{a^2 - x^2}\, dx = \frac{\pi \sin(a)}{a}, \quad PV \int_{-\infty}^{\infty} \frac{\sin(x)}{a^2 - x^2}\, dx = 0. \tag{1.10.16}$$

□

• **Example 1.10.3**

Let us show that

$$\int_{-\infty}^{\infty} \frac{\sin(x)}{x}\, dx = \pi. \tag{1.10.17}$$

Consider the integral

$$\oint_C \frac{e^{iz}}{z}\, dz, \tag{1.10.18}$$

where the closed contour C consists of the real axis from $-R$ to R and a semicircle in the upper half of the z-plane where this segment is its diameter. Because the integrand has a pole at $z = 0$, which lies on the contour, we modify C by making an indentation of radius ϵ at $z = 0$. See Figure 1.10.2. Because e^{iz}/z is analytic along C,

$$\int_0^{\pi} e^{iR\cos(\theta) - R\sin(\theta)} i\, d\theta + \int_{-R}^{-\epsilon} \frac{e^{ix}}{x}\, dx + \int_{C_1} \frac{e^{iz}}{z}\, dz + \int_{\epsilon}^{R} \frac{e^{ix}}{x}\, dx = 0. \tag{1.10.19}$$

Since $e^{-R\sin(\theta)} < e^{-R\theta}$ for $0 < \theta < \pi$,

$$\left| \int_0^{\pi} e^{iR\cos(\theta) - R\sin(\theta)} i\, d\theta \right| \le \int_0^{\pi} e^{-R\theta}\, d\theta = \frac{1 - e^{-\pi R}}{R}, \tag{1.10.20}$$

which tends to zero as $R \to \infty$. Therefore,

$$\int_{-\infty}^{-\epsilon} \frac{e^{ix}}{x}\, dx + \int_{\epsilon}^{\infty} \frac{e^{ix}}{x}\, dx = -\int_{C_1} \frac{e^{iz}}{z}\, dz. \tag{1.10.21}$$

Now,

$$\int_{C_1} \frac{e^{iz}}{z}\,dz = \int_{C_1} \frac{dz}{z} + i\int_{C_1} dz - \int_{C_1} \frac{z}{2}\,dz + \cdots = -\pi i \tag{1.10.22}$$

in the limit $\epsilon \to 0$ because $z = \epsilon e^{\theta i}$. Consequently, in the limit of $\epsilon \to 0$,

$$PV \int_{-\infty}^{\infty} \frac{e^{ix}}{x}\,dx = \pi. \tag{1.10.23}$$

Upon separating the real and imaginary parts, we obtain

$$PV \int_{-\infty}^{\infty} \frac{\cos(x)}{x}\,dx = 0, \qquad \int_{-\infty}^{\infty} \frac{\sin(x)}{x}\,dx = \pi. \tag{1.10.24}$$

Problems

1. Noting that

$$\int_0^{\theta-\epsilon} \frac{d\varphi}{\cos(\varphi)-\cos(\theta)} = \frac{1}{\sin(\theta)} \ln\left|\frac{\sin\left[\frac{1}{2}(\theta+\varphi)\right]}{\sin\left[\frac{1}{2}(\theta-\varphi)\right]}\right|_0^{\theta-\epsilon},$$

and

$$\int_{\theta+\epsilon}^{\pi} \frac{d\varphi}{\cos(\varphi)-\cos(\theta)} = \frac{1}{\sin(\theta)} \ln\left|\frac{\sin\left[\frac{1}{2}(\theta+\varphi)\right]}{\sin\left[\frac{1}{2}(\theta-\varphi)\right]}\right|_{\theta+\epsilon}^{\pi},$$

show that

$$PV \int_0^{\pi} \frac{d\varphi}{\cos(\varphi)-\cos(\theta)} = 0, \qquad 0 < \theta < \pi.$$

2. Show that

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

Step 1: Show that

$$\int_{-\infty}^{\infty} \frac{\cos(\pi x/2)}{x^2-1}\,dx = \Re\left[PV \int_{-\infty}^{\infty} \frac{e^{i\pi x/2}}{x^2-1}\,dx\right].$$

Step 2: Consider now the integral $\oint_C e^{i\pi z/2}\,dz/(z^2-1)$, where the closed contour C consists of the real axis from $-R$ to R plus a semicircle of radius R in the upper half of the z-plane. See Figure 1.10.1. Because the integrand has poles at $z = \pm 1$ which lie on the contour, we modify C by making an indentation of radius η above $z = -1$ and another indentation of radius ϵ above $z = 1$. Why is this closed contour integral equal to zero?

Step 3: Show that this contour integral is given by

$$\lim_{R\to\infty} \int_0^{\pi} \frac{e^{i\pi R\cos(\theta)/2 - R\pi\sin(\theta)/2}}{R^2e^{2\theta i}-1} iRe^{\theta i}\,d\theta + \int_{C_1} \frac{e^{i\pi z/2}}{z^2-1}\,dz + \int_{C_2} \frac{e^{i\pi z/2}}{z^2-1}\,dz$$
$$+ \lim_{\eta\to 0} \int_{-R}^{-1-\eta} \frac{e^{i\pi x/2}}{x^2-1}\,dx + \lim_{\epsilon,\eta\to 0} \int_{-1+\eta}^{1-\epsilon} \frac{e^{i\pi x/2}}{x^2-1}\,dx + \lim_{\epsilon\to 0} \int_{1+\epsilon}^{R} \frac{e^{i\pi x/2}}{x^2-1}\,dx = 0,$$

where C_1 and C_2 denote the integrals around the indentations at -1 and 1, respectively.

Step 4: Show that the first term on the left side tends to zero as $R \to \infty$. Why?

Step 5: Taking $z = -1 + \eta e^{\theta i}$ along C_1, where θ decreases from π to 0, show that

$$\int_{C_1} \frac{e^{i\pi z/2}}{z^2 - 1}\,dz = \lim_{\eta\to 0}\int_{\pi}^{0} \frac{\exp\left[i\pi(-1+\eta e^{\theta i})/2\right]}{-2\eta e^{\theta i} + \eta^2 e^{2\theta i}}\, i\eta e^{\theta i}\,d\theta = \frac{\pi}{2}.$$

Step 6: Similarly, show that along C_2,

$$\int_{C_2} \frac{e^{i\pi z/2}}{z^2 - 1}\,dz = \lim_{\epsilon\to 0}\int_{\pi}^{0} \frac{\exp\left[i\pi(1+\epsilon e^{\theta i})/2\right]}{2\epsilon e^{\theta i} + \epsilon^2 e^{2\theta i}}\, i\epsilon e^{\theta i}\,d\theta = \frac{\pi}{2}.$$

Step 7: Using Steps 1 through 6, obtain the final result.

3. Show that

$$\int_{-\infty}^{\infty} \frac{e^{ax} - e^{bx}}{1 - e^{x}}\,dx = \pi[\cot(a\pi) - \cot(b\pi)], \qquad 0 < a, b < 1.$$

Step 1: Consider the integral $\oint_C \left(e^{az} - e^{bz}\right) dz/(1 - e^{z})$, where the closed contour C consists of the rectangular box with vertices at $(-R, 0)$, $(R, 0)$, $(-R, \pi)$ and (R, π), and a semicircular indentation C_ϵ at the origin. Show that this closed integral equals zero. Why?

Step 2: Show that this closed integral may be rewritten,

$$\lim_{R\to\infty}\int_{R+\pi i}^{-R+\pi i} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz + \lim_{R\to\infty}\int_{-R+\pi i}^{-R} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz + \lim_{R\to\infty,\epsilon\to 0}\int_{-R}^{-\epsilon} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz$$
$$+ \int_{C_\epsilon} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz + \lim_{R\to\infty,\epsilon\to 0}\int_{\epsilon}^{R} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz + \lim_{R\to\infty}\int_{R}^{R+\pi i} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz = 0.$$

Step 3: Show that

$$\begin{aligned}\int_{C_\epsilon} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz &= \lim_{\epsilon\to 0}\int_{\pi}^{0} \frac{1 + a\epsilon e^{\theta i} + a^2\epsilon^2 e^{2\theta i}/2 + \cdots - 1 - b\epsilon e^{\theta i} - b^2\epsilon^2 e^{2\theta i}/2 - \cdots}{1 - 1 - \epsilon e^{\theta i} - \epsilon^2 e^{2\theta i}/2 - \cdots} i\epsilon e^{\theta i}\,d\theta \\ &= 0.\end{aligned}$$

Step 4: Show that

$$\lim_{R\to\infty}\int_{-R+\pi i}^{-R} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz = \lim_{R\to\infty}\int_{\pi}^{0} \frac{e^{-aR}e^{ayi} - e^{-bR}e^{byi}}{1 - e^{-R}e^{yi}}\, i\,dy = 0,$$

and

$$\lim_{R\to\infty}\int_{R}^{R+\pi i} \frac{e^{az} - e^{bz}}{1 - e^{z}}\,dz = \lim_{R\to\infty}\int_{0}^{\pi} \frac{e^{aR}e^{ayi} - e^{bR}e^{byi}}{1 - e^{R}e^{yi}}\, i\,dy = 0,$$

if $0 < a, b < 1$.

Step 5: Show that

$$\int_{-\infty}^{\infty} \frac{e^{ax} - e^{bx}}{1 - e^x}\,dx = \int_{-\infty}^{\infty} \frac{e^{ax}e^{a\pi i}}{1 + e^x}\,dx - \int_{-\infty}^{\infty} \frac{e^{bx}e^{b\pi i}}{1 + e^x}\,dx$$
$$= \frac{\pi e^{a\pi i}}{\sin(a\pi)} - \frac{\pi e^{b\pi i}}{\sin(b\pi)} = \pi\left[\cot(a\pi) + i\right] - \pi\left[\cot(b\pi) + i\right].$$

4. Show[13] that

$$\int_{-\infty}^{\infty} \frac{1 - \cos[2a(x+\zeta)]}{(x+\zeta)^2(x^2+\alpha^2)}\,dx = \frac{\pi}{\alpha(\zeta^2+\alpha^2)^2}\big\{2a\alpha(\zeta^2+\alpha^2) + (\zeta^2-\alpha^2)$$
$$- e^{-2a\alpha}\left[(\zeta^2-\alpha^2)\cos(2a\zeta) + 2\alpha\zeta\sin(2a\zeta)\right]\big\},$$

where a, α, and ζ are real.

Step 1: Show that

$$\left(\int_{C_\infty} + \int_{-R}^{-\zeta-\epsilon} + \int_{C_\epsilon} + \int_{-\zeta+\epsilon}^{R}\right) f(z)\,dz = 2\pi i\,\mathrm{Res}[f(z); i\alpha],$$

where C_∞ denotes the semicircular contour of infinite radius, C_ϵ is the semicircular indentation above $z = -\zeta$ and

$$f(z) = \frac{1 - e^{2ia(z+\zeta)}}{(z+\zeta)^2(z^2+\alpha^2)}.$$

Step 2: Taking the limit of $R \to \infty$ and $\epsilon \to 0$, show that

$$\int_{-\infty}^{\infty} f(x)\,dx = 2\pi i\,\mathrm{Res}[f(z); i\alpha] + \pi i\,\mathrm{Res}[f(z); -\zeta].$$

Step 3: Show that

$$\mathrm{Res}[f(z); i\alpha] = \frac{\zeta^2 - \alpha^2 - e^{-2a\alpha}[(\zeta^2-\alpha^2)\cos(2a\zeta) + 2\alpha\zeta\sin(2a\zeta)]}{2i\alpha(\zeta^2+\alpha^2)^2}$$
$$- \frac{2\alpha\zeta + e^{-2a\alpha}[(\zeta^2-\alpha^2)\sin(2a\zeta) - 2\alpha\zeta\cos(2a\zeta)]}{2\alpha(\zeta^2+\alpha^2)^2}$$

and

$$\mathrm{Res}[f(z); -\zeta] = \lim_{z\to-\zeta} \frac{d}{dz}\left[\frac{1 - e^{2ia(z+\zeta)}}{z^2+\alpha^2}\right] = -\frac{2ia}{\zeta^2+\alpha^2}.$$

Step 4: Use the results from Steps 1 through 3 to obtain the desired result.

5. Show that

$$PV \int_{-\infty}^{\infty} \frac{\cos(mx)}{x-a}\,dx = -\pi\sin(ma), \qquad \text{and} \qquad PV \int_{-\infty}^{\infty} \frac{\sin(mx)}{x-a}\,dx = \pi\cos(ma),$$

[13] Ko, S. H., and A. H. Nuttall, 1991: Analytical evaluation of flush-mounted hydrophone array response to the Corcos turbulent wall pressure spectrum. *J. Acoust. Soc. Am.*, **90**, 579–588.

where $m > 0$ and a is real.

Step 1: Using the complex function $e^{imz}/(z-a)$ and a closed contour similar to that shown in Figure 1.10.2, show that

$$\left(\int_{C_\infty} + \int_{-R}^{a-\epsilon} + \int_{C_\epsilon} + \int_{a+\epsilon}^{R} \right) \frac{e^{imz}}{z-a}\, dz = 0.$$

Why? Here C_∞ denotes the semicircular contour of infinite radius and C_ϵ is the semicircular indentation above $z = a$.

Step 2: Taking the limit of $R \to \infty$ and $\epsilon \to 0$, show that

$$PV \int_{-\infty}^{\infty} \frac{e^{imx}}{x-a}\, dx = \pi i \operatorname{Res}\left[\frac{e^{imz}}{z-a}; a \right] = \pi i e^{ima}.$$

Step 3: Complete the derivation by taking the real and imaginary parts of the equation in Step 2.

6. Show that

$$PV \int_{-\infty}^{\infty} \frac{x e^{xi}}{x^2 - \pi^2}\, dx = -\pi i, \qquad \text{and} \qquad PV \int_{-\infty}^{\infty} \frac{e^{imx}}{(x-1)(x-3)}\, dx = \frac{\pi i}{2}\left(e^{3mi} - e^{mi} \right),$$

where $m > 0$.

Step 1: Show that

$$\left(\int_{C_\infty} + \int_{-R}^{-\pi-\epsilon} + \int_{C_{\epsilon_1}} + \int_{-\pi+\epsilon}^{\pi-\epsilon} + \int_{C_{\epsilon_2}} + \int_{\pi+\epsilon}^{R} \right) \frac{z e^{iz}}{z^2 - \pi^2}\, dz = 0$$

Why? Here C_∞ denotes the semicircular contour of infinite radius and C_{ϵ_1} and C_{ϵ_2} are semicircular indentations above $z = -\pi$ and $z = \pi$.

Step 2: Taking the limit of $R \to \infty$ and $\epsilon \to 0$, show that

$$\begin{aligned} PV \int_{-\infty}^{\infty} \frac{x e^{ix}}{x^2 - \pi^2}\, dx &= \pi i \operatorname{Res}\left[\frac{z\, e^{iz}}{z^2 - \pi^2}; -\pi \right] + \pi i \operatorname{Res}\left[\frac{z\, e^{iz}}{z^2 - \pi^2}; \pi \right] \\ &= \tfrac{1}{2}\pi i e^{-\pi i} + \tfrac{1}{2}\pi i e^{\pi i} = -\pi i. \end{aligned}$$

Step 3: To prove the second relationship, show that

$$\left(\int_{C_\infty} + \int_{-R}^{1-\epsilon} + \int_{C_{\epsilon_1}} + \int_{1+\epsilon}^{3-\epsilon} + \int_{C_{\epsilon_2}} + \int_{3+\epsilon}^{R} \right) \frac{e^{imz}}{(z-1)(z-3)}\, dz = 0.$$

Why? Here C_∞ denotes the semicircular contour of infinite radius and C_{ϵ_1} and C_{ϵ_2} are semicircular indentations above $z = 1$ and $z = 3$.

Step 4: Taking the limit of $R \to \infty$ and $\epsilon \to 0$, show that

$$\begin{aligned} PV \int_{-\infty}^{\infty} \frac{e^{imx}}{(x-1)(x-3)}\, dx &= \pi i \operatorname{Res}\left[\frac{e^{imz}}{(z-1)(z-3)}; 1 \right] + \pi i \operatorname{Res}\left[\frac{e^{imz}}{(z-1)(z-3)}; 3 \right] \\ &= -\tfrac{1}{2}\pi i e^{mi} + \tfrac{1}{2}\pi i e^{3mi} = \frac{\pi i}{2}\left(e^{3mi} - e^{mi} \right). \end{aligned}$$

7. Redo Example 1.10.3, except the contour is now a rectangle with vertices at $\pm R$ and $\pm R + Ri$ indented at the origin.

Step 1: Show that along the left side,

$$\left| \int_{-R}^{-R+Ri} \frac{e^{iz}}{z} \, dz \right| \le \int_0^R \frac{e^{-y}}{\sqrt{R^2+y^2}} \, dy < \frac{1}{R} \int_0^R e^{-y} \, dy = \frac{1}{R} \left(1 - e^{-R}\right),$$

which tends to zero as $R \to \infty$.

Step 2: Show that along the top,

$$\left| \int_{R}^{R+Ri} \frac{e^{iz}}{z} \, dz \right| \le \int_0^R \frac{e^{-y}}{\sqrt{R^2+y^2}} \, dy,$$

which also tends to zero as $R \to \infty$. Why?

Step 3: Show that along the right side,

$$\left| \int_{-R+Ri}^{R+Ri} \frac{e^{iz}}{z} \, dz \right| \le 2e^{-R} \int_0^R \frac{dx}{\sqrt{R^2+x^2}} = 2\ln\left(1+\sqrt{2}\right) e^{-R},$$

which tends to zero as $R \to \infty$. Why?

Step 4: Just as in the case of the semicircle close contour, we only have an integration along the real axis. Do this to complete the problem.

8. Let us show[14] that

$$G(\alpha) = PV \int_{-1}^{1} \frac{dx}{(x+\alpha)\sqrt{1-x^2}} = \begin{cases} \dfrac{\alpha\pi}{|\alpha|\sqrt{\alpha^2-1}}, & |\alpha| > 1, \\ 0, & |\alpha| < 1. \end{cases}$$

Step 1: Using the transformation $2ix = z - z^{-1}$, show that

$$2i \, dx = \frac{z^2+1}{z^2} \, dz, \qquad 1 - x^2 = \tfrac{1}{4} \left(z + \frac{1}{z}\right)^2,$$

$$\sqrt{1-x^2} = \frac{1}{2}\left(\frac{z^2+1}{z}\right), \quad \text{and} \quad x + \alpha = \frac{1}{2i}\left(z - \frac{1}{z}\right) + \alpha = \frac{1}{2i}\left(\frac{z^2 + 2i\alpha z - 1}{z}\right).$$

Substitute these results into the original integral to find $G(\alpha)$ as a contour integration on the unit circle.

Step 2: For $|\alpha| < 1$, we have two singularities within the contours located at $z = \pm\sqrt{1-\alpha^2} - \alpha i$. In that case, show that $G(\alpha) = 0$.

Step 3: If $\alpha > 1$, there is a single singularity within the contours and it is located at $z = i\sqrt{\alpha^2-1} - \alpha i$. Show that $G(\alpha) = \pi/\sqrt{\alpha^2-1}$.

---

[14] Ott, E., T. M. Antonsen, and R. V. Lovelace, 1977: Theory of foil-less diode generation of intense relativistic electron beams. *Phys. Fluids*, **20**, 1180–1184.

Step 4: Finally, if $\alpha < -1$, there is a single singularity within the contours and it is located at $z = -i\sqrt{\alpha^2 - 1} - \alpha i$. Show that $G(\alpha) = -\pi/\sqrt{\alpha^2 - 1}$.

9. Let the function $f(z)$ possess a simple pole with a residue $\text{Res}[f(z); c]$ on a simply closed contour C. If C is indented at c, show that the integral of $f(z)$ around the indentation tends to $-\text{Res}[f(z); c]\alpha i$ as the radius of the indentation tends to zero, α being the internal angle between the two parts of C meeting at c.

1.11 CONFORMAL MAPPING

Conformal mapping is a powerful technique for finding solutions, or for simplifying the process of finding solutions, to Laplace's differential equation in two dimensions. This method involves introducing two complex variables: $z = x + iy$ and $\tau = \rho + i\sigma$. These two complex variables are related to each other via the mapping $z = f(\tau)$. Under this mapping the Argand diagram for the z-variable is mapped into one for the τ-variable. In certain cases, for example $\tau = \sqrt{z}$, the complex z-plane may only map into a portion of the τ-plane. In other cases, say $\tau = z + 3i$, the complete z-plane would be mapped into the complete τ-plane.

Once we map the original domain into a simpler geometry (a half-plane, circle or square), how do we find the solution? There are several techniques available. One method, for example, recalls that the real and imaginary parts of an analytic function satisfy Laplace's equation. Therefore, if we could construct an analytic function whose real or imaginary parts satisfy the boundary conditions in the new domain, we would have the solution in the τ-plane. Then we could use the transformation to obtain the solution in the original z-plane.

What types of functions $f(z)$ are useful? Consider an arbitrary point z_0 in the complex z-plane. Assuming that $f'(z_0) \neq 0$, a straightforward transformation yields

$$\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)_{z_0} = |f'(z_0)|^2 \left(\frac{\partial^2 v}{\partial \rho^2} + \frac{\partial^2 v}{\partial \sigma^2}\right)_{\tau_0}, \tag{1.11.1}$$

where $u(x, y)$ and $v(\rho, \sigma)$ are solutions to Laplace's equation in the z and τ planes, respectively. Thus, $f(z)$ must be analytic.

• Example 1.11.1

In their study of magnetic recording, Curland and Judy[15] modeled the ring heads as two semi-infinite regions located below the x-axis and running to the right of $x = a/2$ and to the left of $x = -a/2$. See Figure 1.11.1.

From symmetry we need only consider the half-space $x > 0$. Consequently, the new boundary consists of the four line segments: AB, BC, CD and DE. If we require that the point D in the τ-plane lies at $\tau = 1$, we shall show in Example 1.11.7 that the desired conformal mapping is

$$z = \frac{a}{\pi}\left[\sqrt{\tau - 1} - \frac{i}{2}\log\left(\frac{1 - i\sqrt{\tau - 1}}{1 + i\sqrt{\tau - 1}}\right)\right] + \frac{a}{2}. \tag{1.11.2}$$

[15] Curland, N., and J. H. Judy, 1986: Calculation of exact ring head fields using conformal mapping. *IEEE Trans. Magnet.*, **MAG-22**, 1901–1903.

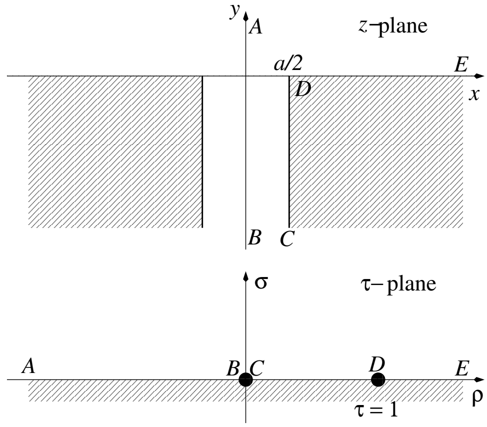


Figure 1.11.1: The conformal mapping used to find the fields of a semi-infinite ring head with a finite gap of width a. The potential on the right pole face equals 1 while the potential of the left pole face equals -1. In the z-plane the point A is located at $(0, \infty)$ while point B is located at $(0, -\infty)$. Because of symmetry the potential along the center of the gap AB equals 0.

A useful method for illustrating this conformal mapping is to draw lines of constant ρ and σ in the z-plane. See Figure 1.11.2. This figure shows the local orthogonality between lines of constant ρ and σ.

The greatest difficulty in creating this figure was computing τ for a given z. This was done using the Newton-Raphson method. Starting at the top of the domain, the first guess there was given by $\tau = 1 + \pi^2 z^2$. Marching downward, the τ from the previous grid point was used for the initial guess. The corresponding MATLAB script is as follows:

```
clear; delta = 0.01; % resolution of the grid

for jj = 1:201
for ii = 1:201
  XX(jj,ii) = delta*ii; YY(jj,ii) = delta*(jj-101);
  RHO(jj,ii) = NaN; SIGMA(jj,ii) = NaN;
end; end

% code for the domain x,y > 0

for jj = 1:100
  y = 1 - delta*(jj-1);
for ii = 1:201
  x = delta*ii; z = complex(x,y);
  if (jj == 1) tau = 1+pi*pi*z*z; else tau = TAU(ii); end
  for icount = 1:10
    temp1 = sqrt(tau-1);
    temp2 = temp1 - 0.5*i*log(1-i*temp1) + 0.5*i*log(1+i*temp1);
    ff = temp2/pi + 0.5 - z; deriv = temp1 /(2*pi*tau);
    temp3 = ff/deriv; tau = tau - temp3; % Newton-Raphson method
  end
  TAU(ii) = tau; RHO(202-jj,ii) = real(tau);
  SIGMA(202-jj,ii) = imag(tau);
end; end
```

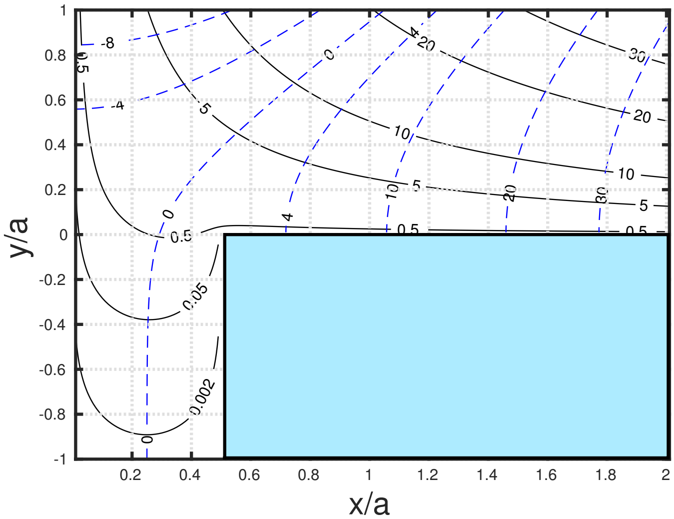


Figure 1.11.2: Lines of constant ρ (dashed lines) and σ (solid lines) given by the conformal mapping expressed by Equation 1.11.2.

```
% code for the domain 0 < x < 1/2 and y < 0

for jj = 1:101
  y = delta - delta*jj;
for ii = 1:49
  x = delta*ii; z = complex(x,y);
  tau = TAU(ii); % first guess
  for icount = 1:10
    temp1 = sqrt(tau-1);
    temp2 = temp1 - 0.5*i*log(1-i*temp1) + 0.5*i*log(1+i*temp1);
    ff = temp2/pi + 0.5 - z; deriv = temp1 /(2*pi*tau);
    temp3 = ff/deriv; tau = tau - temp3; % Newton-Raphson method
  end
  TAU(ii) = tau; RHO(102-jj,ii) = real(tau);
  SIGMA(102-jj,ii) = imag(tau);
end; end

% plot the conformal mapping Equation 1.11.2

figure
[C,h] = contour(XX,YY,SIGMA,[0.002,0.05,0.5,5,10,20,30],'k');
clabel(C,h,'FontSize',10,'Color','k','Rotation',0)
xlabel('x/a','FontSize',20); ylabel('y/a','FontSize',20);
hold on
v = [-8,-4,0,4,10,20,30];
[C,h] = contour(XX,YY,RHO,v,'--b');
clabel(C,h,'FontSize',10,'Color','b','Rotation',0)
```

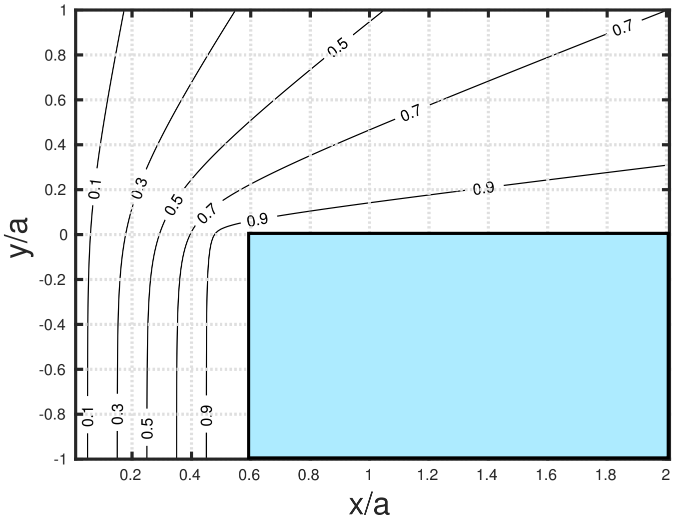


Figure 1.11.3: The solution to Laplace's equation when the left boundary is held at 0 while the left and top sides of the shaded rectangle are held at 1. This figure shows only a portion of the domain $x > 0$ and $|y| < \infty$.

Now that we can transform between the z-plane and the τ-plane, and vice versa, let us turn our attention to finding the solution to Laplace's equation in the τ-plane. There the solution equals 1 for $\rho > 0$ and 0 for $\rho < 0$ along $\sigma = 0$.

Consider now the analytic function (except at the branch point $\tau = 0$)

$$f(\tau) = i - \log(\tau)/\pi. \tag{1.11.3}$$

A quick check (using $\tau = re^{i\theta}$) shows that the *imaginary* part of $f(\tau)$, $v(r, \theta) = 1 - \theta/\pi$, satisfies Laplace's equation and the boundary conditions. Thus, constructing the solution is as follows: For a given x and y, we use our MATLAB code to compute τ. Substituting that τ into Equation 1.11.3 we compute $f(\tau)$. Taking the imaginary part, we have the solution at x and y. Figure 1.11.3 illustrates the solution for the domain $0 < x < 2$ and $-1 < y < 1$. □

In summary, conformal mapping allowed us to transform the original domain into one (an upper half-plane) where we could construct another analytic function whose imaginary part satisfied Laplace's equation and the boundary conditions. A natural question is what do we do if we cannot find this analytic function in the τ-plane? The next example shows an alternative approach.

• Example 1.11.2

For our second example of conformal mapping, consider $\tau = \sqrt{z^2 + a^2}$. To illustrate this mapping we have constructed two Argand diagrams; one is for the z-plane while the second is for the τ-plane. Figure 1.11.4 shows how a particular boundary in the z-plane maps into the τ-plane. The advantage here is that the infinitely thin filament or peg located at $z = 0$ is completely eliminated in the τ-plane.

One source of concern is the presence of the square root; for any value of z we would have two possible solutions. We make the mapping unique by requiring that $\Im(\tau) \geq 0$.

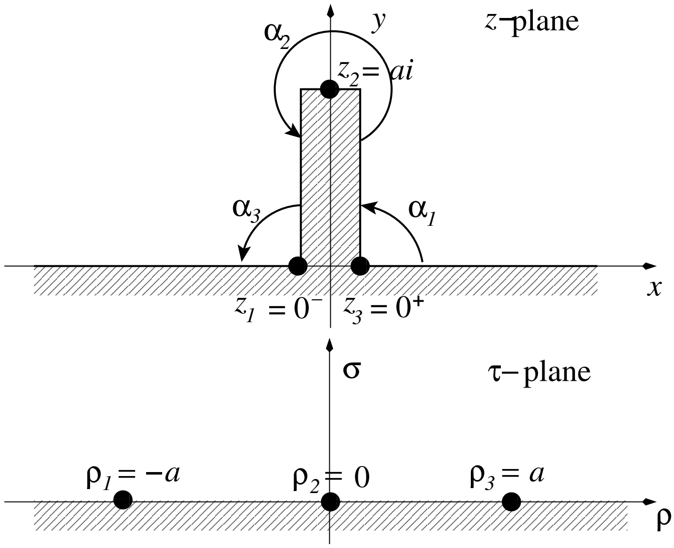


Figure 1.11.4: The conformal mapping between the z-plane and τ-plane achieved by the conformal mapping $\tau = \sqrt{z^2 + a^2}$.

To better understand this transformation, Figure 1.11.5 illustrates various lines of constant $\Re(\tau/a)$ and $\Im(\tau/a)$ as a function of x/a and y/a. This figure was constructed using the MATLAB code:

```
clear;

% compute τ for various values of z

for jj = 1:40
  y = 0.05 * jj;
for ii = 1:42
  x = 0.05 * (ii-21.5); z = x + i*y; tau(ii,jj) = sqrt(z*z+a*a);
  if (imag(tau(ii,jj)) <= 0) tau(ii,jj) = -tau(ii,jj); end
  X(ii,jj) = x; Y(ii,jj) = y;
  IM(ii,jj) = imag(tau(ii,jj)); REAL(ii,jj) = real(tau(ii,jj));
end; end

% plot the conformal mapping Equation τ = √(z^2 + a^2)

figure
[C,h] = contour(X,Y,IM,[0.1,0.25,0.5,0.75,1,1.5,2],'k');
clabel(C,'FontSize',10,'Color','k','Rotation',0)
xlabel('x','FontSize',20); ylabel('y','FontSize',20);
hold on
v = [-1,-0.5,-0.25,-0.01,0.01,0.25,0.5,1];
[C,h] = contour(X,Y,REAL,v,'--b');
clabel(C,'manual','FontSize',10,'Color','b','Rotation',0)
```

As $y \to \infty$, lines of constant $\Im(\tau/a)$ become parallel to the boundary $y = 0$. Only for smaller values of y, and as we approach the peg at $x = 0$, do these lines deviate strongly

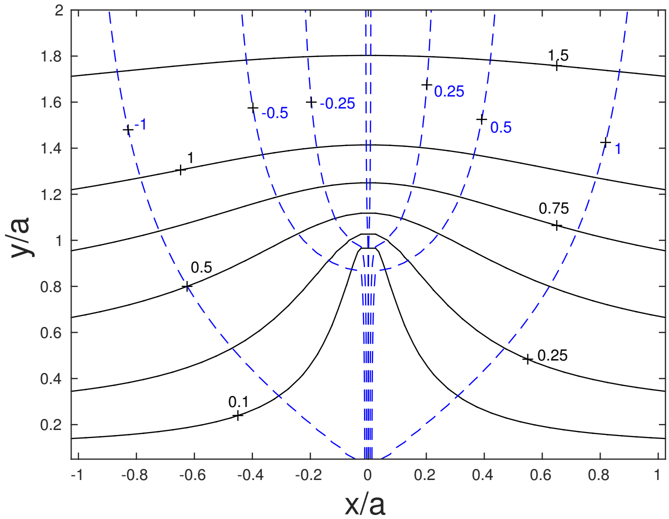


Figure 1.11.5: Lines of constant $\Re(\tau/a)$ (dashed line) and $\Im(\tau/a)$ (solid lines) as a function of x and y for the conformal mapping $\tau = \sqrt{z^2 + a^2}$.

from the horizontal as they pass over the obstacle. The smaller the value of $\Im(\tau/a)$ the more they conform to the shape of the obstacle.

The behavior of lines of constant $\Re(\tau/a)$ are more difficult to understand. There are two general classes, depending upon whether the absolute value of $\Re(\tau/a)$ is less or greater than 1. When $|\Re(\tau/a)| > 1$ they are clearly orthogonal to constant lines of $\Im(\tau/a)$. Positive values of $\Re(\tau/a)$ exist for $x > 0$ while negative values occur when $x < 0$. $|\Re(\tau/a)| < 1$ for $y \geq a$.

This example has two interesting aspects to it. The first is the presence of the square root. The second involves how we will find the solution to Laplace's equation in the τ-plane.

Let us assume that in the original z-plane the solution equals zero along the entire boundary except along the "peg." There, the solution equals 1. In the τ-plane the solution equals zero along the entire boundary *except* for the segment $-a < \rho < a$, where $\sigma = 0$, along which the solution equals 1. Instead of finding an analytic function whose real or imaginary part satisfies this boundary condition, we employ *Poisson's integral formula*[16] for the half-plane $y > 0$ or *Schwarz integral formula*:[17]

$$u(x, y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y f(t)}{(x-t)^2 + y^2}\, dt.$$

In the present case, we find that

$$u(\rho, \sigma) = \frac{1}{\pi} \int_{-a}^{a} \frac{\sigma}{\sigma^2 + (\xi - \rho)^2}\, d\xi \tag{1.11.4}$$

$$= \frac{1}{\pi} \left[\tan^{-1}\left(\frac{a - \rho}{\sigma} \right) + \tan^{-1}\left(\frac{a + \rho}{\sigma} \right) \right]. \tag{1.11.5}$$

---

[16] Poisson, S. D., 1823: Suite du mémoire sur les intégrales définies et sur la sommation des séries. *J. École Polytech.*, **19**, 404–509. See pg. 462.

[17] Schwarz, H. A., 1870: Über die Integration der partiellen Differentialgleichung $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$ für die Fläche eines Kreises. *Vierteljahrsschr. Naturforsch. Ges. Zürich*, **15**, 113–128.

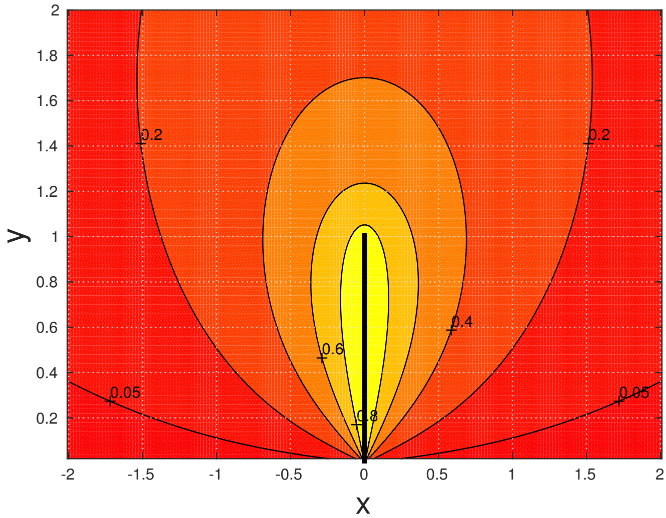


Figure 1.11.6: The solution of Laplace's equation when the solution (potential) along the boundary equals zero except along the peg located at $x = 0$. There the solution (potential) equals one.

Given Equation 1.11.5 we can compute the solution as follows: For a specific value of x and y, we find the corresponding value of ρ and σ. Equation 1.11.5 gives us the solution to Laplace's equation at that point and the corresponding x and y. The MATLAB code is:

```
clear; a = 1;

for jj = 1:100
  y = 0.02 * jj;
for ii = 1:202
  x = 0.02 * (ii-101.5); z = x + i*y; tau = sqrt(z*z+a*a);
  if (imag(tau) <= 0) tau = -tau; end
  sigma = imag(tau); rho = real(tau);
  X(ii,jj) = x; Y(ii,jj) = y;
% Equation 1.11.5
  arg1 = (a-rho)/sigma; arg2 = (a+rho)/sigma;
  T(ii,jj) = (atan(arg1)+atan(arg2)) / pi;
end; end

% plot the solution to Laplace's equation

figure
[C,h] = contourf(X,Y,T,[0,0.05,0.2,0.4,0.6,0.8],'k');
colormap autumn
clabel(C,'FontSize',10,'Color','k','Rotation',0)
xlabel('x','FontSize',20); ylabel('y','FontSize',20);
```

Figure 1.11.6 illustrates this solution. □

So far we have not presented a strategy for finding our conformal mappings. One method would be to simply experiment with transforms that had been used in similar problems. Fortunately, during the 1860s, two German mathematicians, E. B. Christoffel[18] (1829–1900) and H. A. Schwarz[19] (1843–1921), developed a very popular method of mapping a polygon into a half plane. Example 1.11.1 illustrated one of their transforms. Indeed, if we imagine that the boundary of the polygon is constructed from a thin wire, the purpose of the Schwarz-Christoffel transformation is to unbend the corners so that the wire becomes straight.

Our derivation begins by considering a mapping $z = f(\tau)$ where

$$\frac{dz}{d\tau} = C(\tau - \rho_1)^{k_1}(\tau - \rho_2)^{k_2} \cdots (\tau - \rho_n)^{k_n}, \qquad \textbf{(1.11.6)}$$

and $\rho_1, \rho_2, \ldots, \rho_n$ are any n points arranged in order along the real axis in the τ-plane such that $\rho_1 < \rho_2 < \ldots < \rho_n$. Here the k_i's are real constants and C is a real or complex constant. By taking the logarithm of both sides of Equation 1.11.6 we find that

$$\log\left(\frac{dz}{d\tau}\right) = \log(C) + k_1 \log(\tau - \rho_1) + k_2 \log(\tau - \rho_2) + \cdots + k_n \log(\tau - \rho_n). \qquad \textbf{(1.11.7)}$$

We have assumed that the principal value[20] of each logarithm is taken. The local magnification factor of the mapping from the τ-plane to the z-plane equals $dz/d\tau$, while the angle of $dz/d\tau$ gives the angle through which a small portion of the mapped curve in the τ-plane is rotated by the mapping. This angle is given by

$$\measuredangle\left(\frac{dz}{d\tau}\right) = \measuredangle(C) + k_1 \measuredangle(\tau - \rho_1) + k_2 \measuredangle(\tau - \rho_2) + \cdots + k_n \measuredangle(\tau - \rho_n). \qquad \textbf{(1.11.8)}$$

Equation 1.11.8 follows by first taking the imaginary part of Equation 1.11.7 and then noting that $\measuredangle(C) = \Im[\log(C)]$.

Let the point $(\rho, \sigma) = (-\infty, 0)$ in the τ-plane be mapped into the point z^* in the z-plane. See Figure 1.11.7. If we consider the image of a point ρ as it moves to the right along the negative real axis in the τ-plane, then all of the $\rho - \rho_i$ are real and negative as long as $\rho < \rho_1$. Hence the angles for all of the $\rho - \rho_i$ are constant and equal to π in Equation 1.11.8. Therefore, this equation simplifies to

$$\measuredangle\left(\frac{dz}{d\tau}\right) = \measuredangle(C) + (k_1 + k_2 + \cdots + k_n)\pi. \qquad \textbf{(1.11.9)}$$

Thus the portion of the ρ axis to the left of the point ρ_1 is mapped into a straight line segment, making the angle defined by Equation 1.11.9 with the real axis in the z-plane, and extending from z^* to z_1 the image of $\rho - \rho_1$.

Now as the point ρ crosses the point ρ_1 on the real axis, the real number $\rho - \rho_1$ becomes positive so that its angle abruptly changes from π to 0. Hence $\measuredangle(dz/d\tau)$ abruptly decreases by an amount $k_1\pi$ and then remains constant as τ travels from ρ_1 to ρ_2. It follows that

---

[18] Christoffel, E. B., 1868: Sul problema delle temperature stazionarie e la rappresentazione di una data superficie. *Ann. Mat. Pura Appl., Series 2*, **1**, 89–103; Christoffel, E. B., 1870: Sopra un problema proposto da Dirichlet. *Ann. Mat. Pura Appl., Series 2*, **4**, 1–9.

[19] Schwarz, H. A., 1868: Über einige Abbildungsaufgaben. *J. Reine Angew. Math.*, **70**, 105–120.

[20] For the complex number $z = re^{\theta i}$, $r \neq 0$, the principal value of the logarithm is $\log(z) = \ln(r) + \theta i$, where θ must lie between 0 and 2π.

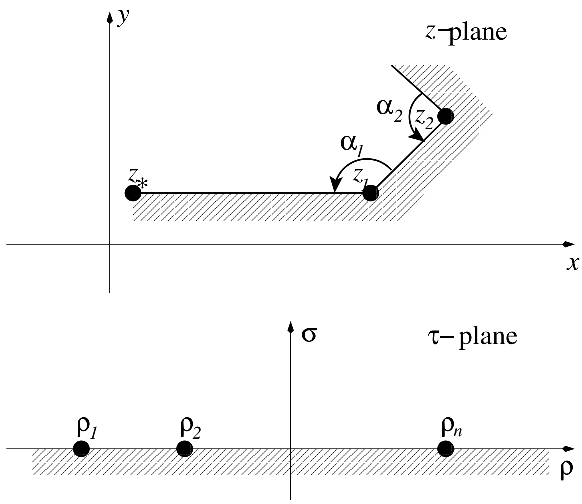


Figure 1.11.7: Diagram used in the derivation of the Schwarz-Christoffel method.

the image of the segment $(\rho_1\rho_2)$ in the z-plane makes an angle of $-k_1\pi$ with the segment (z^*z_1).

Proceeding in this way, we see that each segment (ρ_n, ρ_{n+1}) is mapped into a line segment (z_n, z_{n+1}) in the z-plane, making the angle of $-k_n\pi$ with the segment previously mapped. Thus, if the interior angle of the resultant polynomial contour at the point z_n is to have the magnitude α_n, we must set $\pi - \alpha_n = -k_n\pi$, or $k_n = \alpha_n/\pi - 1$ in Equation 1.11.6. After an integration, we then conclude that the mapping

$$z = C \int^{\tau} (\eta - \rho_1)^{k_1} (\eta - \rho_2)^{k_2} \cdots (\eta - \rho_n)^{k_n}\, d\eta + K, \qquad \textbf{(1.11.10)}$$

where the arbitrary complex constants C and K map the real axis $\sigma = 0$ of the τ-plane into a polynomial boundary in the z-plane in such a way that the vertices $z_1, z_2, \ldots, z_n$ with interior angles $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the images of the points $\rho_1, \rho_2, \ldots, \rho_n$.

For the final segment $\tau - \rho > \rho_n$ the numbers $\tau - \rho_i$ are all real, positive, and equal to zero, so that this segment is rotated through the angle

$$\measuredangle(dz/d\tau) = \measuredangle(C), \qquad \rho > \rho_n. \qquad \textbf{(1.11.11)}$$

For a closed polynomial the sum of the interior angles is

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = (n-2)\pi. \qquad \textbf{(1.11.12)}$$

Therefore,

$$k_1 + k_2 + \cdots + k_n = \frac{(n-2)\pi}{\pi} - n = -2. \qquad \textbf{(1.11.13)}$$

Thus, according to Equations 1.11.8 and 1.11.11, the two infinite segments of the line $\sigma = 0$ are rotated through the angle $\measuredangle(C) - 2\pi$ and $\measuredangle(C)$, as is clearly necessary for a closed figure.

What roles do C and K play? Because C is often complex, this constant introduces any necessary magnification and rotation of the transformation so that *any* prescribed polynomial in the z-plane is made to correspond point by point to the real axis $\sigma = 0$ in

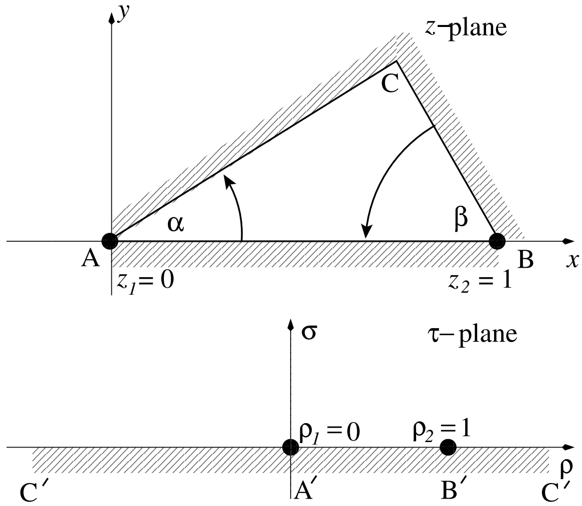


Figure 1.11.8: The complex z- and τ-planes used in Example 1.11.4.

the τ-plane. In fact, this correspondence can be set up in infinitely many ways, in that three of the numbers $\rho_1, \rho_2, \ldots, \rho_n$ can be determined arbitrarily. Finally, the mapping can be shown to establish a one-to-one correspondence between points in the interior of the polygon in the z-plane and points in the *upper half* of the τ-plane.

• Example 1.11.3

Let us derive the conformal mapping used in Example 1.11.2. Referring back to Figure 1.11.4, we see that $\alpha_1 = \pi/2$, $k_1 = -1/2$, and $\rho_1 = -a$ at $z_1 = 0^-$; $\alpha_2 = 2\pi$, $k_2 = 1$, and $\rho_2 = 0$ at $z_2 = ai$; and $\alpha_3 = \pi/2$, $k_3 = -1/2$, and $\rho_3 = a$ at $z_3 = 0^+$. Therefore, from Equation 1.11.6,

$$\frac{dz}{d\tau} = C(\tau + a)^{-1/2}\tau(\tau - a)^{-1/2} = C\frac{\tau}{\sqrt{\tau^2 - a^2}}. \tag{1.11.14}$$

Integrating this differential equation,

$$z = C\sqrt{\tau^2 - a^2} + K. \tag{1.11.15}$$

Because the point $\rho_1 = -a$ corresponds to $z = 0^-$, $K = 0$. Similarly, at $\rho_2 = 0$, we have that

$$ai = C\sqrt{-a^2}, \qquad \text{or} \qquad C = 1. \tag{1.11.16}$$

Therefore, the conformal mapping is given by $z = \sqrt{\tau^2 - a^2}$, or $\tau = \sqrt{z^2 + a^2}$. □

• Example 1.11.4

Consider the triangle ABC located in the z-plane as shown on Figure 1.11.8. Here we desire to map the *interior* space of this triangle into the upper half of the τ-plane. At point C, points along the boundary and to the left of C are to be mapped out to $-\infty$ in the τ-plane while points along the boundary and to the right of C are mapped to $+\infty$.

From Equation 1.11.6 we have that

$$\frac{dz}{d\tau} = C'\tau^{\alpha/\pi-1}(\tau - 1)^{\beta/\pi-1} = C\tau^{\alpha/\pi-1}(1 - \tau)^{\beta/\pi-1}. \tag{1.11.17}$$

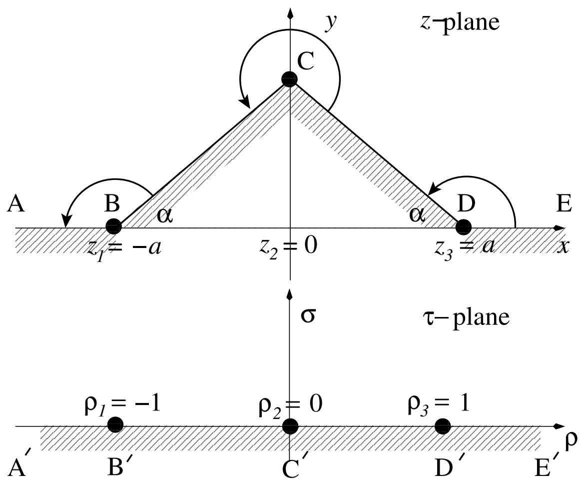


Figure 1.11.9: The complex z- and τ-planes used in Example 1.11.5.

Integrating this differential equation,

$$z = C \int^{\tau} \eta^{\alpha/\pi - 1}(1-\eta)^{\beta/\pi - 1}\, d\eta + K. \tag{1.11.18}$$

Because we want the points $\tau = 0$ and $z = 0$ to correspond to each other, $K = 0$. On the other hand, if we wish $\tau = 1$ and $z = 1$ to correspond, Equation 1.11.18 yields

$$C \int_0^1 \eta^{\alpha/\pi - 1}(1-\eta)^{\beta/\pi - 1}\, d\eta = C\, \frac{\Gamma(\alpha/\pi)\Gamma(\beta/\pi)}{\Gamma[(\alpha+\beta)/\pi]} = 1, \tag{1.11.19}$$

where $\Gamma(\cdot)$ is the gamma function defined by

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t}\, dt. \tag{1.11.20}$$

Consequently,

$$C = \frac{\Gamma[(\alpha+\beta)/\pi]}{\Gamma(\alpha/\pi)\Gamma(\beta/\pi)}, \tag{1.11.21}$$

and

$$z = \frac{\Gamma[(\alpha+\beta)/\pi]}{\Gamma(\alpha/\pi)\Gamma(\beta/\pi)} \int_0^{\tau} \eta^{\alpha/\pi - 1}(1-\eta)^{\beta/\pi - 1}\, d\eta. \tag{1.11.22}$$

A noteworthy aspect of this example is that the conformal mapping is given by an integral and not some analytic expression. □

• Example 1.11.5

Consider the domain lying in the upper half of the z-plane except for a triangular section BCD shown in Figure 1.11.9. We wish to construct the Schwarz-Christoffel transformation that maps this domain into the upper half of the τ-plane. From Equation 1.11.6 we have that

$$\frac{dz}{d\tau} = C'(\tau+1)^{(\pi-\alpha)/\pi - 1}\tau^{(\pi+2\alpha)/\pi - 1}(\tau-1)^{(\pi-\alpha)/\pi - 1} \tag{1.11.23}$$

$$= C' \frac{\tau^{2\alpha/\pi}}{(\tau^2-1)^{\alpha/\pi}} = C \frac{\tau^{2\alpha/\pi}}{(1-\tau^2)^{\alpha/\pi}}. \tag{1.11.24}$$

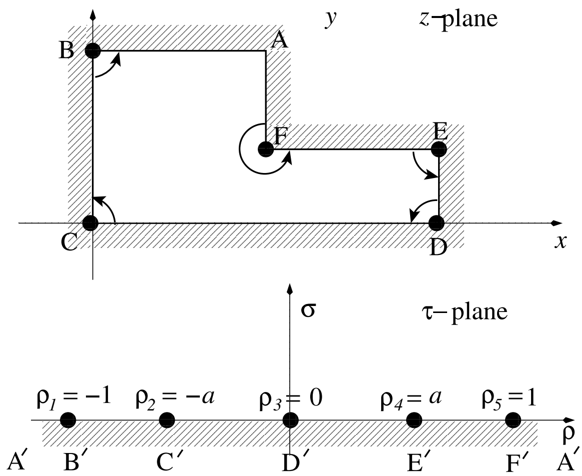


Figure 1.11.10: The complex z- and τ-planes used in Example 1.11.6 with $a < 1$.

Integrating this differential equation,

$$z = C \int_0^\tau \frac{\eta^{2\alpha/\pi}}{(1-\eta^2)^{\alpha/\pi}}\, d\eta + K. \tag{1.11.25}$$

If we want the point $\tau = 0$ to correspond to the point $z = ki$, then $K = ki$. On the other hand, if the point $\tau = 1$ corresponds to $z = a$, then

$$a = C \int_0^1 \frac{\eta^{2\alpha/\pi}}{(1-\eta^2)^{\alpha/\pi}}\, d\eta + ki. \tag{1.11.26}$$

Solving for C,

$$C = \frac{\sqrt{\pi}(a - ki)}{\Gamma\left(\alpha/\pi + \frac{1}{2}\right)\Gamma(1-\alpha/\pi)}. \tag{1.11.27}$$

Therefore, the final answer is

$$z = \frac{\sqrt{\pi}(a - ki)}{\Gamma\left(\alpha/\pi + \frac{1}{2}\right)\Gamma(1-\alpha/\pi)} \int_0^\tau \frac{\eta^{2\alpha/\pi}}{(1-\eta^2)^{\alpha/\pi}}\, d\eta + ki. \tag{1.11.28}$$

□

• Example 1.11.6

Consider the domain within the L-shaped boundary shown in Figure 1.11.10. We wish to construct the Schwarz-Christoffel transform that maps the interior into the upper half of the τ-plane. Note that we broke the boundary in such a manner that points slightly to the left of point A are mapped to $-\infty$ while points slightly below the point A are mapped to $+\infty$.

Because $a < 1$, Equation 1.11.6 gives

$$\frac{dz}{d\tau} = C(\tau+1)^{-1/2}(\tau+a)^{-1/2}\tau^{-1/2}(\tau-a)^{-1/2}(\tau-1)^{1/2}. \tag{1.11.29}$$

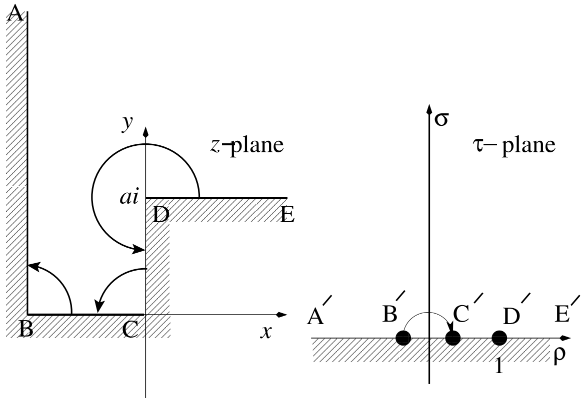


Figure 1.11.11: The complex z- and τ-planes used in Example 1.11.7.

Integrating this differential equation,

$$z = C\int_0^\tau \frac{(\eta-1)\,d\eta}{\eta\sqrt{(\eta^2-1)(\eta^2-a^2)}} + K = \frac{C}{a}\int_0^\tau \frac{(\eta-1)\,d\eta}{\eta\sqrt{(1-\eta^2)(1-p^2\eta^2)}} + K, \qquad (\mathbf{1.11.30})$$

where $p^2 = 1/a^2$. To compute C and K, we would need further information. □

• **Example 1.11.7**

Let us derive the conformal mapping, Equation 1.11.2, used in Example 1.11.1. The $z-$ and $\tau-$planes are shown in Figure 1.11.11. From this figure we see that $\alpha_1 = 3\pi/2$, $\alpha_2 = \pi/2$, $\alpha_3 = \pi/2$, $\rho_1 = 1$, $\rho_2 = 0^-$, and $\rho_3 = 0^+$. This yields

$$\frac{dz}{d\tau} = K(\tau-1)^{(3\pi)/(2\pi)-1}(\tau-0^-)^{(\pi)/(2\pi)-1}(\tau-0^+)^{(\pi)/(2\pi)-1} = K\frac{\sqrt{\tau-1}}{\tau}. \qquad (\mathbf{1.11.31})$$

Integrating Equation 1.11.31, we find that

$$z = 2K\left[\sqrt{\tau-1} - \arctan\left(\sqrt{\tau-1}\right)\right] + C = 2K\left[\sqrt{\tau-1} + \frac{i}{2}\log\left(\frac{1+i\sqrt{\tau-1}}{1-i\sqrt{\tau-1}}\right)\right] + C. \qquad (\mathbf{1.11.32})$$

Because at $\tau = 1$, $z = a/2$, we have $C = a/2$.

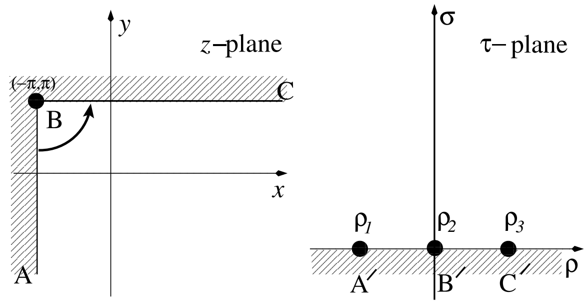
The computation of K is more complicated. Referring to Figure 1.11.11, we note that

$$\int_B^C dz = \int_{B'}^{C'} K\frac{\sqrt{\tau-1}}{\tau}\,d\tau. \qquad (\mathbf{1.11.33})$$

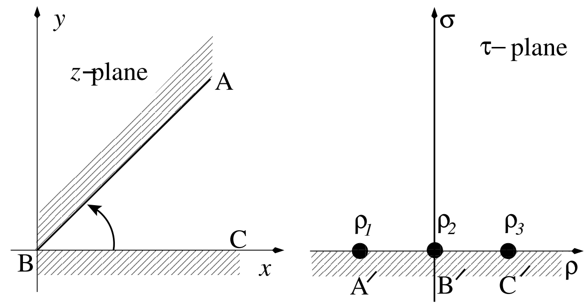
Setting $\tau = r\,e^{\theta i}$ with $r \to 0$, Equation 1.11.34 becomes

$$\frac{a}{2} = K\lim_{r\to 0}\int_\pi^0 \frac{\sqrt{r\,e^{\theta i}-1}}{r\,e^{\theta i}}\,ir\,e^{\theta i}\,d\theta = K\pi. \qquad (\mathbf{1.11.34})$$

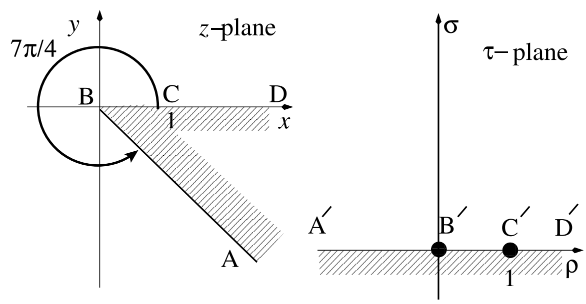
Thus $K = a/(2\pi)$ and we recover Equation 1.11.2.



Problem 3



Problem 4



Problem 5

Problems

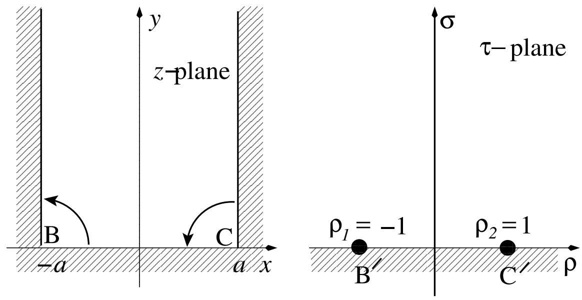
1. Verify that the function $\tau = e^z$ maps the strip $0 < \Im(z) < \pi$ into the half-plane $\Im(\tau) > 0$.

2. Verify that the function $\tau^2 = 1 - e^z$ maps the strip $-\pi < \Im(z) < \pi$, except for the negative real axis, into the upper half of the τ-plane.

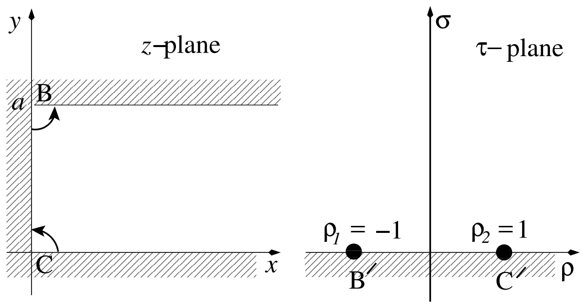
3. Use the Schwarz-Christoffel method to find the conformal mapping that maps the quarter plane $x > -\pi$, $y < \pi$ into the upper half of the τ-plane. We require that the point $(-\pi, \pi)$ in the z-plane maps to the point $(0, 0)$ in the τ-plane.

4. Use the Schwarz-Christoffel method to find the conformal mapping that maps the sector lying between the x-axis and the line $\theta = \pi/3$ into the upper half of the τ-plane. We require that the point $(0, 0)$ in the z-plane maps to the point $(0, 0)$ in the τ-plane.

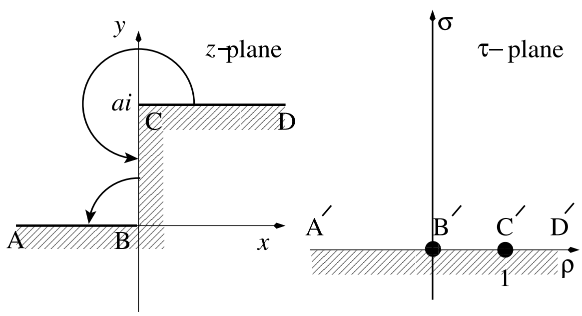
5. Use the Schwarz-Christoffel method to find the conformal mapping that maps the portion of the z-plane defined by $0 < r < \infty$, $0 < \theta < 7\pi/4$ into the upper half of the τ-plane. We



Problem 6



Problem 7



Problem 8

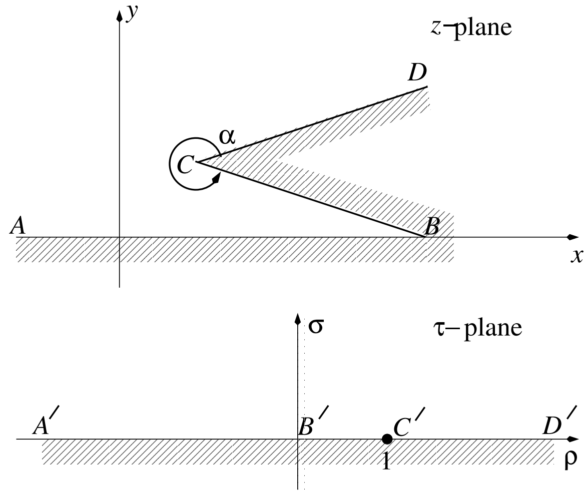
require that the points $(0, 0)$ and $(1, 0)$ in the z-plane map to the points $(0, 0)$ and $(1, 0)$ in the τ-plane, respectively.

6. Use the Schwarz-Christoffel method to find the conformal mapping that maps the domain $|x| < a$, $0 < y$ into the upper half of the τ-plane. Let the point $(-a, 0)$ become the point $(-1, 0)$ while the point $(a, 0)$ becomes the point $(1, 0)$.

7. Use the Schwarz-Christoffel method to find the conformal mapping that maps the region $x > 0$, $0 < y < a$ into the upper half of the τ-plane. We require that the point $(0, a)$ maps to $(-1, 0)$ in the τ-plane while the point $(0, 0)$ maps to $(1, 0)$ in the τ-plane.

8. Use the Schwarz-Christoffel method to find the conformal mapping that maps the region shown in the figure into the upper half of the τ-plane. We require that the points $(0, 0)$ and $(0, a)$ in the z-plane map to the points $(0, 0)$ and $(1, 0)$ in the τ-plane, respectively.

9. Construct a transform between a z-plane which has a barrier that runs parallel to the x-axis from $z = L + \pi L\, i$ to $\infty + \pi L\, i$ and a τ-plane that has no barrier.



Problem 9a

Step 1: Begin by using the Schwarz-Christoffel method to show that the conformal mapping pictured in Figure 9a is given by

$$\frac{dz}{d\tau} = C\tau^{k_1}(\tau - 1)^{k_2},$$

where $k_1 = -\alpha/(2\pi)$ and $k_2 = \alpha/\pi - 1$.

Step 2: Next, consider the limit as the points B and D in the z-plane in Figure 9a move out to infinity (so that $\alpha \to 2\pi$) and we obtain Figure 9b. Consequently, the transform approaches

$$\frac{dz}{d\tau} = C\frac{\tau - 1}{\tau},$$

or

$$z = C[\tau - \log(\tau)] + K.$$

Here we have taken the principal branch of the logarithm so that $\log(z) = \ln(|z|) + i\theta$ where $0 \le \theta \le \pi$. (We do not require that $0 \le \theta < 2\pi$ because we are always in the upper half-plane.)

Step 3: Following Example 1.11.7, consider the area around $\tau = 0$. Show that

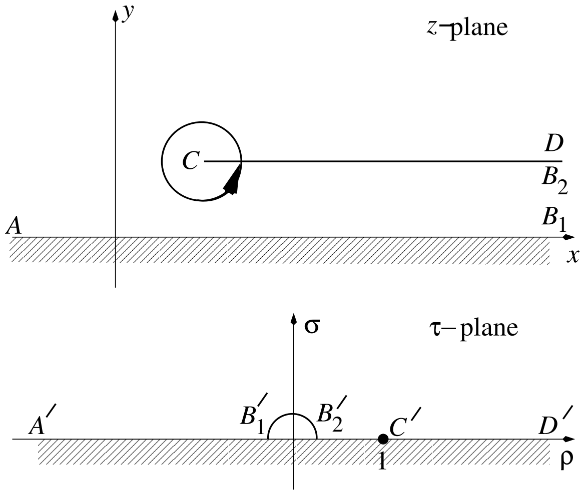
$$dz \approx -C\frac{d\tau}{\tau} = -iC\,d\theta,$$

where $\tau = r\,e^{\theta i}$. Integrating from point B_1' to point B_2', show that $C = L$.

Step 4: To compute K, note that if the point C, located at $z = L + \pi Li$, corresponds to the point C', located at $\tau = 1$, then $K = \pi Li$.

10. Use conformal mapping to solve Laplace's equation for the infinite strip $-\infty < x < \infty$, $0 \le y \le \pi$. The solution equals zero everywhere along the boundary *except* for $x > 0$, $y = 0$, where $u(x, 0) = 1$.

Step 1: Consider the mapping $\tau = e^z$. Show that $\rho = e^x \cos(y)$ and $\sigma = e^x \sin(y)$. In particular, $(\infty, \pi) \to (-\infty, 0)$, $(0, \pi) \to (-1, 0)$, $(-\infty, y) \to (0, 0)$, $(0, 0) \to (1, 0)$, and $(\infty, 0) \to (\infty, 0)$.



Problem 9b

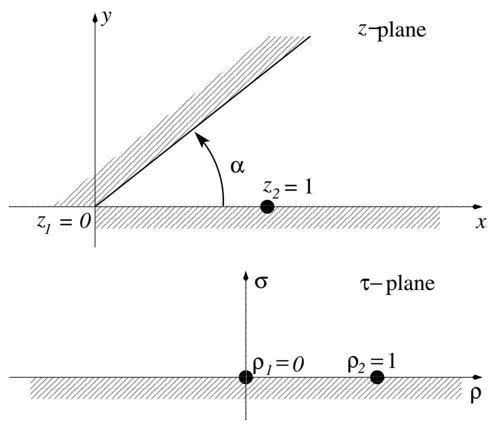


Figure 1.11.12: The conformal mapping between the z-plane and τ-plane achieved by the conformal mapping $\tau = z^{\pi/\alpha}$.

Step 2: Using Poisson's integral formula for the upper half-plane, show that

$$u(\rho, \sigma) = \frac{1}{\pi}\left[\frac{\pi}{2} - \tan^{-1}\left(\frac{1-x}{y}\right)\right] = 1 - \frac{1}{\pi}\tan^{-1}\left(\frac{y}{x-1}\right).$$

Step 3: Show that

$$u(x, y) = 1 - \frac{1}{\pi}\tan^{-1}\left[\frac{e^x \sin(y)}{e^x \cos(y) - 1}\right].$$

11. Use conformal mapping to solve Laplace's equation for a pie-shaped sector in the first quadrant. See Figure 1.11.12. The solution equals zero along the entire boundary except for $0 < x < 1$ where it equals one.

Step 1: Show that the mapping $z = \tau^{\alpha/\pi}$ or $\tau = z^{\pi/\alpha}$ maps the pie-shaped sector into the half-plane $\Im(\tau) > 0$. See Figure 1.11.12.

Step 2: Using Poisson's integral formula for the upper half-plane, show that

$$u(\rho,\sigma) = \frac{1}{\pi}\cot^{-1}\left(\frac{\rho^2+\sigma^2-\rho}{\sigma}\right).$$

Step 3: Show that

$$u(r,\theta) = \frac{1}{\pi}\cot^{-1}\left[\frac{r^{\pi/\alpha}-\cos(\pi\theta/\alpha)}{\sin(\pi\theta/\alpha)}\right],$$

where $x = r\cos(\theta)$ and $y = r\sin(\theta)$.

12. Use conformal mapping to solve Laplace's equation for the semi-infinite strip $0 \le x \le a$, $0 \le y < \infty$, where $u(x,0) = 1$, $0 \le x \le a$, and $u(0,y) = u(a,y) = 0$, $0 \le y < \infty$.

Step 1: Consider the mapping $\tau = -\cos(\pi z/a)$. Show that

$$\rho = -\cos(\pi x/a)\cosh(\pi y/a), \quad \text{and} \quad \sigma = \sin(\pi x/a)\sinh(\pi y/a).$$

In particular, $(0,\infty) \to (-\infty,0)$, $(0,0) \to (-1,0)$, $(a/2,0) \to (0,0)$, $(a,0) \to (1,0)$, and $(a,\infty) \to (\infty,0)$.

Step 2: Using Poisson's integral formula for the upper half-plane, show that

$$u(\rho,\sigma) = \frac{1}{\pi}\cot^{-1}\left(\frac{\rho^2+\sigma^2-1}{2\sigma}\right).$$

Step 3: Show that

$$\begin{aligned} u(x,y) &= \frac{1}{\pi}\cot^{-1}\left\{\left[\sinh^2\left(\frac{\pi y}{a}\right) - \sin^2\left(\frac{\pi x}{a}\right)\right] \Big/ 2\sin\left(\frac{\pi x}{a}\right)\sinh\left(\frac{\pi y}{a}\right)\right\} \\ &= \frac{2}{\pi}\tan^{-1}\left[\sin\left(\frac{\pi x}{a}\right) \Big/ \sinh\left(\frac{\pi y}{a}\right)\right]. \end{aligned}$$

Step 4: In the case that boundary conditions read $u(0,y) = u(a,y) = 1$ for $0 \le y < \infty$ and $u(x,0) = 0$ for $0 \le x \le a$, how could you use the solution in Step 3 to solve this new problem?

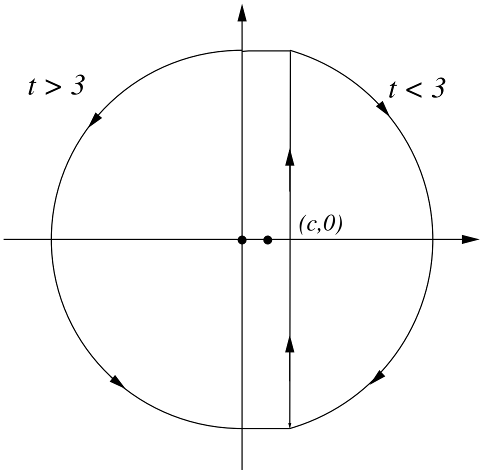
Further Readings

Ablowitz, M. J., and A. S. Fokas, 2003: *Complex Variables: Introduction and Applications*. Cambridge University Press, 660 pp. Covers a wide variety of topics, including complex numbers, analytic functions, singularities, conformal mapping and the Riemann-Hilbert problem.

Carrier, G. F., M. Krook, and C. E. Pearson, 1966: *Functions of a Complex Variable: Theory and Technique*. McGraw-Hill Book Co., 438 pp. Graduate-level textbook.

Churchill, R. V., 1960: *Complex Variables and Applications*. McGraw-Hill Book Co., 297 pp. Classic textbook.

Flanigan, F. J., 1983: *Complex Variables*. Dover, 364 pp. A crystal clear exposition and emphasis on an intuitive understanding of complex analysis.



Chapter 2
Advanced Transform Methods

In their course work, most engineering students are introduced to the concept of the Fourier and Laplace transforms. The presentations are limited because the student has not studied complex variables. Having presented this topic in the previous chapter, the reader is ready to deepen his/her ability to use these transform methods.

This chapter deals with two important aspects of transform methods. In the past you may have inverted Fourier and Laplace transforms using partial fractions, tables and some general properties of the transform. Often these techniques fail and here we show how the power of complex variables can overcome these difficulties.

The reason that Laplace transforms are taught to engineers is their ability to solve ordinary differential equations. When it comes to partial differential equations the student is only taught one method: separation of variables. In Sections 2.4 through 2.6 we show how Laplace transforms can be used to solve the wave, heat, and Laplace equations.

2.1 INVERSION OF FOURIER TRANSFORMS BY CONTOUR INTEGRATION

Although we may find the inverse by direct integration or partial fractions, in many instances the Fourier transform does not lend itself to these techniques. On the other hand, if we view the inverse Fourier transform as a line integral along the real axis in the complex ω-plane, then some of the techniques that we developed in Chapter 1 can be applied to this problem. To this end, we rewrite the inversion integral for the Fourier transform as

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{it\omega}\,d\omega = \frac{1}{2\pi}\oint_C F(z)e^{itz}\,dz - \frac{1}{2\pi}\int_{C_R} F(z)e^{itz}\,dz, \qquad (\mathbf{2.1.1})$$

where C denotes a closed contour consisting of the entire real axis plus a new contour C_R that joins the point $(\infty, 0)$ to $(-\infty, 0)$. There are countless possibilities for C_R. For example, it could be the loop $(\infty, 0)$ to (∞, R) to $(-\infty, R)$ to $(-\infty, 0)$ with $R > 0$. However, any choice of C_R must be such that we can compute $\int_{C_R} F(z)e^{itz}\,dz$. When we take that constraint into account, the number of acceptable contours decreases to just a few. The best is given by *Jordan's lemma*.[1]

Jordan's lemma: *Suppose that, on a circular arc C_R with radius R and center at the origin, $f(z) \to 0$ uniformly as $R \to \infty$. Then*

$$(1) \qquad \lim_{R\to\infty} \int_{C_R} f(z)e^{imz}\,dz = 0, \qquad (m > 0) \tag{2.1.2}$$

if C_R lies in the first and/or second quadrant;

$$(2) \qquad \lim_{R\to\infty} \int_{C_R} f(z)e^{-imz}\,dz = 0, \qquad (m > 0) \tag{2.1.3}$$

if C_R lies in the third and/or fourth quadrant;

$$(3) \qquad \lim_{R\to\infty} \int_{C_R} f(z)e^{mz}\,dz = 0, \qquad (m > 0) \tag{2.1.4}$$

if C_R lies in the second and/or third quadrant; and

$$(4) \qquad \lim_{R\to\infty} \int_{C_R} f(z)e^{-mz}\,dz = 0, \qquad (m > 0) \tag{2.1.5}$$

if C_R lies in the first and/or fourth quadrant.

Technically, only (1) is actually Jordan's lemma, while the remaining points are variations.

Proof: We shall prove the first part; the remaining portions follow by analog. We begin by noting that

$$|I_R| = \left| \int_{C_R} f(z)e^{imz}\,dz \right| \le \int_{C_R} |f(z)| \left|e^{imz}\right| |dz|. \tag{2.1.6}$$

Now

$$|dz| = R\,d\theta, \quad |f(z)| \le M_R, \tag{2.1.7}$$

$$\left|e^{imz}\right| = \left|\exp(imRe^{\theta i})\right| = |\exp\{imR[\cos(\theta) + i\sin(\theta)]\}| = e^{-mR\sin(\theta)}. \tag{2.1.8}$$

Therefore,

$$|I_R| \le RM_R \int_{\theta_0}^{\theta_1} \exp[-mR\sin(\theta)]\,d\theta, \tag{2.1.9}$$

[1] Jordan, C., 1894: *Cours D'Analyse de l'École Polytechnique. Vol. 2*. Gauthier-Villars, pp. 285–286. See also Whittaker, E. T., and G. N. Watson, 1963: *A Course of Modern Analysis*. Cambridge University Press, p. 115.

where $0 \leq \theta_0 < \theta_1 \leq \pi$. Because the integrand is positive, the right side of Equation 2.1.9 is largest if we take $\theta_0 = 0$ and $\theta_1 = \pi$. Then

$$|I_R| \leq RM_R \int_0^{\pi} e^{-mR\sin(\theta)}\,d\theta = 2RM_R \int_0^{\pi/2} e^{-mR\sin(\theta)}\,d\theta. \tag{2.1.10}$$

We cannot evaluate the integrals in Equation 2.1.10 as they stand. However, because $\sin(\theta) \geq 2\theta/\pi$ if $0 \leq \theta \leq \pi/2$, we can bound the value of the integral by

$$|I_R| \leq 2RM_R \int_0^{\pi/2} e^{-2mR\theta/\pi}\,d\theta = \frac{\pi}{m} M_R \left(1 - e^{-mR}\right). \tag{2.1.11}$$

If $m > 0$, $|I_R|$ tends to zero with M_R as $R \to \infty$. □

Consider now the following inversions of Fourier transforms:

- **Example 2.1.1**

For our first example we find the inverse for

$$F(\omega) = \frac{1}{\omega^2 - 2ib\omega - a^2 - b^2}, \qquad a, b > 0. \tag{2.1.12}$$

From the inversion integral,

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\omega}}{\omega^2 - 2ib\omega - a^2 - b^2}\,d\omega, \tag{2.1.13}$$

or

$$f(t) = \frac{1}{2\pi} \oint_C \frac{e^{itz}}{z^2 - 2ibz - a^2 - b^2}\,dz - \frac{1}{2\pi} \int_{C_R} \frac{e^{itz}}{z^2 - 2ibz - a^2 - b^2}\,dz, \tag{2.1.14}$$

where C denotes a closed contour consisting of the entire real axis plus C_R. Because $f(z) = 1/(z^2 - 2ibz - a^2 - b^2)$ tends to zero uniformly as $|z| \to \infty$ and $m = t$, the second integral in Equation 2.1.14 vanishes by Jordan's lemma if C_R is a semicircle of infinite radius in the upper half of the z-plane when $t > 0$ and a semicircle in the lower half of the z-plane when $t < 0$.

Next we must find the location and nature of the singularities. They are located at

$$z^2 - 2ibz - a^2 - b^2 = 0, \qquad \text{or} \qquad z = \pm a + bi. \tag{2.1.15}$$

Therefore we can rewrite Equation 2.1.14 as

$$f(t) = \frac{1}{2\pi} \oint_C \frac{e^{itz}}{(z - a - bi)(z + a - bi)}\,dz. \tag{2.1.16}$$

Thus, all of the singularities are simple poles.

Consider now $t > 0$. As stated earlier, we close the line integral with an infinite semicircle in the upper half-plane. See Figure 2.1.1. Inside this closed contour there are two singularities: $z = \pm a + bi$. For these poles,

$$\operatorname{Res}\left(\frac{e^{itz}}{z^2 - 2ibz - a^2 - b^2}; a + bi\right) = \lim_{z \to a+bi} \frac{(z - a - bi)e^{itz}}{(z - a - bi)(z + a - bi)} \tag{2.1.17}$$

$$= \frac{e^{iat}e^{-bt}}{2a} = \frac{e^{-bt}}{2a}[\cos(at) + i\sin(at)], \tag{2.1.18}$$

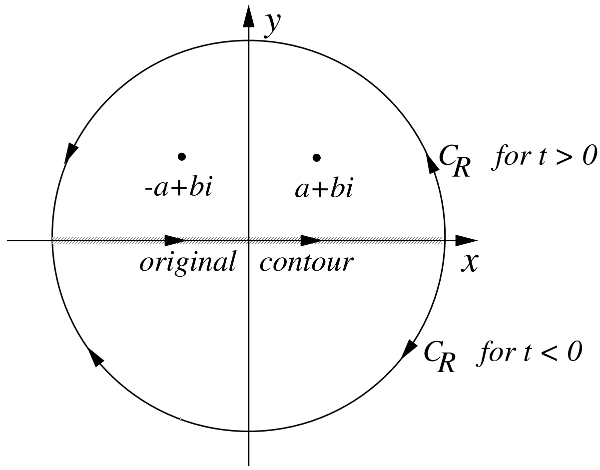


Figure 2.1.1: Contour used to find the inverse of the Fourier transform, Equation 2.1.12. The contour C consists of the line integral along the real axis plus C_R.

where we used Euler's formula to eliminate e^{iat}. Similarly,

$$\operatorname{Res}\left(\frac{e^{itz}}{z^2 - 2ibz - a^2 - b^2}; -a + bi\right) = -\frac{e^{-bt}}{2a}[\cos(at) - i\sin(at)]. \quad (\mathbf{2.1.19})$$

Consequently, the inverse Fourier transform follows from Equation 2.1.16 after applying the residue theorem, and equals

$$f(t) = -\frac{e^{-bt}}{2a}\sin(at) \quad (\mathbf{2.1.20})$$

for $t > 0$.

For $t < 0$, the semicircle is in the lower half-plane because the contribution from the semicircle vanishes as $R \to \infty$. Because there are no singularities within the closed contour, $f(t) = 0$. Therefore, we can write in general that

$$f(t) = -\frac{e^{-bt}}{2a}\sin(at)H(t). \quad (\mathbf{2.1.21})$$

□

• Example 2.1.2

Let us find the inverse of the Fourier transform

$$F(\omega) = \frac{e^{-\omega i}}{\omega^2 + a^2}, \quad (\mathbf{2.1.22})$$

where a is real and positive.

From the inversion integral,

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{i(t-1)\omega}}{\omega^2 + a^2}\,d\omega = \frac{1}{2\pi}\oint_C \frac{e^{i(t-1)z}}{z^2 + a^2}\,dz - \frac{1}{2\pi}\int_{C_R}\frac{e^{i(t-1)z}}{z^2 + a^2}\,dz, \quad (\mathbf{2.1.23})$$

where C denotes a closed contour consisting of the entire real axis plus C_R. The contour C_R is determined by Jordan's lemma because $1/(z^2+a^2) \to 0$ uniformly as $|z| \to \infty$. Since $m = t-1$, the semicircle C_R of infinite radius lies in the upper half-plane if $t > 1$ and in the lower half-plane if $t < 1$. Thus, if $t > 1$,

$$f(t) = \frac{1}{2\pi}(2\pi i)\text{Res}\left[\frac{e^{i(t-1)z}}{z^2+a^2}; ai\right] = \frac{e^{-a(t-1)}}{2a}, \qquad \textbf{(2.1.24)}$$

whereas for $t < 1$,

$$f(t) = \frac{1}{2\pi}(-2\pi i)\text{Res}\left[\frac{e^{i(t-1)z}}{z^2+a^2}; -ai\right] = \frac{e^{a(t-1)}}{2a}. \qquad \textbf{(2.1.25)}$$

The minus sign in front of the $2\pi i$ arises from the clockwise direction or negative sense of the contour. We can write the inverse as the single expression

$$f(t) = \frac{e^{-a|t-1|}}{2a}. \qquad \textbf{(2.1.26)}$$

□

• **Example 2.1.3**

Let us evaluate the integral

$$\int_{-\infty}^{\infty} \frac{\cos(kx)}{x^2+a^2}\,dx, \qquad \textbf{(2.1.27)}$$

where $a, k > 0$.

We begin by noting that

$$\int_{-\infty}^{\infty} \frac{\cos(kx)}{x^2+a^2}\,dx = \Re\left(\int_{-\infty}^{\infty} \frac{e^{ikx}}{x^2+a^2}\,dx\right) = \Re\left(\int_{C_1} \frac{e^{ikz}}{z^2+a^2}\,dz\right), \qquad \textbf{(2.1.28)}$$

where C_1 denotes a line integral along the real axis from $-\infty$ to ∞. A quick check shows that the integrand of the right side of Equation 2.1.28 satisfies Jordan's lemma. Therefore,

$$\int_{-\infty}^{\infty} \frac{e^{ikx}}{x^2+a^2}\,dx = \oint_C \frac{e^{ikz}}{z^2+a^2}\,dz = 2\pi i\ \text{Res}\left(\frac{e^{ikz}}{z^2+a^2}; ai\right) \qquad \textbf{(2.1.29)}$$

$$= 2\pi i \lim_{z\to ai} \frac{(z-ai)e^{ikz}}{z^2+a^2} = \frac{\pi}{a}e^{-ka}, \qquad \textbf{(2.1.30)}$$

where C denotes the closed infinite semicircle in the upper half-plane. Taking the real and imaginary parts of Equation 2.1.30,

$$\int_{-\infty}^{\infty} \frac{\cos(kx)}{x^2+a^2}\,dx = \frac{\pi}{a}e^{-ka} \quad \text{and} \quad \int_{-\infty}^{\infty} \frac{\sin(kx)}{x^2+a^2}\,dx = 0. \qquad \textbf{(2.1.31)}$$

□

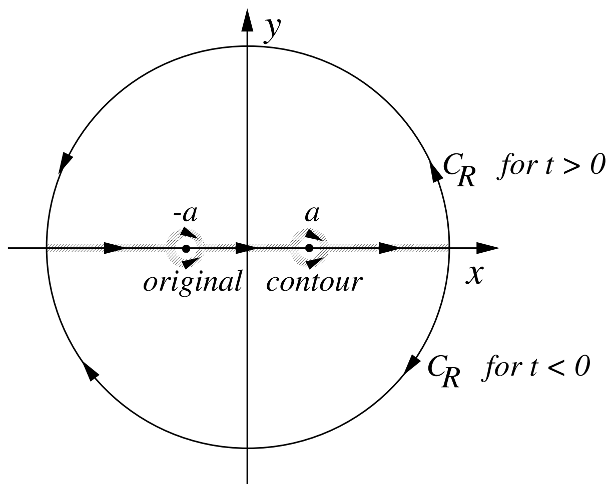


Figure 2.1.2: Contour used in Example 2.1.32.

• **Example 2.1.4**

Let us now invert the Fourier transform $F(\omega) = 2a/(a^2 - \omega^2)$, where a is real. The interesting aspect of this problem is the presence of singularities at $\omega = \pm a$ that lie *along* the contour of integration. How do we use contour integration to compute

$$f(t) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{e^{it\omega}}{a^2 - \omega^2} \, d\omega? \tag{2.1.32}$$

The answer to this question involves the concept of Cauchy principal value integrals, which allows us to extend the conventional definition of integrals to include integrands that become infinite at a finite number of points. See Section 1.10. Thus, by treating Equation 2.1.32 as a Cauchy principal value integral, we again convert it into a closed contour integration by closing the line integration along the real axis as shown in Figure 2.1.2. The semicircles at infinity vanish by Jordan's lemma and

$$f(t) = \frac{a}{\pi} \oint_C \frac{e^{itz}}{a^2 - z^2} \, dz. \tag{2.1.33}$$

For $t > 0$,

$$f(t) = -\frac{2\pi i a}{\pi} \frac{1}{2} \text{Res}\left[\frac{e^{itz}}{z^2 - a^2}; -a\right] - \frac{2\pi i a}{\pi} \frac{1}{2} \text{Res}\left[\frac{e^{itz}}{z^2 - a^2}; a\right]. \tag{2.1.34}$$

We have the factor $\frac{1}{2}$ because we are only passing over the "top" of the singularity at $z = a$ and $z = -a$. Computing the residues and simplifying the results, we obtain $f(t) = \sin(at)$.

Similarly, when $t < 0$,

$$f(t) = \frac{2\pi i a}{\pi} \frac{1}{2} \text{Res}\left[\frac{e^{itz}}{z^2 - a^2}; -a\right] + \frac{2\pi i a}{\pi} \frac{1}{2} \text{Res}\left[\frac{e^{itz}}{z^2 - a^2}; a\right] = -\sin(at). \tag{2.1.35}$$

These results can be collapsed down to the single expression $f(t) = \operatorname{sgn}(t)\sin(at)$. □

• Example 2.1.5

An additional benefit of understanding inversion by the residue method is the ability to qualitatively anticipate the inverse by knowing the location of the poles of $F(\omega)$. This intuition is important because many engineering analyses discuss stability and performance entirely in terms of the properties of the system's Fourier transform. In Figure 2.1.3 we graphed the location of the poles of $F(\omega)$ and the corresponding $f(t)$. The student should go through the mental exercise of connecting the two pictures. □

• Example 2.1.6

So far, we used only the first two points of Jordan's lemma. In this example[2] we illustrate how the remaining two points may be applied.

Consider the contour integral

$$\oint_C \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right]dz,$$

where $c > 0$ and β, τ are real. Let us evaluate this contour integral where the contour is shown in Figure 2.1.4.

From the residue theorem,

$$\begin{aligned}
&\oint_C \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right]dz \\
&\quad= 2\pi i\sum_{n=1}^{\infty}\operatorname{Res}\left\{\cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right];n\right\} \\
&\quad+ 2\pi i\operatorname{Res}\left\{\cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right];\frac{|\tau|+\beta i}{2\pi}\right\} \\
&\quad+ 2\pi i\operatorname{Res}\left\{\cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right];\frac{|\tau|-\beta i}{2\pi}\right\}.
\end{aligned}\tag{2.1.36}$$

Now

$$\operatorname{Res}\left\{\cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right];n\right\}$$

$$= \lim_{z\to n}\frac{(z-n)\cos(\pi z)}{\sin(\pi z)}\lim_{z\to n}\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right]\tag{2.1.37}$$

$$= \frac{1}{\pi}\left[\frac{e^{-nc}}{(\tau+2n\pi)^2+\beta^2}+\frac{e^{-nc}}{(\tau-2n\pi)^2+\beta^2}\right],\tag{2.1.38}$$

$$\operatorname{Res}\left\{\cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right];\frac{|\tau|+\beta i}{2\pi}\right\}$$

$$= \lim_{z\to(|\tau|+\beta i)/2\pi}\frac{\cot(\pi z)}{4\pi^2}\left[\frac{(z-|\tau|-\beta i)e^{-cz}}{(z+\tau/2\pi)^2+\beta^2/4\pi^2}+\frac{(z-|\tau|-\beta i)e^{-cz}}{(z-\tau/2\pi)^2+\beta^2/4\pi^2}\right]\tag{2.1.39}$$

$$= \frac{\cot(|\tau|/2+\beta i/2)\exp(-c|\tau|/2\pi)[\cos(c\beta/2\pi)-i\sin(c\beta/2\pi)]}{4\pi\beta i},\tag{2.1.40}$$

---

[2] See Hsieh, T. C., and R. Greif, 1972: Theoretical determination of the absorption coefficient and the total band absorptance including a specific application to carbon monoxide. *Int. J. Heat Mass Transfer*, **15**, 1477–1487.

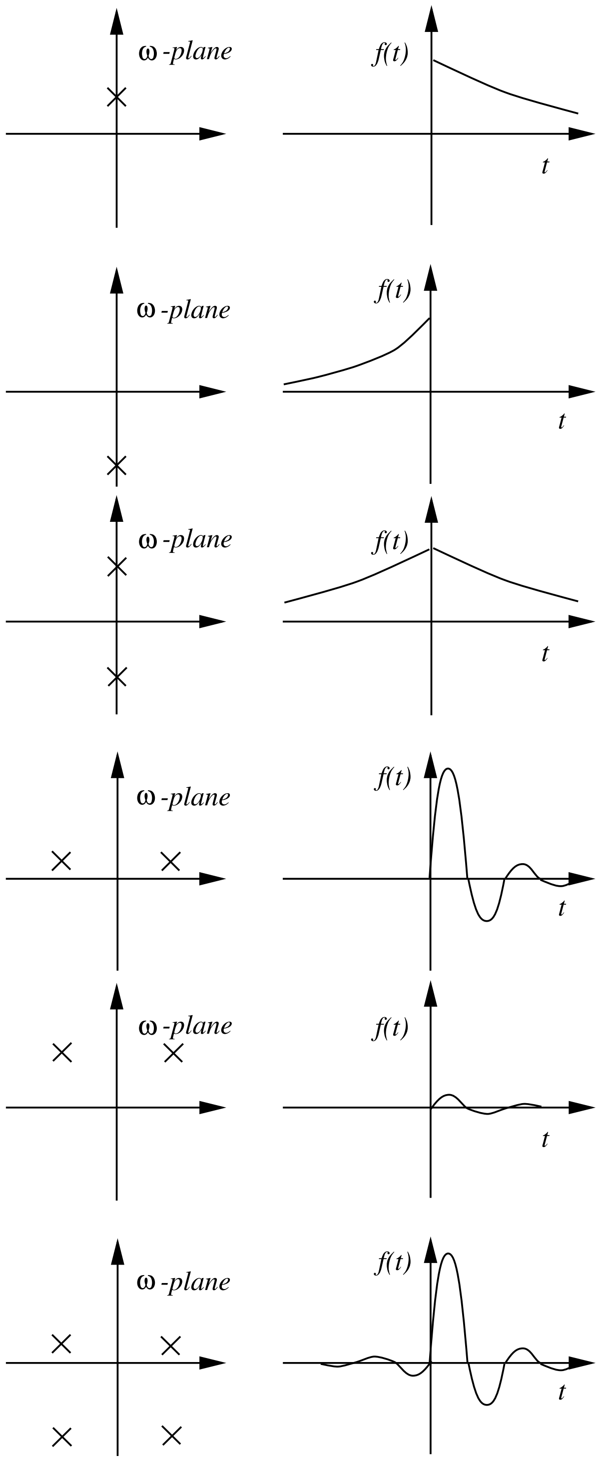


Figure 2.1.3: The correspondence between the location of the simple poles of the Fourier transform $F(\omega)$ and the behavior of $f(t)$.

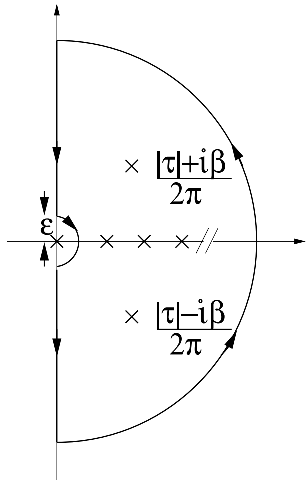


Figure 2.1.4: Contour used in Example 2.1.6.

and

$$\operatorname{Res}\left\{\cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right];\frac{|\tau|-\beta i}{2\pi}\right\}$$

$$= \lim_{z\to(|\tau|-\beta i)/2\pi}\frac{\cot(\pi z)}{4\pi^2}\left[\frac{(z-|\tau|+\beta i)e^{-cz}}{(z+\tau/2\pi)^2+\beta^2/4\pi^2}+\frac{(z-|\tau|+\beta i)e^{-cz}}{(z-\tau/2\pi)^2+\beta^2/4\pi^2}\right] \quad \textbf{(2.1.41)}$$

$$= \frac{\cot(|\tau|/2-\beta i/2)\exp(-c|\tau|/2\pi)[\cos(c\beta/2\pi)+i\sin(c\beta/2\pi)]}{-4\pi\beta i}. \quad \textbf{(2.1.42)}$$

Therefore,

$$\oint_C \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right]dz$$

$$= 2i\sum_{n=1}^{\infty}\left[\frac{e^{-nc}}{(\tau+2n\pi)^2+\beta^2}+\frac{e^{-nc}}{(\tau-2n\pi)^2+\beta^2}\right]$$

$$+ \frac{i}{2\beta}\frac{e^{i|\tau|}+e^{\beta}}{e^{i|\tau|}-e^{\beta}}e^{-c|\tau|/2\pi}[\cos(c\beta/2\pi)-i\sin(c\beta/2\pi)]$$

$$- \frac{i}{2\beta}\frac{e^{i|\tau|}+e^{-\beta}}{e^{i|\tau|}-e^{-\beta}}e^{-c|\tau|/2\pi}[\cos(c\beta/2\pi)+i\sin(c\beta/2\pi)] \quad \textbf{(2.1.43)}$$

$$= 2i\sum_{n=1}^{\infty}\left[\frac{e^{-nc}}{(\tau+2n\pi)^2+\beta^2}+\frac{e^{-nc}}{(\tau-2n\pi)^2+\beta^2}\right]$$

$$- \frac{i}{\beta}\frac{\sinh(\beta)\cos(c\beta/2\pi)+\sin(|\tau|)\sin(c\beta/2\pi)}{\cosh(\beta)-\cos(\tau)}e^{-c|\tau|/2\pi}, \quad \textbf{(2.1.44)}$$

where $\cot(\alpha) = i(e^{2i\alpha}+1)/(e^{2i\alpha}-1)$, and we made extensive use of Euler's formula.

Let us now evaluate the contour integral by direct integration. The contribution from the integration along the semicircle at infinity vanishes according to Jordan's lemma. Indeed, that is why this particular contour was chosen. Therefore,

$$\begin{aligned}
&\oint_C \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz \\
&\quad = \int_{i\infty}^{i\epsilon} \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz \\
&\quad + \int_{C_\epsilon} \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz \\
&\quad + \int_{-i\epsilon}^{-i\infty} \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz. \qquad (\mathbf{2.1.45})
\end{aligned}$$

Now, because $z = iy$,

$$\begin{aligned}
&\int_{i\infty}^{i\epsilon} \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz \\
&\quad = \int_{\infty}^{\epsilon} \coth(\pi y)\left[\frac{e^{-icy}}{(\tau+2\pi iy)^2+\beta^2}+\frac{e^{-icy}}{(\tau-2\pi iy)^2+\beta^2}\right] dy \qquad (\mathbf{2.1.46}) \\
&\quad = -2\int_{\epsilon}^{\infty} \frac{\coth(\pi y)(\tau^2+\beta^2-4\pi^2y^2)e^{-icy}}{(\tau^2+\beta^2-4\pi^2y^2)^2+16\pi^2\tau^2y^2}\, dy, \qquad (\mathbf{2.1.47})
\end{aligned}$$

$$\begin{aligned}
&\int_{-i\epsilon}^{-i\infty} \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz \\
&\quad = \int_{-\epsilon}^{-\infty} \coth(\pi y)\left[\frac{e^{-icy}}{(\tau+2\pi iy)^2+\beta^2}+\frac{e^{-icy}}{(\tau-2\pi iy)^2+\beta^2}\right] dy \qquad (\mathbf{2.1.48}) \\
&\quad = 2\int_{\epsilon}^{\infty} \frac{\coth(\pi y)(\tau^2+\beta^2-4\pi^2y^2)e^{icy}}{(\tau^2+\beta^2-4\pi^2y^2)^2+16\pi^2\tau^2y^2}\, dy, \qquad (\mathbf{2.1.49})
\end{aligned}$$

and

$$\begin{aligned}
&\int_{C_\epsilon} \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz \\
&\quad = \int_{\pi/2}^{-\pi/2} \left[\frac{1}{\pi\epsilon e^{\theta i}}-\frac{\pi\epsilon e^{\theta i}}{3}-\cdots\right] \epsilon i e^{\theta i}\, d\theta \\
&\qquad \times \left[\frac{\exp(-c\epsilon e^{\theta i})}{(\tau+2\pi\epsilon e^{\theta i})^2+\beta^2}+\frac{\exp(-c\epsilon e^{\theta i})}{(\tau-2\pi\epsilon e^{\theta i})^2+\beta^2}\right]. \qquad (\mathbf{2.1.50})
\end{aligned}$$

In the limit of $\epsilon \to 0$,

$$\begin{aligned}
&\oint_C \cot(\pi z)\left[\frac{e^{-cz}}{(\tau+2\pi z)^2+\beta^2}+\frac{e^{-cz}}{(\tau-2\pi z)^2+\beta^2}\right] dz \\
&\quad = 4i\int_0^{\infty} \frac{\coth(\pi y)(\tau^2+\beta^2-4\pi^2y^2)\sin(cy)}{(\tau^2+\beta^2-4\pi^2y^2)^2+16\pi^2\tau^2y^2}\, dy - \frac{2i}{\tau^2+\beta^2} \qquad (\mathbf{2.1.51}) \\
&\quad = 2i\sum_{n=1}^{\infty}\left[\frac{e^{-nc}}{(\tau+2n\pi)^2+\beta^2}+\frac{e^{-nc}}{(\tau-2n\pi)^2+\beta^2}\right] \\
&\quad - \frac{i}{\beta}\frac{\sinh(\beta)\cos(c\beta/2\pi)+\sin(|\tau|)\sin(c\beta/2\pi)}{\cosh(\beta)-\cos(\tau)}e^{-c|\tau|/2\pi}, \qquad (\mathbf{2.1.52})
\end{aligned}$$

or

$$
\begin{aligned}
4\int_0^\infty & \frac{\coth(\pi y)(\tau^2+\beta^2-4\pi^2y^2)\sin(cy)}{(\tau^2+\beta^2-4\pi^2y^2)^2+16\pi^2\tau^2y^2}\,dy \\
&= 2\sum_{n=1}^\infty\left[\frac{e^{-nc}}{(\tau+2n\pi)^2+\beta^2}+\frac{e^{-nc}}{(\tau-2n\pi)^2+\beta^2}\right] \qquad \textbf{(2.1.53)}\\
&\quad -\frac{1}{\beta}\frac{\sinh(\beta)\cos(c\beta/2\pi)+\sin(|\tau|)\sin(c\beta/2\pi)}{\cosh(\beta)-\cos(\tau)}e^{-c|\tau|/2\pi}+\frac{2}{\tau^2+\beta^2}.
\end{aligned}
$$

If we let $y = x/2\pi$,

$$
\begin{aligned}
\frac{\beta}{\pi}\int_0^\infty & \frac{\coth(x/2)(\tau^2+\beta^2-x^2)\sin(cx/2\pi)}{(\tau^2+\beta^2-x^2)^2+4\tau^2x^2}\,dx \\
&= 2\beta\sum_{n=1}^\infty\left[\frac{e^{-nc}}{(\tau+2n\pi)^2+\beta^2}+\frac{e^{-nc}}{(\tau-2n\pi)^2+\beta^2}\right] \qquad \textbf{(2.1.54)}\\
&\quad -\frac{\sinh(\beta)\cos(c\beta/2\pi)+\sin(|\tau|)\sin(c\beta/2\pi)}{\cosh(\beta)-\cos(\tau)}e^{-c|\tau|/2\pi}+\frac{2\beta}{\tau^2+\beta^2}.
\end{aligned}
$$

Problems

By taking the appropriate closed contour, find the inverse of the following Fourier transforms $F(\omega)$ by contour integration. The parameter a is real and positive.

1. $\dfrac{1}{\omega^2+a^2}$ 2. $\dfrac{\omega}{\omega^2+a^2}$ 3. $\dfrac{\omega}{(\omega^2+a^2)^2}$ 4. $\dfrac{\omega^2}{(\omega^2+a^2)^2}$

5. $\dfrac{1}{\omega^2-3i\omega-3}$ 6. $\dfrac{1}{(\omega-ia)^{2n+2}}$ 7. $\dfrac{\omega^2}{(\omega^2-1)^2+4a^2\omega^2}$ 8. $\dfrac{3}{(2-\omega i)(1+\omega i)}$

Then check your answer using MATLAB.

9. Find the inverse of $F(\omega) = \cos(\omega)/(\omega^2+a^2)$, $a > 0$, by first rewriting the transform as

$$F(\omega) = \frac{e^{i\omega}}{2(\omega^2+a^2)} + \frac{e^{-i\omega}}{2(\omega^2+a^2)}$$

and then using the residue theorem on each term.

10. Find[3] the inverse Fourier transform for

$$F_\pm(\omega) = \frac{e^{\pm i\omega}}{(\omega-ai)\left(R^2e^{\omega i}-e^{-\omega i}\right)} = \frac{e^{\pm i\omega-i\omega}}{(\omega-ai)\left(R^2-e^{-2\omega i}\right)},$$

where $a > 0$ and $R > 1$.

[3] See Scharstein, R. W., 1992: Transient electromagnetic plane wave reflection from a dielectric slab. *IEEE Trans. Educ.*, **35**, 170–175.

Step 1: Show that

$$f_{\pm}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i(t-1\pm1)\omega}}{(\omega - ai)(R^2 - e^{-2\omega i})} \, d\omega.$$

Step 2: Show that the singularities consist of simple poles at $z = ai$ and $z_n = \pm n\pi + i\ln(R)$, where $n = 0, \pm1, \pm2, \pm3, \ldots$.

Step 3: For $t > 0$ show that

$$f_{+}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\omega}}{(\omega - ai)(R^2 - e^{-2\omega i})} \, d\omega.$$

and we must close the contour with an infinite semi-circle in the top half-plane.

Step 4: Show that

$$\int_{-\infty}^{\infty} \frac{e^{it\omega}}{(\omega - ai)(R^2 - e^{-2\omega i})} \, d\omega = 2\pi i \operatorname{Res}\left[\frac{e^{itz}}{(z - ai)(R^2 - e^{-2zi})}; ai\right] + 2\pi i \sum_{n=-\infty}^{\infty} \operatorname{Res}\left[\frac{e^{itz}}{(z - ai)(R^2 - e^{-2zi})}; z_n\right].$$

where

$$\operatorname{Res}\left[\frac{e^{itz}}{(z - ai)(R^2 - e^{-2zi})}; ai\right] = \frac{e^{-at}}{R^2 - e^{2a}},$$

and

$$\operatorname{Res}\left[\frac{e^{itz}}{(z - ai)(R^2 - e^{-2zi})}; z_n\right] = \frac{R^{-t}e^{\pm in\pi t}}{2iR^2\left\{\pm n\pi + [\ln(R) - a]\, i\right\}},$$

so that

$$f_{+}(t) = \frac{i\, e^{-at}}{R^2 - e^{2a}} + \frac{1}{2R^{t+2}} \sum_{n=-\infty}^{\infty} \frac{e^{in\pi t}}{n\pi + [\ln(R) - a]\, i}.$$

Step 5: For the case $t < 0$, show that we close the contour with an infinite semi-circle in the bottom half-plane to compute $f_{+}(t)$.

Step 6: Compute the residues of the enclosed singularities in Step 5 and show that $f_{+}(t) = 0$. Why?

Step 7: Show that $f_{+}(t)$ equals

$$f_{+}(t) = \frac{i\, e^{-at}}{R^2 - e^{2a}} H(t) + \frac{H(t)}{2R^{t+2}} \sum_{n=-\infty}^{\infty} \frac{e^{in\pi t}}{n\pi + [\ln(R) - a]\, i}$$

at any time t.

Step 8: For $F_{-}(\omega)$, show that

$$f_{-}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i(t-2)\omega}}{(\omega - ai)(R^2 - e^{-2\omega i})} \, d\omega.$$

Step 9: For $t > 2$, show that we close the contour with an infinite semi-circle in the top half-plane.

Step 10: Compute the residue of the enclosed singularities and show that

$$\int_{-\infty}^{\infty} \frac{e^{i(t-2)\omega}}{(\omega - ai)(R^2 - e^{-2\omega i})}\, d\omega = 2\pi i \ \mathrm{Res}\left[\frac{e^{i(t-2)z}}{(z-ai)(R^2 - e^{-2zi})}; ai\right] + 2\pi i \sum_{n=-\infty}^{\infty} \mathrm{Res}\left[\frac{e^{i(t-2)z}}{(z-ai)(R^2 - e^{-2zi})}; z_n\right].$$

where

$$\mathrm{Res}\left[\frac{e^{i(t-2)z}}{(z-ai)(R^2 - e^{-2zi})}; ai\right] = \frac{e^{-at}}{R^2 e^{-2a} - 1},$$

and

$$\mathrm{Res}\left[\frac{e^{i(t-2)z}}{(z-ai)(R^2 - e^{-2zi})}; z_n\right] = \frac{R^{-t+2} e^{\pm in\pi t}}{2iR^2\left\{\pm n\pi + [\ln(R) - a]\, i\right\}}.$$

so that

$$f_-(t) = \frac{i\, e^{-at}}{R^2 e^{-2a} - 1} + \frac{1}{2R^t} \sum_{n=-\infty}^{\infty} \frac{e^{in\pi t}}{n\pi + [\ln(R) - a]\, i}.$$

Step 11: For $t < 2$ we must close the contour with an infinite semi-circle in the bottom half-plane. Compute the residue of the enclosed singularities and show that $f_-(t) = 0$. Why?

Step 12: Show that the final answer for $f_-(t)$ at any time t is

$$f_-(t) = \frac{i\, e^{-at}}{R^2 e^{-2a} - 1} H(t-2) + \frac{H(t-2)}{2R^t} \sum_{n=-\infty}^{\infty} \frac{e^{in\pi t}}{n\pi + [\ln(R) - a]\, i}.$$

11. During the solution of the heat equation, Taitel et al.[4] inverted the Fourier transform

$$F(\omega) = \frac{\cosh(y\sqrt{\omega^2+1}\,)}{\sqrt{\omega^2+1}\, \sinh(p\sqrt{\omega^2+1}/2)},$$

where y and p are real.

Step 1: From the definition of the Fourier transform, show that

$$f(t) = \frac{1}{2\pi} \oint_C \frac{\cosh\left(y\sqrt{z^2+1}\,\right) e^{izt}}{\sqrt{z^2+1}\ \sinh\left(p\sqrt{z^2+1}/2\right)} dz,$$

where we have closed the line integral with an infinite semicircle in the upper half-plane if $t > 0$. For $t < 0$ we close the contour in the lower half-plane.

Step 2: For $t > 0$, show that the enclosed singularities are simple poles that are located at $z = i$ and $p\sqrt{z^2+1} = 2n\pi i$, $n = 0, 1, 2, \ldots$, or $z_n = i\sqrt{1 + 4n^2\pi^2/p^2}$.

Step 3: Show that

$$\mathrm{Res}\left[\frac{\cosh\left(y\sqrt{z^2+1}\,\right) e^{izt}}{\sqrt{z^2+1}\ \sinh\left(p\sqrt{z^2+1}/2\right)}; i\right] = \frac{e^{-t}}{ip},$$

[4] Taitel, Y., M. Bentwich, and A. Tamir, 1973: Effects of upstream and downstream boundary conditions on heat (mass) transfer with axial diffusion. *Int. J. Heat Mass Transfer*, **16**, 359–369.

and

$$\operatorname{Res}\left[\frac{\cosh\left(y\sqrt{z^2+1}\,\right)e^{izt}}{\sqrt{z^2+1}\ \sinh\left(p\sqrt{z^2+1}/2\right)};z_n\right]=\frac{2\cos(2n\pi y/p)\exp(-\sqrt{1+4n^2\pi^2/p^2}\ t)}{ip(-1)^n\sqrt{1+4n^2\pi^2/p^2}}.$$

Step 4: For $t>0$, show that

$$f(t)=\frac{e^{-t}}{p}+\frac{2}{p}\sum_{n=1}^{\infty}\frac{(-1)^n\cos(2n\pi y/p)\exp(-\sqrt{1+4n^2\pi^2/p^2}\,t)}{\sqrt{1+4n^2\pi^2/p^2}}.$$

Step 5: For $t<0$, show that the enclosed singularities are simple poles located at $z=-i$ and $z_n=-i\sqrt{1+4n^2\pi^2/p^2}$.

Step 6: Show that

$$\operatorname{Res}\left[\frac{\cosh\left(y\sqrt{z^2+1}\,\right)e^{izt}}{\sqrt{z^2+1}\ \sinh\left(p\sqrt{z^2+1}/2\right)};-i\right]=-\frac{e^t}{ip},$$

and

$$\operatorname{Res}\left[\frac{\cosh\left(y\sqrt{z^2+1}\,\right)e^{izt}}{\sqrt{z^2+1}\ \sinh\left(p\sqrt{z^2+1}/2\right)};z_n\right]=-\frac{2\cos(2n\pi y/p)\exp(\sqrt{1+4n^2\pi^2/p^2}\ t)}{ip(-1)^n\sqrt{1+4n^2\pi^2/p^2}}.$$

Step 7: For $t<0$, show that

$$f(t)=\frac{e^{t}}{p}+\frac{2}{p}\sum_{n=1}^{\infty}\frac{(-1)^n\cos(2n\pi y/p)\exp(\sqrt{1+4n^2\pi^2/p^2}\,t)}{\sqrt{1+4n^2\pi^2/p^2}}.$$

Step 8: Show that we write the results from Step 4 and Step 7 as

$$f(t)=\frac{e^{-|t|}}{p}+\frac{2}{p}\sum_{n=1}^{\infty}\frac{(-1)^n}{\sqrt{1+4n^2\pi^2/p^2}}\cos\left(\frac{2n\pi y}{p}\right)e^{-\sqrt{1+4n^2\pi^2/p^2}\,|t|}.$$

In this case, our time variable t was their spatial variable $x-\xi$.

12. Find the inverse of the Fourier transform

$$F(\omega)=\left[\cos\left\{\frac{\omega L}{\beta[1+i\gamma\ \operatorname{sgn}(\omega)]}\right\}\right]^{-1},$$

where L, β, and γ are real and positive and $\operatorname{sgn}(z)=1$ if $\Re(z)>0$ and -1 if $\Re(z)<0$.

Step 1: From the definition of the Fourier transform, show that

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{it\omega}}{\cos\left\{\frac{\omega L}{\beta[1+i\gamma\operatorname{sgn}(\omega)]}\right\}}\,d\omega=\frac{1}{2\pi}\oint_C\frac{e^{itz}}{\cos\left\{\frac{zL}{\beta[1+i\gamma\operatorname{sgn}(z)]}\right\}}\,dz.$$

Step 2: Show that the integral has simple poles at $z_{n\pm}=\pm(2n-1)\beta\pi/(2L)+(2n-1)i\beta\gamma\pi/(2L)$, where $n=1,2,3,\ldots$

Step 3: For $t > 0$, use the residue theorem and show that

$$f(t) = i\left[\sum_{n=1}^{\infty} \mathrm{Res}\left(\frac{e^{itz}}{\cos\left\{\frac{zL}{\beta[1+i\gamma\mathrm{sgn}(z)]}\right\}}; z_{n+}\right) + \mathrm{Res}\left(\frac{e^{itz}}{\cos\left\{\frac{zL}{\beta[1+i\gamma\mathrm{sgn}(z)]}\right\}}; z_{n-}\right)\right],$$

where

$$\mathrm{Res}\left(\frac{e^{itz}}{\cos\left\{\frac{zL}{\beta[1+i\gamma\mathrm{sgn}(z)]}\right\}}; z_{n\pm}\right) = \pm\frac{(-1)^n\beta}{L}[1 + i\gamma\mathrm{sgn}(z_n)]e^{-(2n-1)\beta\gamma\pi t/(2L)\pm(2n-1)\beta\pi i t/(2L)}.$$

Step 4: For $t < 0$, show that $f(t) = 0$. Why?

Step 5: Show that we can summarize the results from Step 3 and Step 4 by

$$f(t) = \frac{2\beta}{L}H(t)\sum_{n=1}^{\infty}(-1)^{n+1}e^{-(2n-1)\beta\gamma\pi t/2L}\left\{\gamma\cos[(2n-1)\beta\pi t/2L] + \sin[(2n-1)\beta\pi t/2L]\right\}.$$

Use the residue theorem to verify the following integrals:

13. $\displaystyle\int_{-\infty}^{\infty} \frac{\sin(x)}{x^2 + 4x + 5}\,dx = -\frac{\pi}{e}\sin(2)$

14. $\displaystyle\int_{0}^{\infty} \frac{\cos(x)}{(x^2 + 1)^2}\,dx = \frac{\pi}{2e}$

15 $\displaystyle\int_{-\infty}^{\infty} \frac{x\sin(ax)}{x^2 + 4}\,dx = \pi e^{-2a}$

16. $\displaystyle\int_{0}^{\infty} \frac{x^2\cos(ax)}{(x^2 + b^2)^2}\,dx = \frac{\pi}{4b}(1 - ab)e^{-ab}$

where $a, b > 0$.

17. The concept of forced convection is normally associated with heat streaming through a duct or past an obstacle. Bentwich[5] showed that a similar transport can exist when convection results from a wave traveling through an essentially stagnant fluid. In the process of computing the amount of heating, he proved the following identity:

$$\int_{-\infty}^{\infty} \frac{\cosh(hx) - 1}{x\,\sinh(hx)}\cos(ax)\,dx = \ln[\coth(|a|\pi/h)], \qquad h > 0.$$

Confirm his result.

Step 1:

$$\int_{-\infty}^{\infty} \frac{\cosh(hx) - 1}{x\,\sinh(hx)}\cos(ax)\,dx = \Re\left(\oint_C \frac{\cosh(hz) - 1}{z\,\sinh(hz)}\,e^{aiz}\,dz\right),$$

if $a > 0$ and C is a semicircle of infinite radius in the upper half-plane.

Step 2: Within the contour, show that there is a removal singularity at $z = 0$ and simple poles at $hz_n = n\pi i$ with $n = 1, 2, 3, \ldots$.

---

[5] Bentwich, M., 1966: Convection enforced by surface and tidal waves. *Int. J. Heat Mass Transfer*, **9**, 663–670.

Step 3: Show that

$$\oint_C \frac{\cosh(hz)-1}{z\sinh(hz)}\, e^{aiz}\, dz = 2\pi i \sum_{n=1}^{\infty} \operatorname{Res}\left[\frac{\cosh(hz)-1}{z\sinh(hz)}\, e^{aiz}; \frac{n\pi i}{h}\right]$$

with

$$\operatorname{Res}\left[\frac{\cosh(hz)-1}{z\sinh(hz)}\, e^{aiz}; \frac{n\pi i}{h}\right] = \frac{1-(-1)^n}{n\pi i} e^{-n\pi a/h}.$$

Step 4: Show that

$$\int_{-\infty}^{\infty} \frac{\cosh(hx)-1}{x\sinh(hx)} \cos(ax)\, dx = 4\sum_{m=1}^{\infty} \frac{\exp[-(2m-1)\pi a/h]}{2m-1} = \ln[\coth(\pi a/h)].$$

Step 5: Redo the analysis if we replace a by $-a$. Reconcile your results with those given by Bentwich.

2.2 INVERSION OF LAPLACE TRANSFORMS BY CONTOUR INTEGRATION

Partial fractions and convolution are two common methods for finding the inverse of the Laplace transform $F(s)$. In many instances these methods fail simply because of the complexity of the transform to be inverted. In this section we shall show how we can invert transforms through the powerful method of contour integration. Of course, the student must be proficient in the use of complex variables.

Consider the piece-wise differentiable function $f(x)$, which vanishes for $x < 0$. We can express the function $e^{-cx} f(x)$ by the complex Fourier representation of

$$f(x)e^{-cx} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} \left[\int_0^{\infty} e^{-ct} f(t) e^{-i\omega t}\, dt\right] d\omega, \qquad \textbf{(2.2.1)}$$

for any value of the real constant c, where the integral

$$I = \int_0^{\infty} e^{-ct} |f(t)|\, dt \qquad \textbf{(2.2.2)}$$

exists. By multiplying both sides of Equation 2.2.1 by e^{cx} and bringing it inside the first integral,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{(c+\omega i)x} \left[\int_0^{\infty} f(t) e^{-(c+\omega i)t}\, dt\right] d\omega. \qquad \textbf{(2.2.3)}$$

With the substitution $z = c + \omega i$, where z is a new, complex variable of integration,

$$f(x) = \frac{1}{2\pi i} \int_{c-\infty i}^{c+\infty i} e^{zx} \left[\int_0^{\infty} f(t) e^{-zt}\, dt\right] dz. \qquad \textbf{(2.2.4)}$$

The quantity inside the square brackets is the Laplace transform $F(z)$. Therefore, we can express $f(t)$ in terms of its transform by the complex contour integral

$$\boxed{f(t) = \frac{1}{2\pi i} \int_{c-\infty i}^{c+\infty i} F(z) e^{tz}\, dz.} \qquad \textbf{(2.2.5)}$$

An outstanding mathematician at Cambridge University at the turn of the twentieth century, Thomas John I'Anson Bromwich (1875–1929) came to Heaviside's operational calculus through his interest in divergent series. Beginning a correspondence with Heaviside, Bromwich was able to justify operational calculus through the use of contour integrals by 1915. After his premature death, individuals such as J. R. Carson and Sir H. Jeffreys brought Laplace transforms to the increasing attention of scientists and engineers. (Portrait courtesy of the Royal Society of London.)

This line integral, the *Bromwich integral*,[6] runs along the line $x = c$ parallel to the imaginary axis and c units to the right of it, the so-called *Bromwich contour.* We select the value of c sufficiently large so that the integral, Equation 2.2.2, exists; subsequent analysis shows that this occurs when c is larger than the real part of any of the singularities of $F(z)$.

We must now evaluate the contour integral. Because of the power of the *residue* theorem in complex variables, the contour integral is usually transformed into a closed contour through the use of *Jordan's lemma.* See Section 2.1, Equations 2.1.4 and Equation 2.1.5. The following examples will illustrate the proper use of Equation 2.2.5.

• **Example 2.2.1**

Let us invert

$$F(s) = \frac{e^{-3s}}{s^2(s-1)}. \tag{2.2.6}$$

[6] Bromwich, T. J. I'A., 1916: Normal coordinates in dynamical systems. *Proc. London Math. Soc., Ser. 2*, **15**, 401–448.

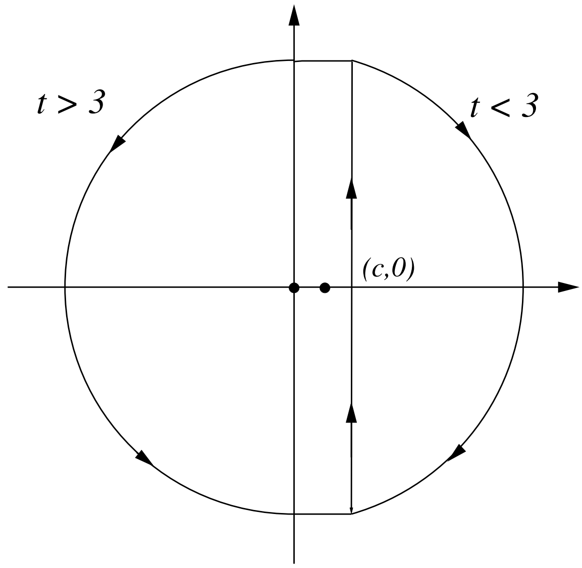


Figure 2.2.1: Contours used in the inversion of Equation 2.2.6.

From Bromwich's integral,

$$f(t) = \frac{1}{2\pi i} \int_{c-\infty i}^{c+\infty i} \frac{e^{(t-3)z}}{z^2(z-1)}\, dz = \frac{1}{2\pi i} \oint_C \frac{e^{(t-3)z}}{z^2(z-1)}\, dz - \frac{1}{2\pi i} \int_{C_R} \frac{e^{(t-3)z}}{z^2(z-1)}\, dz, \qquad (\mathbf{2.2.7})$$

where C_R is a semicircle of infinite radius in either the right or left half of the z-plane and C is the closed contour that includes C_R and Bromwich's contour. See Figure 2.2.1.

Our first task is to choose an appropriate contour so that the integral along C_R vanishes. By Jordan's lemma, this requires a semicircle in the right half-plane if $t - 3 < 0$ and a semicircle in the left half-plane if $t - 3 > 0$. Consequently, by considering these two separate cases, we force the second integral in Equation 2.2.7 to zero and the inversion simply equals the closed contour.

Consider the case $t < 3$ first. Because Bromwich's contour lies to the right of any singularities, there are no singularities within the closed contour and $f(t) = 0$.

Consider now the case $t > 3$. Within the closed contour in the left half-plane, there is a second-order pole at $z = 0$ and a simple pole at $z = 1$. Therefore,

$$f(t) = \mathrm{Res}\left[\frac{e^{(t-3)z}}{z^2(z-1)}; 0\right] + \mathrm{Res}\left[\frac{e^{(t-3)z}}{z^2(z-1)}; 1\right], \qquad (\mathbf{2.2.8})$$

where

$$\mathrm{Res}\left[\frac{e^{(t-3)z}}{z^2(z-1)}; 0\right] = \lim_{z\to 0} \frac{d}{dz}\left[z^2 \frac{e^{(t-3)z}}{z^2(z-1)}\right] = \lim_{z\to 0} \left[\frac{(t-3)e^{(t-3)z}}{z-1} - \frac{e^{(t-3)z}}{(z-1)^2}\right] = 2 - t, \qquad (\mathbf{2.2.9})$$

and

$$\mathrm{Res}\left[\frac{e^{(t-3)z}}{z^2(z-1)}; 1\right] = \lim_{z\to 1} (z-1)\frac{e^{(t-3)z}}{z^2(z-1)} = e^{t-3}. \qquad (\mathbf{2.2.10})$$

Taking our earlier results into account, the inverse equals

$$f(t) = \left[e^{t-3} - (t-3) - 1\right] H(t-3), \qquad (\mathbf{2.2.11})$$

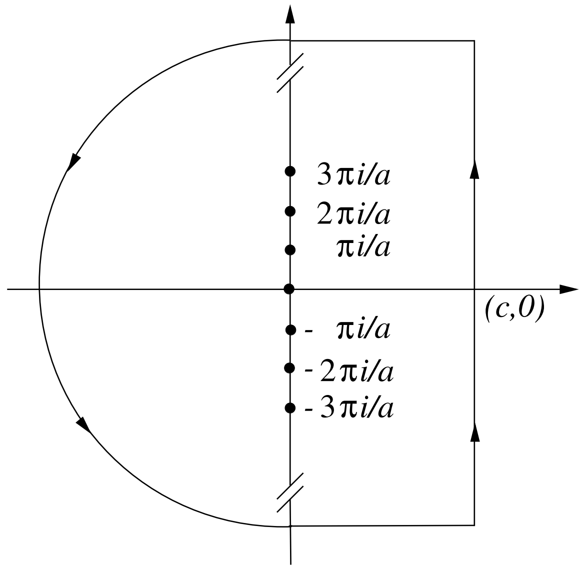


Figure 2.2.2: Contours used in the inversion of Equation 2.2.12.

which we would have obtained from the second shifting theorem and tables. □

• Example 2.2.2

For our second example of the inversion of Laplace transforms by complex integration, let us find the inverse of

$$F(s) = \frac{1}{s\sinh(as)}, \tag{2.2.12}$$

where a is real. From Bromwich's integral,

$$f(t) = \frac{1}{2\pi i}\int_{c-\infty i}^{c+\infty i} \frac{e^{tz}}{z\sinh(az)}\,dz. \tag{2.2.13}$$

Here c is greater than the real part of any of the singularities in Equation 2.2.12. Using the infinite product for the hyperbolic sine,[7]

$$\frac{e^{tz}}{z\sinh(az)} = \frac{e^{tz}}{az^2[1+a^2z^2/\pi^2][1+a^2z^2/(4\pi^2)][1+a^2z^2/(9\pi^2)]\cdots}. \tag{2.2.14}$$

Thus, we have a second-order pole at $z = 0$ and simple poles at $z_n = \pm n\pi i/a$, where $n = 1, 2, 3, \ldots$.

We can convert the line integral Equation 2.2.13, with the Bromwich contour lying parallel and slightly to the right of the imaginary axis, into a closed contour using Jordan's lemma through the addition of an infinite semicircle joining $i\infty$ to $-i\infty$, as shown in Figure 2.2.2. We now apply the residue theorem. For the second-order pole at $z = 0$,

[7] Gradshteyn, I. S., and I. M. Ryzhik, 1965: *Table of Integrals, Series and Products*. Academic Press, Section 1.431, Formula 2.

$$\operatorname{Res}\left[\frac{e^{tz}}{z\sinh(az)};0\right] = \frac{1}{1!}\lim_{z\to 0}\frac{d}{dz}\left[\frac{(z-0)^2 e^{tz}}{z\sinh(az)}\right] = \lim_{z\to 0}\frac{d}{dz}\left[\frac{ze^{tz}}{\sinh(az)}\right] \tag{2.2.15}$$

$$= \lim_{z\to 0}\left[\frac{e^{tz}}{\sinh(az)} + \frac{zte^{tz}}{\sinh(az)} - \frac{az\cosh(az)e^{tz}}{\sinh^2(az)}\right] = \frac{t}{a} \tag{2.2.16}$$

after using $\sinh(az) = az + O(z^3)$. For the simple poles $z_n = \pm n\pi i/a$,

$$\operatorname{Res}\left[\frac{e^{tz}}{z\sinh(az)};z_n\right] = \lim_{z\to z_n}\frac{(z-z_n)e^{tz}}{z\sinh(az)} = \lim_{z\to z_n}\frac{e^{tz}}{\sinh(az)+az\cosh(az)} \tag{2.2.17}$$

$$= \frac{\exp(\pm n\pi it/a)}{(-1)^n(\pm n\pi i)}, \tag{2.2.18}$$

because $\cosh(\pm n\pi i) = \cos(n\pi) = (-1)^n$. Thus, summing up all of the residues gives

$$f(t) = \frac{t}{a} + \sum_{n=1}^{\infty}\frac{(-1)^n\exp(n\pi it/a)}{n\pi i} - \sum_{n=1}^{\infty}\frac{(-1)^n\exp(-n\pi it/a)}{n\pi i} \tag{2.2.19}$$

$$= \frac{t}{a} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(n\pi t/a). \tag{2.2.20}$$

□

In addition to computing the inverse of Laplace transforms, Bromwich's integral places certain restrictions on $F(s)$ in order that an inverse exists. If α denotes the minimum value that c may possess, the restrictions are threefold.[8] First, $F(z)$ must be analytic in the half-plane $x \geq \alpha$, where $z = x + iy$. Second, in the same half-plane it must behave as z^{-k}, where $k > 1$. Finally, $F(x)$ must be real when $x \geq \alpha$.

• Example 2.2.3

Is the function $\sin(s)/(s^2+4)$ a proper Laplace transform? Although the function satisfies the first and third criteria listed in the previous paragraph on the half-plane $x > 2$, the function becomes unbounded as $y \to \pm\infty$ for any fixed $x > 2$. Thus, $\sin(s)/(s^2+4)$ cannot be a Laplace transform. □

• Example 2.2.4

An additional benefit of understanding inversion by the residue method is the ability to qualitatively anticipate the inverse by knowing the location of the poles of $F(s)$. This intuition is important because many engineering analyses discuss stability and performance entirely in terms of the properties of the system's Laplace transform. In Figure 2.2.3 we have graphed the location of the poles of $F(s)$ and the corresponding $f(t)$. The student should go through the mental exercise of connecting the two pictures.

---

[8] For the proof, see Churchill, R. V., 1972: *Operational Mathematics*. McGraw-Hill, Section 67.

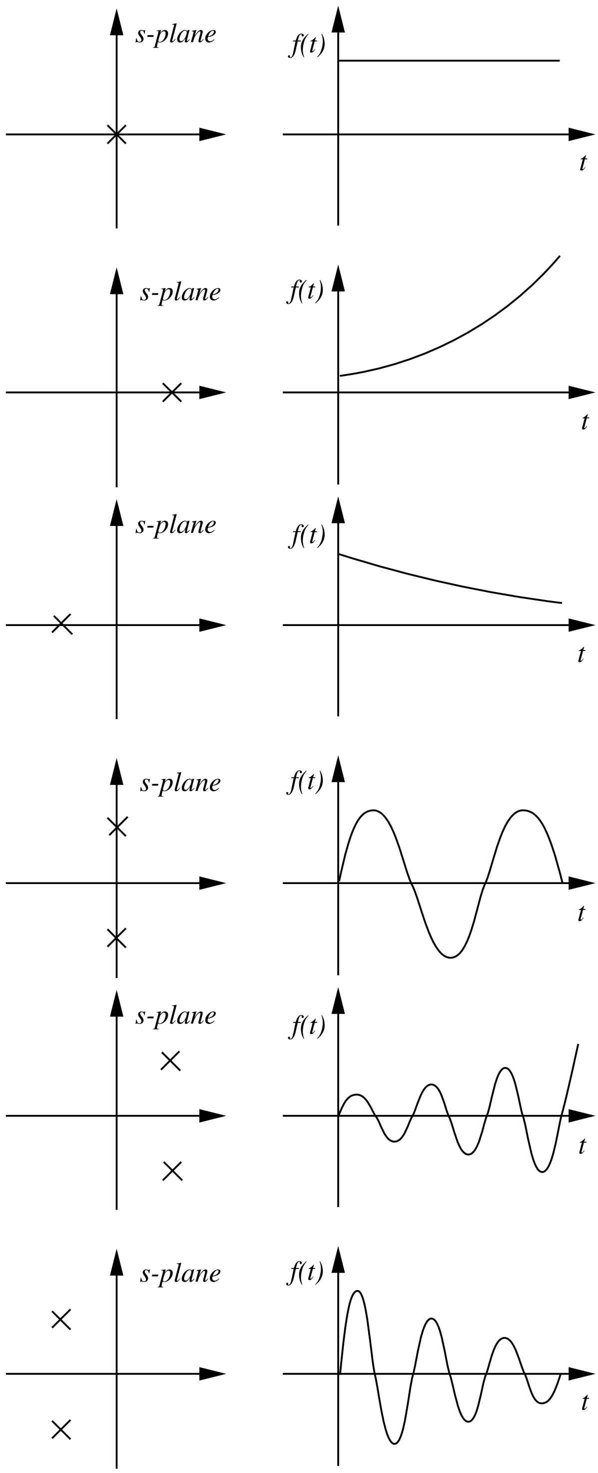


Figure 2.2.3: The correspondence between the location of the simple poles of the Laplace transform $F(s)$ and the behavior of $f(t)$.

Problems

Use Bromwich's integral to invert the following Laplace transforms $F(s)$:

1. $\dfrac{s+1}{(s+2)^2(s+3)}$ 2. $\dfrac{1}{s^2(s+a)^2}$ 3. $\dfrac{1}{s(s-2)^3}$ 4. $\dfrac{1}{s(s+a)^2(s^2+b^2)}$ 5. $\dfrac{e^{-s}}{s^2(s+2)}$

6. Use Bromwich's integral to invert

$$F(s) = \frac{1}{s(1+e^{-as})}.$$

Step 1: Show that the singularities are all simple poles and are located at $z = 0$ and $z_n = \pm(2n-1)\pi i/a$, where $n = 1, 2, 3, \ldots$.

Step 2: Show that the corresponding residues are

$$\operatorname{Res}\left[\frac{e^{tz}}{z(1+e^{-az})}; 0\right] = \frac{1}{2}, \quad \text{and} \quad \operatorname{Res}\left[\frac{e^{tz}}{z(1+e^{-az})}; z_n\right] = \pm\frac{\exp[\pm(2n-1)\pi it/a]}{(2n-1)\pi i}.$$

Step 3: Show that the inverse Laplace transform equals

$$f(t) = \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sin[(2n-1)\pi t/a]}{(2n-1)\pi}.$$

7. Use Bromwich's integral to invert

$$F(s) = \frac{1}{(s+b)\cosh(as)}.$$

Step 1: Show that the singularities are all simple poles and are located at $z = -b$ and $z_n = \pm(2n-1)\pi i/(2a)$, where $n = 1, 2, 3, \ldots$ because $\cosh(az) = \cos(iaz) = 0$.

Step 2: Show that the corresponding residues are

$$\operatorname{Res}\left[\frac{e^{tz}}{(z+b)\cosh(az)}; -b\right] = \frac{e^{-bt}}{\cosh(ab)},$$

and

$$\operatorname{Res}\left[\frac{e^{tz}}{(z+b)\cosh(az)}; z_n\right] = \pm\frac{\exp[\pm(2n-1)\pi it/(2a)]}{ai[b\pm(2n-1)\pi i/(2a)]\sin[(2n-1)\pi/2]}.$$

Step 3: Show that the inverse Laplace transform equals

$$f(t) = \frac{e^{-bt}}{\cosh(ab)} - 8ab\sum_{n=1}^{\infty}(-1)^n\frac{\sin[(2n-1)\pi t/(2a)]}{4a^2b^2+(2n-1)^2\pi^2}$$
$$+ 4\sum_{n=1}^{\infty}(-1)^n\frac{(2n-1)\pi\cos[(2n-1)\pi t/(2a)]}{4a^2b^2+(2n-1)^2\pi^2}.$$

8. Use Bromwich's integral to invert

$$F(s) = \frac{1}{s(1 - e^{-as})}.$$

Step 1: Show that the singularities are all simple poles and are located at $z = 0$ and $z_n = \pm 2n\pi i/a$, where $n = 1, 2, 3, \ldots$.

Step 2: Show that the corresponding residues are

$$\operatorname{Res}\left[\frac{e^{tz}}{z(1 - e^{-az})}; 0\right] = \frac{t}{a} + \frac{1}{2}, \qquad \text{and} \qquad \operatorname{Res}\left[\frac{e^{tz}}{z(1 - e^{-az})}; z_n\right] = \pm\frac{\exp(\pm 2n\pi it/a)}{2n\pi i}.$$

Hint: Near $z = 0$, show that

$$\frac{e^{tz}}{z(1 - e^{-az})} = \frac{1 + tz + \cdots}{az^2(1 - az/2 + \cdots)} = \frac{1}{az^2}\left(1 + tz + \frac{az}{2} + \cdots\right)$$

and read off the residue from the Laurent expansion.

Step 3: Show that the inverse Laplace transform equals

$$f(t) = \frac{t}{a} + \frac{1}{2} + \frac{1}{\pi}\sum_{n=1}^{\infty} \frac{\sin(2n\pi t/a)}{n}.$$

9. Consider a function $f(t)$ that has the Laplace transform $F(z)$, which is analytic in the half-plane $\Re(z) > s_0$. Can we use this knowledge to find $g(t)$, whose Laplace transform $G(z)$ equals $F[\varphi(z)]$, where $\varphi(z)$ is also analytic for $\Re(z) > s_0$? The answer to this question leads to the Schouten[9]–Van der Pol[10] theorem.

Step 1: Show that the following relationships hold true:

$$G(z) = F[\varphi(z)] = \int_0^\infty f(\tau) e^{-\varphi(z)\tau}\, d\tau, \qquad \text{and} \qquad g(t) = \frac{1}{2\pi i}\int_{c-\infty i}^{c+\infty i} F[\varphi(z)] e^{tz}\, dz.$$

Step 2: Using the results from Step 1, show that

$$g(t) = \int_0^\infty f(\tau)\left[\frac{1}{2\pi i}\int_{c-\infty i}^{c+\infty i} e^{-\varphi(z)\tau} e^{tz}\, dz\right] d\tau.$$

This is the Schouten-Van der Pol theorem.

Step 3: If $G(z) = F(\sqrt{z}\,)$ show that

$$g(t) = \frac{1}{2\sqrt{\pi t^3}}\int_0^\infty \tau f(\tau) \exp\left(-\frac{\tau^2}{4t}\right) d\tau.$$

---

[9] Schouten, J. P., 1935: A new theorem in operational calculus together with an application of it. *Physica*, **2**, 75–80.

[10] Van der Pol, B., 1934: A theorem on electrical networks with applications to filters. *Physica*, **1**, 521–530.

Hint: Do not evaluate the contour integral. Instead, ask yourself: What function of time has a Laplace transform that equals $e^{-\varphi(z)\tau}$, where τ is a parameter? Then use tables.

2.3 INTEGRAL EQUATIONS

An *integral equation* contains the dependent variable under an integral sign. The convolution theorem provides an excellent tool for solving a very special class of these equations, *the Volterra equation of the second kind*:[11]

$$f(t) - \int_0^t K[t, x, f(x)]\, dx = g(t), \qquad 0 \le t \le T. \tag{2.3.1}$$

These equations appear in history-dependent problems, such as epidemics,[12] vibration problems,[13] and viscoelasticity.[14]

• Example 2.3.1

Let us find $f(t)$ from the integral equation

$$f(t) = 4t - 3\int_0^t f(x)\sin(t-x)\, dx. \tag{2.3.2}$$

The integral in Equation 2.3.2 is such that we can use the convolution theorem to find its Laplace transform. Then, because $\mathcal{L}[\sin(t)] = 1/(s^2+1)$, the convolution theorem yields

$$\mathcal{L}\left[\int_0^t f(x)\sin(t-x)\, dx\right] = \frac{F(s)}{s^2+1}. \tag{2.3.3}$$

Therefore, the Laplace transform converts Equation 2.3.2 into

$$F(s) = \frac{4}{s^2} - \frac{3F(s)}{s^2+1}. \tag{2.3.4}$$

Solving for $F(s)$,

$$F(s) = \frac{4(s^2+1)}{s^2(s^2+4)}. \tag{2.3.5}$$

By partial fractions, or by inspection,

$$F(s) = \frac{1}{s^2} + \frac{3}{s^2+4}. \tag{2.3.6}$$

---

[11] Fock, V., 1924: Über eine Klasse von Integralgleichungen. *Math. Z.*, **21**, 161–173; Koizumi, S., 1931: On Heaviside's operational solution of a Volterra's integral equation when its nucleus is a function of $(x-\xi)$. *Philos. Mag., Ser. 7*, **11**, 432–441.

[12] Wang, F. J. S., 1978: Asymptotic behavior of some deterministic epidemic models. *SIAM J. Math. Anal.*, **9**, 529–534.

[13] Lin, S. P., 1975: Damped vibration of a string. *J. Fluid Mech.*, **72**, 787–797.

[14] Rogers, T. G., and E. H. Lee, 1964: The cylinder problem in viscoelastic stress analysis. *Q. Appl. Math.*, **22**, 117–131.

Therefore, inverting term by term,

$$f(t) = t + \tfrac{3}{2}\sin(2t). \tag{2.3.7}$$

Note that the integral equation

$$f(t) = 4t - 3\int_0^t f(t-x)\sin(x)\,dx \tag{2.3.8}$$

also has the same solution. □

• Example 2.3.2

Let us solve the equation

$$f'(t) + \alpha^2 \int_0^t f(\tau)\,d\tau = B - C\cos(\omega t), \qquad f(0) = 0. \tag{2.3.9}$$

Again the integral is one of the convolution type; it differs from the previous example in that it includes a derivative. Taking the Laplace transform of Equation 2.3.9,

$$sF(s) - f(0) + \frac{\alpha^2 F(s)}{s} = \frac{B}{s} - \frac{sC}{s^2+\omega^2}. \tag{2.3.10}$$

Because $f(0) = 0$, Equation 2.3.10 simplifies to

$$(s^2+\alpha^2)F(s) = B - \frac{Cs^2}{s^2+\omega^2}. \tag{2.3.11}$$

Solving for $F(s)$,

$$F(s) = \frac{B}{s^2+\alpha^2} - \frac{Cs^2}{(s^2+\alpha^2)(s^2+\omega^2)}. \tag{2.3.12}$$

Using partial fractions to invert Equation 2.3.12,

$$f(t) = \left(\frac{B}{\alpha} + \frac{\alpha C}{\omega^2-\alpha^2}\right)\sin(\alpha t) - \frac{\omega C}{\omega^2-\alpha^2}\sin(\omega t). \tag{2.3.13}$$

□

• Example 2.3.3

Let us solve[15] the integral equation

$$f(t) = \frac{a}{2(1+2a)}\int_0^t f(t-x)f(x)\,dx + e^{-t}. \tag{2.3.14}$$

[15] Hounslow, M. J., 1990: A discretized population balance for continuous systems at steady state. *AICHE J.*, **36**, 106–116.

Taking the Laplace transform of Equation 2.3.14, we obtain

$$F(s) = \frac{a\,F^2(s)}{2(1+2a)} + \frac{1}{s+1}. \tag{2.3.15}$$

Solving for $F(s)$ so that $F(s) \to 0$ as $s \to \infty$, we have

$$F(s) = \frac{2a+1}{a} - \frac{2a+1}{a}\sqrt{\frac{(2a+1)(s+1)-2a}{(2a+1)(s+1)}} = \frac{2a+1}{a} - \frac{\sqrt{2a+1}}{a}\sqrt{\frac{(2a+1)s+1}{s+1}}. \tag{2.3.16}$$

Taking the inverse of Equation 2.3.16,

$$f(t) = \frac{2a+1}{a}\,\delta(t) - \frac{\sqrt{2a+1}}{a} g(t), \tag{2.3.17}$$

where $g(t)$ is the inverse of the Laplace transform $G(s)$,

$$G(s) = \sqrt{\frac{(2a+1)s+1}{s+1}} = \sqrt{2a+1}\,\frac{s+1/(1+2a)}{\sqrt{s+1}\sqrt{s+1/(2a+1)}} \tag{2.3.18}$$

$$= \sqrt{2a+1}\,sH(s) + \frac{H(s)}{\sqrt{2a+1}} \tag{2.3.19}$$

and

$$H(s) = \frac{1}{\sqrt{s+1}\sqrt{s+1/(2a+1)}}. \tag{2.3.20}$$

Taking the inverse of $H(s)$, we find that

$$h(t) = \exp\left(-\frac{a+1}{2a+1}t\right) I_0\left(\frac{at}{2a+1}\right) \tag{2.3.21}$$

and

$$h'(t) = -\frac{a+1}{2a+1}\exp\left(-\frac{a+1}{2a+1}t\right) I_0\left(\frac{at}{2a+1}\right) + \frac{a}{2a+1}\exp\left(-\frac{a+1}{2a+1}t\right) I_1\left(\frac{at}{2a+1}\right), \tag{2.3.22}$$

where $I_0(\cdot)$ and $I_1(\cdot)$ are modified Bessel functions of the first kind.

Because $sH(s) = \mathcal{L}[h'(t)] + h(0)$ and $h(0) = 1$, $h'(t) = \mathcal{L}^{-1}[sH(s)] - \delta(t)$ or $\mathcal{L}^{-1}[sH(s)] = h'(t) + \delta(t)$. Then,

$$g(t) = \sqrt{2a+1}\,\left[h'(t) + \frac{h(t)}{2a+1} + \delta(t)\right] \tag{2.3.23}$$

$$= \sqrt{2a+1}\left[\delta(t) + \frac{a}{2a+1}\exp\left(-\frac{a+1}{2a+1}t\right) I_1\left(\frac{at}{2a+1}\right) - \frac{a}{2a+1}\exp\left(-\frac{a+1}{2a+1}t\right) I_0\left(\frac{at}{2a+1}\right)\right]. \tag{2.3.24}$$

Finally, substituting Equation 2.3.24 into Equation 2.3.17,

$$f(t) = \exp\left(-\frac{a+1}{2a+1}t\right)\left[I_0\left(\frac{at}{2a+1}\right) - I_1\left(\frac{at}{2a+1}\right)\right]. \tag{2.3.25}$$

Problems

Solve the following integral equations:

1. $f(t) = 1 + 2\int_0^t f(t-x)e^{-2x}\,dx$

2. $f(t) = 1 + \int_0^t f(x)\sin(t-x)\,dx$

3. $f(t) = t + \int_0^t f(t-x)e^{-x}\,dx$

4. $f(t) = 4t^2 - \int_0^t f(t-x)e^{-x}\,dx$

5. $f(t) = t^3 + \int_0^t f(x)\sin(t-x)\,dx$

6. $f(t) = 8t^2 - 3\int_0^t f(x)\sin(t-x)\,dx$

7. $f(t) = t^2 - 2\int_0^t f(t-x)\sinh(2x)\,dx$

8. $f(t) = 1 + 2\int_0^t f(t-x)\cos(x)\,dx$

9. $f(t) = e^{2t} + 2\int_0^t f(t-x)\cos(x)\,dx$

10. $f(t) = t^2 + \int_0^t f(x)\sin(t-x)\,dx$

11. $f(t) = e^{-t} - 2\int_0^t f(x)\cos(t-x)\,dx$

12. $f(t) + \int_0^t f(x)(t-x)\,dx = t$

13. $f(t) = 6t + 4\int_0^t f(x)(t-x)^2\,dx$

14. $f(t) = a\sqrt{t} - \int_0^t \frac{f(t-x)}{\sqrt{x}}\,dx$

15. Solve the following equation for $f(t)$ with the condition that $f(0) = 4$:

$$f'(t) = t + \int_0^t f(t-x)\cos(x)\,dx.$$

16. Solve the following equation for $f(t)$ with the condition that $f(0) = 0$:

$$f'(t) = \sin(t) + \int_0^t f(t-x)\cos(x)\,dx.$$

17. During a study of nucleation involving idealized active sites along a boiling surface, Marto and Rohsenow[16] solved the integral equation

$$A = B\sqrt{t} + C\int_0^t \frac{x'(\tau)}{\sqrt{t-\tau}}\,d\tau$$

to find the position $x(t)$ of the liquid/vapor interface. If A, B, and C are constants and $x(0) = 0$, find the solution for them.

18. Solve the following equation for $x(t)$ with the condition that $x(0) = 0$:

$$x(t) + t = \frac{1}{c\sqrt{\pi}}\int_0^t \frac{x'(\tau)}{\sqrt{t-\tau}}\,d\tau,$$

where c is constant.

Step 1: Show that

$$X(s) = -\frac{c}{s^2(c-\sqrt{s}\,)} = -\frac{c(c+\sqrt{s}\,)}{s^2(c^2-s)}.$$

---

[16] Marto, P. J., and W. M. Rohsenow, 1966: Nucleate boiling instability of alkali metals. *J. Heat Transfer*, **88**, 183–193.

Step 2: Use partial fractions to show that

$$X(s) = -\frac{1}{c^2}\left(1 + \frac{\sqrt{s}}{c}\right)\left(\frac{c^2}{s^2} + \frac{1}{s} - \frac{1}{s - c^2}\right).$$

Step 3: Show that

$$x(t) = \frac{1}{c^2}\left\{e^{c^2 t}\left[1 + \operatorname{erf}\left(c\sqrt{t}\right)\right] - c^2 t - 1 - 2c\sqrt{\frac{t}{\pi}}\right\}.$$

19. During a study of the temperature $f(t)$ of a heat reservoir attached to a semi-infinite heat-conducting rod, Huber[17] solved the integral equation

$$f'(t) = \alpha - \frac{\beta}{\sqrt{\pi}}\int_0^t \frac{f'(\tau)}{\sqrt{t-\tau}}\,d\tau,$$

where α and β are constants and $f(0) = 0$. Find $f(t)$ for him.

Step 1: Show that

$$\begin{aligned} F(s) &= \frac{\alpha}{s^{3/2}\left(s^{1/2} + \beta\right)} = \frac{\alpha}{s\left(s - \beta^2\right)} - \frac{\alpha\beta}{s^{3/2}\left(s - \beta^2\right)} \\ &= \frac{\alpha}{\beta^2(s - \beta^2)} - \frac{\alpha}{\beta^2 s} - \frac{\alpha\beta}{s^{3/2}\left(s - \beta^2\right)}. \end{aligned}$$

Step 2: Taking the inverse term by term, show that

$$\begin{aligned} f(t) &= \frac{\alpha}{\beta^2}\left(e^{\beta^2 t} - 1 - \frac{4e^{\beta^2 t}}{\sqrt{\pi}}\int_0^{\beta\sqrt{t}} e^{-x^2} x^2\,dx\right) \\ &= \frac{\alpha}{\beta^2}\left(e^{\beta^2 t} - 1 + \frac{2\beta\sqrt{t}}{\sqrt{\pi}} - \frac{2e^{\beta^2 t}}{\sqrt{\pi}}\int_0^{\beta\sqrt{t}} e^{-x^2}\,dx\right). \end{aligned}$$

20. During the solution of a diffusion problem, Zhdanov, Chikhachev, and Yavlinskii[18] solved an integral equation similar to

$$\int_0^t f(\tau)\left[1 - \operatorname{erf}\left(a\sqrt{t-\tau}\,\right)\right]\,d\tau = at,$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-y^2}\,dy$ is the error function. What should they have found?

Step 1: Show that

$$F(s) = \frac{a}{s} + \frac{a^3}{s^2} + \frac{a^2}{s\sqrt{s + a^2}} + \frac{a^4}{s^2\sqrt{s + a^2}}.$$

---

[17] Huber, A., 1934: Eine Methode zur Bestimmung der Wärme- und Temperaturleitfähigkeit. *Monatsh. Math. Phys.*, **41**, 35–42.

[18] Zhdanov, S. K., A. S. Chikhachev, and Yu. N. Yavlinskii, 1976: Diffusion boundary-value problem for regions with moving boundaries and conservation of particles. *Sov. Phys. Tech. Phys.*, **21**, 883–884.

Step 2: Show that

$$\mathcal{L}\left[t\ \mathrm{erf}(a\sqrt{t}\,) - \frac{1}{2a^2}\mathrm{erf}(a\sqrt{t}\,) + \frac{\sqrt{t}}{a\sqrt{\pi}}e^{-a^2t}\right]$$
$$= -\frac{d}{ds}\left[\frac{a}{s\sqrt{s+a^2}}\right] - \frac{1}{2as\sqrt{s+a^2}} + \frac{1}{2a\,(s+a^2)^{3/2}} = \frac{a}{s^2\sqrt{s+a^2}}.$$

Step 3: Taking the inverse of Step 1 term by term, show that

$$f(t) = a + a^2t + \tfrac{1}{2}a\,\mathrm{erf}(\sqrt{at}\,) + a^3t\,\mathrm{erf}(\sqrt{at}\,) + \frac{a^2\sqrt{t}}{\sqrt{\pi}}e^{-a^2t}.$$

21. The *Laguerre polynomial*[19]

$$y(t) = L_n(t) = \frac{e^t}{n!}\frac{d^n}{dt^n}\left(t^ne^{-t}\right), \qquad n = 0, 1, 2, 3, \ldots$$

satisfies the ordinary differential equation

$$ty'' + (1-t)y' + ny = (ty')' - ty' + ny = 0,$$

with $y(0) = 1$ and $y'(0) = -n$.

Step 1: Using the properties that $\mathcal{L}[f^{(n)}(t)] = s^nF(s) - s^{n-1}f(0) - \cdots - sf^{(n-2)}(0) - f^{(n-1)}(0)$ and Equation $\mathcal{L}[t\,f(t)] = -F'(s)$, show that the Laplace transformed version of this differential equation is

$$Y'(s) = \frac{n+1-s}{s(s-1)}Y(s) = \frac{n}{s-1}Y(s) - \frac{n+1}{s}Y(s),$$

where $Y(s)$ is the Laplace transform of $y(t)$.

Step 2: Using the property that $\mathcal{L}[t\,f(t)] = -F'(s)$ and the convolution theorem, show that Laguerre polynomials are the solution to the integral equation

$$ty(t) = (n+1)\int_0^t y(\tau)\,d\tau - ne^t\int_0^t y(\tau)\,e^{-\tau}\,d\tau.$$

2.4 THE SOLUTION OF THE WAVE EQUATION BY USING LAPLACE TRANSFORMS

The solution of linear partial differential equations by Laplace transforms is the most commonly employed analytic technique after separation of variables. Because the transform consists solely of an integration with respect to time, the transform $U(x,s)$ of the solution of the wave equation $u(x,t)$ is

$$U(x,s) = \int_0^\infty u(x,t)e^{-st}\,dt, \qquad \textbf{(2.4.1)}$$

[19] See Section 5.3 in Andrews, L. C., 1985: *Special Functions for Engineers and Applied Mathematicians*. MacMillan, 357 pp.

assuming that the wave equation only varies in a single spatial variable x and time t.

Partial derivatives involving time have transforms similar to those that we encountered in the case of functions of a single variable. They include

$$\mathcal{L}[u_t(x,t)] = sU(x,s) - u(x,0), \tag{2.4.2}$$

and

$$\mathcal{L}[u_{tt}(x,t)] = s^2U(x,s) - su(x,0) - u_t(x,0). \tag{2.4.3}$$

These transforms introduce the initial conditions via $u(x,0)$ and $u_t(x,0)$. On the other hand, derivatives involving x become

$$\mathcal{L}[u_x(x,t)] = \frac{d}{dx}\{\mathcal{L}[u(x,t)]\} = \frac{dU(x,s)}{dx}, \tag{2.4.4}$$

and

$$\mathcal{L}[u_{xx}(x,t)] = \frac{d^2}{dx^2}\{\mathcal{L}[u(x,t)]\} = \frac{d^2U(x,s)}{dx^2}. \tag{2.4.5}$$

Because the transformation eliminates the time variable, only $U(x,s)$ and its derivatives remain in the equation. Consequently, we transform the partial differential equation into a boundary-value problem involving an ordinary differential equation. Because this equation is often easier to solve than a partial differential equation, the use of Laplace transforms considerably simplifies the original problem. Of course, the Laplace transforms must exist for this technique to work.

The following schematic summarizes the Laplace transform method:

In the following examples, we illustrate transform methods by solving the classic equation of telegraphy as it applies to a uniform transmission line. The line has a resistance R, an inductance L, a capacitance C, and a leakage conductance G per unit length. We denote the current in the direction of positive x by I; V is the voltage drop across the transmission line at the point x. The dependent variables I and V are functions of both distance x along the line and time t.

To derive the differential equations that govern the current and voltage in the line, consider the points A at x and B at $x+\Delta x$ in Figure 2.4.1. The current and voltage at A are $I(x,t)$ and $V(x,t)$; at B, $I + \frac{\partial I}{\partial x}\Delta x$ and $V + \frac{\partial V}{\partial x}\Delta x$. Therefore, the voltage drop from A to B is $-\frac{\partial V}{\partial x}\Delta x$ and the current in the line is $I + \frac{\partial I}{\partial x}\Delta x$. Neglecting terms that are proportional to $(\Delta x)^2$,

$$\left(L\frac{\partial I}{\partial t} + RI\right)\Delta x = -\frac{\partial V}{\partial x}\Delta x. \tag{2.4.6}$$

The voltage drop over the parallel portion HK of the line is V while the current in this portion of the line is $-\frac{\partial I}{\partial x}\Delta x$. Thus,

$$\left(C\frac{\partial V}{\partial t} + GV\right)\Delta x = -\frac{\partial I}{\partial x}\Delta x. \tag{2.4.7}$$

Therefore, the differential equations for I and V are

$$L\frac{\partial I}{\partial t} + RI = -\frac{\partial V}{\partial x}, \tag{2.4.8}$$

and

$$C\frac{\partial V}{\partial t} + GV = -\frac{\partial I}{\partial x}. \tag{2.4.9}$$

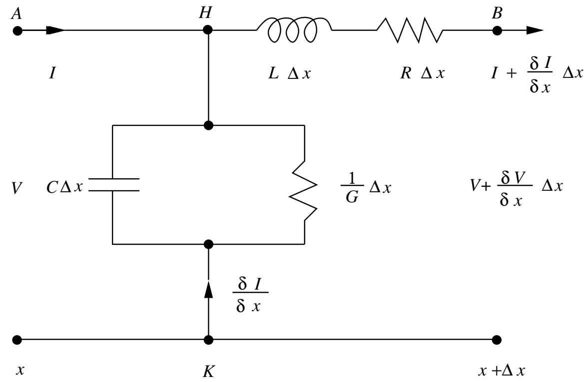


Figure 2.4.1: Schematic of a uniform transmission line.

Turning to the initial conditions, we solve these simultaneous partial differential equations with the initial conditions

$$I(x, 0) = I_0(x), \tag{2.4.10}$$

and

$$V(x, 0) = V_0(x) \tag{2.4.11}$$

for $0 < t$. There are also boundary conditions at the ends of the line; we will introduce them for each specific problem. For example, if the line is short-circuited at $x = a$, $V = 0$ at $x = a$; if there is an open circuit at $x = a$, $I = 0$ at $x = a$.

To solve Equation 2.4.8 and Equation 2.4.9 by Laplace transforms, we take the Laplace transform of both sides of these equations, which yields

$$(Ls + R)\overline{I}(x, s) = -\frac{d\overline{V}(x, s)}{dx} + LI_0(x), \tag{2.4.12}$$

and

$$(Cs + G)\overline{V}(x, s) = -\frac{d\overline{I}(x, s)}{dx} + CV_0(x). \tag{2.4.13}$$

Eliminating $\overline{I}$ gives an ordinary differential equation in $\overline{V}$

$$\frac{d^2\overline{V}}{dx^2} - q^2\overline{V} = L\frac{dI_0(x)}{dx} - C(Ls + R)V_0(x), \tag{2.4.14}$$

where $q^2 = (Ls + R)(Cs + G)$. After finding $\overline{V}$, we may compute $\overline{I}$ from

$$\overline{I} = -\frac{1}{Ls + R}\frac{d\overline{V}}{dx} + \frac{LI_0(x)}{Ls + R}. \tag{2.4.15}$$

At this point we treat several classic cases.

• Example 2.4.1: The semi-infinite transmission line

We consider the problem of a semi-infinite line $0 < x$ with no initial current and charge. The end $x = 0$ has a constant voltage E for $0 < t$.

In this case,

$$\frac{d^2\overline{V}}{dx^2} - q^2\overline{V} = 0, \qquad 0 < x. \tag{2.4.16}$$

The boundary conditions at the ends of the line are

$$V(0,t) = E, \qquad 0 < t, \tag{2.4.17}$$

and $V(x,t)$ is finite as $x \to \infty$. The transform of these boundary conditions is

$$\overline{V}(0,s) = E/s, \qquad \text{and} \qquad \lim_{x\to\infty} \overline{V}(x,s) \to 0. \tag{2.4.18}$$

The general solution of Equation 2.4.16 is

$$\overline{V}(x,s) = Ae^{-qx} + Be^{qx}. \tag{2.4.19}$$

The requirement that $\overline{V}$ remains finite as $x \to \infty$ forces $B = 0$. The boundary condition at $x = 0$ gives $A = E/s$. Thus,

$$\overline{V}(x,s) = \frac{E}{s}\exp\left[-\sqrt{(Ls+R)(Cs+G)}\,x\right]. \tag{2.4.20}$$

We discuss the general case later. However, for the so-called "lossless" line, where $R = G = 0$,

$$\overline{V}(x,s) = \frac{E}{s}\exp(-sx/c), \tag{2.4.21}$$

where $c = 1/\sqrt{LC}$. Consequently,

$$V(x,t) = EH\left(t - \frac{x}{c}\right), \tag{2.4.22}$$

where $H(t)$ is Heaviside's step function. The physical interpretation of this solution is as follows: $V(x,t)$ is zero up to the time x/c, at which time a wave traveling with speed c from $x = 0$ would arrive at the point x. $V(x,t)$ has the constant value E afterwards.

For the so-called "distortionless" line,[20] $R/L = G/C = \rho$,

$$V(x,t) = Ee^{-\rho x/c}H\left(t - \frac{x}{c}\right). \tag{2.4.23}$$

In this case, the disturbance not only propagates with velocity c but also attenuates as we move along the line.

Suppose now, that instead of applying a constant voltage E at $x = 0$, we apply a time-dependent voltage, $f(t)$. The only modification is that in place of Equation 2.4.20,

$$\overline{V}(x,s) = F(s)e^{-qx}. \tag{2.4.24}$$

In the case of the distortionless line, $q = (s+\rho)/c$, this becomes

$$\overline{V}(x,s) = F(s)e^{-(s+\rho)x/c} \tag{2.4.25}$$

and

$$V(x,t) = e^{-\rho x/c}f\left(t - \frac{x}{c}\right)H\left(t - \frac{x}{c}\right). \tag{2.4.26}$$

---

[20] Prechtl and Schürhuber (Prechtl, A., and R. Schürhuber, 2000: Nonuniform distortionless transmission lines. *Electr. Eng. [Berlin]*, **82**, 127–134) generalized this problem to nonuniform transmission lines.

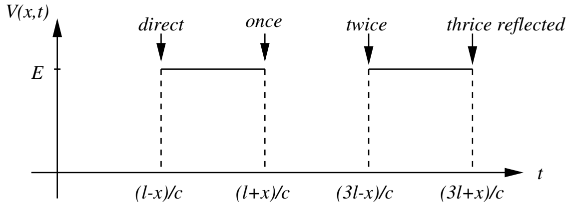


Figure 2.4.2: The voltage within a lossless, finite transmission line of length l as a function of time t.

Thus, our solution shows that the voltage at x is zero up to the time x/c. Afterwards $V(x,t)$ follows the voltage at $x = 0$ with a time lag of x/c and decreases in magnitude by $e^{-\rho x/c}$.□

• Example 2.4.2: The finite transmission line

We now discuss the problem of a finite transmission line $0 < x < l$ with zero initial current and charge. We ground the end $x = 0$ and maintain the end $x = l$ at constant voltage E for $0 < t$.

The transformed partial differential equation becomes

$$\frac{d^2\overline{V}}{dx^2} - q^2\overline{V} = 0, \qquad 0 < x < l. \tag{2.4.27}$$

The boundary conditions are

$$V(0,t) = 0, \quad \text{and} \quad V(l,t) = E, \qquad 0 < t. \tag{2.4.28}$$

The Laplace transform of these boundary conditions is

$$\overline{V}(0,s) = 0, \quad \text{and} \quad \overline{V}(l,s) = E/s. \tag{2.4.29}$$

The solution of Equation 2.4.27 that satisfies the boundary conditions is

$$\overline{V}(x,s) = \frac{E\sinh(qx)}{s\sinh(ql)}. \tag{2.4.30}$$

Let us rewrite Equation 2.4.30 in a form involving negative exponentials and expand the denominator by the binomial theorem,

$$\overline{V}(x,s) = \frac{E}{s}e^{-q(l-x)}\frac{1-e^{-2qx}}{1-e^{-2ql}} \tag{2.4.31}$$

$$= \frac{E}{s}e^{-q(l-x)}(1-e^{-2qx})\left(1+e^{-2ql}+e^{-4ql}+\cdots\right) \tag{2.4.32}$$

$$= \frac{E}{s}\left[e^{-q(l-x)} - e^{-q(l+x)} + e^{-q(3l-x)} - e^{-q(3l+x)} + \cdots\right]. \tag{2.4.33}$$

In the special case of the lossless line where $q = s/c$,

$$\overline{V}(x,s) = \frac{E}{s}\left[e^{-s(l-x)/c} - e^{-s(l+x)/c} + e^{-s(3l-x)/c} - e^{-s(3l+x)/c} + \cdots\right], \tag{2.4.34}$$

or

$$V(x,t) = E\left[H\left(t - \frac{l-x}{c}\right) - H\left(t - \frac{l+x}{c}\right) + H\left(t - \frac{3l-x}{c}\right) - H\left(t - \frac{3l+x}{c}\right) + \cdots\right]. \tag{2.4.35}$$

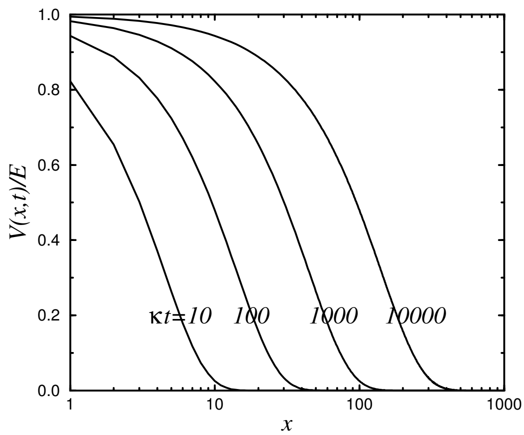


Figure 2.4.3: The voltage within a submarine cable as a function of distance for various values of κt.

We illustrate Equation 2.4.35 in Figure 2.4.2. The voltage at x is zero up to the time $(l - x)/c$, at which time a wave traveling directly from the end $x = l$ would reach the point x. The voltage then has the constant value E up to the time $(l + x)/c$, at which time a wave traveling from the end $x = l$ and reflected back from the end $x = 0$ would arrive. From this time up to the time of arrival of a twice-reflected wave, it has the value zero, and so on. □

• Example 2.4.3: The semi-infinite transmission line reconsidered

In the first example, we showed that the transform of the solution for the semi-infinite line is

$$\overline{V}(x,s) = \frac{E}{s} e^{-qx}, \tag{2.4.36}$$

where $q^2 = (Ls + R)(Cs + G)$. In the case of a lossless line ($R = G = 0$), we found traveling wave solutions.

In this example, we shall examine the case of a submarine cable,[21] where $L = G = 0$. In this special case,

$$\overline{V}(x,s) = \frac{E}{s} e^{-x\sqrt{s/\kappa}}, \tag{2.4.37}$$

where $\kappa = 1/(RC)$. From a table of Laplace transforms,[22] we can immediately invert Equation 2.4.37 and find that

$$V(x,t) = E \operatorname{erfc}\left(\frac{x}{2\sqrt{\kappa t}}\right), \tag{2.4.38}$$

where erfc(·) is the complementary error function. Unlike the traveling wave solution, the voltage diffuses into the cable as time increases. We illustrate Equation 2.4.38 in Figure 2.4.3. □

---

[21] First solved by Thomson, W., 1855: On the theory of the electric telegraph. *Proc. R. Soc. London, Ser. A*, **7**, 382–399.

[22] See Churchill, R. V., 1972: *Operational Mathematics*. McGraw-Hill Book, Section 27.

• **Example 2.4.4: A short-circuited, finite transmission line**

Let us find the voltage of a lossless transmission line of length l that initially has the constant voltage E. At $t = 0$, we ground the line at $x = 0$ while we leave the end $x = l$ insulated.

The transformed partial differential equation now becomes

$$\frac{d^2\overline{V}}{dx^2} - \frac{s^2}{c^2}\overline{V} = -\frac{sE}{c^2}, \qquad (\mathbf{2.4.39})$$

where $c = 1/\sqrt{LC}$. The boundary conditions are

$$\overline{V}(0, s) = 0, \qquad (\mathbf{2.4.40})$$

and

$$\overline{I}(l, s) = -\frac{1}{Ls}\frac{d\overline{V}(l, s)}{dx} = 0 \qquad (\mathbf{2.4.41})$$

from Equation 2.4.15.

The solution to this boundary-value problem is

$$\overline{V}(x, s) = \frac{E}{s} - \frac{E\,\cosh[s(l-x)/c]}{s\,\cosh(sl/c)}. \qquad (\mathbf{2.4.42})$$

The first term on the right side of Equation 2.4.42 is easy to invert and the inversion equals E. The second term is much more difficult to handle. We will use Bromwich's integral.

In Section 2.2 we showed that

$$\mathcal{L}^{-1}\left\{\frac{\cosh[s(l-x)/c]}{s\,\cosh(sl/c)}\right\} = \frac{1}{2\pi i}\int_{c-\infty i}^{c+\infty i}\frac{\cosh[z(l-x)/c]e^{tz}}{z\,\cosh(zl/c)}\,dz. \qquad (\mathbf{2.4.43})$$

To evaluate this integral, we must first locate and then classify the singularities. Using the product formula for the hyperbolic cosine,

$$\frac{\cosh[z(l-x)/c]}{z\,\cosh(zl/c)} = \frac{[1+\frac{4z^2(l-x)^2}{c^2\pi^2}][1+\frac{4z^2(l-x)^2}{9c^2\pi^2}]\cdots}{z[1+\frac{4z^2l^2}{c^2\pi^2}][1+\frac{4z^2l^2}{9c^2\pi^2}]\cdots}. \qquad (\mathbf{2.4.44})$$

This shows that we have an infinite number of simple poles located at $z = 0$, and $z_n = \pm(2n-1)\pi ci/(2l)$, where $n = 1, 2, 3, \ldots$. Therefore, Bromwich's contour can lie along, and just to the right of, the imaginary axis. By Jordan's lemma we close the contour with a semicircle of infinite radius in the left half of the complex plane. Computing the residues,

$$\mathrm{Res}\left\{\frac{\cosh[z(l-x)/c]e^{tz}}{z\,\cosh(zl/c)}; 0\right\} = \lim_{z\to 0}\frac{\cosh[z(l-x)/c]e^{tz}}{\cosh(zl/c)} = 1, \qquad (\mathbf{2.4.45})$$

and

$$\mathrm{Res}\left\{\frac{\cosh[z(l-x)/c]e^{tz}}{z\,\cosh(zl/c)}; z_n\right\} = \lim_{z\to z_n}\frac{(z-z_n)\cosh[z(l-x)/c]e^{tz}}{z\,\cosh(zl/c)} \qquad (\mathbf{2.4.46})$$

$$= \frac{\cosh[(2n-1)\pi(l-x)i/(2l)]\exp[\pm(2n-1)\pi cti/(2l)]}{[(2n-1)\pi i/2]\,\sinh[(2n-1)\pi i/2]} \qquad (\mathbf{2.4.47})$$

$$= \frac{2(-1)^n}{(2n-1)\pi}\cos\left[\frac{(2n-1)\pi(l-x)}{2l}\right]\exp\left[\pm\frac{(2n-1)\pi cti}{2l}\right]. \qquad (\mathbf{2.4.48})$$

Summing the residues and using the relationship that $\cos(t) = (e^{ti} + e^{-ti})/2$,

$$V(x,t) = E - E\left\{1 - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n-1}\cos\left[\frac{(2n-1)\pi(l-x)}{2l}\right]\cos\left[\frac{(2n-1)\pi ct}{2l}\right]\right\} \tag{2.4.49}$$

$$= \frac{4E}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n-1}\cos\left[\frac{(2n-1)\pi(l-x)}{2l}\right]\cos\left[\frac{(2n-1)\pi ct}{2l}\right]. \tag{2.4.50}$$

An alternative to contour integration is to rewrite Equation 2.4.42 as

$$\overline{V}(x,s) = \frac{E}{s}\left\{1 - \frac{e^{-sx/c}\left[1 + e^{-2s(l-x)/c}\right]}{1 + e^{-2sl/c}}\right] \tag{2.4.51}$$

$$= \frac{E}{s}\left[1 - e^{-sx/c} - e^{-s(2l-x)/c} + e^{-s(2l+x)/c} + \cdots\right] \tag{2.4.52}$$

so that

$$V(x,t) = E\left[1 - H\left(t - \frac{x}{c}\right) - H\left(t - \frac{2l-x}{c}\right) + H\left(t - \frac{2l+x}{c}\right) + \cdots\right]. \tag{2.4.53}$$

□

• Example 2.4.5: The general solution of the equation of telegraphy

In this example we solve the equation of telegraphy without any restrictions on R, C, G, or L. We begin by eliminating the dependent variable $I(x,t)$ from the set of equations, Equation 2.4.8 and Equation 2.4.9. This yields

$$CL\frac{\partial^2 V}{\partial t^2} + (GL + RC)\frac{\partial V}{\partial t} + RGV = \frac{\partial^2 V}{\partial x^2}. \tag{2.4.54}$$

We next take the Laplace transform of Equation 2.4.54 assuming that $V(x,0) = f(x)$, and $V_t(x,0) = g(x)$. The transformed version of Equation 2.4.54 is

$$\frac{d^2\overline{V}}{dx^2} - [CLs^2 + (GL + RC)s + RG]\overline{V} = -CLg(x) - (CLs + GL + RC)f(x), \tag{2.4.55}$$

or

$$\frac{d^2\overline{V}}{dx^2} - \frac{(s+\rho)^2 - \sigma^2}{c^2}\overline{V} = -\frac{g(x)}{c^2} - \left(\frac{s}{c^2} + \frac{2\rho}{c^2}\right)f(x), \tag{2.4.56}$$

where $c^2 = 1/LC$, $\rho = c^2(RC + GL)/2$, and $\sigma = c^2(RC - GL)/2$.

We solve Equation 2.4.56 by Fourier transforms with the requirement that the solution dies away as $|x| \to \infty$. The most convenient way of expressing this solution is the convolution product

$$\overline{V}(x,s) = \left[\frac{g(x)}{c} + \left(\frac{s}{c} + \frac{2\rho}{c}\right)f(x)\right] * \frac{\exp[-|x|\sqrt{(s+\rho)^2 - \sigma^2}/c]}{2\sqrt{(s+\rho)^2 - \sigma^2}}. \tag{2.4.57}$$

From a table of Laplace transforms,

$$\mathcal{L}^{-1}\left[\frac{\exp\left(-b\sqrt{s^2 - a^2}\,\right)}{\sqrt{s^2 - a^2}}\right] = I_0\left(a\sqrt{t^2 - b^2}\,\right)H(t - b), \tag{2.4.58}$$

where $b > 0$ and $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. Therefore, by the first shifting theorem,

$$\mathcal{L}^{-1}\left\{\frac{\exp\left[-|x|\sqrt{(s+\rho)^2-\sigma^2}/c\right]}{\sqrt{(s+\rho)^2-\sigma^2}}\right\} = e^{-\rho t} I_0\left[\sigma\sqrt{t^2-(x/c)^2}\right] H\left(t-\frac{|x|}{c}\right). \tag{2.4.59}$$

Using Equation 2.4.59 to invert Equation 2.4.57, we have that

$$\begin{aligned} V(x,t) &= \tfrac{1}{2c}e^{-\rho t}g(x) * I_0\left[\sigma\sqrt{t^2-(x/c)^2}\right] H(t-|x|/c) \\ &+ \tfrac{1}{2c}e^{-\rho t}f(x) * \frac{\partial}{\partial t}\left\{I_0[\sigma\sqrt{t^2-(x/c)^2}]\right\} H(t-|x|/c) \\ &+ \tfrac{\rho}{c}e^{-\rho t}f(x) * I_0\left[\sigma\sqrt{t^2-(x/c)^2}\right] H(t-|x|/c) \\ &+ \tfrac{1}{2}e^{-\rho t}[f(x+ct)+f(x-ct)]. \end{aligned} \tag{2.4.60}$$

The last term in Equation 2.4.60 arises from noting that $sF(s) = \mathcal{L}[f(t)] + f(0)$. If we explicitly write out the convolution, the final form of the solution is

$$\begin{aligned} V(x,t) &= \tfrac{1}{2}e^{-\rho t}[f(x+ct)+f(x-ct)] \\ &+ \tfrac{1}{2c}e^{-\rho t}\int_{x-ct}^{x+ct}[g(\eta)+2\rho f(\eta)] I_0\left[\sigma\sqrt{c^2t^2-(x-\eta)^2}\Big/c\right] d\eta \\ &+ \tfrac{1}{2c}e^{-\rho t}\int_{x-ct}^{x+ct} f(\eta)\frac{\partial}{\partial t}\left\{I_0\left[\sigma\sqrt{c^2t^2-(x-\eta)^2}\Big/c\right]\right\} d\eta. \end{aligned} \tag{2.4.61}$$

There is a straightforward physical interpretation of the first line of Equation 2.4.61. It represents damped progressive waves; one is propagating to the right and the other to the left. In addition to these progressive waves, there is a contribution from the integrals, even after the waves pass. These integrals include all of the points where $f(x)$ and $g(x)$ are nonzero within a distance ct from the point in question. This effect persists through all time, although dying away, and constitutes a residue or tail. Figure 2.4.4 illustrates this for $\rho = 0.1$, $\sigma = 0.2$, and $c = 1$. This figure was obtained using the MATLAB script:

```
% initialize parameters in calculation
clear; dx = 0.1; dt = 0.5; rho_over_c = 0.1; sigma_over_c = 0.2;
X=[-10:dx:10]; T = [0:dt:10]; % compute locations of x and t
for j=1:length(T); t = T(j);
for i=1:length(X); x = X(i);
  XX(i,j) = x; TT(i,j) = t; deta_i = 0.05 % set up grid
% compute characteristics x+ct and x-ct
  characteristic_1 = x - t; characteristic_2 = x + t;
% compute first term in Equation 2.4.61
  F = inline('stepfun(x,-1.0001)-stepfun(x,1.0001)');
  u(i,j ) = F(characteristic_1) + F(characteristic_2);
% find the upper and lower limits of the integration
  upper = characteristic_2; lower = characteristic_1;
  if t > 0 & upper > -1 & lower < 1
  if upper > 1 upper = 1; end
  if lower < -1 lower = -1; end
```

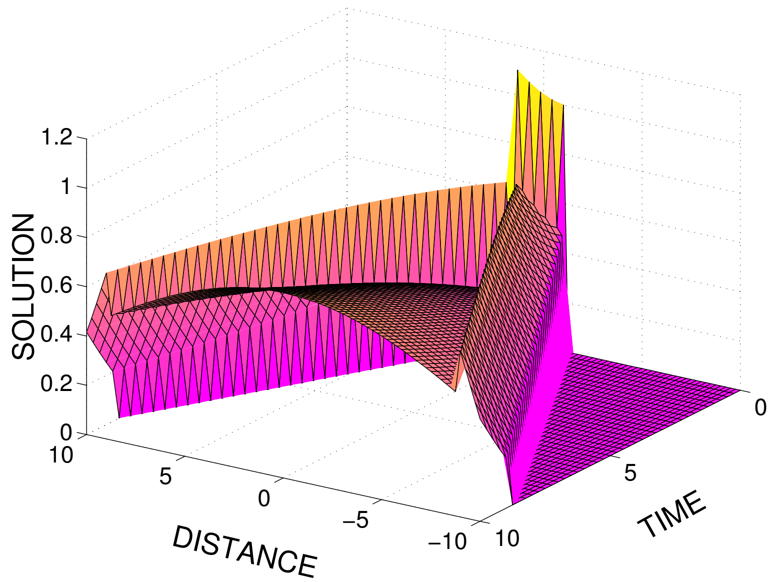


Figure 2.4.4: The evolution of the voltage with time given by the general equation of telegraphy for initial conditions and parameters stated in the text.

```
% set up parameters needed for integration
  interval = upper-lower;
  NN = interval / deta_i;
  if mod(NN,2) > 0 NN = NN + 1; end;
  deta = interval / NN;
% compute integrals in Equation 2.4.61 by Simpson's rule
% sum1 deals with the first integral while sum2 is the second
  sum1 = 0; sum2 = 0; eta = lower;
  for k = 0:2:NN-2
    arg = sigma_over_c * sqrt(t*t-(x-eta)*(x-eta));
    sum1 = sum1 + besseli(0,arg);
    if (arg == 0)
      sum2 = sum2 + 0.5 * sigma_over_c * t;
    else
      sum2 = sum2 + t * besseli(1,arg) / arg; end
    eta = eta + deta;
    arg = sigma_over_c * sqrt(t*t-(x-eta)*(x-eta));
    sum1 = sum1 + 4*besseli(0,arg);
    if (arg == 0)
      sum2 = sum2 + 4 * 0.5 * sigma_over_c * t;
    else
      sum2 = sum2 + 4 * t * besseli(1,arg) / arg; end
    eta = eta + deta;
    arg = sigma_over_c * sqrt(t*t-(x-eta)*(x-eta));
    sum1 = sum1 + besseli(0,arg);
    if (arg == 0)
      sum2 = sum2 + 0.5 * sigma_over_c * t;
```

```
    else
      sum2 = sum2 + t * besseli(1,arg) / arg; end
    end
  u(i,j) = u(i,j) + 2 * rho_over_c * deta * sum1 / 3 ...
         + sigma_over_c * deta * sum2 / 3;
  end
% multiply final answer by damping coefficient
  u(i,j) = 0.5 * exp(-rho_over_c * t) * u(i,j);
end;end;
% plot results
mesh(XX,TT,real(u)); colormap spring;
xlabel('DISTANCE','Fontsize',20); ylabel('TIME','Fontsize',20)
zlabel('SOLUTION','Fontsize',20)
```

We evaluated the integrals by Simpson's rule for the initial conditions $f(x) = H(x+1) - H(x-1)$, and $g(x) = 0$. If there was no loss, then two pulses would propagate to the left and right. However, with resistance and leakage the waves leave a residue after their leading edge has passed.

Problems

1. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad 0 < t,$$

with the boundary conditions $u(0,t) = u(1,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = 0$, $u_t(x,0) = 1$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - s^2U(x,s) = -1, \qquad 0 < x < 1,$$

subject to the boundary conditions $U(0,s) = U(1,s) = 1$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{1-\cosh(sx)}{s^2} + \frac{[\cosh(s)-1]\sinh(sx)}{s^2\,\sinh(s)}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s_n = \pm n\pi i$ with $n = 1, 2, 3, \ldots$ and a removable pole at $s = 0$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{4}{\pi^2}\sum_{m=1}^{\infty} \frac{\sin[(2m-1)\pi x]\sin[(2m-1)\pi t]}{(2m-1)^2}.$$

2. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad 0 < t,$$

with the boundary conditions $u(0,t) = u_x(1,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = 0$, $u_t(x,0) = x$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - s^2U(x,s) = -x, \qquad 0 < x < 1,$$

with the boundary condition $U(0,s) = U'(1,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{xs\,\cosh(s) - \sinh(sx)}{s^3\,\cosh(s)}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s_n = \pm(2n-1)\pi i/2$ with $n = 1, 2, 3, \ldots$ and a removable pole at $s = 0$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{16}{\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{(2n-1)^3}\sin\left[\frac{(2n-1)\pi x}{2}\right]\sin\left[\frac{(2n-1)\pi t}{2}\right].$$

3. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad 0 < t,$$

with the boundary conditions $u(0,t) = u(1,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = \sin(\pi x)$, $u_t(x,0) = -\sin(\pi x)$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - s^2U(x,s) = -s\sin(\pi x) + \sin(\pi x), \qquad 0 < x < 1,$$

with the boundary conditions $U(0,s) = U(1,s) = 0$.

Step 2: Show that the solution to the previous step is $U(x,s) = (s-1)\sin(\pi x)/(s^2+\pi^2)$.

Step 3: Inverting by inspection, show that $u(x,t) = \sin(\pi x)\cos(\pi t) - \sin(\pi x)\sin(\pi t)/\pi$.

4. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < a, \quad 0 < t,$$

with the boundary conditions $u(0,t) = \sin(\omega t)$, $u(a,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = u_t(x,0) = 0$, $0 < x < a$. Assume that $\omega a/c$ is *not* an integer multiple of π. Why?

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s^2}{c^2}U(x,s) = 0, \qquad 0 < x < a,$$

with the boundary condition $U(0,s) = \omega/(s^2+\omega^2)$ and $U(a,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{\omega \ \sinh[s(a-x)/c]}{(s^2+\omega^2) \ \sinh(sa/c)}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s = \pm\omega i$ and $s_n = \pm n\pi ci/a$ with $n = 1,2,3,\ldots$ and a removable pole at $s = 0$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{\sin[\omega(a-x)/c]}{\sin(\omega a/c)} \sin(\omega t) - \frac{2\omega a}{c} \sum_{n=1}^{\infty} \frac{\sin(n\pi x/a)}{n^2\pi^2 - a^2\omega^2/c^2} \sin\left(\frac{n\pi ct}{a}\right).$$

5. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t,$$

with the boundary conditions $u_x(0,t) = -f(t)$, $u_x(L,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = u_t(x,0) = 0$, $0 < x < L$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s^2}{c^2}U(x,s) = 0, \qquad 0 < x < L,$$

with the boundary conditions $U'(0,s) = -F(s)$ and $U'(L,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{cF(s)\cosh[s(L-x)/c]}{s \ \sinh(sL/c)}.$$

Step 3: Replacing sinh and cosh by their definitions and expanding the denominator as a geometric series, show that

$$U(x,s) = \frac{cF(s)}{s}\left[e^{-sx/c} + e^{-s(2L-x)/c}\right]\left(1 + e^{-2sL/c} + e^{-4sL/c} + \cdots\right).$$

Step 4: Multiplying everything out and inverting term by term, show that

$$\begin{aligned} u(x,t) = {} & c\sum_{n=0}^{\infty} f(t - x/c - 2nL/c)H(t - x/c - 2nL/c) \\ & + c\sum_{m=1}^{\infty} f(t + x/c - 2mL/c)H(t + x/c - 2mL/c). \end{aligned}$$

6. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} - q'(t), \qquad a < x < b, \quad 0 < t,$$

with the boundary conditions $u(a,t) = u_x(b,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = 0$, $u_t(x,0) = -q(0)$, $a < x < b$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$c^2 \frac{d^2 U(x,s)}{dx^2} - s^2 U(x,s) = sQ(s), \qquad a < x < b,$$

with the boundary conditions $U(a,s) = U'(b,s) = 0$.

Step 2: Show that the eigenfunctions $\sin[k_n(x-a)]$, where $k_n = (2n+1)\pi/[2(b-a)]$ and $n = 0, 1, 2, \ldots$, satisfy the boundary conditions.

Step 3: Expand the right side of the differential equation using an eigenfunction expansion consisting of $\sin[k_n(x-a)]$. Show that

$$sQ(s) = \frac{4sQ(s)}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin\left[\frac{(2n+1)\pi(x-a)}{2(b-a)}\right].$$

Step 4: Assuming that

$$U(x,s) = \sum_{n=0}^{\infty} A_n \sin\left[\frac{(2n+1)\pi(x-a)}{2(b-a)}\right],$$

show by direct substitution that

$$A_n = -\frac{4sc^2 Q(s)}{\pi(2n+1)} \left[s^2 + \frac{(2n+1)^2\pi^2 c^2}{4(b-a)^2}\right]^{-1}.$$

Step 5: Invert $U(x,s)$ term by term and show that

$$u(x,t) = -\frac{4c^2}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin\left[\frac{(2n+1)\pi(x-a)}{2(b-a)}\right] \int_0^t q(\tau) \cos\left[\frac{(2n+1)\pi c(t-\tau)}{2(b-a)}\right] d\tau.$$

7. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = te^{-x}, \qquad 0 < x < \infty, \quad 0 < t,$$

with the boundary conditions $u(0,t) = 1 - e^{-t}$, $\lim_{x\to\infty} |u(x,t)| \sim x^n$, n finite, $0 < t$, and the initial conditions $u(x,0) = 0$, $u_t(x,0) = x$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2 U(x,s)}{dx^2} - s^2 U(x,s) = -x - \frac{e^{-x}}{s^2}, \qquad 0 < x < \infty,$$

with the boundary conditions

$$U(0,s) = \frac{1}{s} - \frac{1}{s+1} \quad \text{and} \quad \lim_{x\to\infty} |U(x,s)| \sim x^n.$$

Step 2: Show that the solution to the previous step is

$$U(x,s) = \left(\frac{1}{s} - \frac{1}{s+1} + \frac{1}{s^2} - \frac{1}{s^2-1}\right) e^{-sx} + \frac{x}{s^2} - \frac{e^{-x}}{s^2} + \frac{e^{-x}}{s^2-1}.$$

Step 3: Inverting term by term, show that

$$u(x,t) = xt - te^{-x} + \sinh(t)e^{-x} + \left[1 - e^{-(t-x)} + t - x - \sinh(t-x)\right] H(t-x).$$

8. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = xe^{-t}, \qquad 0 < x < \infty, \quad 0 < t,$$

with the boundary conditions $u(0,t) = \cos(t)$, $\lim_{x\to\infty} |u(x,t)| \sim x^n$, n finite, $0 < t$, and the initial conditions $u(x,0) = 1$, $u_t(x,0) = 0$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - s^2U(x,s) = -s - \frac{x}{s+1}, \qquad 0 < x < \infty,$$

with the boundary conditions $U(0,s) = s/(s^2+1)$ and $\lim_{x\to\infty} |U(x,s)| \sim x^n$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \left(\frac{s}{s^2+1} - \frac{1}{s}\right) e^{-sx} + \frac{1}{s} + \frac{x}{s^2} - \frac{x}{s} + \frac{x}{s+1}.$$

Step 3: Inverting term by term, show that $u(x,t) = 1 + xt - x + xe^{-t} + [\cos(t-x) - 1]H(t-x)$.

9. Use transform methods to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t,$$

with the boundary conditions

$$u(0,t) = 0, \qquad \frac{\partial^2 u(L,t)}{\partial t^2} + \frac{k}{m}\frac{\partial u(L,t)}{\partial x} = g, \qquad 0 < t,$$

and the initial conditions $u(x,0) = u_t(x,0) = 0$, $0 < x < L$, where k, m, and g are constants.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - s^2U(x,s) = 0, \qquad 0 < x < L,$$

with the boundary conditions $U(0,s) = 0$ and $s^2U(L,s) + \omega^2U'(L,s) = g/s$, where $\omega^2 = k/m$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{g\ \sinh(sx)}{s^2[s\ \sinh(sL) + \omega^2 \cosh(sL)]}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s = 0$ and $s_n = \pm\lambda_n i$, where $\lambda_n = \omega^2 \cot(\lambda_n L)$ with $n = 1, 2, 3, \ldots$.

Step 4: Using Bromwich's integral, show that

$$u(x,t) = \frac{gx}{\omega^2} - \frac{2g\omega^2}{L}\sum_{n=1}^{\infty}\frac{\sin(\lambda_n x)\cos(\lambda_n t)}{\lambda_n^2(\omega^4 + \omega^2/L + \lambda_n^2)\sin(\lambda_n L)}.$$

10. Use transform methods[23] to solve the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial}{\partial x}\left(x\frac{\partial u}{\partial x}\right), \qquad 0 < x < 1, \quad 0 < t,$$

with the boundary conditions $\lim_{x\to 0}|u(x,t)| < \infty$ and $u(1,t) = A\sin(\omega t)$, $0 < t$, and the initial conditions $u(x,0) = u_t(x,0) = 0$, $0 < x < 1$. Assume that $2\omega \neq c\beta_n$, where $J_0(\beta_n) = 0$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d}{dx}\left[x\frac{dU(x,s)}{dx}\right] - \frac{s^2}{c^2}U(x,s) = 0, \qquad 0 < x < 1.$$

with the boundary conditions $\lim_{x\to 0}|U(x,s)| < \infty$ and $U(1,s) = A\omega/(s^2+\omega^2)$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{A\omega}{s^2+\omega^2}\frac{I_0(2s\sqrt{x}/c)}{I_0(2s/c)}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s = \pm\omega i$ and $s_n = \pm c\beta_n i/2$, where $J_0(\beta_n) = 0$, $n = 1, 2, 3, \ldots$

[23] Suggested by a problem solved by Brown, J., 1975: Stresses in towed cables during re-entry. *J. Spacecr. Rockets*, **12**, 524–527.

Step 4: Using Bromwich's integral, show that

$$u(x,t) = A\frac{J_0(2\omega\sqrt{x}/c)}{J_0(2\omega/c)}\sin(\omega t) + Ac\omega\sum_{n=1}^{\infty}\frac{J_0(\beta_n\sqrt{x}\,)\sin(\beta_n ct/2)}{(\omega^2 - c^2\beta_n^2/4)J_1(\beta_n)}.$$

11. A lossless transmission line of length ℓ has a constant voltage E applied to the end $x = 0$ while we insulate the other end $[V_x(\ell, t) = 0]$. Find the voltage at any point on the line if the initial current and charge are zero.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s^2}{c^2}U(x,s) = 0, \qquad 0 < x < \ell,$$

with the boundary conditions $U(0, s) = E/s$ and $U'(\ell, s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{E\ \cosh[s(\ell - x)/c]}{s\ \cosh(s\ell/c)}.$$

Step 3: Show that $U(x, s)$ has simple poles at $s = 0$ and $s_n = \pm(2n - 1)c\pi i/(2\ell)$ with $n = 1, 2, 3, \ldots$

Step 4: Using Bromwich's integral, show that

$$u(x,t) = E - \frac{4E}{\pi}\sum_{n=1}^{\infty}\frac{1}{2n-1}\sin\left[\frac{(2n-1)\pi x}{2\ell}\right]\cos\left[\frac{(2n-1)c\pi t}{2\ell}\right].$$

Step 5: An alternative approach is to replace the hyperbolic functions with their exponential definitions. Then,

$$\begin{aligned} U(x,s) = {} & \frac{E}{s}\left[e^{-sx/c} - e^{-s(x+2\ell)/c} + e^{-s(x+4\ell)/c} - \cdots\right] \\ & + \frac{E}{s}\left[e^{-s(2\ell-x)/c} - e^{-s(4\ell-x)/c} + e^{-s(6\ell-x)/c} - \cdots\right] \end{aligned}$$

after using the summation rule for the geometric series. Take the inverse by inspection and show that

$$u(x,t) = E\sum_{n=0}^{\infty}(-1)^n H\left(t - \frac{x+2n\ell}{c}\right) + E\sum_{n=0}^{\infty}(-1)^n H\left\{t - \frac{[(2n+2)\ell - x]}{c}\right\}.$$

12. Solve the equation of telegraphy without leakage

$$\frac{\partial^2 u}{\partial x^2} = CR\frac{\partial u}{\partial t} + CL\frac{\partial^2 u}{\partial t^2}, \qquad 0 < x < \ell, \quad 0 < t,$$

subject to the boundary conditions $u(0,t) = 0$, $u(\ell,t) = E$, $0 < t$, and the initial conditions $u(x,0) = u_t(x,0) = 0$, $0 < x < \ell$. Assume that $4\pi^2 L/CR^2\ell^2 > 1$. Why?

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s^2}{c^2}U(x,s) = 0, \qquad 0 < x < \ell,$$

with the boundary conditions $U(0,s) = E/s$ and $U'(\ell,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{E\ \cosh[s(\ell - x)/c]}{s\ \cosh(s\ell/c)}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s = 0$ and $s_n = \pm(2n-1)c\pi i/(2\ell)$ with $n = 1, 2, 3, \ldots$

Step 4: Using Bromwich's integral, show that

$$\frac{u(x,t)}{E} = \frac{x}{\ell} - \frac{2}{\pi}e^{-t/2T}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin\left(\frac{n\pi x}{\ell}\right)\left[\frac{\sin(t\sqrt{n^2\delta^2-1}/2T\,)}{\sqrt{n^2\delta^2-1}} + \cos\left(t\sqrt{n^2\delta^2-1}/2T\,\right)\right].$$

13. The pressure and velocity oscillations from water hammer in a pipe without friction[24] are given by the equations

$$\frac{\partial p}{\partial t} = -\rho c^2\frac{\partial u}{\partial x}, \qquad \text{and} \qquad \frac{\partial u}{\partial t} = -\frac{1}{\rho}\frac{\partial p}{\partial x},$$

where $p(x,t)$ denotes the pressure perturbation, $u(x,t)$ is the velocity perturbation, c is the speed of sound in water, and ρ is the density of water. These two first-order partial differential equations can be combined to yield

$$\frac{\partial^2 p}{\partial t^2} = c^2\frac{\partial^2 p}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t.$$

Find the solution to this partial differential equation if $p(0,t) = p_0$, and $u(L,t) = 0$, and the initial conditions are $p(x,0) = p_0$, $p_t(x,0) = 0$, and $u(x,0) = u_0$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2P(x,s)}{dx^2} - \frac{s^2}{c^2}P(x,s) = -\frac{s}{c^2}p_0, \qquad 0 < x < L,$$

with the boundary conditions $P(0,s) = p_0/s$ and $P'(L,s) = \rho u_0$.

Step 2: Show that the solution to the previous step is

$$P(x,s) = \frac{p_0}{s} + \frac{\rho u_0 c\ \sinh(sx/c)}{s\ \cosh(sL/c)}.$$

[24] See Rich, G. R., 1945: Water-hammer analysis by the Laplace-Mellin transformation. *Trans. ASME*, **67**, 361–376.

Step 3: Show that $P(x,s)$ has simple poles at $s_n = \pm(2n-1)c\pi i/(2L)$ with $n = 1, 2, 3, \ldots$ and a removable singularity at $s = 0$.

Step 4: Using Bromwich's integral, show that

$$p(x,t) = p_0 - \frac{4\rho u_0 c}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1} \sin\left[\frac{(2n-1)\pi x}{2L}\right] \sin\left[\frac{(2n-1)c\pi t}{2L}\right].$$

14. Use Laplace transforms to solve the wave equation[25]

$$\frac{\partial^2 u}{\partial t^2} = c^2 \left(\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} - \frac{2u}{r^2} \right), \qquad a < r < \infty, \quad 0 < t,$$

subject to the boundary conditions that

$$u(a,t) = A\left(1 - e^{-ct/a}\right) H(t), \quad \lim_{r\to\infty} u(r,t) \to 0, \qquad 0 < t,$$

and the initial conditions that $u(r,0) = u_t(r,0) = 0$, $a < r < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(r,s)}{dr^2} + \frac{2}{r}\frac{dU(r,s)}{dr} - \frac{2}{r^2}U(r,s) - \frac{s^2}{c^2}U(r,s) = 0, \qquad a < r < \infty,$$

with the boundary condition

$$U(a,s) = \frac{A}{s} - \frac{A}{s + c/a}, \qquad \lim_{r\to\infty} |U(r,s)| < \infty.$$

Step 2: Show that the solution to the previous step is

$$U(r,s) = A\left[\frac{a^2}{sr^2} - \frac{a^2}{r^2(s+c/a)} - \frac{c}{a}\left(\frac{a^2}{r^2} - \frac{a}{r}\right)\frac{1}{(s+c/a)^2}\right] e^{-s(r-a)/c}.$$

Step 3: Use tables and the second shifting theorem to show that

$$u(r,t) = A\left\{\frac{a^2}{r^2} - \left[\frac{a^2}{r^2} + \frac{c\tau}{a}\left(\frac{a^2}{r^2} - \frac{a}{r}\right)\right] e^{-c\tau/a}\right\} H(\tau),$$

where $\tau = t - (r-a)/c$.

15. Use Laplace transforms to solve the wave equation[26]

$$\frac{\partial^2 (ru)}{\partial t^2} = c^2 \frac{\partial^2 (ru)}{\partial r^2}, \qquad a < r < \infty, \quad 0 < t,$$

[25] Wolf, J. P., and G. R. Darbre, 1986: Time-domain boundary element method in visco-elasticity with application to a spherical cavity. *Soil Dynam. Earthq. Eng.*, **5**, 138–148.

[26] Originally solved using Fourier transforms by Sharpe, J. A., 1942: The production of elastic waves by explosion pressures. I. Theory and empirical field observations. *Geophysics*, **7**, 144–154.

subject to the boundary conditions that

$$-\rho c^2\left(\frac{\partial^2 u}{\partial r^2}+\frac{2}{3r}\frac{\partial u}{\partial r}\right)\bigg|_{r=a}=p_0 e^{-\alpha t}H(t),\qquad \lim_{r\to\infty}u(r,t)\to 0,\qquad 0<t,$$

where $\alpha>0$, and the initial conditions that $u(r,0)=u_t(r,0)=0$, $a<r<\infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2[rU(r,s)]}{dr^2}-\frac{s^2}{c^2}[rU(r,s)]=0,\qquad a<r<\infty.$$

with the boundary condition

$$-\rho c^2\left[\frac{d^2U(a,s)}{dr^2}+\frac{2}{3a}\frac{dU(a,s)}{dr}\right]=\frac{p_0}{s+\alpha}\quad\text{and}\quad \lim_{r\to\infty}|U(r,s)|<\infty.$$

Step 2: Show that the solution to the previous step is

$$U(r,s)=-\frac{ap_0\exp[-s(r-a)/c]}{\rho r(s+\alpha)[s^2+4sc/(3a)+4c^2/(3a^2)]}.$$

Step 3: Show that $U(r,s)$ has three simple poles, $s=-\alpha$ and $s=-\beta/\sqrt{2}\pm\beta i$, where $\beta=2\sqrt{2}c/(3a)$.

Step 4: Use Bromwich's integral and show that

$$u(r,t)=\frac{ap_0}{\rho r[(\beta/\sqrt{2}-\alpha)^2+\beta^2]}\left\{e^{-\beta\tau/\sqrt{2}}\left[\left(\frac{1}{\sqrt{2}}-\frac{\alpha}{\beta}\right)\sin(\beta\tau)+\cos(\beta\tau)\right]-e^{-\alpha\tau}\right\}H(\tau),$$

where $\tau=t-(r-a)/c$.

16. Consider a vertical rod or column of length L that is supported at both ends. The elastic waves that arise when the support at the bottom is suddenly removed are governed by the wave equation[27]

$$\frac{\partial^2 u}{\partial t^2}=c^2\frac{\partial^2 u}{\partial x^2}+g,\qquad 0<x<L,\quad 0<t,$$

where g denotes the gravitational acceleration, $c^2=E/\rho$, E is Young's modulus, and ρ is the mass density. Find the wave solution if the boundary conditions are $u_x(0,t)=u_x(L,t)=0$, $0<t$, and the initial conditions are

$$u(x,0)=-\frac{gx^2}{2c^2},\qquad \frac{\partial u(x,0)}{\partial t}=0,\qquad 0<x<L.$$

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2}-\frac{s^2}{c^2}U(x,s)=\frac{sgx^2}{2c^4}-\frac{g}{sc^2},\qquad 0<x<L,$$

[27] See Hall, L. H., 1953: Longitudinal vibrations of a vertical column by the method of Laplace transform. *Am. J. Phys.*, **21**, 287–292.

with $U'(0,s) = U'(L,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{gL\cosh(sx/c)}{cs^2\sinh(sL/c)} - \frac{gx^2}{2sc^2}.$$

Step 3: Show that $U(x,s)$ has poles that are located at $s = 0$ and $s_n = \pm n\pi ci/L$, where $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{gt^2}{2} - \frac{gL^2}{6c^2} - \frac{2gL^2}{c^2\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos\left(\frac{n\pi x}{L}\right)\cos\left(\frac{n\pi ct}{L}\right).$$

17. Use Laplace transforms to solve the hyperbolic equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} + 1 = 0, \qquad 0 < x < 1, \quad 0 < t,$$

subject to the boundary conditions that $u_x(0,t) = 0$, $u_x(1,t) = 1$, $0 < t$, and the initial conditions that $u(x,0) = u_t(x,0) = 0$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - s^2U(x,s) = \frac{1}{s} + x^2 - 1, \qquad 0 < x < 1,$$

with $U'(0,s) = 0$ and $U'(1,s) = 1/s$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{1-x^2}{s^2} - \frac{1}{s^3} - \frac{2}{s^4} + \frac{\cosh(sx)}{s^2\sinh(s)} + \frac{2\cosh(sx)}{s^3\sinh(s)}.$$

Step 3: Show that $U(x,s)$ has poles that are located at $s = 0$ and $z_n = \pm n\pi i$, where $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{2t}{3} + \frac{x^2}{2} - \frac{1}{6} - 2\sum_{n=1}^{\infty}(-1)^n\cos(n\pi x)\left[\frac{\cos(n\pi t)}{n^2\pi^2} + \frac{2\sin(n\pi t)}{n^3\pi^3}\right].$$

18. Solve the telegraph-like equation[28]

$$\frac{\partial^2 u}{\partial t^2} + k\frac{\partial u}{\partial t} = c^2\left(\frac{\partial^2 u}{\partial x^2} + \alpha\frac{\partial u}{\partial x}\right), \qquad 0 < x < \infty, \quad 0 \le t,$$

---

[28] See Abbott, M. R., 1959: The downstream effect of closing a barrier across an estuary with particular reference to the Thames. *Proc. R. Soc. London, Ser. A*, **251**, 426–439.

subject to the boundary conditions $u_x(0,t) = -u_0\delta(t)$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 \le t$, and the initial conditions $u(x,0) = u_0$, $u_t(x,0) = 0$, $0 < x < \infty$, with $\alpha c > k$. Here $\delta(t)$ denotes the Dirac delta function.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} + \alpha\frac{dU(x,s)}{dx} - \left(\frac{s^2+ks}{c^2}\right)U(x,s) = -\left(\frac{s+k}{c^2}\right)u_0, \qquad 0 < x < \infty,$$

with $U'(0,s) = -u_0$, and $\lim_{x\to\infty} U(x,s) \to 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{u_0}{s} + u_0 e^{-\alpha x/2}\,\frac{\exp\left[-x\sqrt{\left(s+\frac{k}{2}\right)^2 + a^2/c}\right]}{\frac{\alpha}{2} + \sqrt{(s+\frac{k}{2})^2 + a^2/c}},$$

where $4a^2 = \alpha^2c^2 - k^2 > 0$.

Step 3: Using the first and second shifting theorems and the property that

$$F\left(\sqrt{s^2+a^2}\right) = \mathcal{L}\left[f(t) - a\int_0^t \frac{J_1\left(a\sqrt{t^2-\tau^2}\,\right)}{\sqrt{t^2-\tau^2}}\tau f(\tau)\,d\tau\right],$$

show that

$$u(x,t) = u_0 + u_0 c e^{-kt/2} H(t - x/c)\left[e^{-\alpha ct/2} - a\int_{x/c}^t \frac{J_1\left(a\sqrt{t^2-\tau^2}\,\right)}{\sqrt{t^2-\tau^2}}\tau e^{-\alpha c\tau/2} d\tau\right].$$

19. As an electric locomotive travels down a track at the speed V, the pantograph (the metallic framework that connects the overhead power lines to the locomotive) pushes up the line with a force P. Let us find the behavior[29] of the overhead wire as a pantograph passes between two supports of the electrical cable that are located a distance L apart. We model this system as a vibrating string with a point load:

$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2} + \frac{P}{\rho V}\,\delta\left(t - \frac{x}{V}\right), \qquad 0 < x < L, \quad 0 < t.$$

Let us assume that the wire is initially at rest [$u(x,0) = u_t(x,0) = 0$ for $0 < x < L$] and fixed at both ends [$u(0,t) = u(L,t) = 0$ for $0 < t$].

Step 1: Take the Laplace transform of the partial differential equation and show that

$$s^2U(x,s) = c^2\frac{d^2U(x,s)}{dx^2} + \frac{P}{\rho V}e^{-xs/V}, \qquad 0 < x < L.$$

---

[29] See Oda, O., and Y. Ooura, 1976: Vibrations of catenary overhead wire. *Q. Rep., (Tokyo) Railway Tech. Res. Inst.*, **17**, 134–135.

Step 2: Solve the ordinary differential equation in Step 1 as a Fourier half-range sine series

$$U(x,s) = \sum_{n=1}^{\infty} B_n(s) \sin\left(\frac{n\pi x}{L}\right),$$

where

$$B_n(s) = \frac{2P\beta_n}{\rho L(\beta_n^2 - \alpha_n^2)} \left[\frac{1}{s^2+\alpha_n^2} - \frac{1}{s^2+\beta_n^2}\right] \left[1 - (-1)^n e^{-Ls/V}\right],$$

$\alpha_n = n\pi c/L$ and $\beta_n = n\pi V/L$. This solution satisfies the boundary conditions.

Step 3: By inverting the solution in Step 2, show that

$$\begin{aligned} u(x,t) = {} & \frac{2P}{\rho L} \sum_{n=1}^{\infty} \left[\frac{\sin(\beta_n t)}{\alpha_n^2 - \beta_n^2} - \frac{V}{c}\,\frac{\sin(\alpha_n t)}{\alpha_n^2 - \beta_n^2}\right] \sin\left(\frac{n\pi x}{L}\right) \\ & - \frac{2P}{\rho L} H\left(t - \frac{L}{V}\right) \sum_{n=1}^{\infty} (-1)^n \sin\left(\frac{n\pi x}{L}\right) \left\{\frac{\sin[\beta_n(t - L/V)]}{\alpha_n^2 - \beta_n^2} - \frac{V}{c}\,\frac{\sin[\alpha_n(t-L/V)]}{\alpha_n^2 - \beta_n^2}\right\}, \end{aligned}$$

or

$$\begin{aligned} u(x,t) = {} & \frac{2P}{\rho L} \sum_{n=1}^{\infty} \left[\frac{\sin(\beta_n t)}{\alpha_n^2 - \beta_n^2} - \frac{V}{c}\,\frac{\sin(\alpha_n t)}{\alpha_n^2 - \beta_n^2}\right] \sin\left(\frac{n\pi x}{L}\right) \\ & - \frac{2P}{\rho L} H\left(t - \frac{L}{V}\right) \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right) \left\{\frac{\sin(\beta_n t)}{\alpha_n^2 - \beta_n^2} - \frac{V}{c}(-1)^n \frac{\sin[\alpha_n(t-L/V)]}{\alpha_n^2 - \beta_n^2}\right\}. \end{aligned}$$

The first term in both summations represents the static uplift on the line; this term disappears after the pantograph passes. The second term in both summations represents the vibrations excited by the traveling force. Even after the pantograph passes, they continue to exist.

20. Solve the wave equation

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial r^2} - \frac{1}{r}\frac{\partial u}{\partial r} + \frac{u}{r^2} = \frac{\delta(r-\alpha)}{\alpha^2}, \qquad 0 \le r < a, \quad 0 < t,$$

where $0 < \alpha < a$, subject to the boundary conditions $\lim_{r\to 0} |u(r,t)| < \infty$, $u_r(a,t) + h\,u(a,t)/a = 0$, $0 < t$, and the initial conditions $u(r,0) = u_t(r,0) = 0$, $0 \le r < a$.

Step 1: Take the Laplace transform of the partial differential equation and show that

$$\frac{d^2U(r,s)}{dr^2} + \frac{1}{r}\frac{dU(r,s)}{dr} - \left(\frac{s^2}{c^2} + \frac{1}{r^2}\right) U(r,s) = -\frac{\delta(r-\alpha)}{s\alpha^2}, \qquad 0 \le r < a,$$

with $\lim_{r\to 0} |U(r,s)| < \infty$ and $U'(a,s) + \frac{h}{a}U(a,s) = 0$.

Step 2: Show that the Dirac delta function can be reexpressed as the Fourier-Bessel series

$$\delta(r-\alpha) = \frac{2\alpha}{a^2} \sum_{n=1}^{\infty} \frac{\beta_n^2\, J_1(\beta_n \alpha/a)}{(\beta_n^2 + h^2 - 1) J_1^2(\beta_n)} J_1(\beta_n r/a), \qquad 0 \le r < a,$$

where β_n is the nth root of $\beta J_1'(\beta) + h\, J_1(\beta) = \beta J_0(\beta) + (h-1)J_1(\beta) = 0$ and $J_0(\cdot), J_1(\cdot)$ are the zeroth and first-order Bessel functions of the first kind, respectively.

Step 3: Show that the solution to the ordinary differential equation in Step 1 is

$$U(r,s) = \frac{2}{\alpha}\sum_{n=1}^{\infty} \frac{J_1(\beta_n\alpha/a)J_1(\beta_n r/a)}{(\beta_n^2+h^2-1)\,J_1^2(\beta_n)}\left[\frac{1}{s} - \frac{s}{s^2+c^2\beta_n^2/a^2}\right].$$

Note that this solution satisfies the boundary conditions.

Step 4: Taking the inverse of the Laplace transform in Step 3, show that the solution to the partial differential equation is

$$u(r,t) = \frac{2}{\alpha}\sum_{n=1}^{\infty} \frac{J_1(\beta_n\alpha/a)J_1(\beta_n r/a)}{(\beta_n^2+h^2-1)\,J_1^2(\beta_n)}\left[1-\cos\left(\frac{c\beta_n t}{a}\right)\right].$$

21. Solve the hyperbolic equation

$$\frac{\partial^2 u}{\partial x \partial t} + u = 0, \qquad 0 < x, t,$$

subject to the boundary conditions $u(0,t) = e^{-t}$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and $u(x,0) = 1$, $\lim_{t\to\infty} |u(x,t)| < Me^{kt}$, $0 < k, M, x, t$.

Step 1: Take the Laplace transform of the partial differential equation and show that

$$s\frac{dU(x,s)}{dx} + U = 0, \qquad 0 < x < \infty,$$

with $U(0,s) = 1/(s+1)$ and $\lim_{x\to\infty} U(x,s) \to 0$.

Step 2: Show that

$$U(x,s) = \frac{e^{-x/s}}{s+1} = \frac{e^{-x/s}}{s} - \frac{e^{-x/s}}{s(s+1)}.$$

Step 3: Using tables and the convolution theorem, show that the solution is

$$u(x,t) = J_0(2\sqrt{xt}\,) - e^{-t}\int_0^t e^{\tau} J_0(2\sqrt{x\tau}\,)\,d\tau,$$

where $J_0(\cdot)$ is the Bessel function of the first kind and order zero.

22. Solve the hyperbolic equation

$$\frac{\partial^2 u}{\partial x \partial t} + a\frac{\partial u}{\partial t} + b\frac{\partial u}{\partial x} = 0, \qquad 0 < a, b, x, t,$$

subject to the boundary conditions $u(0,t) = e^{ct}$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and the initial conditions $u(x,0) = 1$, $\lim_{t\to\infty} |u(x,t)| < Me^{kt}$, $0 < k, M, t, x$.

Step 1: Take the Laplace transform of the partial differential equation and show that

$$(s+b)\frac{dU(x,s)}{dx} + asU = a, \qquad 0 < x < \infty,$$

with $U(0,s) = 1/(s-c)$ and $\lim_{x\to\infty} U(x,s) \to 0$.

Step 2: Show that

$$U(x,s) = \frac{1}{s} + \frac{c\,e^{-ax}}{s(s-c)} \exp\left(\frac{bx}{s+b}\right).$$

Step 3: Using tables, the first shifting theorem, and the convolution theorem, show that the solution is

$$u(x,t) = 1 + c\,e^{ct-ax} \int_0^t e^{-(b+c)\tau} I_0\left(2\sqrt{bx\tau}\right)\,d\tau,$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero.

2.5 THE SOLUTION OF THE HEAT EQUATION BY USING LAPLACE TRANSFORMS

In the previous section we showed that we can solve the wave equation by the method of Laplace transforms. This is also true for the heat equation. Once again, we take the Laplace transform with respect to time. From the definition of Laplace transforms,

$$\mathcal{L}[u(x,t)] = U(x,s), \tag{2.5.1}$$

$$\mathcal{L}[u_t(x,t)] = sU(x,s) - u(x,0), \tag{2.5.2}$$

and

$$\mathcal{L}[u_{xx}(x,t)] = \frac{d^2U(x,s)}{dx^2}. \tag{2.5.3}$$

We next solve the resulting ordinary differential equation, known as the *auxiliary equation*, along with the corresponding Laplace transformed boundary conditions. The initial condition gives us the value of $u(x,0)$. The final step is the inversion of the Laplace transform $U(x,s)$. We typically use the inversion integral.

• Example 2.5.1

To illustrate these concepts, we solve a heat conduction problem[30] in a plane slab of thickness $2L$. Initially the slab has a constant temperature of unity. For $0 < t$, we allow both faces of the slab to radiatively cool in a medium that has a temperature of zero.

If $u(x,t)$ denotes the temperature, a^2 is the thermal diffusivity, h is the relative emissivity, t is the time, and x is the distance perpendicular to the face of the slab and measured from the middle of the slab, then the governing equation is

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}, \qquad -L < x < L, \quad 0 < t, \tag{2.5.4}$$

with the initial condition

$$u(x,0) = 1, \qquad -L < x < L, \tag{2.5.5}$$

and boundary conditions

$$\frac{\partial u(L,t)}{\partial x} + hu(L,t) = 0, \quad \frac{\partial u(-L,t)}{\partial x} + hu(-L,t) = 0, \quad 0 < t. \tag{2.5.6}$$

---

[30] Goldstein, S., 1932: The application of Heaviside's operational method to the solution of a problem in heat conduction. *Z. Angew. Math. Mech.*, **12**, 234–243.

Taking the Laplace transform of Equation 2.5.4 and substituting the initial condition,

$$a^2 \frac{d^2U(x,s)}{dx^2} - sU(x,s) = -1. \tag{2.5.7}$$

If we write $s = a^2q^2$, Equation 2.5.7 becomes

$$\frac{d^2U(x,s)}{dx^2} - q^2U(x,s) = -\frac{1}{a^2}. \tag{2.5.8}$$

From the boundary conditions, $U(x,s)$ is an even function in x and we may conveniently write the solution as

$$U(x,s) = \frac{1}{s} + A\cosh(qx). \tag{2.5.9}$$

From Equation 2.5.6,

$$qA\sinh(qL) + \frac{h}{s} + hA\cosh(qL) = 0, \tag{2.5.10}$$

and

$$U(x,s) = \frac{1}{s} - \frac{h\cosh(qx)}{s[q\sinh(qL) + h\cosh(qL)]}. \tag{2.5.11}$$

The inverse of $U(x,s)$ consists of two terms. The first term is simply unity. We will invert the second term by contour integration.

We begin by examining the nature and location of the singularities in the second term. Using the product formulas for the hyperbolic cosine and sine functions, the second term equals

$$\frac{h\left(1+\frac{4q^2x^2}{\pi^2}\right)\left(1+\frac{4q^2x^2}{9\pi^2}\right)\cdots}{s\left[q^2L\left(1+\frac{q^2L^2}{\pi^2}\right)\left(1+\frac{q^2L^2}{4\pi^2}\right)\cdots + h\left(1+\frac{4q^2L^2}{\pi^2}\right)\left(1+\frac{4q^2L^2}{9\pi^2}\right)\cdots\right]}. \tag{2.5.12}$$

Because $q^2 = s/a^2$, Equation 2.5.12 shows that we do not have any $\sqrt{s}$ in the transform and we need not concern ourselves with branch points and cuts. Furthermore, we have only simple poles: one located at $s = 0$ and the others where

$$q\sinh(qL) + h\cosh(qL) = 0. \tag{2.5.13}$$

If we set $q = i\lambda$, Equation 2.5.13 becomes

$$h\cos(\lambda L) - \lambda\sin(\lambda L) = 0, \qquad \text{or} \qquad \lambda L\tan(\lambda L) = hL. \tag{2.5.14}$$

From Bromwich's integral,

$$\mathcal{L}^{-1}\left\{\frac{h\cosh(qx)}{s[q\sinh(qL) + h\cosh(qL)]}\right\} = \frac{1}{2\pi i}\oint_C \frac{h\cosh(qx)e^{tz}}{z[q\sinh(qL) + h\cosh(qL)]}\,dz, \tag{2.5.15}$$

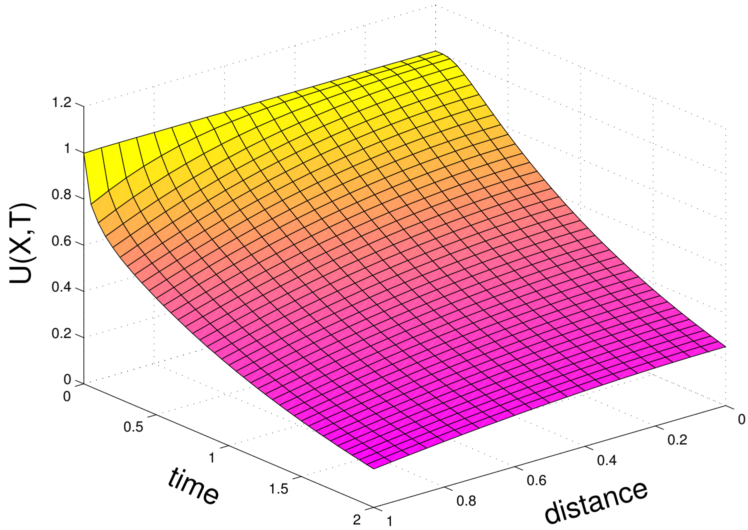


Figure 2.5.1: The temperature within the portion of a slab $0 < x/L < 1$ at various times a^2t/L^2 if the faces of the slab radiate to free space at temperature zero and the slab initially has the temperature 1. The parameter $hL = 1$.

where $q = z^{1/2}/a$ and the closed contour C consists of Bromwich's contour plus a semicircle of infinite radius in the left half of the z-plane. The residue at $z = 0$ is 1 while at $z_n = -a^2\lambda_n^2$,

$$\operatorname{Res}\left\{\frac{h\cosh(qx)e^{tz}}{z[q\sinh(qL)+h\cosh(qL)]}; z_n\right\} = \lim_{z\to z_n} \frac{h(z+a^2\lambda_n^2)\cosh(qx)e^{tz}}{z[q\sinh(qL)+h\cosh(qL)]} \tag{2.5.16}$$

$$= \lim_{z\to z_n} \frac{h\cosh(qx)e^{tz}}{z[(1+hL)\sinh(qL)+qL\cosh(qL)]/(2a^2q)} \tag{2.5.17}$$

$$= \frac{2ha^2\lambda_n i\cosh(i\lambda_n x)\exp(-\lambda_n^2a^2t)}{(-a^2\lambda_n^2)[(1+hL)i\sin(\lambda_nL)+i\lambda_nL\cos(\lambda_nL)]} \tag{2.5.18}$$

$$= -\frac{2h\cos(\lambda_n x)\exp(-a^2\lambda_n^2t)}{\lambda_n[(1+hL)\sin(\lambda_nL)+\lambda_nL\cos(\lambda_nL)]}. \tag{2.5.19}$$

Therefore, the inversion of $U(x,s)$ is

$$u(x,t) = 1 - \left\{1 - 2h\sum_{n=1}^{\infty}\frac{\cos(\lambda_n x)\exp(-a^2\lambda_n^2t)}{\lambda_n[(1+hL)\sin(\lambda_nL)+\lambda_nL\cos(\lambda_nL)]}\right\}, \tag{2.5.20}$$

or

$$u(x,t) = 2h\sum_{n=1}^{\infty}\frac{\cos(\lambda_n x)\exp(-a^2\lambda_n^2t)}{\lambda_n[(1+hL)\sin(\lambda_nL)+\lambda_nL\cos(\lambda_nL)]}. \tag{2.5.21}$$

We can further simplify Equation 2.5.21 by using $h/\lambda_n = \tan(\lambda_nL)$. This yields $hL = \lambda_nL\tan(\lambda_nL)$. Substituting these relationships into Equation 2.5.21 and simplifying,

$$u(x,t) = 2\sum_{n=1}^{\infty}\frac{\sin(\lambda_nL)\cos(\lambda_n x)\exp(-a^2\lambda_n^2t)}{\lambda_nL+\sin(\lambda_nL)\cos(\lambda_nL)}. \tag{2.5.22}$$

Figure 2.5.1 illustrates Equation 2.5.22. It was created using the MATLAB script

```
clear
hL = 1; m = 0; M = 100; dx = 0.05; dt = 0.05;
% create initial guess at zero_n
zero = zeros(length(M));
for n = 1:10000
  k1 = 0.1*n; k2 = 0.1*(n+1);
  prod = k1 * tan(k1); y1 = hL - prod; y2 = hL - k2 * tan(k2);
  if (y1*y2 <= 0 & prod < 2 & m < M) m = m+1; zero(m) = k1; end;
end;
% use Newton-Raphson method to improve values of zero_n
for n = 1:M; for k = 1:10
  f = hL - zero(n) * tan(zero(n));
  fp = - tan(zero(n)) - zero(n) * sec(zero(n))^2;
  zero(n) = zero(n) - f / fp;
end; end;
% compute Fourier coefficients
for m = 1:M
  a(m) = 2 * sin(zero(m)) / (zero(m) + sin(zero(m))*cos(zero(m)));
end
% compute grid and initialize solution
X = [0:dx:1]; T = [0:dt:2];
u = zeros(length(T),length(X));
XX = repmat(X,[length(T) 1]); TT = repmat(T',[1 length(X)]);
% compute solution from Equation 2.5.22
for m = 1:M
  u = u + a(m) * cos(zero(m)*XX) .* exp(-zero(m)*zero(m)*TT);
end
surf(XX,TT,u)
xlabel('distance','Fontsize',20); ylabel('time','Fontsize',20)
zlabel('U(X,T)','Fontsize',20)
```
□

• Example 2.5.2: Heat dissipation in disc brakes

Disc brakes consist of two blocks of frictional material known as pads that press against each side of a rotating annulus, usually made of a ferrous material. In this problem we determine the transient temperatures reached in a disc brake during a single brake application.[31] If we ignore the errors introduced by replacing the cylindrical portion of the drum by a rectangular plate, we can model our disc brakes as a one-dimensional solid, which friction heats at both ends. Assuming symmetry about $x = 0$, the boundary condition there is $u_x(0, t) = 0$. To model the heat flux from the pads, we assume a uniform disc deceleration that generates heat from the frictional surfaces at the rate $N(1 - Mt)$, where M and N are experimentally determined constants.

If $u(x, t)$, κ, and a^2 denote the temperature, thermal conductivity, and diffusivity of the rotating annulus, respectively, then the heat equation is

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t, \tag{2.5.23}$$

[31] From Newcomb, T. P., 1958: The flow of heat in a parallel-faced infinite solid. *Brit. J. Appl. Phys.*, **9**, 370–372. See also Newcomb, T. P., 1958/59: Transient temperatures in brake drums and linings. *Proc. Inst. Mech. Eng., Auto. Div.*, 227–237; Newcomb, T. P., 1959: Transient temperatures attained in disk brakes. *Brit. J. Appl. Phys.*, **10**, 339–340.

with the boundary conditions

$$\frac{\partial u(0,t)}{\partial x} = 0, \quad \kappa \frac{\partial u(L,t)}{\partial x} = N(1 - Mt), \qquad 0 < t. \tag{2.5.24}$$

The boundary condition at $x = L$ gives the frictional heating of the disc pads.

Introducing the Laplace transform of $u(x,t)$, defined as

$$U(x,s) = \int_0^\infty u(x,t) e^{-st}\, dt, \tag{2.5.25}$$

the equation to be solved becomes

$$\frac{d^2U}{dx^2} - \frac{s}{a^2} U = 0, \tag{2.5.26}$$

subject to the boundary conditions that

$$\frac{dU(0,s)}{dx} = 0, \quad \text{and} \quad \frac{dU(L,s)}{dx} = \frac{N}{\kappa}\left(\frac{1}{s} - \frac{M}{s^2}\right). \tag{2.5.27}$$

The solution of Equation 2.5.26 is

$$U(x,s) = A\cosh(qx) + B\sinh(qx), \tag{2.5.28}$$

where $q = s^{1/2}/a$. Using the boundary conditions, the solution becomes

$$U(x,s) = \frac{N}{\kappa}\left(\frac{1}{s} - \frac{M}{s^2}\right) \frac{\cosh(qx)}{q\sinh(qL)}. \tag{2.5.29}$$

It now remains to invert the transform, Equation 2.5.29. We will invert $\cosh(qx)/[sq\sinh(qL)]$; the inversion of the second term follows by analog.

Our first concern is the presence of $s^{1/2}$ because this is a multivalued function. However, when we replace the hyperbolic cosine and sine functions with their Taylor expansions, $\cosh(qx)/[sq\sinh(qL)]$ contains only powers of s and is, in fact, a single-valued function.

From Bromwich's integral,

$$\mathcal{L}^{-1}\left[\frac{\cosh(qx)}{sq\sinh(qL)}\right] = \frac{1}{2\pi i}\int_{c-\infty i}^{c+\infty i} \frac{\cosh(qx)e^{tz}}{zq\sinh(qL)}\, dz, \tag{2.5.30}$$

where $q = z^{1/2}/a$. Just as in the previous example, we replace the hyperbolic cosine and sine with their product expansion to determine the nature of the singularities. The point $z = 0$ is a second-order pole. The remaining poles are located where $z_n^{1/2} L/a = n\pi i$, or $z_n = -n^2\pi^2a^2/L^2$, where $n = 1, 2, 3, \ldots$. We have chosen the positive sign because $z^{1/2}$ must be single-valued; if we had chosen the negative sign, the answer would have been the same. Our expansion also shows that the poles are simple.

Having classified the poles, we now close Bromwich's contour, which lies slightly to the right of the imaginary axis, with an infinite semicircle in the left half-plane, and use the residue theorem. The values of the residues are

$$\operatorname{Res}\left[\frac{\cosh(qx)e^{tz}}{zq\sinh(qL)};0\right] = \frac{1}{1!}\lim_{z\to 0}\frac{d}{dz}\left\{\frac{(z-0)^2\cosh(qx)e^{tz}}{zq\sinh(qL)}\right\} \tag{2.5.31}$$

$$= \lim_{z\to 0}\frac{d}{dz}\left\{\frac{z\cosh(qx)e^{tz}}{q\sinh(qL)}\right\} \tag{2.5.32}$$

$$= \frac{a^2}{L}\lim_{z\to 0}\frac{d}{dz}\left\{\frac{z\left[1+\frac{zx^2}{2!a^2}+\cdots\right]\left[1+tz+\frac{t^2z^2}{2!}+\cdots\right]}{z+\frac{L^2z^2}{3!a^2}+\cdots}\right\} \tag{2.5.33}$$

$$= \frac{a^2}{L}\lim_{z\to 0}\frac{d}{dz}\left\{1+tz+\frac{zx^2}{2a^2}-\frac{zL^2}{3!a^2}+\cdots\right\} \tag{2.5.34}$$

$$= \frac{a^2}{L}\left\{t+\frac{x^2}{2a^2}-\frac{L^2}{6a^2}\right\}, \tag{2.5.35}$$

and

$$\operatorname{Res}\left[\frac{\cosh(qx)e^{tz}}{zq\sinh(qL)};z_n\right] = \left[\lim_{z\to z_n}\frac{\cosh(qx)}{zq}e^{tz}\right]\left[\lim_{z\to z_n}\frac{z-z_n}{\sinh(qL)}\right] \tag{2.5.36}$$

$$= \lim_{z\to z_n}\frac{\cosh(qx)e^{tz}}{zq\cosh(qL)L/(2a^2q)} \tag{2.5.37}$$

$$= \frac{\cosh(n\pi xi/L)\exp(-n^2\pi^2a^2t/L^2)}{(-n^2\pi^2a^2/L^2)\cosh(n\pi i)L/(2a^2)} \tag{2.5.38}$$

$$= -\frac{2L(-1)^n}{n^2\pi^2}\cos(n\pi x/L)e^{-n^2\pi^2a^2t/L^2}. \tag{2.5.39}$$

When we sum all of the residues from both inversions, the solution is

$$\begin{aligned} u(x,t) = {} & \frac{a^2N}{\kappa L}\left\{t+\frac{x^2}{2a^2}-\frac{L^2}{6a^2}\right\} - \frac{2LN}{\kappa\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos(n\pi x/L)e^{-n^2\pi^2a^2t/L^2} \\ & - \frac{a^2NM}{\kappa L}\left\{\frac{t^2}{2}+\frac{tx^2}{2a^2}-\frac{tL^2}{6a^2}+\frac{x^4}{24a^4}-\frac{x^2L^2}{12a^4}+\frac{7L^4}{360a^4}\right\} \\ & - \frac{2L^3NM}{a^2\kappa\pi^4}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^4}\cos(n\pi x/L)e^{-n^2\pi^2a^2t/L^2}. \end{aligned} \tag{2.5.40}$$

Figure 2.3.2 shows the temperature in the brake lining at various places within the lining $[x' = x/L]$ if $a^2 = 3.3 \times 10^{-3}$ cm^2/sec, $\kappa = 1.8 \times 10^{-3}$ cal/(cm sec°C), $L = 0.48$ cm, and $N = 1.96$ cal/(cm^2 sec). Initially the frictional heating results in an increase in the disc brake's temperature. As time increases, the heating rate decreases and radiative cooling becomes sufficiently large that the temperature begins to fall. □

• Example 2.5.3

In the previous example we showed that Laplace transforms are particularly useful when the boundary conditions are time dependent. Consider now the case when one of the boundaries is moving.

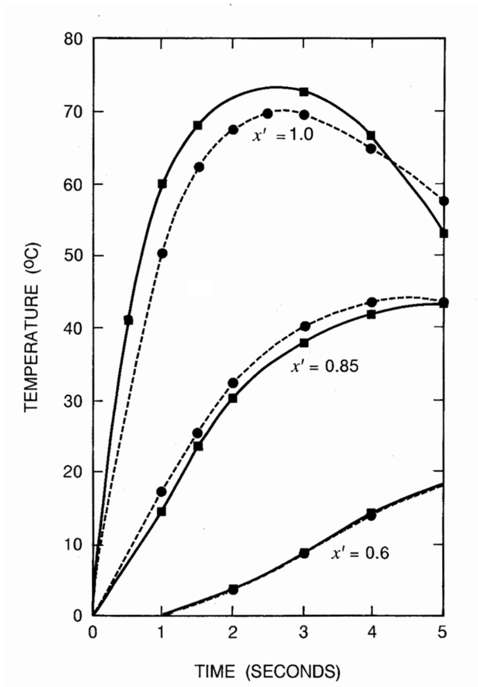


Figure 2.5.2: Typical curves of transient temperature at different locations in a brake lining. Circles denote computed values while squares are experimental measurements. (From Newcomb, T. P., 1958: The flow of heat in a parallel-faced infinite solid. *Brit. J. Appl. Phys.*, **9**, 372 with permission.)

We wish to solve the heat equation

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}, \qquad \beta t < x < \infty, \quad 0 < t, \tag{2.5.41}$$

subject to the boundary conditions

$$u(x,t)\big|_{x=\beta t} = f(t), \quad \text{and} \quad \lim_{x\to\infty} u(x,t) \to 0, \quad 0 < t, \tag{2.5.42}$$

and the initial condition

$$u(x,0) = 0, \qquad 0 < x < \infty. \tag{2.5.43}$$

This type of problem arises in combustion problems where the boundary moves due to the burning of the fuel.

We begin by introducing the coordinate $\eta = x - \beta t$. Then the problem can be reformulated as

$$\frac{\partial u}{\partial t} - \beta \frac{\partial u}{\partial \eta} = a^2 \frac{\partial^2 u}{\partial \eta^2}, \quad 0 < \eta < \infty, \quad 0 < t, \tag{2.5.44}$$

subject to the boundary conditions

$$u(0,t) = f(t), \quad \lim_{\eta\to\infty} u(\eta,t) \to 0, \qquad 0 < t, \tag{2.5.45}$$

and the initial condition

$$u(\eta,0) = 0, \qquad 0 < \eta < \infty. \tag{2.5.46}$$

Taking the Laplace transform of Equation 2.5.44, we have that

$$\frac{d^2U(\eta,s)}{d\eta^2} + \frac{\beta}{a^2}\frac{dU(\eta,s)}{d\eta} - \frac{s}{a^2}U(\eta,s) = 0, \tag{2.5.47}$$

with

$$U(0,s) = F(s), \quad \text{and} \quad \lim_{\eta\to\infty} U(\eta,s) \to 0. \tag{2.5.48}$$

The solution to Equation 2.5.47 and Equation 2.5.48 is

$$U(\eta,s) = F(s)\exp\left(-\frac{\beta\eta}{2a^2} - \frac{\eta}{a}\sqrt{s + \frac{\beta^2}{4a^2}}\right). \tag{2.5.49}$$

Because

$$\mathcal{L}[\Phi(\eta,t)] = \exp\left(-\frac{\eta}{a}\sqrt{s + \frac{\beta^2}{4a^2}}\right), \tag{2.5.50}$$

where

$$\Phi(\eta,t) = \tfrac{1}{2}\left[e^{-\beta\eta/2a^2}\operatorname{erfc}\left(\frac{\eta}{2a\sqrt{t}} - \frac{\beta\sqrt{t}}{2a}\right) + e^{\beta\eta/2a^2}\operatorname{erfc}\left(\frac{\eta}{2a\sqrt{t}} + \frac{\beta\sqrt{t}}{2a}\right)\right], \tag{2.5.51}$$

and

$$\operatorname{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}}\int_0^x e^{-\eta^2}\,d\eta, \tag{2.5.52}$$

we have by the convolution theorem that

$$u(\eta,t) = e^{-\beta\eta/2a^2}\int_0^t f(t-\tau)\Phi(\eta,\tau)\,d\tau, \tag{2.5.53}$$

or

$$u(x,t) = e^{-\beta(x-\beta t)/2a^2}\int_0^t f(t-\tau)\Phi(x-\beta\tau,\tau)\,d\tau. \tag{2.5.54}$$

Problems

1. Solve the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - a^2(u - T_0), \qquad 0 < x < 1, \quad 0 < t,$$

subject to the boundary conditions $u_x(0,t) = u_x(1,t) = 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - (s + a^2)U(x,s) = -\frac{a^2T_0}{s}, \qquad 0 < x < 1,$$

subject to the boundary conditions $U'(0,s) = U'(1,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = T_0\left(\frac{1}{s} - \frac{1}{s+a^2}\right).$$

Step 3: Invert $U(x,s)$ and show that $u(x,t) = T_0\left(1 - e^{-a^2t}\right)$.

2. Solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad 0 < t,$$

subject to the boundary conditions $u_x(0,t) = 0$, $u(1,t) = t$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = 0, \qquad 0 < x < 1,$$

subject to the boundary conditions $U'(0,s) = 0$ and $U(1,s) = 1/s^2$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{\cosh(x\sqrt{s}\,)}{s^2\,\cosh(\sqrt{s}\,)}.$$

Step 3: Show that $U(x,s)$ has a second-order pole at $s = 0$ and simple poles at $s_n = -(2n-1)^2\pi^2/4$ with $\sqrt{z_n} = (2n-1)\pi i/2$, where $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = t + \frac{1}{2}(x^2 - 1) - \frac{16}{\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n-1)^3}\cos\left[\frac{(2n-1)\pi x}{2}\right]\exp\left[-\frac{(2n-1)^2\pi^2 t}{4}\right].$$

3. Solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad 0 < t,$$

subject to the boundary conditions $u(0,t) = 0$, $u(1,t) = 1$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = 0, \qquad 0 < x < 1,$$

subject to the boundary conditions $U(0,s) = 0$ and $U(1,s) = 1/s$.

Step 2: Show that the solution to the previous step is $U(x,s) = \sinh(x\sqrt{s}\,)/[s\,\sinh(\sqrt{s}\,)]$.

Step 3: Show that $U(x,s)$ has simple poles at $s = 0$ and $s_n = -n^2\pi^2$ with $\sqrt{z_n} = n\pi i$, where $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = x + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(n\pi x)e^{-n^2\pi^2 t}.$$

4. Solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad -\tfrac{1}{2} < x < \tfrac{1}{2}, \quad 0 \le t,$$

subject to the boundary conditions $u_x\left(-\frac{1}{2}, t\right) = 0$, $u_x\left(\frac{1}{2}, t\right) = \delta(t)$, $0 \le t$, and the initial condition $u(x,0) = 0$, $-\frac{1}{2} < x < \frac{1}{2}$. Here $\delta(t)$ is the Dirac delta function.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = 0, \qquad -\frac{1}{2} < x < \frac{1}{2},$$

subject to the boundary conditions $U'\left(-\frac{1}{2}, s\right) = 0$ and $U'\left(\frac{1}{2}, s\right) = 1$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{\cosh\left[(x+\frac{1}{2})\sqrt{s}\,\right]}{\sqrt{s}\,\sinh(\sqrt{s}\,)}.$$

Step 3: Replacing the hyperbolic functions by their exponential definition, show that

$$U(x,s) = \frac{1}{\sqrt{s}}\left\{\exp\left[\left(x-\tfrac{1}{2}\right)\sqrt{s}\,\right] + \exp\left[-\left(x+\tfrac{3}{2}\right)\sqrt{s}\,\right]\right\}\left(1 + e^{-2\sqrt{s}} + e^{-4\sqrt{s}} + \cdots\right).$$

Step 4: Taking the inverse of $U(x,s)$ term by term, show that

$$u(x,t) = \frac{1}{\sqrt{\pi\,t}}\sum_{n=0}^{\infty}\left\{\exp\left[-\frac{\left(2n+\frac{1}{2}-x\right)^2}{4t}\right] + \exp\left[-\frac{\left(2n+\frac{3}{2}+x\right)^2}{4t}\right]\right\}.$$

Step 5: Show that $U(x,s)$ has simple poles at $s = 0$ and $s_n = -n^2\pi^2$ with $\sqrt{z_n} = n\pi i$, where $n = 1, 2, 3, \ldots$.

Step 6: Use Bromwich's integral and show that

$$u(x,t) = 1 + 2\sum_{n=1}^{\infty}(-1)^n\cos\left[n\pi\left(x+\tfrac{1}{2}\right)\right]e^{-n^2\pi^2 t}.$$

5. Solve

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 1, \qquad 0 < x < 1, \quad 0 < t,$$

subject to the boundary conditions $u(0,t) = u(1,t) = 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = -\frac{1}{s}, \qquad 0 < x < 1,$$

subject to the boundary conditions $U(0,s) = U(1,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{1 - \cosh(x\sqrt{s}\,)}{s^2} - \frac{[1 - \cosh(\sqrt{s}\,)]\sinh(x\sqrt{s}\,)}{s^2 \sinh(\sqrt{s}\,)}.$$

Step 3: Show that $U(x,s)$ has a second-order pole at $s = 0$ and simple poles at $s_n = -n^2\pi^2$ with $\sqrt{z_n} = n\pi i$, where $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{x(1-x)}{2} - \frac{4}{\pi^3}\sum_{m=1}^{\infty} \frac{\sin[(2m-1)\pi x]}{(2m-1)^3} e^{-(2m-1)^2\pi^2 t}.$$

6. Solve[32]

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < \infty, \quad 0 < t,$$

subject to the boundary conditions $u(0,t) = 1$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s}{a^2}U(x,s) = 0, \qquad 0 < x < \infty,$$

subject to the boundary conditions $U(0,s) = 1/s$ and $\lim_{x\to\infty} |U(x,s)| < \infty$.

Step 2: Show that the solution to the previous step is $U(x,s) = e^{-x\sqrt{s}/a}/s$.

Step 3: From an extensive table of inverses, show that $u(x,t) = \operatorname{erfc}\big[x/(2a\sqrt{t}\,)\big]$, where $\operatorname{erfc}(\cdot)$ is the complementary error function.

7. Solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < \infty, \quad 0 < t,$$

subject to the boundary conditions $u_x(0,t) = 1$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = 0, \qquad 0 < x < \infty,$$

[32] If $u(x,t)$ denotes the Eulerian velocity of a viscous fluid in the half space $x > 0$ and parallel to the wall located at $x = 0$, then this problem was first solved by Stokes, G. G., 1850: On the effect of the internal friction of fluids on the motions of pendulums. *Proc. Cambridge Philos. Soc.*, **9, Part II**, [8]–[106].

subject to the boundary conditions $U'(0,s) = 1/s$ and $\lim_{x\to\infty} |U(x,s)| < \infty$.

Step 2: Show that the solution to the previous step is $U(x,s) = -e^{-x\sqrt{s}}/s^{3/2}$.

Step 3: From an extensive table of inverses, show that

$$u(x,t) = x\,\mathrm{erfc}\left(\frac{x}{2\sqrt{t}}\right) - 2\sqrt{\frac{t}{\pi}}\exp\left(-\frac{x^2}{4t}\right),$$

where erfc$(\cdot)$ is the complementary error function.

8. Solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < \infty, \quad 0 < t,$$

subject to the boundary conditions $u(0,t) = 1$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and the initial condition $u(x,0) = e^{-x}$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = -e^{-x}, \qquad 0 < x < \infty,$$

subject to the boundary conditions $U(0,s) = 1/s$ and $\lim_{x\to\infty} |U(x,s)| < \infty$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{e^{-x}}{s-1} + \left(\frac{1}{s} - \frac{1}{s-1}\right)e^{-x\sqrt{s}}.$$

Step 3: From an extensive table of inverses, show that

$$u(x,t) = e^{t-x} + \mathrm{erfc}\left(\frac{x}{2\sqrt{t}}\right) - \tfrac{1}{2}e^t\left[e^{-x}\mathrm{erfc}\left(\frac{x}{2\sqrt{t}} - \sqrt{t}\right) + e^x\mathrm{erfc}\left(\frac{x}{2\sqrt{t}} + \sqrt{t}\right)\right],$$

where erfc$(\cdot)$ is the complementary error function.

9. Solve

$$\frac{\partial u}{\partial t} = a^2\left[\frac{\partial^2 u}{\partial x^2} + (1+\delta)\frac{\partial u}{\partial x} + \delta u\right], \qquad 0 < x < \infty, \quad 0 < t,$$

where δ is a constant, subject to the boundary conditions $u(0,t) = u_0$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} + (1+\delta)\frac{dU(x,s)}{dx} + \left(\delta - \frac{s}{a^2}\right)U(x,s) = 0, \qquad 0 < x < \infty,$$

subject to the boundary conditions $U(0,s) = u_0/s$ and $\lim_{x\to\infty} |U(x,s)| < \infty$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{u_0}{s}\exp\left[-\frac{(1+\delta)x}{2} - \frac{x}{a}\sqrt{\frac{a^2(1-\delta)^2}{4} + s}\right].$$

Step 3: From an extensive table of inverses, show that

$$u(x,t) = \frac{u_0}{2}e^{-\delta x}\operatorname{erfc}\left(\frac{x}{2a\sqrt{t}} + \frac{a(1-\delta)\sqrt{t}}{2}\right) + \frac{u_0}{2}e^{-x}\operatorname{erfc}\left(\frac{x}{2a\sqrt{t}} - \frac{a(1-\delta)\sqrt{t}}{2}\right).$$

10. During their modeling of a chemical reaction with a back reaction, Agmon et al.[33] solved

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < \infty, \quad 0 < t,$$

subject to the boundary conditions

$$\kappa_d + a^2 u_x(0,t) + a^2\kappa_d\int_0^t u_x(0,\tau)\,d\tau = \kappa_r u(0,t), \quad \lim_{x\to\infty} u(x,t) \to 0, \qquad 0 < t,$$

and the initial condition $u(x,0) = 0$, $0 < x < \infty$, where κ_d and κ_r denote the intrinsic dissociation and recombination rate coefficients, respectively. What should they have found?

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s}{a^2}U(x,s) = 0, \qquad 0 < x < \infty,$$

subject to the boundary conditions $\lim_{x\to\infty}|U(x,s)| < \infty$ and

$$\kappa_d + (s+\kappa_d)a^2U'(0,s) = s\kappa_r U(0,s).$$

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{2\kappa_d\exp(-x\sqrt{s}/a)}{a\Delta\sqrt{s}}\left(\frac{1}{2a\sqrt{s}+\kappa_r-\Delta} - \frac{1}{2a\sqrt{s}+\kappa_r+\Delta}\right),$$

where $\Delta \equiv \sqrt{\kappa_r^2 - 4a^2\kappa_d}$.

Step 3: From an extensive table of inverses, show that

$$u(x,t) = \frac{\kappa_d}{\Delta}e^{-x^2/(4a^2t)}\left[e^{x_-^2}\operatorname{erfc}(x_-) - e^{x_+^2}\operatorname{erfc}(x_+)\right],$$

where $x_\pm = [x + (\kappa_r \pm \Delta)a^2t]/(2a\sqrt{t})$ and $\operatorname{erfc}(\cdot)$ is the complementary error function.

11. Solve[34]

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - \beta u, \qquad 0 < x < \infty, \quad 0 < t,$$

subject to the boundary conditions $\rho u(0,t) - u_x(0,t) = e^{(\sigma^2-\beta)t}$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < \infty$, where β, ρ, and σ are constants and $\sigma \neq \rho$.

---

[33] Agmon, N., E. Pines, and D. Huppert, 1988: Germinate recombination in proton-transfer reactions. II. Comparison of diffusional and kinetic schemes. *J. Chem. Phys.*, **88**, 5631–5638.

[34] Saidel, G. M., E. D. Morris, and G. M. Chisolm, 1987: Transport of macromolecules in arterial wall *in vivo*: A mathematical model and analytic solutions. *Bull. Math. Biol.*, **49**, 153–169.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - (s+\beta)U(x,s) = 0, \qquad 0 < x < \infty,$$

subject to the boundary conditions

$$\rho U(0,s) - U'(0,s) = \frac{1}{s+\beta-\sigma^2}, \qquad \lim_{x\to\infty} |U(x,s)| < \infty.$$

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{\exp(-x\sqrt{s+\beta}\,)}{(s+\beta-\sigma^2)(\rho+\sqrt{s+\beta}\,)}.$$

Step 3: Using partial fractions, show that

$$\begin{aligned} U(x,s) &= \frac{e^{-x\sqrt{s'}}}{(s'+\sigma^2)(\sqrt{s'}+\rho)} = \frac{e^{-x\sqrt{s'}}}{(\sqrt{s'}+\sigma)(\sqrt{s'}-\sigma)(\sqrt{s'}+\rho)} \\ &= \frac{e^{-x\sqrt{s'}}}{(\rho^2-\sigma^2)(\sqrt{s'}+\rho)} + \frac{e^{-x\sqrt{s'}}}{2\sigma(\rho+\sigma)(\sqrt{s'}-\sigma)} - \frac{e^{-x\sqrt{s'}}}{2\sigma(\rho-\sigma)(\sqrt{s'}+\sigma)}, \end{aligned}$$

where $s' = s + \beta$.

Step 4: Using the first shifting theorem and the fact that

$$\mathcal{L}^{-1}\left(\frac{e^{-k\sqrt{s}}}{a+\sqrt{s}}\right) = \frac{1}{\sqrt{\pi t}}\exp\left(-\frac{k^2}{4t}\right) - ae^{ak}e^{a^2t}\operatorname{erfc}\left(a\sqrt{t} + \frac{k}{2\sqrt{t}}\right),$$

show that

$$\begin{aligned} u(x,t) = {} & \tfrac{1}{2}e^{\sigma^2 t-\beta t}\left[\frac{e^{-\sigma x}}{\rho+\sigma}\operatorname{erfc}\left(\frac{x}{2\sqrt{t}} - \sigma\sqrt{t}\right) + \frac{e^{\sigma x}}{\rho-\sigma}\operatorname{erfc}\left(\frac{x}{2\sqrt{t}} + \sigma\sqrt{t}\right)\right] \\ & - \frac{\rho}{\rho^2-\sigma^2}e^{\rho x+\rho^2 t-\beta t}\operatorname{erfc}\left(\frac{x}{2\sqrt{t}} + \rho\sqrt{t}\right). \end{aligned}$$

12. Solve

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2} + Ae^{-kx}, \qquad 0 < x < \infty, \quad 0 < t,$$

subject to the boundary conditions $u_x(0,t) = 0$, $\lim_{x\to\infty} u(x,t) = u_0$, $0 < t$, and the initial condition $u(x,0) = u_0$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s}{a^2}U(x,s) = -\frac{u_0}{a^2} - \frac{A}{a^2\,s}e^{-kx}, \qquad 0 < x < \infty,$$

subject to the boundary conditions $U'(0,s) = 0$ and $\lim_{x\to\infty} U(x,s) = u_0/s$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{u_0}{s} + \frac{Ae^{-kx}}{a^2k^2}\left(\frac{1}{s-a^2k^2} - \frac{1}{s}\right) + \frac{Ae^{-qx}}{aks\sqrt{s}} - \frac{Ae^{-qx}}{ak\sqrt{s}(s-a^2k^2)},$$

where $q = \sqrt{s}/a$.

Step 3: Using the convolution theorem,

$$u(x,t) = u_0 + \frac{Ae^{-kx}}{a^2k^2}\left(e^{a^2k^2t} - 1\right) + \frac{A}{ak}\left[2\sqrt{\frac{t}{\pi}}\exp\left(-\frac{x^2}{4a^2t}\right) - \frac{x}{a}\operatorname{erfc}\left(\frac{x}{2a\sqrt{t}}\right)\right]$$
$$- \frac{Ae^{a^2k^2t}}{ak}\int_0^t e^{-a^2k^2\tau}\exp\left(-\frac{x^2}{4a^2\tau}\right)\frac{d\tau}{\sqrt{\pi\tau}}.$$

13. Solve

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2} - P, \qquad 0 < x < L, \quad 0 < t,$$

subject to the boundary conditions $u(0,t) = t$, $u(L,t) = 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < L$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s}{a^2}U(x,s) = \frac{P}{sa^2}, \qquad 0 < x < L,$$

subject to the boundary conditions $U(0,s) = 1/s^2$ and $U(L,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{P}{s^2}\left[\frac{\sinh(qx)}{\sinh(qL)} - 1\right] + (P+1)\frac{\sinh[q(L-x)]}{s^2\sinh(qL)},$$

where $q = \sqrt{s}/a$.

Step 3: Show that $U(x,s)$ has a second-order pole at $s = 0$ and simple poles at $s_n = -n^2\pi^2a^2/L^2$ and $\sqrt{s_n} = n\pi ai/L$, where $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{t(L-x)}{L} + \frac{Px(x-L)}{2a^2} - \frac{x(x-L)(x-2L)}{6a^2L}$$
$$- \frac{2PL^2}{a^2\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3}\sin\left(\frac{n\pi x}{L}\right)\exp\left(-\frac{a^2n^2\pi^2t}{L^2}\right)$$
$$+ \frac{2(P+1)L^2}{a^2\pi^3}\sum_{n=1}^{\infty}\frac{1}{n^3}\sin\left(\frac{n\pi x}{L}\right)\exp\left(-\frac{a^2n^2\pi^2t}{L^2}\right).$$

14. Solve

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2} + ku, \qquad 0 < x < L, \quad 0 < k, t,$$

subject to the boundary conditions $u(0,t) = u(L,t) = T_0$, $0 < t$, and the initial condition $u(x,0) = T_0$, $0 < x < L$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s-k}{a^2}U(x,s) = -\frac{T_0}{a^2}, \qquad 0 < x < L,$$

subject to the boundary conditions $U(0,s) = U(L,s) = T_0/s$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{T_0}{s-k} - \frac{kT_0}{s(s-k)}\frac{\sinh(qx) + \sinh[q(L-x)]}{\sinh(qL)}.$$

where $q = \sqrt{s-k}/a$.

Step 3: Show that $U(x,s)$ has simple poles at $s=0$, $s=k$, and $s_n = k - n^2\pi^2a^2/L^2$, where $n = 1, 2, 3, \ldots$. Note here that $q_n = n\pi i/L$.

Step 4: Use Bromwich's integral and show that

$$\begin{aligned} u(x,t) &= \frac{T_0\cos[(L/2-x)\sqrt{k/a^2}]}{\cos(L\sqrt{k/a^2}/2)} \\ &\quad + \frac{4kT_0}{\pi}\sum_{m=1}^{\infty}\frac{\sin[(2m-1)\pi x/L]}{(2m-1)[k-(2m-1)^2\pi^2a^2/L^2]}e^{kt-(2m-1)^2\pi^2a^2t/L^2} \\ &= \frac{4T_0}{\pi}\sum_{m=1}^{\infty}\frac{1}{2m-1}\left[\frac{\kappa_m}{\kappa_m-k} - \left(\frac{k}{\kappa_m-k}\right)e^{kt-\kappa_m t}\right]\sin\left[\frac{(2m-1)\pi x}{L}\right], \end{aligned}$$

where $\kappa_m = (2m-1)^2\pi^2a^2/L^2$.

15. An electric fuse protects electrical devices by using resistance heating to melt an enclosed wire when excessive current passes through it. A knowledge of the distribution of temperature along the wire is important in the design of the fuse. If the temperature rises to the melting point only over a small interval of the element, the melt will produce a small gap, resulting in an unnecessary prolongation of the fault and a considerable release of energy. Therefore, the desirable temperature distribution should melt most of the wire. For this reason, Guile and Carne[35] solved the heat conduction equation

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2} + q(1+\alpha u), \qquad -L < x < L, \quad 0 < t,$$

to understand the temperature structure within the fuse just before meltdown. The second term on the right side of the heat conduction equation gives the resistance heating, which is assumed to vary linearly with temperature. If the terminals at $x = \pm L$ remain at a constant temperature, which we can take to be zero, the boundary conditions are $u(-L,t) = u(L,t) = 0$, $0 < t$. The initial condition is $u(x,0) = 0$, $-L < x < L$. Find the temperature field as a function of the parameters a, q, and α.

---

[35] Guile, A. E., and E. B. Carne, 1954: An analysis of an analogue solution applied to the heat conduction problem in a cartridge fuse. *AIEE Trans., Part 1*, **72**, 861–868.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} + \frac{\alpha q - s}{a^2}U(x,s) = -\frac{q}{a^2 s}, \qquad 0 < x < L,$$

subject to the boundary conditions $U(-L,s) = U(L,s) = 0$.

Step 2: Show that the solution to the previous step is

$$sU(x,s) = \frac{q}{s-\alpha q} - \frac{q\cosh(x\sqrt{s-\alpha q}/a)}{(s-\alpha q)\cosh(L\sqrt{s-\alpha q}/a)}.$$

Step 3: Show that $U(x,s)$ has a removable singularity at $s = \alpha q$ and simple poles at $s_n = \alpha q - (2n-1)^2\pi^2a^2/4L^2$ with $\sqrt{s_n - \alpha q} = (2n-1)\pi a i/2L$, where $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u_t(x,t) = -\frac{4q}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{2n-1}\cos[(2n-1)\pi x/2L]\exp[\alpha qt - (2n-1)^2\pi^2a^2t/4L^2].$$

Step 5: Integrate $u_t(x,t)$ with respect to time and obtain

$$u(x,t) = \frac{4q}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n\cos[(2n-1)\pi x/2L]}{(2n-1)[\alpha q - (2n-1)^2\pi^2a^2/4L^2]}\left\{1 - \exp[\alpha qt - (2n-1)^2\pi^2a^2t/4L^2]\right\},$$

where the constant of integration ensures that $u(x,0) = 0$.

16. Solve[36]

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}, \qquad 0 \le r < 1, \quad 0 < t,$$

subject to the boundary conditions $\lim_{r\to 0}|u(r,t)| < \infty$, $u_r(1,t) = 1$, $0 < t$, and the initial condition $u(r,0) = 0$, $0 \le r < 1$.

Step 1: Introduce the new variable $v(r,t) = r\,u(r,t)$ and show that the problem becomes

$$\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial r^2}, \qquad 0 \le r < 1, 0 < t,$$

with the boundary conditions $\lim_{r\to 0} v(r,t) \to 0$ and

$$\frac{\partial v(1,t)}{\partial r} - v(1,t) = 1, \qquad t > 0$$

and the initial condition $v(r,0) = 0, 0 \le r < 1$.

Step 2: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2V(r,s)}{dr^2} - sV(r,s) = 0, \qquad 0 \le r < 1,$$

[36] See Reismann, H., 1962: Temperature distribution in a spinning sphere during atmospheric entry. *J. Aerosp. Sci.*, **29**, 151–159.

subject to the boundary conditions $\lim_{r\to 0} V(r,s) \to 0$ and $V'(1,s) - V(1,s) = 1/s$.

Step 3: Show that the solution to the previous step is

$$V(r,s) = \frac{\sinh(r\sqrt{s}\,)}{s\,[\sqrt{s}\cosh(\sqrt{s}\,) - \sinh(\sqrt{s}\,)]}.$$

Step 4: Show that $V(x,s)$ has a second-order pole at $s=0$ and simple poles where $\sqrt{s_n} = i\lambda_n$, $s_n = -\lambda_n^2$ and $\tan(\lambda_n) = \lambda_n$, $n = 1, 2, 3, \ldots$.

Step 5: Use Bromwich's integral and show that

$$u(r,t) = \frac{r^2}{2} + 3t - \frac{3}{10} - \frac{2}{r}\sum_{n=1}^{\infty} \frac{\sin(\lambda_n r)}{\lambda_n^2 \sin(\lambda_n)} e^{-\lambda_n^2 t},$$

where $\tan(\lambda_n) = \lambda_n$.

17. Solve[37]

$$\frac{\partial u}{\partial t} = a^2\left(\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}\right) + q(t) = \frac{a^2}{r}\frac{\partial^2 (ru)}{\partial r^2} + q(t), \qquad b < r < \infty, \quad 0 < t,$$

subject to the boundary conditions

$$\frac{\partial u(b,t)}{\partial r} = u(b,t), \qquad \lim_{r\to\infty} u(r,t) = u_0 + \int_0^t q(\tau)\,d\tau, \qquad 0 < t,$$

and the initial condition $u(r,0) = u_0$, $b < r < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2[rU(r,s)]}{dr^2} - \frac{s}{a^2} rU(r,s) = -\frac{r[Q(s) + u_0]}{a^2}, \qquad b < r < \infty,$$

subject to the boundary conditions $U'(b,s) = U(b,s)$ and $\lim_{r\to\infty} U(r,s) = [u_0 + Q(s)]/s$.

Step 2: Show that the solution to the previous step is

$$U(r,s) = \frac{u_0 + Q(s)}{s} - \frac{b\,Q(s)}{s(q + 1/\beta)}\frac{e^{-q(r-b)}}{r} - \frac{b\,u_0}{s(q+1/\beta)}\frac{e^{-q(r-b)}}{r},$$

where $q = \sqrt{s}/a$ and $\beta = b/(1+b)$.

Step 3: Because

$$\mathcal{L}^{-1}\left[\frac{e^{-\sqrt{s}\,(r-b)/a}}{s(\sqrt{s}/a + 1/\beta)}\right] = \beta\left[\operatorname{erfc}\left(\frac{r-b}{2a\sqrt{t}}\right) - \exp\left(\frac{r-b}{\beta} + \frac{a^2 t}{\beta^2}\right)\operatorname{erfc}\left(\frac{a\sqrt{t}}{\beta} + \frac{r-b}{2a\sqrt{t}}\right)\right],$$

---

[37] See Frisch, H. L, and F. C. Collins, 1952: Diffusional processes in the growth of aerosol particles. *J. Chem. Phys.*, **20**, 1797–1803.

show that

$$u(r,t) = u_0\left[1 - \frac{b-\beta}{r} f(r,t)\right] + \int_0^t \left[1 - \frac{b-\beta}{r} f(r,t-\tau)\right] q(\tau)\, d\tau,$$

where

$$f(r,t) = \operatorname{erfc}\left(\frac{r-b}{2a\sqrt{t}}\right) - \exp\left(\frac{r-b}{\beta} + \frac{a^2 t}{\beta^2}\right) \operatorname{erfc}\left(\frac{a\sqrt{t}}{\beta} + \frac{r-b}{2a\sqrt{t}}\right).$$

18. Consider[38] a viscous fluid located between two fixed walls $x = \pm L$. At $x = 0$ we introduce a thin, infinitely long rigid barrier of mass m per unit area and let it fall under the force of gravity, which points in the direction of positive x. We wish to find the velocity of the fluid $u(x,t)$. The fluid is governed by the partial differential equation

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t,$$

subject to the boundary conditions $u(L,t) = 0$ and $u_t(0,t) - 2\mu u_x(0,t)/m = g$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < L$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2 U(x,s)}{dx^2} - \frac{s}{\nu} U(x,s) = 0, \qquad 0 < x < L,$$

subject to the boundary conditions $U(L,s) = 0$ and $sU(0,s) - 2\mu U'(0,s)/m = g/s$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{g \sinh[(L-x)\sqrt{s/\nu}\,]}{s[s\,\sinh(L\sqrt{s/\nu}\,) + 2\mu\sqrt{s}\,\cosh(L\sqrt{s/\nu}\,)/(m\sqrt{\nu}\,)]}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s = 0$ and $s_n = -\nu\lambda_n^2/L^2$, where $\lambda_n \tan(\lambda_n) = 2\mu L/(m\nu) \equiv k$ and $n = 1, 2, 3, \ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{mg(L-x)}{2\mu} - \frac{4g\mu L^3}{m\nu^2} \sum_{n=1}^{\infty} \frac{\sin[\lambda_n(L-x)/L]\exp(-\nu\lambda_n^2 t/L^2)}{\lambda_n^2[\lambda_n^2 + k(1+k)]\sin(\lambda_n)}.$$

19. Consider[39] a viscous fluid located between two fixed walls $x = \pm L$. At $x = 0$ we introduce a thin, infinitely long rigid barrier of mass m per unit area. The barrier is acted upon by an elastic force in such a manner that it would vibrate with a frequency ω if the

---

[38] See Havelock, T. H., 1921: The solution of an integral equation occurring in certain problems of viscous fluid motion. *Philos. Mag., Ser. 6*, **42**, 620–628.

[39] See Havelock, T. H., 1921: On the decay of oscillation of a solid body in a viscous fluid. *Philos. Mag., Ser. 6*, **42**, 628–634.

liquid were absent. We wish to find the barrier's deviation from equilibrium, $y(t)$. The fluid is governed by the partial differential equation

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t.$$

The boundary conditions are $u(L,t) = my''(t) - 2\mu u_x(0,t) + m\omega^2 y(t) = 0$, $y'(t) = u(0,t)$, $0 < t$, and the initial conditions are $u(x,0) = 0$, $0 < x < L$, $y(0) = A$, and $y'(0) = 0$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s}{\nu}U(x,s) = 0, \qquad 0 < x < L,$$

subject to the boundary conditions $U(L,s) = 0$ and $ms^2Y(s) - 2\mu U'(0,s) + m\omega^2 Y(s) = -msA$ and $sY(s) - A = U(0,s)$.

Step 2: Show that $U(x,s) = B\sinh\left[\sqrt{s/\nu}(L-x)\right]$.

Step 3: Show that at $x = 0$,

$$ms^2Y(s) + 2\mu B\sqrt{\frac{s}{\nu}}\cosh\left[L\sqrt{\frac{s}{\nu}}\,\right] + m\omega^2 Y(s) = msA$$

and

$$sY(s) - A = B\sinh\left[L\sqrt{\frac{s}{\nu}}\,\right].$$

Step 4: Eliminating B in Step 3, show that

$$Y(s) = A\frac{ms + 2\mu\sqrt{s/\nu}\coth\left(L\sqrt{s/\nu}\right)}{ms^2 + 2\mu s\sqrt{s/\nu}\coth\left(L\sqrt{s/\nu}\right) + m\omega^2}.$$

Step 5: Show that $Y(s)$ has simple poles at λ_n which are the roots of

$$\lambda_n^2 + 2\mu\lambda_n^{3/2}\coth\left(L\sqrt{\lambda_n/\nu}\,\right)/(m\sqrt{\nu}\,) + \omega^2 = 0, \qquad n = 1, 2, 3, \ldots.$$

Step 6: Use Bromwich's integral and show that

$$y(t) = \frac{4\mu A\omega^2}{mL}\sum_{n=1}^{\infty}\frac{\lambda_n e^{\lambda_n t}}{\lambda_n^4 - (\frac{2\mu}{mL})(1 + \frac{2\mu L}{m\nu})\lambda_n^3 + 2\omega^2\lambda_n^2 + \frac{6\omega^2\mu}{mL}\lambda_n + \omega^4}.$$

20. Solve[40]

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t,$$

[40] See McCarthy, T. A., and H. J. Goldsmid, 1970: Electro-deposited copper in bismuth telluride. *J. Phys. D*, **3**, 697–706.

subject to the boundary conditions $u_x(0,t) = 0$, $a^2u_x(L,t) + \alpha u(L,t) = F$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < L$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - \frac{s}{a^2}U(x,s) = 0, \qquad 0 < x < L,$$

subject to the boundary conditions $U'(0,s) = 0$ and $a^2U'(L,s) + \alpha U(L,s) = F/s$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{F\,\cosh(qx)}{s[a^2q\sinh(qL) + \alpha\cosh(qL)]},$$

where $q = \sqrt{s}/a$.

Step 3: Show that $U(x,s)$ has simple poles at $s = 0$ and $s_n = -a^2\lambda_n^2/L^2$, where λ_n is the nth root of $\lambda\tan(\lambda) = \alpha L/a^2$, $q_n = i\lambda_n/L$, and $n = 1,2,3,\ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{F}{\alpha}\left\{1 - 2hL\sum_{n=1}^{\infty}\frac{\cos(\lambda_n x/L)\exp(-a^2\lambda_n^2 t/L^2)}{[hL(1+hL)+\lambda_n^2)\cos(\lambda_n)}\right\},$$

where $h = \alpha/a^2$.

21. Solve

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 \le x < 1, \quad 0 \le t,$$

subject to the boundary conditions $u(0,t) = 0$ and $3a\,[u_x(1,t) - u(1,t)] + u_t(1,t) = \delta(t)$, $0 \le t$, and the initial condition $u(x,0) = 0$, $0 \le x < 1$. Here $\delta(t)$ denotes the Dirac delta function.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - sU(x,s) = 0, \qquad 0 < x < 1,$$

subject to the boundary conditions $U(0,s) = 0$ and $3a[U'(1,s) - U(1,s)] + sU(1,s) = 1$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{\sinh(x\sqrt{s}\,)}{3a\,[\sqrt{s}\cosh(\sqrt{s}\,) - \sinh(\sqrt{s}\,)] + s\sinh(\sqrt{s}\,)}.$$

Step 3: Show that $U(x,s)$ has simple poles at $s = 0$ and $s_n = -\lambda_n^2$ where $\lambda_n\cot(\lambda_n) = (3a + \lambda_n^2)/3a$, $n = 1,2,3,\ldots$.

Step 4: Use Bromwich's integral and show that

$$u(x,t) = \frac{x}{a+1} + 2\sum_{n=1}^{\infty}\frac{\sin(\lambda_n x)\exp(-\lambda_n^2 t)}{[3a + 3 + \lambda_n^2/(3a)]\sin(\lambda_n)}.$$

22. Solve[41] the partial differential equation

$$\frac{\partial u}{\partial t} + V\frac{\partial u}{\partial x} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad 0 < t,$$

where V is a constant, subject to the boundary conditions $u(0,t) = 1$, $u_x(1,t) = 0$, $0 < t$, and the initial condition $u(x,0) = 0$, $0 < x < 1$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U(x,s)}{dx^2} - V\frac{dU(x,s)}{dx} - sU(x,s) = 0, \qquad 0 < x < 1,$$

subject to the boundary conditions $U(0,s) = 1/s$ and $U'(1,s) = 0$.

Step 2: Show that the solution to the previous step is

$$\begin{aligned} U(x,s) &= e^{Vx/2}\frac{\mu\cosh[\mu(1-x)] + (V/2)\sinh[\mu(1-x)]}{s\,[\mu\cosh(\mu) + (V/2)\sinh(\mu)]} \\ &= e^{Vx/2}\frac{\sqrt{s'}\cosh\left[(1-x)\sqrt{s'}\,\right] + (V/2)\sinh\left[(1-x)\sqrt{s'}\,\right]}{(s'-V^2/4)\left[\sqrt{s'}\cosh\left(\sqrt{s'}\,\right) + (V/2)\sinh\left(\sqrt{s'}\,\right)\right]}, \end{aligned}$$

where $\mu = \sqrt{s + V^2/4}$ and $s' = s + V^2/4$.

Step 3: Show that $U(x,s)$ has simple poles at $s' = V^2/4$ and $s'_n = -\lambda_n^2$ with $\lambda_n \cot(\lambda_n) = -V/2$, where $n = 1, 2, 3, \ldots$.

Step 4: Using Bromwich's integral, show that

$$\begin{aligned} u(x,t) = 1 &- 2e^{Vx/2 - V^2t/4} \\ &\times \sum_{n=1}^{\infty} \frac{\lambda_n\{(V/2)\sin[\lambda_n(1-x)] + \lambda_n\cos[\lambda_n(1-x)]\}e^{-\lambda_n^2 t}}{(\lambda_n^2 + V^2/4)[\lambda_n\sin(\lambda_n) - (1+V/2)\cos(\lambda_n)]}. \end{aligned}$$

23. Solve[42] the partial differential equation

$$\frac{\partial^2 u}{\partial x\partial t} + a\frac{\partial u}{\partial t} + b\frac{\partial u}{\partial x} = 0, \qquad 0 < x < \infty, \quad 0 < a, b, t,$$

subject to the boundary conditions $u(0,t) = 1$, $\lim_{x\to\infty} u(x,t) \to 0$, $0 < t$, and the initial condition $u_x(x,0) + au(x,0) = 0$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and

$$(s+b)U'(x,s) + asU(x,s) = 0, \qquad 0 < x < \infty,$$

---

[41] See Yoo, H., and E.-T. Pak, 1996: Analytical solutions to a one-dimensional finite-domain model for stratified thermal storage tanks. *Sol. Energy*, **56**, 315–322.

[42] See Liaw, C. H., J. S. P. Wang, R. A. Greenhorn, and K. C. Chao, 1979: Kinetics of fixed-bed absorption: A new solution. *AICHE J.*, **25**, 376–381.

subject to the boundary conditions $\lim_{x\to 0} |U(x,s)| < \infty$ and $U(0,s) = 1/s$.

Step 2: Show that the solution to the previous step is

$$U(x,s) = \frac{1}{s}\exp\left(-\frac{asx}{s+b}\right).$$

Step 3: Because

$$e^{-c\xi} = 1 - c\int_0^{\xi} e^{-c\eta}\,d\eta,$$

show that

$$U(x,s) = \frac{1}{s} - \int_0^{ax} e^{-\eta}\exp\left(\frac{b\eta}{s+b}\right)\frac{d\eta}{s+b}.$$

Step 4: By inverting $U(x,s)$ term by term and using the first shifting theorem,

$$u(x,t) = 1 - e^{-bt}\int_0^{ax} e^{-\eta} I_0\left(2\sqrt{bt\eta}\right)\,d\eta.$$

24. Solve

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) - \frac{\partial u}{\partial t} = \delta(t), \qquad 0 \le r < a, \quad 0 \le t,$$

subject to the boundary conditions $\lim_{r\to 0} |u(r,t)| < \infty$, $u(a,t) = 0$, $0 < t$, and the initial condition $u(r,0) = 0$, $0 \le r < a$, where $\delta(t)$ is the Dirac delta function. Note that $J_n(iz) = i^n I_n(z)$ and $I_n(iz) = i^n J_n(z)$ for all complex z.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dU(r,s)}{dr}\right] - sU(r,s) = 1, \qquad 0 \le r < a,$$

subject to the boundary conditions $\lim_{r\to 0} |U(r,s)| < \infty$ and $U(a,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(r,s) = \frac{I_0(r\sqrt{s}\,) - I_0(a\sqrt{s}\,)}{s\ I_0(a\sqrt{s}\,)}.$$

Step 3: Show that $U(r,s)$ has a removable singularity at $s = 0$ and simple poles at $s_n = -k_n^2/a^2$, where $J_0(k_n) = 0$ and $n = 1, 2, 3, \ldots$.

Step 4: Using Bromwich's integral, show that

$$u(r,t) = -2\sum_{n=1}^{\infty} \frac{J_0(k_n r/a)}{k_n\, J_1(k_n)} e^{-k_n^2 t/a^2}.$$

25. Solve

$$\frac{\partial u}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + H(t), \qquad 0 \le r < a, \quad 0 < t,$$

subject to the boundary conditions $\lim_{r\to 0}|u(r,t)| < \infty$, $u(a,t) = 0$, $0 < t$, and the initial condition $u(r,0) = 0$, $0 \le r < a$. Note that $J_n(iz) = i^n I_n(z)$ and $I_n(iz) = i^n J_n(z)$ for all complex z.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dU(r,s)}{dr}\right] - sU(r,s) = -\frac{1}{s}, \qquad 0 \le r < a,$$

subject to the boundary conditions $\lim_{r\to 0}|U(r,s)| < \infty$ and $U(a,s) = 0$.

Step 2: Show that the solution to the previous step is

$$U(r,s) = \frac{I_0(a\sqrt{s}\,) - I_0(r\sqrt{s}\,)}{s^2\ I_0(a\sqrt{s})}.$$

Step 3: Show that $U(r,s)$ has simple poles at $s = 0$ and $s_n = -k_n^2/a^2$ where $J_0(k_n) = 0$ and $n = 1, 2, 3, \ldots$.

Step 4: Using Bromwich's integral, show that

$$u(r,t) = \frac{a^2 - r^2}{4} - 2a^2 \sum_{n=1}^{\infty} \frac{J_0(k_n r/a)}{k_n^3\ J_1(k_n)} e^{-k_n^2 t/a^2}.$$

26. Solve

$$\frac{\partial u}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right), \qquad 0 \le r < a, \quad 0 < t,$$

subject to the boundary conditions $\lim_{r\to 0}|u(r,t)| < \infty$, $u(a,t) = e^{-t/\tau_0}$, $0 < t$, and the initial condition $u(r,0) = 1$, $0 \le r < a$. Note that $J_n(iz) = i^n I_n(z)$ and $I_n(iz) = i^n J_n(z)$ for all complex z.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dU(r,s)}{dr}\right] - sU(r,s) = -1, \qquad 0 \le r < a,$$

subject to the boundary conditions $\lim_{r\to 0}|U(r,s)| < \infty$ and $U(a,s) = 1/(s + 1/\tau_0)$.

Step 2: Show that the solution to the previous step is

$$U(r,s) = \frac{1}{s} + \left(\frac{1}{s + 1/\tau_0} - \frac{1}{s}\right)\frac{I_0(r\sqrt{s}\,)}{I_0(a\sqrt{s}\,)}.$$

Step 3: Show that $U(r,s)$ has simple poles at $s = 0$, $s = -1/\tau_0$, and $s_n = -k_n^2/a^2$, where $J_0(k_n) = 0$ and $n = 1, 2, 3, \ldots$.

Step 4: Using Bromwich's integral, show that

$$\begin{aligned} u(r,t) &= \frac{J_0(r\sqrt{1/\tau_0}\,)}{J_0(a\sqrt{1/\tau_0}\,)} e^{-t/\tau_0} + 2a^2 \sum_{n=1}^{\infty} \frac{J_0(k_n r/a)}{k_n(a^2 - k_n^2\tau_0)\, J_1(k_n)} e^{-k_n^2 t/a^2} \\ &= e^{-t/\tau_0} + 2a^2 \sum_{n=1}^{\infty} \frac{J_0(k_n r/a)}{k_n(a^2 - k_n^2\tau_0)\, J_1(k_n)} \left(e^{-k_n^2 t/a^2} - e^{-t/\tau_0}\right). \end{aligned}$$

27. Solve

$$\frac{\partial u}{\partial t} = a^2 \left(\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} \right), \qquad 0 \le r < b, \quad 0 < t,$$

subject to the boundary conditions

$$\lim_{r\to 0} |u(r,t)| < \infty, \quad u(b,t) = kt, \qquad 0 < t,$$

and the initial condition $u(r,0) = 0$, $0 \le r < b$. Note that $J_n(iz) = i^n I_n(z)$ and $I_n(iz) = i^n J_n(z)$ for all complex z.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{1}{r}\frac{d}{dr}\left[r \frac{dU(r,s)}{dr} \right] - \frac{s}{a^2} U(r,s) = 0, \qquad 0 \le r < b,$$

subject to the boundary conditions $\lim_{r\to 0} |U(r,s)| < \infty$ and $U(b,s) = k/s^2$.

Step 2: Show that the solution to the previous step is

$$U(r,s) = \frac{k\, I_0(r\sqrt{s}/a)}{s^2\, I_0(b\sqrt{s}/a)}.$$

Step 3: Show that $U(r,s)$ has a second-order pole at $z = 0$ and simple poles at $i\kappa_n = \sqrt{z_n}/a$ or $z_n = -a^2\kappa_n^2$, where $J_0(\kappa_n b) = 0$ and $n = 1, 2, 3, \ldots$.

Step 4: Using Bromwich's integral, show that

$$u(r,t) = k \left[t - \frac{b^2 - r^2}{4a^2} + \frac{2}{a^2 b} \sum_{n=1}^{\infty} \frac{J_0(\kappa_n r)}{\kappa_n^3 J_1(\kappa_n b)} \right].$$

28. Solve the nonhomogeneous heat equation for the spherical shell[43]

$$\frac{\partial u}{\partial t} = a^2 \left(\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{A}{r^4} \right), \qquad \alpha < r < \beta, \quad 0 < t,$$

subject to the boundary conditions $u_r(\alpha,t) = u(\beta,t) = 0$, $0 < t$, and the initial condition $u(r,0) = 0$, $\alpha < r < \beta$.

Step 1: By introducing $v(r,t) = r\,u(r,t)$, show that the problem simplifies to

$$\frac{\partial v}{\partial t} = a^2 \left(\frac{\partial^2 v}{\partial r^2} + \frac{A}{r^3} \right), \qquad \alpha < r < \beta, \quad 0 < t,$$

subject to the boundary conditions $v_r(\alpha,t) - v(\alpha,t)/\alpha = v(\beta,t) = 0$, $0 < t$, and the initial condition $v(r,0) = 0$, $\alpha < r < \beta$.

---

[43] See Malkovich, R. Sh., 1977: Heating of a spherical shell by a radial current. *Sov. Phys. Tech. Phys.*, **22**, 636.

Step 2: Taking the Laplace transform of the differential equation and boundary conditions in Step 1, show that

$$\frac{d^2V(r,s)}{dr^2} - \frac{s}{a^2}V(r,s) = -\frac{A}{sr^3}, \qquad \alpha < r < \beta,$$

along with $V'(\alpha,s) + V(\alpha,s)/\alpha = V(\beta,s) = 0$.

Step 3: Using the method of variation of parameters, show that the particular solution to Step 2 is $V_p(r,s) = u_1(r,s)\cosh(qr) + u_2(r,s)\sinh(qr)$, where

$$u_1(r,s) = \frac{A}{sq}\int_\beta^r \frac{\sinh(q\tau)}{\tau^3}\,d\tau, \qquad u_2(r,s) = -\frac{A}{sq}\int_\beta^r \frac{\cosh(q\tau)}{\tau^3}\,d\tau, \qquad \text{and} \qquad q = \sqrt{s}/a.$$

Step 4: Show that the general solution to Step 2 is

$$V(r,s) = C\sinh[q(r-\beta)] - \frac{A}{sq}\int_\beta^r \frac{\sinh[q(r-\tau)]}{\tau^3}\,d\tau.$$

This solution satisfies $V(\beta,s) = 0$.

Step 5: Use the remaining boundary condition and show that

$$U(r,s) = \frac{A}{srq}\left\{\frac{\sinh[q(\beta-r)]}{\alpha q\,\cosh(q\ell) + \sinh(q\ell)}\int_0^\ell \frac{\alpha q\,\cosh(q\eta) + \sinh(q\eta)}{(\alpha+\eta)^3}\,d\eta - \int_0^{\beta-r}\frac{\sinh(q\eta)}{(r+\eta)^3}\,d\eta\right\},$$

where $\ell = \beta - \alpha$. Note: $V(r,s) = r\,U(r,s)$.

Step 6: Show that $U(r,s)$ has simple poles at $s = 0$ and $s_n = -a^2\gamma_n^2$, where γ_n is the nth root of $\alpha\gamma_n\cos(\gamma_n\ell) + \sin(\gamma_n\ell) = 0$, $n = 1, 2, 3, \ldots$.

Step 7: Use Bromwich's integral and show that

$$\begin{aligned} u(r,t) = A\Biggl\{&\left(\frac{1}{r} - \frac{1}{\beta}\right)\left[\frac{1}{\alpha} - \frac{1}{2}\left(\frac{1}{r} + \frac{1}{\beta}\right)\right] \\ &- \frac{2\alpha^2}{r\ell^2}\sum_{n=0}^\infty \frac{\sin[\gamma_n(\beta-r)]\exp(-a^2\gamma_n^2 t)}{\sin^2(\gamma_n\ell)(\beta + \alpha^2\ell\gamma_n^2)}\int_0^1 \frac{\sin(\gamma_n\ell\eta)}{(\delta-\eta)^3}\,d\eta\Biggr\}, \end{aligned}$$

where γ_n is the nth root of $\alpha\gamma + \tan(\ell\gamma) = 0$, and $\delta = 1 + \alpha/\ell$.

2.6 THE SOLUTION OF LAPLACE'S EQUATION BY LAPLACE TRANSFORMS

Laplace transforms are useful in solving Laplace's or Poisson's equation over a semi-infinite strip. The following problem illustrates this technique.

Let us solve Poisson's equation within a semi-infinite circular cylinder

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{\partial^2 u}{\partial z^2} = \frac{2}{b}n(z)\delta(r-b), \quad 0 \le r < a, \quad 0 < z < \infty, \tag{2.6.1}$$

subject to the boundary conditions

$$u(r,0) = 0, \quad \lim_{z\to\infty}|u(r,z)| < \infty, \quad 0 \le r < a, \tag{2.6.2}$$

and

$$u(a, z) = 0, \qquad 0 < z < \infty, \tag{2.6.3}$$

where $0 < b < a$. This problem gives the electrostatic potential within a semi-infinite cylinder of radius a that is grounded and has the charge density of $n(z)$ within an infinitesimally thin shell located at $r = b$.

Because the domain is semi-infinite in the z direction, we introduce the Laplace transform

$$U(r, s) = \int_0^\infty u(r, z)\, e^{-sz}\, dz. \tag{2.6.4}$$

Thus, taking the Laplace transform of Equation 2.6.1, we have that

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dU(r,s)}{dr}\right] + s^2U(r,s) - su(r,0) - u_z(r,0) = \frac{2}{b}N(s)\delta(r-b). \tag{2.6.5}$$

Although $u(r, 0) = 0$, $u_z(r, 0)$ is unknown and we denote its value by $f(r)$. Therefore, Equation 2.6.5 becomes

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dU(r,s)}{dr}\right] + s^2U(r,s) = f(r) + \frac{2}{b}N(s)\delta(r-b), \quad 0 \le r < a, \tag{2.6.6}$$

with $\lim_{r\to 0}|U(r, s)| < \infty$, and $U(a, s) = 0$.

To solve Equation 2.6.6 we first assume that we can rewrite $f(r)$ as the Fourier-Bessel series

$$f(r) = \sum_{n=1}^{\infty} A_n J_0(k_n r/a), \tag{2.6.7}$$

where k_n is the nth root of the $J_0(k) = 0$, and

$$A_n = \frac{2}{a^2 J_1^2(k_n)} \int_0^a f(r)\, J_0(k_n r/a)\, r\, dr. \tag{2.6.8}$$

Similarly, the expansion for the delta function is

$$\delta(r-b) = \frac{2b}{a^2} \sum_{n=1}^{\infty} \frac{J_0(k_n b/a) J_0(k_n r/a)}{J_1^2(k_n)}, \tag{2.6.9}$$

because

$$\int_0^a \delta(r-b) J_0(k_n r/a)\, r\, dr = b\, J_0(k_n b/a). \tag{2.6.10}$$

Why we chose this particular expansion will become apparent shortly.

Thus, Equation 2.6.6 may be rewritten as

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dU(r,s)}{dr}\right] + s^2U(r,s) = \frac{2}{a^2}\sum_{n=1}^{\infty} \frac{2N(s)J_0(k_n b/a) + a_k}{J_1^2(k_n)} J_0(k_n r/a), \tag{2.6.11}$$

where $a_k = \int_0^a f(r)\, J_0(k_n r/a)\, r\, dr$.

The form of the right side of Equation 2.6.11 suggests that we seek solutions of the form

$$U(r, s) = \sum_{n=1}^{\infty} B_n J_0(k_n r/a), \qquad 0 \le r < a. \tag{2.6.12}$$

We now understand why we rewrote the right side of Equation 2.6.6 as a Fourier-Bessel series; the solution $U(r,s)$ automatically satisfies the boundary condition $U(a,s)=0$. Substituting Equation 2.6.12 into Equation 2.6.11, we find that

$$U(r,s) = \frac{2}{a^2}\sum_{n=1}^{\infty} \frac{2N(s)J_0(k_n b/a) + a_k}{(s^2 - k_n^2/a^2)J_1^2(k_n)} J_0(k_n r/a), \qquad 0 \le r < a. \tag{2.6.13}$$

We have not yet determined a_k. Note, however, that in order for the inverse of Equation 2.6.13 *not* to grow as $e^{k_n z/a}$, the numerator must vanish when $s = k_n/a$ and $s = k_n/a$ is a removable pole. Thus,

$$a_k = -2N(k_n/a)J_0(k_n b/a), \tag{2.6.14}$$

and

$$U(r,s) = \frac{4}{a^2}\sum_{n=1}^{\infty} \frac{[N(s) - N(k_n/a)]J_0(k_n b/a)}{(s^2 - k_n^2/a^2)J_1^2(k_n)} J_0(k_n r/a), \qquad 0 \le r < a. \tag{2.6.15}$$

The inverse of $U(r,s)$ then follows directly from simple inversions, the convolution theorem, and the definition of the Laplace transform. The complete solution is

$$\begin{aligned} u(r,z) = {} & \frac{2}{a}\sum_{n=1}^{\infty} \frac{J_0(k_n b/a)J_0(k_n r/a)}{k_n\, J_1^2(k_n)} \\ & \times \left[\int_0^z n(\tau)e^{k_n(z-\tau)/a}\,d\tau - \int_0^z n(\tau)e^{-k_n(z-\tau)/a}\,d\tau \right. \\ & \qquad \left. - \int_0^\infty n(\tau)e^{-k_n\tau/a}e^{k_n z/a}\,d\tau + \int_0^\infty n(\tau)e^{-k_n\tau/a}e^{-k_n z/a}\,d\tau\right] \qquad (2.6.16) \\ = {} & \frac{2}{a}\sum_{n=1}^{\infty} \frac{J_0(k_n b/a)J_0(k_n r/a)}{k_n\, J_1^2(k_n)} \qquad (2.6.17) \\ & \times \left[\int_0^\infty n(\tau)e^{-k_n(z+\tau)/a}\,d\tau - \int_0^z n(\tau)e^{-k_n(z-\tau)/a}\,d\tau - \int_z^\infty n(\tau)e^{-k_n(\tau-z)/a}\,d\tau\right]. \end{aligned}$$

Problems

1. Use Laplace transforms to solve

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \qquad 0 < x < \infty, \quad 0 < y < a,$$

subject to the boundary conditions $u(0,y) = 1$, $\lim_{x\to\infty} |u(x,y)| < \infty$, $0 < y < a$, and $u(x,0) = u(x,a) = 0$, $0 < x < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{d^2U}{dy^2} + s^2 U = s + f(s,y),$$

subject to the boundary conditions $U(s,0) = U(s,a) = 0$.

Step 2: Because

$$1 = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{[1-(-1)^n]}{n}\sin\left(\frac{n\pi y}{a}\right)$$

and expanding $f(s, y)$ in a half-range sine expansion:

$$f(s,y) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi y}{a}\right), \text{ where } A_n = \frac{2}{a}\int_0^a f(s,y)\sin\left(\frac{n\pi y}{a}\right)\,dy,$$

show that the differential equations in Step 1 can be rewritten

$$\frac{d^2U}{dy^2} + s^2 U = \sum_{n=1}^{\infty}\left\{\frac{2s[1-(-1)^n]}{n\pi} + A_n\right\}\sin\left(\frac{n\pi y}{a}\right),$$

Step 3: Show that the solution of the differential equation in Step 2 is

$$U(s,y) = \sum_{n=1}^{\infty}\frac{2s[1-(-1)^n] + n\pi A_n}{n\pi(s^2 - n^2\pi^2/a^2)}\sin\left(\frac{n\pi y}{a}\right).$$

Step 4: For the solution to remain finite as $x \to \infty$, $s = n\pi/a$ cannot be a pole of the transform $U(s, y)$. Show that $A_n = -2[1-(-1)^n]/a$ and

$$U(s,y) = \sum_{n=1}^{\infty}\frac{2[1-(-1)^n]}{n\pi(s + n\pi/a)}\sin\left(\frac{n\pi y}{a}\right).$$

Step 5: Take the inverse of $U(s, y)$ term by term and show that

$$u(x,y) = \frac{4}{\pi}\sum_{m=1}^{\infty}\frac{1}{2m-1}\exp\left[-\frac{(2m-1)\pi x}{a}\right]\sin\left[\frac{(2m-1)\pi y}{a}\right].$$

2. Use Laplace transforms to solve

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{\partial^2 u}{\partial z^2} = 0, \qquad 0 \le r < a, \quad 0 < z < \infty,$$

subject to the boundary conditions $u(r, 0) = 1$, $\lim_{z\to\infty}|u(r, z)| < \infty$ $0 \le r < a$, and $\lim_{r\to 0}|u(r, z)| < \infty$, $u(a, z) = 0$, $0 < z < \infty$.

Step 1: Take the Laplace transform of the partial differential equation and boundary conditions and show that

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{dU(r,s)}{dr}\right] + s^2 U(r,s) = s + f(r), \quad 0 \le r < a,$$

with $|U(0, s)| < \infty$ and $U(a, s) = 0$.

Step 2: Rewrite $f(r)$ as the Fourier-Bessel series:

$$f(r) = \sum_{n=1}^{\infty} A_n J_0(k_n r/a),$$

where k_n is the nth root of the $J_0(k) = 0$ and

$$A_n = \frac{2}{a^2 J_1^2(k_n)} \int_0^a f(r)\, J_0(k_n r/a)\, r\, dr.$$

Step 3: Because

$$1 = 2 \sum_{n=1}^{\infty} \frac{J_0(k_n r/a)}{k_n\, J_1(k_n)}, \qquad 0 \le r < a,$$

show that the differential equation in Step 1 becomes

$$\frac{1}{r}\frac{d}{dr}\left[r \frac{dU(r,s)}{dr}\right] + s^2 U(r,s) = \frac{2}{a^2} \sum_{n=1}^{\infty} \frac{s a^2 J_1(k_n) - k_n a_k}{k_n\, J_1^2(k_n)} J_0\left(\frac{k_n r}{a}\right),$$

where $a_k = \int_0^a f(r) J_0(k_n r/a)\, r\, dr$.

Step 4: Show that the solution to the differential equation is

$$U(r,s) = \frac{2}{a^2} \sum_{n=1}^{\infty} \frac{s a^2 J_1(k_n) - k_n a_k}{(s^2 - k_n^2/a^2) k_n J_1^2(k_n)} J_0(k_n r/a), \qquad 0 \le r < a.$$

Step 5: Because $s = k_n/a$ cannot be a pole of $U(r,s)$, $a_k = a J_1(k_n)$. Therefore,

$$U(r,s) = 2 \sum_{n=1}^{\infty} \frac{J_0(k_n r/a)}{k_n (s + k_n/a) J_1(k_n)}, \quad 0 \le r < a.$$

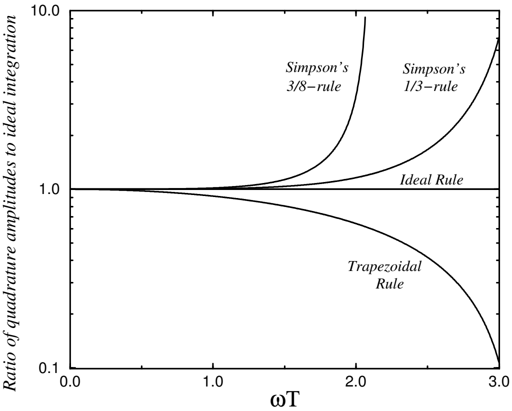
Step 6: Find the inverse of $U(r,s)$ and show that

$$u(r,z) = 2 \sum_{n=1}^{\infty} \frac{J_0(k_n r/a)}{k_n\, J_1(k_n)} e^{-k_n z/a}.$$

Further Readings

Debnath, L., and D. Bhatta, 2015: *Integral Transforms and Their Applications*. CRC Press, 792 pp. A book that covers Laplace, Fourier, z-, Hankel, Mellin, Hilbert and Stieltjes transforms and their application.

Duffy, D. G., 2015: *Transform Methods for Solving Partial Differential Equations*. CRC Press, 728 pp. This book covers the material of this chapter in greater depth.



Chapter 3
The Z-Transform

Since the Second World War, the rise of digital technology has resulted in a corresponding demand for designing and understanding discrete-time (data sampled) systems. These systems are governed by *difference equations* in which members of the sequence y_n are coupled to each other.

One source of difference equations is the numerical evaluation of integrals on a digital computer. Because we can only have values at discrete time points $t_k = kT$ for $k = 0, 1, 2, \ldots$, the value of the integral $y(t) = \int_0^t f(\tau)\,d\tau$ is

$$y(kT) = \int_0^{kT} f(\tau)\,d\tau = \int_0^{(k-1)T} f(\tau)\,d\tau + \int_{(k-1)T}^{kT} f(\tau)\,d\tau \tag{3.0.1}$$

$$= y[(k-1)T] + \int_{(k-1)T}^{kT} f(\tau)\,d\tau = y[(k-1)T] + Tf(kT), \tag{3.0.2}$$

because $\int_{(k-1)T}^{kT} f(\tau)\,d\tau \approx Tf(kT)$. The right side of Equation 3.0.2 is an example of a first-order difference equation because the numerical scheme couples the sequence value $y(kT)$ directly to the previous sequence value $y[(k-1)T]$. If Equation 3.0.2 had contained $y[(k-2)T]$, then it would have been a second-order difference equation, and so forth.

Although we could use the conventional Laplace transform to solve these difference equations, the use of z-transforms can greatly facilitate the analysis, especially when we only desire responses at the sampling instants. Often the entire analysis can be done using only the transforms and the analyst does not actually find the sequence $y(kT)$.

In this chapter we will first define the z-transform and discuss its properties. Then we will show how to find its inverse. Finally, we shall use them to solve difference equations.

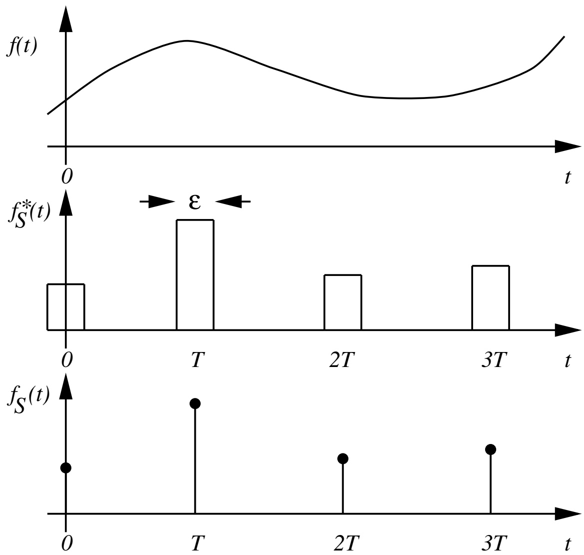


Figure 3.1.1: Schematic of how a continuous function $f(t)$ is sampled by a narrow-width pulse sampler $f_S^*(t)$ and an ideal sampler $f_S(t)$.

3.1 THE RELATIONSHIP OF THE Z-TRANSFORM TO THE LAPLACE TRANSFORM[1]

Let $f(t)$ be a continuous function that an instrument samples every T units of time. We denote this data-sampled function by $f_S^*(t)$. See Figure 3.1.1. Taking ϵ, the duration of an individual sampling event, to be small, we may approximate the narrow-width pulse in Figure 3.1.1 by flat-topped pulses. Then $f_S^*(t)$ approximately equals

$$f_S^*(t) \approx \frac{1}{\epsilon}\sum_{n=0}^{\infty} f(nT)\left[H(t-nT+\epsilon/2) - H(t-nT-\epsilon/2)\right], \tag{3.1.1}$$

if $\epsilon \ll T$.

Clearly the presence of ϵ is troublesome in Equation 3.1.1; it adds one more parameter to our problem. For this reason we introduce the concept of the *ideal sampler*, where the sampling time becomes infinitesimally small so that

$$f_S(t) = \lim_{\epsilon\to 0}\sum_{n=0}^{\infty} f(nT)\left[\frac{H(t-nT+\epsilon/2) - H(t-nT-\epsilon/2)}{\epsilon}\right] = \sum_{n=0}^{\infty} f(nT)\delta(t-nT). \tag{3.1.2}$$

Let us now find the Laplace transform of this data-sampled function. From the linearity property of Laplace transforms,

$$F_S(s) = \mathcal{L}[f_S(t)] = \mathcal{L}\left[\sum_{n=0}^{\infty} f(nT)\delta(t-nT)\right] = \sum_{n=0}^{\infty} f(nT)\mathcal{L}[\delta(t-nT)]. \tag{3.1.3}$$

---

[1] Gera (Gera, A. E., 1999: The relationship between the z-transform and the discrete-time Fourier transform. *IEEE Trans. Auto. Control*, **AC-44**, 370–371) explored the general relationship between the one-sided discrete-time Fourier transform and the one-sided z-transform. See also Naumović, M. B., 2001: Interrelationship between the one-sided discrete-time Fourier transform and one-sided delta transform. *Electr. Engng.*, **83**, 99–101.

Because $\mathcal{L}[\delta(t - nT)] = e^{-nsT}$, Equation 3.1.3 simplifies to

$$F_S(s) = \sum_{n=0}^{\infty} f(nT)e^{-nsT}. \tag{3.1.4}$$

If we now make the substitution that $z = e^{sT}$, then $F_S(s)$ becomes

$$\boxed{F(z) = \mathcal{Z}(f_n) = \sum_{n=0}^{\infty} f_n z^{-n},} \tag{3.1.5}$$

where $F(z)$ is the one-sided z-transform[2] of the sequence $f(nT)$, which we shall now denote by f_n. Here $\mathcal{Z}$ denotes the operation of taking the z-transform while $\mathcal{Z}^{-1}$ represents the inverse z-transformation. We will consider methods for finding the inverse z-transform in Section 3.3.

Just as the Laplace transform was defined by an integration in t, the z-transform is defined by a power series (Laurent series) in z. Consequently, every z-transform has a region of convergence that must be implicitly understood if not explicitly stated. Furthermore, just as the Laplace integral diverged for certain functions, there are sequences where the associated power series diverges and its z-transform does not exist.

Consider now the following examples of how to find the z-transform.

• Example 3.1.1

Given the unit sequence $f_n = 1,\ n \geq 0$, let us find $F(z)$. Substituting f_n into the definition of the z-transform leads to

$$F(z) = \sum_{n=0}^{\infty} z^{-n} = \frac{z}{z-1}, \tag{3.1.6}$$

because $\sum_{n=0}^{\infty} z^{-n}$ is a complex-valued *geometric series* with common ratio z^{-1}. This series converges if $|z^{-1}| < 1$ or $|z| > 1$, which gives the region of convergence of $F(z)$.

MATLAB's symbolic toolbox provides an alternative to the hand computation of the z-transform. In the present case, the command

```
>> syms z; syms n positive
>> ztrans(1,n,z)
```

yields

```
ans =
z/(z-1)
```

□

• Example 3.1.2

Let us find the z-transform of the sequence

$$f_n = e^{-anT}, \qquad n \geq 0, \tag{3.1.7}$$

---

[2] The standard reference is Jury, E. I., 1964: *Theory and Application of the z-Transform Method*. John Wiley & Sons, 330 pp.

for a real and a imaginary.

For a real, substitution of the sequence into the definition of the z-transform yields

$$F(z) = \sum_{n=0}^{\infty} e^{-anT} z^{-n} = \sum_{n=0}^{\infty} \left(e^{-aT} z^{-1}\right)^n. \tag{3.1.8}$$

If $u = e^{-aT} z^{-1}$, then Equation 3.1.8 is a geometric series so that

$$F(z) = \sum_{n=0}^{\infty} u^n = \frac{1}{1-u}. \tag{3.1.9}$$

Because $|u| = e^{-aT}|z^{-1}|$, the condition for convergence is that $|z| > e^{-aT}$. Thus,

$$F(z) = \frac{z}{z - e^{-aT}}, \qquad |z| > e^{-aT}. \tag{3.1.10}$$

For imaginary a, the infinite series in Equation 3.1.8 converges if $|z| > 1$, because $|u| = |z^{-1}|$ when a is imaginary. Thus,

$$F(z) = \frac{z}{z - e^{-aT}}, \qquad |z| > 1. \tag{3.1.11}$$

Although the z-transforms in Equation 3.1.10 and Equation 3.1.11 are the same in these two cases, the corresponding regions of convergence are different. If a is a complex number, then

$$F(z) = \frac{z}{z - e^{-aT}}, \qquad |z| > |e^{-aT}|. \tag{3.1.12}$$

Checking our work using MATLAB, we type the commands:

```
>> syms a z; syms n T positive
>> ztrans(exp(-a*n*T),n,z);
>> simplify(ans)
```

which yields

```
ans =
z*exp(a*T)/(z*exp(a*T)-1)
```

□

• Example 3.1.3

Let us find the z-transform of the sinusoidal sequence

$$f_n = \cos(n\omega T), \qquad n \geq 0. \tag{3.1.13}$$

Substituting Equation 3.1.13 into the definition of the z-transform results in

$$F(z) = \sum_{n=0}^{\infty} \cos(n\omega T) z^{-n}. \tag{3.1.14}$$

From Euler's formula,

$$\cos(n\omega T) = \tfrac{1}{2}(e^{in\omega T} + e^{-in\omega T}), \tag{3.1.15}$$

so that Equation 3.1.14 becomes

$$F(z) = \frac{1}{2}\sum_{n=0}^{\infty}\left(e^{in\omega T}z^{-n} + e^{-in\omega T}z^{-n}\right), \tag{3.1.16}$$

or

$$F(z) = \tfrac{1}{2}\left[\mathcal{Z}(e^{in\omega T}) + \mathcal{Z}(e^{-in\omega T})\right]. \tag{3.1.17}$$

From Equation 3.1.11,

$$\mathcal{Z}(e^{\pm in\omega T}) = \frac{z}{z - e^{\pm i\omega T}}, \qquad |z| > 1. \tag{3.1.18}$$

Substituting Equation 3.1.18 into Equation 3.1.17 and simplifying yields

$$F(z) = \frac{z[z - \cos(\omega T)]}{z^2 - 2z\cos(\omega T) + 1}, \qquad |z| > 1. \tag{3.1.19}$$

□

• Example 3.1.4

Let us find the z-transform for the sequence

$$f_n = \begin{cases} 1, & 0 \le n \le 5, \\ (\frac{1}{2})^n, & 6 \le n. \end{cases} \tag{3.1.20}$$

From the definition of the z-transform,

$$\mathcal{Z}(f_n) = F(z) = \sum_{n=0}^{5} z^{-n} + \sum_{n=6}^{\infty}\left(\frac{1}{2z}\right)^n. \tag{3.1.21}$$

$$= 1 + \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} + \frac{1}{z^4} + \frac{1}{z^5} + \frac{2z}{2z-1}$$
$$- 1 - \frac{1}{2z} - \frac{1}{4z^2} - \frac{1}{8z^3} - \frac{1}{16z^4} - \frac{1}{32z^5} \tag{3.1.22}$$

$$= \frac{2z}{2z-1} + \frac{1}{2z} + \frac{3}{4z^2} + \frac{7}{8z^3} + \frac{15}{16z^4} + \frac{31}{32z^5}. \tag{3.1.23}$$

We could also have obtained Equation 3.1.23 via MATLAB by typing the commands:

```
>> syms z; syms n positive
>> ztrans('1+((1/2)^n-1)*Heaviside(n-6)',n,z)
```

which yields

```
ans =
2*z/(2*z-1)+1/2/z+3/4/z^2+7/8/z^3+15/16/z^4+31/32/z^5
```

□

We summarize some of the more commonly encountered sequences and their transforms in Table 3.1.1 along with their regions of convergence.

• Example 3.1.5

In many engineering studies, the analysis is done entirely using transforms without actually finding any inverses. Consequently, it is useful to compare and contrast how various transforms behave in very simple test problems.

Consider the time function $f(t) = ae^{-at}H(t)$, $a > 0$. Its Laplace and Fourier transform are identical, namely $a/(a + i\omega)$, if we set $s = i\omega$. In Figure 3.1.2 we illustrate its behavior as a function of positive ω.

Let us now generate the sequence of observations that we would measure if we sampled $f(t)$ every T units of time apart: $f_n = ae^{-anT}$. Taking the z-transform of this sequence, it equals $az/\left(z - e^{-aT}\right)$. Recalling that $z = e^{sT} = e^{i\omega T}$, we can also plot this transform as a function of positive ω. For small ω, the transforms agree, but as ω becomes larger they diverge markedly. Why does this occur?

Recall that the z-transform is computed from a sequence comprised of samples from a continuous signal. One very important flaw in sampled data is the possible misrepresentation of high-frequency effects as lower-frequency phenomena. It is this *aliasing* or *folding* effect that we are observing here. Consequently, the z-transform of a sampled record can differ markedly from the corresponding Laplace or Fourier transforms of the continuous record at frequencies above one half of the sampling frequency. This also suggests that care should be exercised in interpolating between sampling instants. Indeed, in those applications where the output between sampling instants is very important, such as in a hybrid mixture of digital and analog systems, we must apply the so-called "modified z-transform."

Problems

From the fundamental definition of the z-transform, find the transform of the following sequences, where $n \geq 0$. Then check your answer using MATLAB.

1. $f_n = \left(\frac{1}{2}\right)^n$

2. $f_n = e^{in\theta}$

3. $f_n = \begin{cases} 1, & 0 \leq n \leq 5, \\ 0, & 5 < n \end{cases}$

4. $f_n = \begin{cases} \left(\frac{1}{2}\right)^n, & n = 0, 1, \ldots, 10 \\ \left(\frac{1}{4}\right)^n, & n \geq 11 \end{cases}$

5. $f_n = \begin{cases} 0, & n = 0, \\ -1, & n = 1, \\ a^n, & n \geq 2 \end{cases}$

3.2 SOME USEFUL PROPERTIES

In principle we could construct any desired transform from the definition of the z-transform. However, there are several general theorems that are much more effective in finding new transforms.

Table 3.1.1: Z-Transforms of Some Commonly Used Sequences

	$\mathbf{f_n},\ \mathbf{n} \geq \mathbf{0}$	$\mathbf{F(z)}$	**Region of convergence**
1.	$f_0 = k = \text{const.}$ $f_n = 0,\ n \geq 1$	k	$\|z\| > 0$
2.	$f_m = k = \text{const.}$ $f_n = 0$, for all values of $n \neq m$	kz^{-m}	$\|z\| > 0$
3.	$k = \text{constant}$	$kz/(z-1)$	$\|z\| > 1$
4.	kn	$kz/(z-1)^2$	$\|z\| > 1$
5.	kn^2	$kz(z+1)/(z-1)^3$	$\|z\| > 1$
6.	ke^{-anT}, a complex	$kz/\left(z - e^{-aT}\right)$	$\|z\| > \|e^{-aT}\|$
7.	kne^{-anT}, a complex	$\frac{kze^{-aT}}{(z-e^{-aT})^2}$	$\|z\| > \|e^{-aT}\|$
8.	$\sin(\omega_0 nT)$	$\frac{z\sin(\omega_0 T)}{z^2-2z\cos(\omega_0 T)+1}$	$\|z\| > 1$
9.	$\cos(\omega_0 nT)$	$\frac{z[z-\cos(\omega_0 T)]}{z^2-2z\cos(\omega_0 T)+1}$	$\|z\| > 1$
10.	$e^{-anT}\sin(\omega_0 nT)$	$\frac{ze^{-aT}\sin(\omega_0 T)}{z^2-2ze^{-aT}\cos(\omega_0 T)+e^{-2aT}}$	$\|z\| > e^{-aT}$
11.	$e^{-anT}\cos(\omega_0 nT)$	$\frac{ze^{-aT}[ze^{aT}-\cos(\omega_0 T)]}{z^2-2ze^{-aT}\cos(\omega_0 T)+e^{-2aT}}$	$\|z\| > e^{-aT}$
12.	α^n , α constant	$z/(z-\alpha)$	$\|z\| > \|\alpha\|$
13.	$n\alpha^n$	$\alpha z/(z-\alpha)^2$	$\|z\| > \|\alpha\|$
14.	$n^2\alpha^n$	$\alpha z(z+\alpha)/(z-\alpha)^3$	$\|z\| > \|\alpha\|$
15.	$\sinh(\omega_0 nT)$	$\frac{z\sinh(\omega_0 T)}{z^2-2z\cosh(\omega_0 T)+1}$	$\|z\| > \cosh(\omega_0 T)$
16.	$\cosh(\omega_0 nT)$	$\frac{z[z-\cosh(\omega_0 T)]}{z^2-2z\cosh(\omega_0 T)+1}$	$\|z\| > \sinh(\omega_0 T)$
17.	$a^n/n!$	$e^{a/z}$	$\|z\| > 0$
18.	$[\ln(a)]^n/n!$	$a^{1/z}$	$\|z\| > 0$

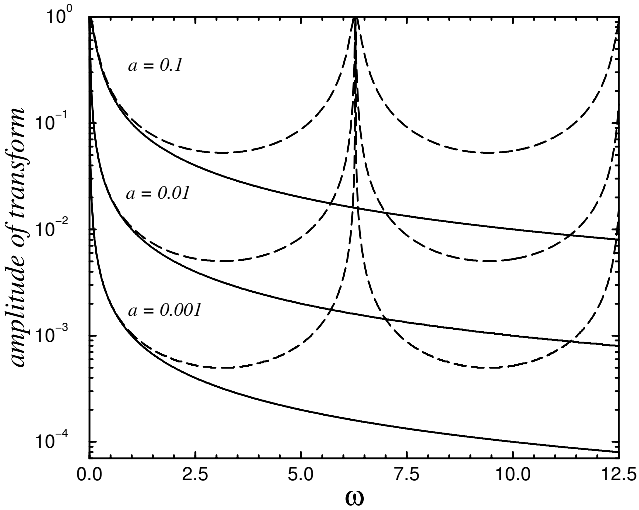


Figure 3.1.2: The amplitude of the Laplace or Fourier transform (solid line) for the function $ae^{-at}H(t)$ and the z-transform (dashed line) for the sequence $f_n = ae^{-anT}$ as a function of frequency ω for various positive values of a and $T = 1$.

Linearity

From the definition of the z-transform, it immediately follows that

$$\text{if} \quad h_n = c_1 f_n + c_2 g_n, \quad \text{then} \quad H(z) = c_1 F(z) + c_2 G(z), \tag{3.2.1}$$

where $F(z) = \mathcal{Z}(f_n)$, $G(z) = \mathcal{Z}(g_n)$, $H(z) = \mathcal{Z}(h_n)$, and c_1, c_2 are arbitrary constants.

Multiplication by an exponential sequence

$$\text{If} \quad g_n = e^{-anT} f_n, \quad n \geq 0, \quad \text{then} \quad G(z) = F(ze^{aT}). \tag{3.2.2}$$

This follows from

$$G(z) = \mathcal{Z}(g_n) = \sum_{n=0}^{\infty} g_n\, z^{-n} = \sum_{n=0}^{\infty} e^{-anT} f_n\, z^{-n} = \sum_{n=0}^{\infty} f_n (ze^{aT})^{-n} = F(ze^{aT}). \tag{3.2.3}$$

This is the z-transform analog to the first shifting theorem in Laplace transforms.

Shifting

The effect of shifting depends upon whether it is to the right or to the left, as Table 3.2.1 illustrates. For the sequence f_{n-2}, no values from the sequence f_n are lost; thus, we anticipate that the z-transform of f_{n-2} only involves $F(z)$. However, in forming the

Table 3.2.1: Examples of Shifting Involving Sequences

n	f_n	f_{n-2}	f_{n+2}
0	1	0	4
1	2	0	8
2	4	1	16
3	8	2	64
4	16	4	128
$\vdots$	$\vdots$	$\vdots$	$\vdots$

sequence f_{n+2}, the first two values of f_n are lost, and we anticipate that the z-transform of f_{n+2} cannot be expressed solely in terms of $F(z)$ but must include those two lost pieces of information.

Let us now confirm these conjectures by finding the z-transform of f_{n+1}, which is a sequence that has been shifted one step to the left. From the definition of the z-transform, it follows that

$$\mathcal{Z}(f_{n+1}) = \sum_{n=0}^{\infty} f_{n+1} z^{-n} = z \sum_{n=0}^{\infty} f_{n+1} z^{-(n+1)} \tag{3.2.4}$$

or

$$\mathcal{Z}(f_{n+1}) = z \sum_{k=1}^{\infty} f_k z^{-k} + z f_0 - z f_0, \tag{3.2.5}$$

where we added zero in Equation 3.2.5. This algebraic trick allows us to collapse the first two terms on the right side of Equation 3.2.5 into one and

$$\mathcal{Z}(f_{n+1}) = zF(z) - zf_0. \tag{3.2.6}$$

In a similar manner, repeated applications of Equation 3.2.6 yield

$$\mathcal{Z}(f_{n+m}) = z^m F(z) - z^m f_0 - z^{m-1} f_1 - \ldots - z f_{m-1}, \tag{3.2.7}$$

where $m > 0$. This shifting operation transforms f_{n+m} into an algebraic expression involving m. Furthermore, we introduced initial sequence values, just as we introduced initial conditions when we took the Laplace transform of the nth derivative of $f(t)$. We will make frequent use of this property in solving difference equations in Section 3.4.

Consider now shifting to the right by the positive integer k,

$$g_n = f_{n-k} H_{n-k}, \quad n \geq 0, \tag{3.2.8}$$

where $H_{n-k} = 0$ for $n < k$ and 1 for $n \geq k$. Then the z-transform of Equation 3.2.8 is

$$G(z) = z^{-k} F(z), \tag{3.2.9}$$

where $G(z) = \mathcal{Z}(g_n)$, and $F(z) = \mathcal{Z}(f_n)$. This follows from

$$G(z) = \sum_{n=0}^{\infty} g_n z^{-n} = \sum_{n=0}^{\infty} f_{n-k} H_{n-k} z^{-n} = z^{-k} \sum_{n=k}^{\infty} f_{n-k} z^{-(n-k)} = z^{-k} \sum_{m=0}^{\infty} f_m z^{-m} = z^{-k} F(z). \tag{3.2.10}$$

This result is the z-transform analog to the second shifting theorem in Laplace transforms.

In symbolic calculations involving MATLAB, the operator H_{n-k} can be expressed by `Heaviside(n-k)`.

Initial-value theorem

The initial value of the sequence f_n, f_0, can be computed from $F(z)$ using the initial-value theorem:

$$f_0 = \lim_{z\to\infty} F(z). \tag{3.2.11}$$

From the definition of the z-transform,

$$F(z) = \sum_{n=0}^{\infty} f_n z^{-n} = f_0 + f_1 z^{-1} + f_2 z^{-2} + \ldots. \tag{3.2.12}$$

In the limit of $z \to \infty$, we obtain the desired result.

Final-value theorem

The value of f_n, as $n \to \infty$, is given by the final-value theorem:

$$f_\infty = \lim_{z\to 1} (z-1)F(z), \tag{3.2.13}$$

where $F(z)$ is the z-transform of f_n.

We begin by noting that

$$\mathcal{Z}(f_{n+1} - f_n) = \lim_{n\to\infty} \sum_{k=0}^{n} (f_{k+1} - f_k) z^{-k}. \tag{3.2.14}$$

Using the shifting theorem on the left side of Equation 3.2.14,

$$zF(z) - zf_0 - F(z) = \lim_{n\to\infty} \sum_{k=0}^{n} (f_{k+1} - f_k) z^{-k}. \tag{3.2.15}$$

Applying the limit as z approaches 1 to both sides of Equation 3.2.15:

$$\lim_{z\to 1} (z-1)F(z) - f_0 = \lim_{n\to\infty} \sum_{k=0}^{n} (f_{k+1} - f_k) \tag{3.2.16}$$

$$= \lim_{n\to\infty} \left[(f_1 - f_0) + (f_2 - f_1) + \ldots + (f_n - f_{n-1}) + (f_{n+1} - f_n) + \ldots\right] \tag{3.2.17}$$

$$= \lim_{n\to\infty} (-f_0 + f_{n+1}) = -f_0 + f_\infty. \tag{3.2.18}$$

Consequently,

$$f_\infty = \lim_{z\to 1} (z-1)F(z). \tag{3.2.19}$$

Note that this limit has meaning only if f_∞ exists. This occurs if $F(z)$ has no second-order or higher poles on the unit circle and no poles outside the unit circle.

Multiplication by n

Given

$$g_n = nf_n, \qquad n \geq 0, \tag{3.2.20}$$

this theorem states that

$$G(z) = -z\frac{dF(z)}{dz}, \tag{3.2.21}$$

where $G(z) = \mathcal{Z}(g_n)$, and $F(z) = \mathcal{Z}(f_n)$.

This follows from

$$G(z) = \sum_{n=0}^{\infty} g_n z^{-n} = \sum_{n=0}^{\infty} nf_n z^{-n} = z\sum_{n=0}^{\infty} nf_n z^{-n-1} = -z\frac{dF(z)}{dz}. \tag{3.2.22}$$

Periodic sequence theorem

Consider the N-periodic sequence:

$$f_n = \{\underbrace{f_0 f_1 f_2 \ldots f_{N-1}}_{\text{first period}} f_0 f_1 \ldots\}, \tag{3.2.23}$$

and the related sequence:

$$x_n = \begin{cases} f_n, & 0 \leq n \leq N-1, \\ 0, & N \leq n. \end{cases} \tag{3.2.24}$$

This theorem allows us to find the z-transform of f_n if we can find the z-transform of x_n via the relationship

$$F(z) = \frac{X(z)}{1 - z^{-N}}, \qquad |z^N| > 1, \tag{3.2.25}$$

where $X(z) = \mathcal{Z}(x_n)$.

This follows from

$$F(z) = \sum_{n=0}^{\infty} f_n z^{-n} = \sum_{n=0}^{N-1} x_n z^{-n} + \sum_{n=N}^{2N-1} x_{n-N} z^{-n} + \sum_{n=2N}^{3N-1} x_{n-2N} z^{-n} + \cdots. \tag{3.2.26}$$

Application of the shifting theorem in Equation 3.2.26 leads to

$$F(z) = X(z) + z^{-N}X(z) + z^{-2N}X(z) + = X(z)\left[1 + z^{-N} + z^{-2N} + \cdots\right]. \tag{3.2.27}$$

Equation 3.2.27 contains an infinite geometric series with common ratio z^{-N}, which converges if $|z^{-N}| < 1$. Thus,

$$F(z) = \frac{X(z)}{1 - z^{-N}}, \qquad |z^N| > 1. \tag{3.2.28}$$

Convolution

Given the sequences f_n and g_n, the convolution product of these two sequences is

$$w_n = f_n * g_n = \sum_{k=0}^{n} f_k g_{n-k} = \sum_{k=0}^{n} f_{n-k} g_k. \tag{3.2.29}$$

Given $F(z)$ and $G(z)$, we then have that $W(z) = F(z)G(z)$.

This follows from

$$W(z) = \sum_{n=0}^{\infty} \left[\sum_{k=0}^{n} f_k g_{n-k} \right] z^{-n} = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} f_k g_{n-k} z^{-n}, \tag{3.2.30}$$

because $g_{n-k} = 0$ for $k > n$. Reversing the order of summation and letting $m = n - k$,

$$W(z) = \sum_{k=0}^{\infty} \sum_{m=-k}^{\infty} f_k g_m z^{-(m+k)} = \left[\sum_{k=0}^{\infty} f_k z^{-k} \right] \left[\sum_{m=0}^{\infty} g_m z^{-m} \right] = F(z)G(z). \tag{3.2.31}$$

We can use MATLAB's command `conv( )`, which multiplies two polynomials to perform discrete convolution as follows:

```
>>x = [1 1 1 1 1 1 1];
>>y = [1 2 4 8 16 32 64];
>>z = conv(x,y)
```

produces

```
z =
    1 3 7 15 31 63 127 126 124 120 112 96 64
```

The first seven values of `z` contain the convolution of the sequence `x` with the sequence `y`.

Consider now the following examples of the properties discussed in this section.

• Example 3.2.1

From

$$\mathcal{Z}(a^n) = \frac{1}{1 - az^{-1}}, \tag{3.2.32}$$

for $n \geq 0$ and $|z| > |a|$, we have that

$$\mathcal{Z}\left(e^{inx}\right) = \frac{1}{1 - e^{ix} z^{-1}}, \tag{3.2.33}$$

and

$$\mathcal{Z}\left(e^{-inx}\right) = \frac{1}{1 - e^{-ix}z^{-1}}, \tag{3.2.34}$$

if $n \geq 0$ and $|z| > 1$. Therefore, the sequence $f_n = \cos(nx)$ has the z-transform

$$F(z) = \mathcal{Z}[\cos(nx)] = \tfrac{1}{2}\mathcal{Z}\left(e^{inx}\right) + \tfrac{1}{2}\mathcal{Z}\left(e^{-inx}\right) \tag{3.2.35}$$

$$= \frac{1}{2}\frac{1}{1 - e^{ix}z^{-1}} + \frac{1}{2}\frac{1}{1 - e^{-ix}z^{-1}} = \frac{1 - \cos(x)z^{-1}}{1 - 2\cos(x)z^{-1} + z^{-2}}. \tag{3.2.36}$$

□

• **Example 3.2.2**

Using the z-transform,

$$\mathcal{Z}(a^n) = \frac{1}{1 - az^{-1}}, \qquad n \geq 0, \tag{3.2.37}$$

we find that

$$\mathcal{Z}(na^n) = -z\frac{d}{dz}\left[\left(1 - az^{-1}\right)^{-1}\right] = (-z)(-1)\left(1 - az^{-1}\right)^{-2}(-a)(-1)z^{-2} \tag{3.2.38}$$

$$= \frac{az^{-1}}{\left(1 - az^{-1}\right)^2} = \frac{az}{(z-a)^2}. \tag{3.2.39}$$

□

• **Example 3.2.3**

Consider $F(z) = 2az^{-1}/(1 - az^{-1})^3$, where $|a| < |z|$ and $|a| < 1$. Here we have that

$$f_0 = \lim_{z\to\infty} F(z) = \lim_{z\to\infty} \frac{2az^{-1}}{(1 - az^{-1})^3} = 0 \tag{3.2.40}$$

from the initial-value theorem. This agrees with the inverse of $F(z)$:

$$f_n = n(n+1)a^n, \qquad n \geq 0. \tag{3.2.41}$$

If the z-transform consists of the ratio of two polynomials, we can use MATLAB to find f_0. For example, if $F(z) = 2z^2/(z-1)^3$, we can find f_0 as follows:

```
>>num = [2 0 0];
>>den = conv([1 -1],[1 -1]);
>>den = conv(den,[1 -1]);
>>initialvalue = polyval(num,1e20) / polyval(den,1e20)
initialvalue =
    2.0000e-20
```

Therefore, $f_0 = 0$. □

• Example 3.2.4

Given the z-transform $F(z) = (1-a)z/[(z-1)(z-a)]$, where $|z| > 1 > a > 0$, then from the final-value theorem we have that

$$\lim_{n\to\infty} f_n = \lim_{z\to 1}(z-1)F(z) = \lim_{z\to 1}\frac{1-a}{1-az^{-1}} = 1. \tag{3.2.42}$$

This is consistent with the inverse transform $f_n = 1 - a^n$ with $n \geq 0$. □

• Example 3.2.5

Using the sequences $f_n = 1$ and $g_n = a^n$, where a is real, verify the convolution theorem.

We first compute the convolution of f_n with g_n, namely

$$w_n = f_n * g_n = \sum_{k=0}^{n} a^k = \frac{1}{1-a} - \frac{a^{n+1}}{1-a}. \tag{3.2.43}$$

Taking the z-transform of w_n,

$$W(z) = \frac{z}{(1-a)(z-1)} - \frac{az}{(1-a)(z-a)} = \frac{z^2}{(z-1)(z-a)} = F(z)G(z), \tag{3.2.44}$$

and the convolution theorem holds true for this special case.

Problems

Use the properties of z-transforms and Table 3.1.1 to find the z-transform of the following sequences. Then check your answer using MATLAB.

1. $f_n = nTe^{-anT}$

2. $f_n = \begin{cases} 0, & n = 0 \\ na^{n-1}, & n \geq 1 \end{cases}$

3. $f_n = \begin{cases} 0, & n = 0 \\ n^2a^{n-1}, & n \geq 1 \end{cases}$

4. $f_n = a^n \cos(n)$

   [Hint : Use $\cos(n) = \frac{1}{2}(e^{in} + e^{-in})$]

5. $f_n = \cos(n-2)H_{n-2}$

6. $f_n = 3 + e^{-2nT}$

7. $f_n = \sin(n\omega_0 T + \theta)$,

8. $f_n = \begin{cases} 0, & n = 0 \\ 1, & n = 1 \\ 2, & n = 2 \\ 1, & n = 3, \end{cases} \qquad f_{n+4} = f_n$

9. $f_n = (-1)^n$

   (Hint: f_n is a periodic sequence.)

10. Using the property stated in Equation 3.2.20 and Equation 3.2.21 *twice*, find the z-transform of $n^2 = n[n(1)^n]$. Then verify your result using MATLAB.

11. Verify the convolution theorem using the sequences $f_n = g_n = 1$. Then check your results using MATLAB.

12. Verify the convolution theorem using the sequences $f_n = 1$ and $g_n = n$. Then check your results using MATLAB.

13. Verify the convolution theorem using the sequences $f_n = g_n = 1/(n!)$. Then check your results using MATLAB. Hint: Use the binomial theorem with $x = 1$ to evaluate the summation.

14. If a is a real number, show that $\mathcal{Z}(a^n f_n) = F(z/a)$, where $\mathcal{Z}(f_n) = F(z)$.

3.3 INVERSE Z-TRANSFORMS

In the previous two sections we dealt with finding the z-transform. In this section we find f_n by inverting the z-transform $F(z)$. There are four methods for finding the inverse: (1) power series, (2) recursion, (3) partial fractions, and (4) the residue method. We will discuss each technique individually. The first three apply only to those functions $F(z)$ that are *rational* functions while the residue method is more general. For symbolic computations with MATLAB, you can use `iztrans`.

Power series

By means of the long-division process, we can always rewrite $F(z)$ as the Laurent expansion:

$$F(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots. \tag{3.3.1}$$

From the definition of the z-transform,

$$F(z) = \sum_{n=0}^{\infty} f_n z^{-n} = f_0 + f_1 z^{-1} + f_2 z^{-2} + \cdots, \tag{3.3.2}$$

the desired sequence f_n is given by a_n.

• Example 3.3.1

Let

$$F(z) = \frac{z+1}{2z-2} = \frac{N(z)}{D(z)}. \tag{3.3.3}$$

Using long division, $N(z)$ is divided by $D(z)$ and we obtain

$$F(z) = \tfrac{1}{2} + z^{-1} + z^{-2} + z^{-3} + z^{-4} + \cdots. \tag{3.3.4}$$

Therefore,

$$a_0 = \tfrac{1}{2},\ a_1 = 1,\ a_2 = 1,\ a_3 = 1,\ a_4 = 1,\ \text{etc.}, \tag{3.3.5}$$

which suggests that $f_0 = \frac{1}{2}$ and $f_n = 1$ for $n \geq 1$ is the inverse of $F(z)$. □

• Example 3.3.2

Let us find the inverse of the z-transform:

$$F(z) = \frac{2z^2 - 1.5z}{z^2 - 1.5z + 0.5}. \tag{3.3.6}$$

By the long-division process, we have that

$$\begin{array}{r|llllllll}
 & 2 & + & 1.5z^{-1} & + & 1.25z^{-2} & + & 1.125z^{-3} & + \cdots \\ \hline
z^2 - 1.5z + 0.5 & 2z^2 & - & 1.5z & & & & & \\
 & 2z^2 & - & 3z & + & 1 & & & \\ \cline{2-6}
 & & & 1.5z & - & 1 & & & \\
 & & & 1.5z & - & 2.25 & + & 0.750z^{-1} & \\ \cline{4-8}
 & & & & & 1.25 & - & 0.750z^{-1} & \\
 & & & & & 1.25 & - & 1.870z^{-1} & + \cdots \\ \cline{6-9}
 & & & & & & & 1.125z^{-1} & + \cdots
\end{array}$$

Thus, $f_0 = 2$, $f_1 = 1.5$, $f_2 = 1.25$, $f_3 = 1.125$, and so forth, or $f_n = 1 + (\frac{1}{2})^n$. In general, this technique only produces numerical values for some of the elements of the sequence. Note also that our long division must always yield the power series Equation 3.3.1 in order for this method to be of any use.

To check our answer using MATLAB, we type the commands:

```
syms z; syms n positive
iztrans((2*z^2 - 1.5*z)/(z^2 - 1.5*z + 0.5),z,n)
```

which yields

```
ans =
1 + (1/2)^n
```

□

Recursive method

An alternative to long division was suggested[3] several years ago. It obtains the inverse recursively.

We begin by assuming that the z-transform is of the form

$$F(z) = \frac{a_0 z^m + a_1 z^{m-1} + a_2 z^{m-2} + \cdots + a_{m-1} z + a_m}{b_0 z^m + b_1 z^{m-1} + b_2 z^{m-2} + \cdots + b_{m-1} z + b_m}, \tag{3.3.7}$$

where some of the coefficients a_i and b_i may be zero and $b_0 \neq 0$. Applying the initial-value theorem,

$$f_0 = \lim_{z\to\infty} F(z) = a_0/b_0. \tag{3.3.8}$$

---

[3] Jury, E. I., 1964: *Theory and Application of the z-Transform Method*. John Wiley & Sons, p. 41; Pierre, D. A., 1963: A tabular algorithm for z-transform inversion. *Control Engng.*, **10(9)**, 110–111; Jenkins, L. B., 1967: A useful recursive form for obtaining inverse z-transforms. *Proc. IEEE*, **55**, 574–575.

Next, we apply the initial-value theorem to $z[F(z) - f_0]$ and find that

$$f_1 = \lim_{z\to\infty} z[F(z) - f_0] \tag{3.3.9}$$

$$= \lim_{z\to\infty} z \frac{(a_0 - b_0 f_0)z^m + (a_1 - b_1 f_0)z^{m-1} + \cdots + (a_m - b_m f_0)}{b_0 z^m + b_1 z^{m-1} + b_2 z^{m-2} + \cdots + b_{m-1} z + b_m} \tag{3.3.10}$$

$$= (a_1 - b_1 f_0)/b_0. \tag{3.3.11}$$

Note that the coefficient $a_0 - b_0 f_0 = 0$ from Equation 3.3.8. Similarly,

$$f_2 = \lim_{z\to\infty} z[zF(z) - zf_0 - f_1] \tag{3.3.12}$$

$$= \lim_{z\to\infty} z \frac{(a_0 - b_0 f_0)z^{m+1} + (a_1 - b_1 f_0 - b_0 f_1)z^m + (a_2 - b_2 f_0 - b_1 f_1)z^{m-1} + \cdots - b_m f_1}{b_0 z^m + b_1 z^{m-1} + b_2 z^{m-2} + \cdots + b_{m-1} z + b_m} \tag{3.3.13}$$

$$= (a_2 - b_2 f_0 - b_1 f_1)/b_0 \tag{3.3.14}$$

because $a_0 - b_0 f_0 = a_1 - b_1 f_0 - f_1 b_0 = 0$. Continuing this process, we finally have that

$$f_n = (a_n - b_n f_0 - b_{n-1} f_1 - \cdots - b_1 f_{n-1})/b_0, \tag{3.3.15}$$

where $a_n = b_n \equiv 0$ for $n > m$.

• Example 3.3.3

Let us redo Example 3.3.2 using the recursive method. Comparing Equation 3.3.7 to Equation 3.3.6, $a_0 = 2$, $a_1 = -1.5$, $a_2 = 0$, $b_0 = 1$, $b_1 = -1.5$, $b_2 = 0.5$, and $a_n = b_n = 0$ if $n \geq 3$. From Equation 3.3.15,

$$f_0 = a_0/b_0 = 2/1 = 2, \tag{3.3.16}$$

$$f_1 = (a_1 - b_1 f_0)/b_0 = [-1.5 - (-1.5)(2)]/1 = 1.5, \tag{3.3.17}$$

$$f_2 = (a_2 - b_2 f_0 - b_1 f_1)/b_0 \tag{3.3.18}$$

$$= [0 - (0.5)(2) - (-1.5)(1.5)]/1 = 1.25, \tag{3.3.19}$$

and

$$f_3 = (a_3 - b_3 f_0 - b_2 f_1 - b_1 f_2)/b_0 \tag{3.3.20}$$

$$= [0 - (0)(2) - (0.5)(1.5) - (-1.5)(1.25)]/1 = 1.125. \tag{3.3.21}$$

□

Partial fraction expansion

One of the popular methods for inverting Laplace transforms is partial fractions. A similar, but slightly different, scheme works here.

• Example 3.3.4

Given $F(z) = z/\left(z^2 - 1\right)$, let us find f_n. The first step is to obtain the partial fraction expansion of $F(z)/z$. Why we want $F(z)/z$ rather than $F(z)$ will be made clear in a moment. Thus,

$$\frac{F(z)}{z} = \frac{1}{(z-1)(z+1)} = \frac{A}{z-1} + \frac{B}{z+1}, \tag{3.3.22}$$

where

$$A = (z-1)\left.\frac{F(z)}{z}\right|_{z=1} = \frac{1}{2}, \qquad \text{and} \qquad B = (z+1)\left.\frac{F(z)}{z}\right|_{z=-1} = -\frac{1}{2}. \tag{3.3.23}$$

Multiplying Equation 3.3.22 by z,

$$F(z) = \frac{1}{2}\left(\frac{z}{z-1} - \frac{z}{z+1}\right). \tag{3.3.24}$$

Next, we find the inverse z-transform of $z/(z-1)$ and $z/(z+1)$ in Table 3.1.1. This yields

$$\mathcal{Z}^{-1}\left(\frac{z}{z-1}\right) = 1, \qquad \text{and} \qquad \mathcal{Z}^{-1}\left(\frac{z}{z+1}\right) = (-1)^n. \tag{3.3.25}$$

Thus, the inverse is

$$f_n = \tfrac{1}{2}\left[1 - (-1)^n\right], \ n \geq 0. \tag{3.3.26}$$

□

From this example it is clear that there are two steps: (1) obtain the partial fraction expansion of $F(z)/z$, and (2) find the inverse z-transform by referring to Table 3.1.1.

• Example 3.3.5

Given $F(z) = 2z^2/[(z+2)(z+1)^2]$, let us find f_n. We begin by expanding $F(z)/z$ as

$$\frac{F(z)}{z} = \frac{2z}{(z+2)(z+1)^2} = \frac{A}{z+2} + \frac{B}{z+1} + \frac{C}{(z+1)^2}, \tag{3.3.27}$$

where

$$A = (z+2)\left.\frac{F(z)}{z}\right|_{z=-2} = -4, \qquad B = \frac{d}{dz}\left.\left[(z+1)^2\frac{F(z)}{z}\right]\right|_{z=-1} = 4, \tag{3.3.28}$$

and

$$C = (z+1)^2\left.\frac{F(z)}{z}\right|_{z=-1} = -2, \tag{3.3.29}$$

so that

$$F(z) = \frac{4z}{z+1} - \frac{4z}{z+2} - \frac{2z}{(z+1)^2}, \tag{3.3.30}$$

or

$$f_n = \mathcal{Z}^{-1}\left[\frac{4z}{z+1}\right] - \mathcal{Z}^{-1}\left[\frac{4z}{z+2}\right] - \mathcal{Z}^{-1}\left[\frac{2z}{(z+1)^2}\right]. \tag{3.3.31}$$

From Table 3.1.1,

$$\mathcal{Z}^{-1}\left(\frac{z}{z+1}\right) = (-1)^n, \qquad \mathcal{Z}^{-1}\left(\frac{z}{z+2}\right) = (-2)^n, \tag{3.3.32}$$

and

$$\mathcal{Z}^{-1}\left[\frac{z}{(z+1)^2}\right] = -\ \mathcal{Z}^{-1}\left[\frac{-z}{(z+1)^2}\right] = -n(-1)^n = n(-1)^{n+1}. \tag{3.3.33}$$

Applying Equation 3.3.32 and Equation 3.3.33 to Equation 3.3.31,

$$f_n = 4(-1)^n - 4(-2)^n + 2n(-1)^n,\ n \geq 0. \tag{3.3.34}$$

□

• **Example 3.3.6**

Given $F(z) = (z^2 + z)/(z-2)^2$, let us determine f_n. Because

$$\frac{F(z)}{z} = \frac{z+1}{(z-2)^2} = \frac{1}{z-2} + \frac{3}{(z-2)^2}, \tag{3.3.35}$$

$$f_n = \mathcal{Z}^{-1}\left[\frac{z}{z-2}\right] + \mathcal{Z}^{-1}\left[\frac{3z}{(z-2)^2}\right]. \tag{3.3.36}$$

Referring to Table 3.1.1,

$$\mathcal{Z}^{-1}\left(\frac{z}{z-2}\right) = 2^n, \qquad \text{and} \qquad \mathcal{Z}^{-1}\left[\frac{3z}{(z-2)^2}\right] = \tfrac{3}{2}n2^n. \tag{3.3.37}$$

Substituting Equation 3.3.37 into Equation 3.3.36 yields

$$f_n = \left(\tfrac{3}{2}n + 1\right)2^n,\ n \geq 0. \tag{3.3.38}$$

□

Residue method

The power series, recursive, and partial fraction expansion methods are rather limited. We now prove that f_n may be computed from the following *inverse integral formula*:

$$f_n = \frac{1}{2\pi i}\oint_C z^{n-1}F(z)\,dz, \quad n \geq 0, \tag{3.3.39}$$

where C is any simple curve, taken in the positive sense, that encloses all of the singularities of $F(z)$. It is readily shown that the power series and partial fraction methods are *special cases* of the residue method.

Proof: Starting with the definition of the z-transform

$$F(z) = \sum_{n=0}^{\infty} f_n z^{-n}, \qquad |z| > R_1, \tag{3.3.40}$$

we multiply Equation 3.3.40 by z^{n-1} and integrating both sides around any contour C that includes all of the singularities:

$$\frac{1}{2\pi i}\oint_C z^{n-1}F(z)\,dz = \sum_{m=0}^{\infty} f_m \frac{1}{2\pi i}\oint_C z^{n-m}\frac{dz}{z}. \tag{3.3.41}$$

Let C be a circle of radius R, where $R > R_1$. Then, changing variables to $z = Re^{i\theta}$, and $dz = iz\,d\theta$,

$$\frac{1}{2\pi i}\oint_C z^{n-m}\frac{dz}{z} = \frac{R^{n-m}}{2\pi}\int_0^{2\pi} e^{i(n-m)\theta}\,d\theta = \begin{cases} 1, & m = n, \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.42}$$

Substituting Equation 3.3.42 into Equation 3.3.41 yields the desired result that

$$\frac{1}{2\pi i}\oint_C z^{n-1}F(z)\,dz = f_n. \tag{3.3.43}$$

□

We can easily evaluate the inversion integral, Equation 3.3.39, using Cauchy's residue theorem.

• Example 3.3.7

Let us find the inverse z-transform of

$$F(z) = \frac{1}{(z-1)(z-2)}. \tag{3.3.44}$$

From the inversion integral,

$$f_n = \frac{1}{2\pi i}\oint_C \frac{z^{n-1}}{(z-1)(z-2)}\,dz. \tag{3.3.45}$$

Clearly the integral has simple poles at $z = 1$ and $z = 2$. However, when $n = 0$ we also have a simple pole at $z = 0$. Thus the cases $n = 0$ and $n > 0$ must be considered separately.

Case 1: $n = 0$. The residue theorem yields

$$f_0 = \operatorname{Res}\left[\frac{1}{z(z-1)(z-2)};0\right] + \operatorname{Res}\left[\frac{1}{z(z-1)(z-2)};1\right] + \operatorname{Res}\left[\frac{1}{z(z-1)(z-2)};2\right]. \tag{3.3.46}$$

Evaluating these residues,

$$\operatorname{Res}\left[\frac{1}{z(z-1)(z-2)};0\right] = \left.\frac{1}{(z-1)(z-2)}\right|_{z=0} = \frac{1}{2}, \tag{3.3.47}$$

$$\operatorname{Res}\left[\frac{1}{z(z-1)(z-2)};1\right] = \left.\frac{1}{z(z-2)}\right|_{z=1} = -1, \tag{3.3.48}$$

and

$$\operatorname{Res}\left[\frac{1}{z(z-1)(z-2)};2\right] = \left.\frac{1}{z(z-1)}\right|_{z=2} = \frac{1}{2}. \tag{3.3.49}$$

Substituting Equation 3.3.47 through Equation 3.3.49 into Equation 3.3.46 yields $f_0 = 0$.

Case 2: $n > 0$. Here we only have contributions from $z = 1$ and $z = 2$.

$$f_n = \operatorname{Res}\left[\frac{z^{n-1}}{(z-1)(z-2)};1\right] + \operatorname{Res}\left[\frac{z^{n-1}}{(z-1)(z-2)};2\right], \quad n > 0, \tag{3.3.50}$$

where

$$\operatorname{Res}\left[\frac{z^{n-1}}{(z-1)(z-2)};1\right] = \left.\frac{z^{n-1}}{z-2}\right|_{z=1} = -1, \tag{3.3.51}$$

and

$$\operatorname{Res}\left[\frac{z^{n-1}}{(z-1)(z-2)};2\right] = \left.\frac{z^{n-1}}{z-1}\right|_{z=2} = 2^{n-1}, \ n > 0. \tag{3.3.52}$$

Thus,

$$f_n = 2^{n-1} - 1, \qquad n > 0. \tag{3.3.53}$$

Combining our results,

$$f_n = \begin{cases} 0, & n = 0, \\ \frac{1}{2}\left(2^n - 2\right), & n > 0. \end{cases} \tag{3.3.54}$$

□

- **Example 3.3.8**

Let us use the inversion integral to find the inverse of

$$F(z) = \frac{z^2 + 2z}{(z-1)^2}. \tag{3.3.55}$$

The inversion theorem gives

$$f_n = \frac{1}{2\pi i}\oint_C \frac{z^{n+1} + 2z^n}{(z-1)^2}\,dz = \operatorname{Res}\left[\frac{z^{n+1} + 2z^n}{(z-1)^2};1\right], \tag{3.3.56}$$

where the pole at $z = 1$ is second order. Consequently, the corresponding residue is

$$\operatorname{Res}\left[\frac{z^{n+1} + 2z^n}{(z-1)^2};1\right] = \left.\frac{d}{dz}\left(z^{n+1} + 2z^n\right)\right|_{z=1} = 3n + 1. \tag{3.3.57}$$

Thus, the inverse z-transform of Equation 3.3.55 is

$$f_n = 3n + 1, \quad n \ge 0. \tag{3.3.58}$$

□

• Example 3.3.9

Let $F(z)$ be a z-transform whose poles lie within the unit circle $|z| = 1$. Then

$$F(z) = \sum_{n=0}^{\infty} f_n z^{-n}, \qquad |z| > 1, \tag{3.3.59}$$

and

$$F(z)F(z^{-1}) = \sum_{n=0}^{\infty} f_n^2 + \sum_{\substack{n=0 \\ n\neq m}}^{\infty} \sum_{m=0}^{\infty} f_m f_n z^{m-n}. \tag{3.3.60}$$

We now multiply both sides of Equation 3.3.60 by z^{-1} and integrate around the unit circle C. Therefore,

$$\oint_{|z|=1} F(z)F(z^{-1})z^{-1}\,dz = \sum_{n=0}^{\infty} \oint_{|z|=1} f_n^2 z^{-1}\,dz + \sum_{\substack{n=0 \\ n\neq m}}^{\infty} \sum_{m=0}^{\infty} f_m f_n \oint_{|z|=1} z^{m-n-1}\,dz, \tag{3.3.61}$$

after interchanging the order of integration and summation. Performing the integration,

$$\sum_{n=0}^{\infty} f_n^2 = \frac{1}{2\pi i} \oint_{|z|=1} F(z)F(z^{-1})z^{-1}\,dz, \tag{3.3.62}$$

which is *Parseval's theorem* for one-sided z-transforms. Recall that there are similar theorems for Fourier series and transforms. □

• Example 3.3.10: Evaluation of partial summations[4]

We begin by noting that

$$S_N = \sum_{n=1}^{N} f_n = \frac{1}{2\pi i} \oint_C F(z) \sum_{n=1}^{N} z^{n-1}\,dz. \tag{3.3.63}$$

Here we employed the inversion integral to replace f_n and reversed the order of integration and summation. This interchange is permissible since we only have a partial summation. Because the summation in Equation 3.3.63 is a geometric series, we have the final result that

$$S_N = \frac{1}{2\pi i} \oint_C \frac{F(z)(z^N - 1)}{z - 1}\,dz. \tag{3.3.64}$$

Therefore, we can use the residue theorem and z-transforms to evaluate partial summations.

---

[4] See Bunch, K. J., W. N. Cain, and R. W. Grow, 1990: The z-transform method of evaluating partial summations in closed form. *J. Phys. A*, **23**, L1213–L1215.

Let us find $S_N = \sum_{n=1}^{N} n^3$. Because $f_n = n^3$, $F(z) = z(z^2+4z+1)/(z-1)^4$. Consequently

$$S_N = \text{Res}\left[\frac{z(z^2+4z+1)(z^N-1)}{(z-1)^5};1\right] = \frac{1}{4!}\frac{d^4}{dz^4}\left[z(z^2+4z+1)(z^N-1)\right]\bigg|_{z=1} \tag{3.3.65}$$

$$= \frac{1}{4!}\frac{d^4}{dz^4}\left(z^{N+3}+4z^{N+2}+z^{N+1}-z^3-4z^2-z\right)\bigg|_{z=1} = \tfrac{1}{4}(N+1)^2N^2. \tag{3.3.66}$$

□

• Example 3.3.11

An additional benefit of understanding inversion by the residue method is the ability to *qualitatively* anticipate the inverse by knowing the location of the poles of $F(z)$. This intuition is important because many engineering analyses discuss stability and performance entirely in terms of the properties of the system's z-transform. In Figure 3.3.1 we graphed the location of the poles of $F(z)$ and the corresponding f_n. The student should go through the mental exercise of connecting the two pictures.

Problems

Use the power series or recursive method to compute the first few values of f_n of the following z-transforms. Then check your answers with MATLAB.

1. $F(z) = \dfrac{0.09z^2+0.9z+0.09}{12.6z^2-24z+11.4}$

2. $F(z) = \dfrac{z+1}{2z^4-2z^3+2z-2}$

3. $F(z) = \dfrac{1.5z^2+1.5z}{15.25z^2-36.75z+30.75}$

4. $F(z) = \dfrac{6z^2+6z}{19z^3-33z^2+21z-7}$

Use partial fractions to find the inverse of the following z-transforms. Then verify your answers with MATLAB.

5. $F(z) = \dfrac{z(z+1)}{(z-1)(z^2-z+1/4)}$

6. $F(z) = \dfrac{(1-e^{-aT})z}{(z-1)(z-e^{-aT})}$

7. $F(z) = \dfrac{z^2}{(z-1)(z-\alpha)}$

8. $F(z) = \dfrac{(2z-a-b)z}{(z-a)(z-b)}$

9. Using the property that the z-transform of $g_n = f_{n-k}H_{n-k}$ if $n \ge 0$ is $G(z) = z^{-k}F(z)$, find the inverse of

$$F(z) = \frac{z+1}{z^{10}(z-1/2)}.$$

Then check your answer with MATLAB.

Use the residue method to find the inverse z-transform of the following z-transforms. Then verify your answer with MATLAB.

10. $F(z) = \dfrac{z^2+3z}{(z-1/2)^3}$

11. $F(z) = \dfrac{z}{(z+1)^2(z-2)}$

12. $F(z) = \dfrac{z}{(z+1)^2(z-1)^2}$

13. $F(z) = e^{a/z}$

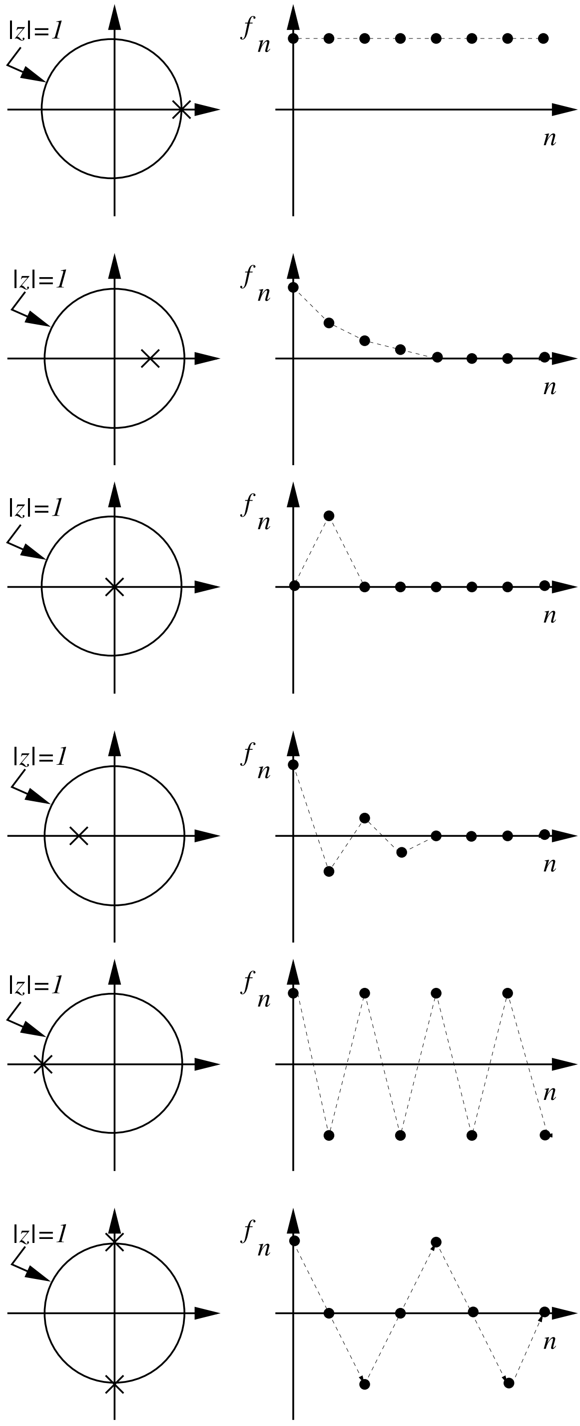


Figure 3.3.1: The correspondence between the location of the simple poles of the z-transform $F(z)$ and the behavior of f_n.

3.4 SOLUTION OF DIFFERENCE EQUATIONS

Having reached the point where we can take a z-transform and then find its inverse, we are ready to use it to solve difference equations. The procedure parallels that of solving ordinary differential equations by Laplace transforms. Essentially we reduce the difference equation to an algebraic problem. We then find the solution by inverting $Y(z)$.

• Example 3.4.1

Let us solve the second-order difference equation

$$2y_{n+2} - 3y_{n+1} + y_n = 5\ 3^n, \ n \geq 0, \tag{3.4.1}$$

where $y_0 = 0$ and $y_1 = 1$.

Taking the z-transform of both sides of Equation 3.4.1, we obtain

$$2\mathcal{Z}(y_{n+2}) - 3\mathcal{Z}(y_{n+1}) + \mathcal{Z}(y_n) = 5\,\mathcal{Z}(3^n). \tag{3.4.2}$$

From the shifting theorem and Table 3.1.1,

$$2z^2Y(z) - 2z^2y_0 - 2zy_1 - 3[zY(z) - zy_0] + Y(z) = \frac{5z}{z-3}. \tag{3.4.3}$$

Substituting $y_0 = 0$ and $y_1 = 1$ into Equation 3.4.3 and simplifying yields

$$(2z-1)(z-1)Y(z) = \frac{z(2z-1)}{z-3}, \qquad \text{or} \qquad Y(z) = \frac{z}{(z-3)(z-1)}. \tag{3.4.4}$$

To obtain y_n from $Y(z)$ we can employ partial fractions or the residue method. Applying partial fractions gives

$$\frac{Y(z)}{z} = \frac{A}{z-1} + \frac{B}{z-3}, \tag{3.4.5}$$

where

$$A = (z-1)\left.\frac{Y(z)}{z}\right|_{z=1} = -\frac{1}{2}, \qquad \text{and} \qquad B = (z-3)\left.\frac{Y(z)}{z}\right|_{z=3} = \frac{1}{2}. \tag{3.4.6}$$

Thus,

$$Y(z) = -\frac{1}{2}\frac{z}{z-1} + \frac{1}{2}\frac{z}{z-3}, \qquad \text{or} \qquad y_n = -\frac{1}{2}\mathcal{Z}^{-1}\left(\frac{z}{z-1}\right) + \frac{1}{2}\mathcal{Z}^{-1}\left(\frac{z}{z-3}\right). \tag{3.4.7}$$

From Equation 3.4.7 and Table 3.1.1,

$$y_n = \tfrac{1}{2}\left(3^n - 1\right), \qquad n \geq 0. \tag{3.4.8}$$

An alternative to this hand calculation is to use MATLAB's `ztrans` and `iztrans` to solve difference equations. In the present case, the MATLAB script would read

```
clear
% define symbolic variables
syms z Y; syms n positive
```

```
% take z-transform of left side of difference equation
LHS = ztrans(2*sym('y(n+2)')-3*sym('y(n+1)')+sym('y(n)'),n,z);
% take z-transform of right side of difference equation
RHS = 5 * ztrans(3^n,n,z);
% set Y for z-transform of y and introduce initial conditions
newLHS = subs(LHS,'ztrans(y(n),n,z)','y(0)','y(1)',Y,0,1);
% solve for Y
Y = solve(newLHS-RHS,Y);
% invert z-transform and find y(n)
y = iztrans(Y,z,n)
```

This script produced

```
y =
-1/2+1/2*3^n
```

Two checks confirm that we have the *correct* solution. First, our solution must satisfy the initial values of the sequence. Computing y_0 and y_1,

$$y_0 = \tfrac{1}{2}(3^0 - 1) = \tfrac{1}{2}(1-1) = 0, \qquad \text{and} \qquad y_1 = \tfrac{1}{2}(3^1 - 1) = \tfrac{1}{2}(3-1) = 1. \tag{3.4.9}$$

Thus, our solution gives the correct initial values.

Our sequence y_n must also satisfy the difference equation. Now

$$y_{n+2} = \tfrac{1}{2}(3^{n+2}-1) = \tfrac{1}{2}(9\,3^n-1), \qquad \text{and} \qquad y_{n+1} = \tfrac{1}{2}(3^{n+1}-1) = \tfrac{1}{2}(3\,3^n-1). \tag{3.4.10}$$

Therefore,

$$2y_{n+2} - 3y_{n+1} + y_n = \left(9 - \tfrac{9}{2} + \tfrac{1}{2}\right) 3^n - 1 + \tfrac{3}{2} - \tfrac{1}{2} = 5\,3^n \tag{3.4.11}$$

and our solution is correct.

Finally, we note that the term $3^n/2$ is necessary to give the right side of Equation 3.4.1; it is the particular solution. The $-1/2$ term is necessary so that the sequence satisfies the initial values; it is the complementary solution. □

• Example 3.4.2

Let us find the y_n in the difference equation

$$y_{n+2} - 2y_{n+1} + y_n = 1, \quad n \geq 0 \tag{3.4.12}$$

with the initial conditions $y_0 = 0$ and $y_1 = 3/2$.

From Equation 3.4.12,

$$\mathcal{Z}(y_{n+2}) - 2\mathcal{Z}(y_{n+1}) + \mathcal{Z}(y_n) = \mathcal{Z}(1). \tag{3.4.13}$$

The z-transform of the left side of Equation 3.4.13 is obtained from the shifting theorem, and Table 3.1.1 yields $\mathcal{Z}(1)$. Thus,

$$z^2Y(z) - z^2y_0 - zy_1 - 2zY(z) + 2zy_0 + Y(z) = \frac{z}{z-1}. \tag{3.4.14}$$

Substituting $y_0 = 0$ and $y_1 = 3/2$ in Equation 3.4.14 and simplifying gives

$$Y(z) = \frac{3z^2 - z}{2(z-1)^3} \qquad \text{or} \qquad y_n = \mathcal{Z}^{-1}\left[\frac{3z^2 - z}{2(z-1)^3}\right]. \tag{3.4.15}$$

We find the inverse z-transform of Equation 3.4.15 by the residue method, or

$$y_n = \frac{1}{2\pi i} \oint_C \frac{3z^{n+1} - z^n}{2(z-1)^3}\, dz = \frac{1}{2!} \frac{d^2}{dz^2} \left[\frac{3z^{n+1}}{2} - \frac{z^n}{2} \right] \bigg|_{z=1} = \tfrac{1}{2}n^2 + n. \qquad \textbf{(3.4.16)}$$

Thus,

$$y_n = \tfrac{1}{2}n^2 + n, \quad n \geq 0. \qquad \textbf{(3.4.17)}$$

Note that $n^2/2$ gives the particular solution to Equation 3.4.12, while n is there so that y_n satisfies the initial conditions. This problem is particularly interesting because our constant forcing produces a response that grows as n^2, just as in the case of resonance in a time-continuous system when a finite forcing such as $\sin(\omega_0 t)$ results in a response whose amplitude grows as t^m. □

• **Example 3.4.3**

Let us solve the difference equation

$$b^2 y_n + y_{n+2} = 0, \qquad \textbf{(3.4.18)}$$

where the initial conditions are $y_0 = b^2$ and $y_1 = 0$.

We begin by taking the z-transform of each term in Equation 3.4.18. This yields

$$b^2 \mathcal{Z}(y_n) + \mathcal{Z}(y_{n+2}) = 0. \qquad \textbf{(3.4.19)}$$

From the shifting theorem, it follows that

$$b^2 Y(z) + z^2 Y(z) - z^2 y_0 - z y_1 = 0. \qquad \textbf{(3.4.20)}$$

Substituting $y_0 = b^2$ and $y_1 = 0$ into Equation 3.4.20,

$$b^2 Y(z) + z^2 Y(z) - b^2 z^2 = 0, \qquad \text{or} \qquad Y(z) = \frac{b^2 z^2}{z^2 + b^2}. \qquad \textbf{(3.4.21)}$$

To find y_n we employ the residue method or

$$y_n = \frac{1}{2\pi i} \oint_C \frac{b^2 z^{n+1}}{(z - ib)(z + ib)}\, dz. \qquad \textbf{(3.4.22)}$$

Thus,

$$y_n = \frac{b^2 z^{n+1}}{z + ib} \bigg|_{z=ib} + \frac{b^2 z^{n+1}}{z - ib} \bigg|_{z=-ib} = \frac{b^{n+2} i^n}{2} + \frac{b^{n+2}(-i)^n}{2} \qquad \textbf{(3.4.23)}$$

$$= \frac{b^{n+2} e^{in\pi/2}}{2} + \frac{b^{n+2} e^{-in\pi/2}}{2} = b^{n+2} \cos\left(\frac{n\pi}{2}\right), \qquad \textbf{(3.4.24)}$$

because $\cos(x) = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$. Consequently, we obtain the desired result that

$$y_n = b^{n+2} \cos\left(\frac{n\pi}{2}\right) \quad \text{for } n \geq 0. \qquad \textbf{(3.4.25)}$$

□

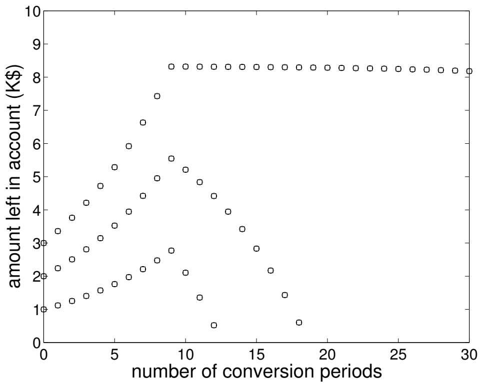


Figure 3.4.1: The amount in a savings account as a function of an annual conversion period when interest is compounded at the annual rate of 12% and $1000 is taken from the account every period starting with period 10.

• Example 3.4.4: Compound interest

Difference equations arise in finance because the increase or decrease in an account occurs in discrete steps. For example, the amount of money in a compound interest savings account after $n + 1$ conversion periods (the time period between interest payments) is

$$y_{n+1} = y_n + ry_n, \tag{3.4.26}$$

where r is the interest rate per conversion period. The second term on the right side of Equation 3.4.32 is the amount of interest paid at the end of each period.

Let us ask a somewhat more difficult question of how much money we will have if we withdraw the amount A at the end of every period starting after the period ℓ. Now the difference equation reads

$$y_{n+1} = y_n + ry_n - AH_{n-\ell-1}. \tag{3.4.27}$$

Taking the z-transform of Equation 3.4.27,

$$zY(z) - zy_0 = (1 + r)Y(z) - \frac{Az^{2-\ell}}{z - 1} \tag{3.4.28}$$

after using Equation 3.2.9 or

$$Y(z) = \frac{y_0 z}{z - (1 + r)} - \frac{Az^{2-\ell}}{(z - 1)[z - (1 + r)]}. \tag{3.4.29}$$

Taking the inverse of Equation 3.4.29,

$$y_n = y_0(1 + r)^n - \frac{A}{r}\left[(1 + r)^{n-\ell+1} - 1\right] H_{n-\ell}. \tag{3.4.30}$$

The first term in Equation 3.4.30 represents the growth of money by compound interest while the second term gives the depletion of the account by withdrawals.

Figure 3.4.1 gives the values of y_n for various starting amounts assuming an annual conversion period with $r = 0.12$, $\ell = 10$ years, and $A = \$1000$. These computations were done in two ways using MATLAB as follows:

```
% load in parameters
clear; r = 0.12; A = 1; k = 0:30;
y = zeros(length(k),3); yanswer = zeros(length(k),3);
% set initial condition
for m=1:3
  y(1,m) = m;
% compute other y values
  for n = 1:30
    y(n+1,m) = y(n,m)+r*y(n,m);
    y(n+1,m) = y(n+1,m)-A*stepfun(n,11);
  end
% now use Equation 3.4.30
  for n = 1:31
    yanswer(n,m) = y(1,m)*(1+r)^(n-1);
    yanswer(n,m) = yanswer(n,m)-A*((1+r)^(n-10)-1)
            *stepfun(n,11)/r;
end; end;
plot(k,y,'o'); hold; plot(k,yanswer,'s');
axis([0 30 0 10])
xlabel('number of conversion periods','Fontsize',20)
ylabel('amount left in account (K$)','Fontsize',20)
```

Figure 3.4.1 shows that if an investor places an initial amount of \$3000 in an account bearing 12% annually, after 10 years he can withdraw \$1000 annually, essentially forever. This is because the amount that he removes every year is replaced by the interest on the funds that remain in the account. □

• Example 3.4.5

Let us solve the following system of difference equations:

$$x_{n+1} = 4x_n + 2y_n, \qquad \text{and} \qquad y_{n+1} = 3x_n + 3y_n, \tag{3.4.31}$$

with the initial values of $x_0 = 0$ and $y_0 = 5$.

Taking the z-transform of Equation 3.4.31,

$$zX(z) - x_0 z = 4X(z) + 2Y(z), \qquad zY(z) - y_0 z = 3X(z) + 3Y(z), \tag{3.4.32}$$

or

$$(z-4)X(z) - 2Y(z) = 0, \qquad 3X(z) - (z-3)Y(z) = -5z. \tag{3.4.33}$$

Solving for $X(z)$ and $Y(z)$,

$$X(z) = \frac{10z}{(z-6)(z-1)} = \frac{2z}{z-6} - \frac{2z}{z-1}, \tag{3.4.34}$$

and

$$Y(z) = \frac{5z(z-4)}{(z-6)(z-1)} = \frac{2z}{z-6} + \frac{3z}{z-1}. \tag{3.4.35}$$

Taking the inverse of Equation 3.4.34 and Equation 3.4.35 term by term,

$$x_n = -2 + 2\ 6^n, \qquad \text{and} \qquad y_n = 3 + 2\ 6^n. \tag{3.4.36}$$

We can also check our work using the MATLAB script

```
clear
% define symbolic variables
syms X Y z; syms n positive
% take z-transform of left side of differential equations
LHS1 = ztrans(sym('x(n+1)')-4*sym('x(n)')-2*sym('y(n)'),n,z);
LHS2 = ztrans(sym('y(n+1)')-3*sym('x(n)')-3*sym('y(n)'),n,z);
% set X and Y for the z-transform of x and y
%     and introduce initial conditions
newLHS1 = subs(LHS1,'ztrans(x(n),n,z)','ztrans(y(n),n,z)',...
        'x(0)','y(0)',X,Y,0,5);
newLHS2 = subs(LHS2,'ztrans(x(n),n,z)','ztrans(y(n),n,z)',...
        'x(0)','y(0)',X,Y,0,5);
% solve for X and Y
[X,Y] = solve(newLHS1,newLHS2,X,Y);
% invert z-transform and find x(n) and y(n)
x = iztrans(X,z,n)
y = iztrans(Y,z,n)
```

This script yields

```
x =
2*6^n-2
y =
2*6^n+3
```

Problems

Solve the following difference equations using z-transforms, where $n \geq 0$. Check your answer using MATLAB.

1. $y_{n+1} - y_n = n^2, \quad y_0 = 1.$

2. $y_{n+2} - 2y_{n+1} + y_n = 0, \quad y_0 = y_1 = 1.$

3. $y_{n+2} - 2y_{n+1} + y_n = 1, \quad y_0 = y_1 = 0.$

4. $y_{n+1} + 3y_n = n, \quad y_0 = 0.$

5. $y_{n+1} - 5y_n = \cos(n\pi), \quad y_0 = 0.$

6. $y_{n+2} - 4y_n = 1, \quad y_0 = 1, y_1 = 0.$

7. $y_{n+2} - \frac{1}{4}y_n = (\frac{1}{2})^n, y_0 = y_1 = 0.$

8. $y_{n+2} - 5y_{n+1} + 6y_n = 0, y_0 = y_1 = 1.$

9. $y_{n+2} - 3y_{n+1} + 2y_n = 1, \quad y_0 = y_1 = 0.$

10. $y_{n+2} - 2y_{n+1} + y_n = 2, \quad y_0 = 0, y_1 = 2.$

11. $x_{n+1} = 3x_n - 4y_n,\ y_{n+1} = 2x_n - 3y_n, \quad x_0 = 3,\ y_0 = 2.$

12. $x_{n+1} = 2x_n - 10y_n,\ y_{n+1} = -x_n - y_n, \quad x_0 = 3,\ y_0 = -2.$

13. $x_{n+1} = x_n - 2y_n,\ y_{n+1} = -6y_n, \quad x_0 = -1,\ y_0 = -7.$

14. $x_{n+1} = 4x_n - 5y_n,\ y_{n+1} = x_n - 2y_n, \quad x_0 = 6,\ y_0 = 2.$

3.5 STABILITY OF DISCRETE-TIME SYSTEMS

When we discussed the solution of ordinary differential equations by Laplace transforms, we introduced the concept of transfer function and impulse response. In the case of discrete-time systems, similar considerations come into play.

Consider the recursive system

$$y_n = a_1 y_{n-1} H_{n-1} + a_2 y_{n-2} H_{n-2} + x_n, \quad n \geq 0, \tag{3.5.1}$$

where H_{n-k} is the unit step function. It equals 0 for $n < k$ and 1 for $n \geq k$. Equation 3.5.1 is called a *recursive system* because future values of the sequence depend upon all of the previous values. At present, a_1 and a_2 are free parameters that we shall vary.

Using Equation 3.2.7,

$$z^2 Y(z) - a_1 z Y(z) - a_2 Y(z) = z^2 X(z), \tag{3.5.2}$$

or

$$G(z) = \frac{Y(z)}{X(z)} = \frac{z^2}{z^2 - a_1 z - a_2}. \tag{3.5.3}$$

As in the case of Laplace transforms, the ratio $Y(z)/X(z)$ is the transfer function. The inverse of the transfer function gives the impulse response for our discrete-time system. This particular transfer function has two poles, namely

$$z_{1,2} = \frac{a_1}{2} \pm \sqrt{\frac{a_1^2}{4} + a_2}. \tag{3.5.4}$$

At this point, we consider three cases.

Case 1: $a_1^2/4 + a_2 < 0$. In this case z_1 and z_2 are complex conjugates. Let us write them as $z_{1,2} = re^{\pm i\omega_0 T}$. Then

$$G(z) = \frac{z^2}{(z - re^{i\omega_0 T})(z - re^{-i\omega_0 T})} = \frac{z^2}{z^2 - 2r\cos(\omega_0 T)z + r^2}, \tag{3.5.5}$$

where $r^2 = -a_2$, and $\omega_0 T = \cos^{-1}(a_1/2r)$. From the inversion integral,

$$g_n = \operatorname{Res}\left[\frac{z^{n+1}}{z^2 - 2r\cos(\omega_0 T)z + r^2}; z_1\right] + \operatorname{Res}\left[\frac{z^{n+1}}{z^2 - 2r\cos(\omega_0 T)z + r^2}; z_2\right], \tag{3.5.6}$$

where g_n denotes the impulse response. Now

$$\operatorname{Res}\left[\frac{z^{n+1}}{z^2 - 2r\cos(\omega_0 T)z + r^2}; z_1\right] = \lim_{z \to z_1} \frac{(z - z_1)z^{n+1}}{(z - z_1)(z - z_2)} \tag{3.5.7}$$

$$= r^n \frac{\exp[i(n+1)\omega_0 T]}{e^{i\omega_0 T} - e^{-i\omega_0 T}} = \frac{r^n \exp[i(n+1)\omega_0 T]}{2i\sin(\omega_0 T)}. \tag{3.5.8}$$

Similarly,

$$\operatorname{Res}\left[\frac{z^{n+1}}{z^2 - 2r\cos(\omega_0 T)z + r^2}; z_2\right] = -\frac{r^n \exp[-i(n+1)\omega_0 T]}{2i\sin(\omega_0 T)}, \tag{3.5.9}$$

and

$$g_n = \frac{r^n \sin[(n+1)\omega_0 T]}{\sin(\omega_0 T)}. \tag{3.5.10}$$

A graph of $\sin[(n+1)\omega_0 T]/\sin(\omega_0 T)$ with respect to n gives a sinusoidal envelope. More importantly, if $|r| < 1$ these oscillations vanish as $n \to \infty$ and the system is stable. On the other hand, if $|r| > 1$ the oscillations grow without bound as $n \to \infty$ and the system is unstable.

Recall that $|r| > 1$ corresponds to poles that lie outside the unit circle while $|r| < 1$ is exactly the opposite. Our example suggests that for discrete-time systems to be stable, all of the poles of the transfer function must lie within the unit circle while an unstable system has at least one pole that lies outside of this circle.

Case 2: $a_1^2/4 + a_2 > 0$. This case leads to two real roots, z_1 and z_2. From the inversion integral, the sum of the residues gives the impulse response

$$g_n = \frac{z_1^{n+1} - z_2^{n+1}}{z_1 - z_2}. \tag{3.5.11}$$

Once again, if the poles lie within the unit circle, $|z_1| < 1$ and $|z_2| < 1$, the system is stable.

Case 3: $a_1^2/4 + a_2 = 0$. This case yields $z_1 = z_2$,

$$G(z) = \frac{z^2}{(z - a_1/2)^2} \qquad \text{and} \qquad g_n = \frac{1}{2\pi i}\oint_C \frac{z^{n+1}}{(z-a_1/2)^2}\,dz = \left(\frac{a_1}{2}\right)^n (n+1). \tag{3.5.12}$$

This system is obviously stable if $|a_1/2| < 1$ and the pole of the transfer function lies within the unit circle.

In summary, finding the transfer function of a discrete-time system is important in determining its stability. Because the location of the poles of $G(z)$ determines the response of the system, a stable system has all of its poles within the unit circle. Conversely, if any of the poles of $G(z)$ lie outside of the unit circle, the system is unstable. Finally, if $\lim_{n\to\infty} g_n = c$, the system is marginally stable. For example, if $G(z)$ has simple poles, some of the poles must lie *on* the unit circle.

• Example 3.5.1

Numerical methods of integration provide some of the simplest, yet most important, difference equations in the literature. In this example,[5] we show how z-transforms can be used to highlight the strengths and weaknesses of such schemes.

Consider the trapezoidal integration rule in numerical analysis. The integral y_n is updated by adding the latest trapezoidal approximation of the continuous curve. Thus, the integral is computed by

$$y_n = \tfrac{1}{2}T(x_n + x_{n-1}H_{n-1}) + y_{n-1}H_{n-1}, \tag{3.5.13}$$

where T is the interval between evaluations of the integrand.

---

[5] See Salzer, J. M., 1954: Frequency analysis of digital computers operating in real time. *Proc. IRE*, **42**, 457–466.

We first determine the stability of this rule because it is of little value if it is not stable. Using Equation 3.2.7, the transfer function is

$$G(z) = \frac{Y(z)}{X(z)} = \frac{T}{2}\left(\frac{z+1}{z-1}\right). \tag{3.5.14}$$

To find the impulse response, we use the inversion integral and find that

$$g_n = \frac{T}{4\pi i}\oint_C z^{n-1}\,\frac{z+1}{z-1}\,dz. \tag{3.5.15}$$

At this point, we must consider two cases: $n = 0$ and $n > 0$. For $n = 0$,

$$g_0 = \frac{T}{2}\operatorname{Res}\left[\frac{z+1}{z(z-1)};0\right] + \frac{T}{2}\operatorname{Res}\left[\frac{z+1}{z(z-1)};1\right] = \frac{T}{2}. \tag{3.5.16}$$

For $n > 0$,

$$g_0 = \frac{T}{2}\operatorname{Res}\left[\frac{z^{n-1}(z+1)}{z-1};1\right] = T. \tag{3.5.17}$$

Therefore, the impulse response for this numerical scheme is $g_0 = \frac{T}{2}$ and $g_n = T$ for $n > 0$. Note that this is a marginally stable system (the solution neither grows nor decays with n) because the pole associated with the transfer function lies *on* the unit circle.

Having discovered that the system is not unstable, let us continue and explore some of its properties. Recall now that $z = e^{sT} = e^{i\omega T}$ if $s = i\omega$. Then the transfer function becomes

$$G(\omega) = \frac{T}{2}\frac{1+e^{-i\omega T}}{1-e^{-i\omega T}} = -\frac{iT}{2}\cot\left(\frac{\omega T}{2}\right). \tag{3.5.18}$$

On the other hand, the transfer function of an ideal integrator is $1/s$ or $-i/\omega$. Thus, the trapezoidal rule has ideal phase but its shortcoming lies in its amplitude characteristic; it lies below the ideal integrator for $0 < \omega T < \pi$. We show this behavior, along with that for Simpson's one-third rule and Simpson's three-eighths rule, in Figure 3.5.1.

Figure 3.5.1 confirms the superiority of Simpson's one-third rule over his three-eighths rule. The figure also shows that certain schemes are better at suppressing noise at higher frequencies, an effect not generally emphasized in numerical calculus but often important in system design. For example, the trapezoidal rule is inferior to all others at low frequencies but only to Simpson's one-third rule at higher frequencies. Furthermore, the trapezoidal rule might actually be preferred, not only because of its simplicity but also because it attenuates at higher frequencies, thereby counteracting the effect of noise. □

• **Example 3.5.2**

Given the transfer function

$$G(z) = \frac{z^2}{(z-1)(z-1/2)}, \tag{3.5.19}$$

is this discrete-time system stable or marginally stable?

This transfer function has two simple poles. The pole at $z = 1/2$ gives rise to a term that varies as $(\frac{1}{2})^n$ in the impulse response, while the $z = 1$ pole gives a constant. Because this constant neither grows nor decays with n, the system is marginally stable. □

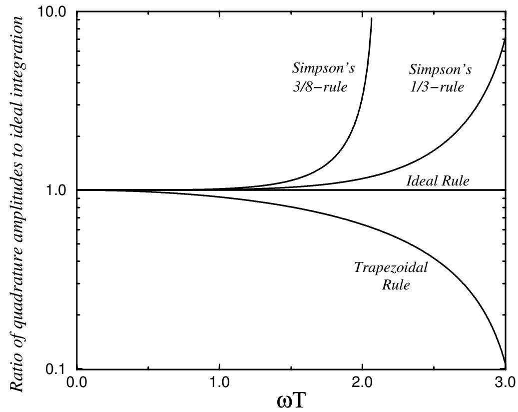


Figure 3.5.1: Comparison of various quadrature formulas by ratios of their amplitudes to that of an ideal integrator. (From Salzer, J. M., 1954: Frequency analysis of digital computers operating in real time. *Proc. IRE*, **42**, p. 463.)

• Example 3.5.3

In most cases the transfer function consists of a ratio of two polynomials. In this case we can use the MATLAB function `filter` to compute the impulse response as follows: Consider the Kronecker delta sequence, $x_0 = 1$, and $x_n = 0$ for $n > 0$. From the definition of the z-transform, $X(z) = 1$. Therefore, if our input into `filter` is the Kronecker delta sequence, the output y_n will be the impulse response since $Y(z) = G(z)$. If the impulse response grows without bound as n increases, the system is unstable. If it goes to zero as n increases, the system is stable. If it remains constant, it is marginally stable.

To illustrate this concept, the following MATLAB script finds the impulse response corresponding to the transfer function, Equation 3.5.19:

```
% enter the coefficients of the numerator
%     of the transfer function, Equation 3.5.19
num = [1 0 0];
% enter the coefficients of the denominator
%     of the transfer function, Equation 3.5.19
den = [1 -1.5 0.5];
% create the Kronecker delta sequence
x = [1 zeros(1,20)];
% find the impulse response
y = filter(num,den,x);
% plot impulse response
plot(y,'o'), axis([0 20 0.5 2.5])
xlabel('n+1','Fontsize',20)
ylabel('impulse response','Fontsize',20)
```

Figure 3.5.2 shows the computed impulse response. The asymptotic limit is two, so the system is marginally stable, as we found before.

We note in closing that the same procedure can be used to find the inverse of *any* z-transform that consists of a ratio of two polynomials. Here we simply set $G(z)$ equal to the given z-transform and perform the same analysis.

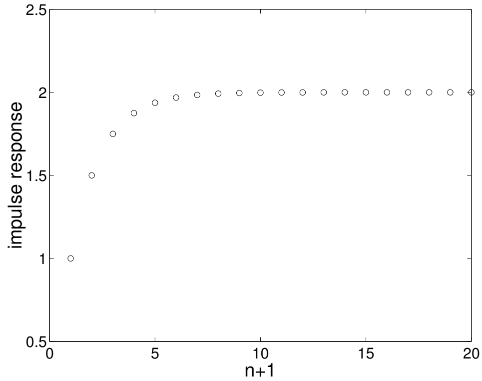


Figure 3.5.2: The impulse response for a discrete system with a transform function given by Equation 3.5.19.

Problems

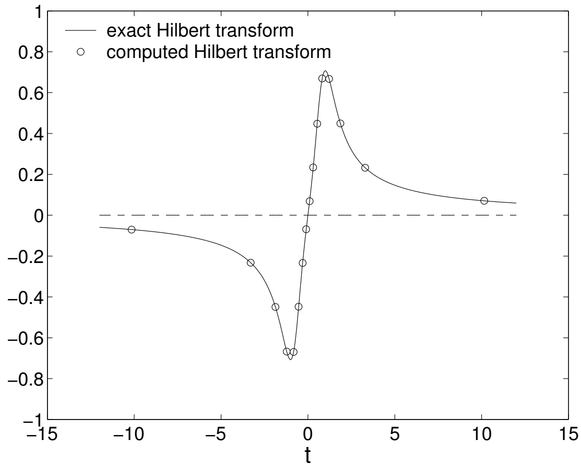
For the following time-discrete systems, find the transfer function and determine whether the systems are unstable, marginally stable, or stable. Check your answer by graphing the impulse response using MATLAB.

1. $y_n = y_{n-1}H_{n-1} + x_n$
2. $y_n = 2y_{n-1}H_{n-1} - y_{n-2}H_{n-2} + x_n$
3. $y_n = 3y_{n-1}H_{n-1} + x_n$
4. $y_n = \frac{1}{4}y_{n-2}H_{n-2} + x_n$

Further Readings

Jury, E. I., 1964: *Theory and Application of the z-Transform Method*. John Wiley & Sons, 330 pp. The classic text on z-transforms.

LePage, W. R., 1980: *Complex Variables and the Laplace Transform for Engineers.* Dover, 483 pp. Chapter 16 is on z-transforms.



Chapter 4
The Hilbert Transform

In addition to the Fourier, Laplace, and z-transforms, there are many other linear transforms that have their own special niche in engineering. Examples include Hankel, Walsh, Radon, and Hartley transforms. In this chapter we consider the *Hilbert transform*, which is a commonly used technique for relating the real and imaginary parts of a spectral response, particularly in communication theory.

We begin our study of Hilbert transforms by first defining them and then exploring their properties. Next, we develop the concept of the analytic signal. Finally, we explore a property of Hilbert transforms that is frequently applied to data analysis: the Kramers-Kronig relationship.

4.1 DEFINITION

In Chapter 3 we motivated the development of z-transforms by exploring the concept of the ideal sampler. In the case of Hilbert transforms, we introduce another fundamental operation, namely *quadrature phase shifting* or the *ideal Hilbert transformer*. This procedure does nothing more than shift the phase of all input frequency components by $-\pi/2$. Hilbert transformers are frequently used in communication systems and signal processing; examples include the generation of single-sideband modulated signals and radar and speech signal processing.

Because a $-\pi/2$ phase shift is equivalent to multiplying the Fourier transform of a signal by $e^{-i\pi/2} = -i$, and because phase shifting must be an odd function of frequency,[1]

[1] For a real function the phase of its Fourier transform must be an odd function of ω.

the transfer function of the phase shifter is $G(\omega) = -i\ \mathrm{sgn}(\omega)$, where $\mathrm{sgn}(\cdot)$ is defined by

$$\mathrm{sgn}(t) = \begin{cases} 1, & t > 0, \\ 0, & t = 0, \\ -1, & t < 0. \end{cases}$$

In other words, if $X(\omega)$ denotes the input spectrum to the phase shifter, the output spectrum must be $-i\ \mathrm{sgn}(\omega)X(\omega)$. If the process is repeated, the total phase shift is $-\pi$, a complete phase reversal of all frequency components. The output spectrum then equals $[-i\ \mathrm{sgn}(\omega)]^2X(\omega) = -X(\omega)$. This agrees with the notion of phase reversal because the output function is $-x(t)$.

Consider now the impulse response of the quadrature phase shifter, $g(t) = \mathcal{F}^{-1}[G(\omega)]$. From the definition of Fourier transforms,

$$\frac{dG}{d\omega} = -i\int_{-\infty}^{\infty} t\, g(t)e^{-i\omega t}\, dt, \tag{4.1.1}$$

and

$$g(t) = \frac{i}{t}\mathcal{F}^{-1}\left(\frac{dG}{d\omega}\right). \tag{4.1.2}$$

Since $G'(\omega) = -2i\delta(\omega)$, the corresponding impulse response is

$$g(t) = \frac{i}{t}\mathcal{F}^{-1}[-2i\delta(\omega)] = \frac{1}{\pi t}. \tag{4.1.3}$$

Consequently, if $x(t)$ is the input to a quadrature phase shifter, the superposition integral gives the output time function as

$$\boxed{\widehat{x}(t) = x(t) * \frac{1}{\pi t} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau}\, d\tau.} \tag{4.1.4}$$

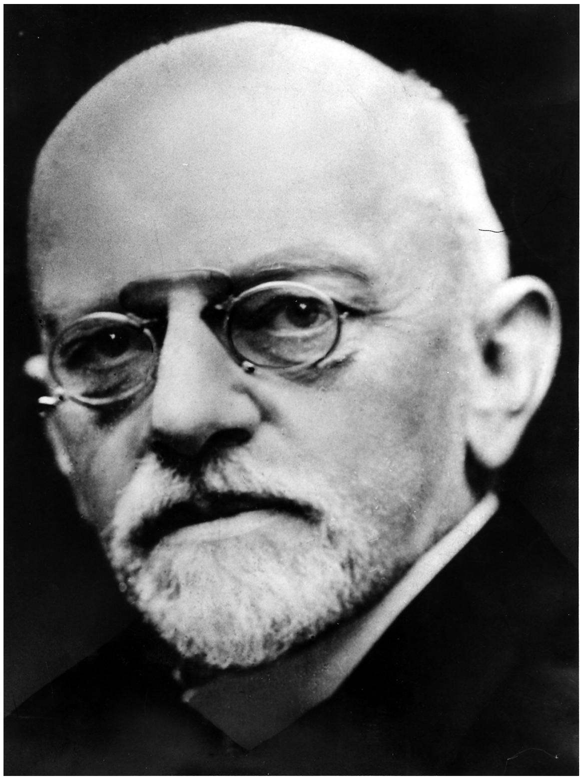
We shall define $\widehat{x}(t)$ as the *Hilbert transform* of $x(t)$, although some authors use the negative of Equation 4.1.4 corresponding to a $+\pi/2$ phase shift. The transform $\widehat{x}(t)$ is also called the *harmonic conjugate* of $x(t)$.

In similar fashion, $\widehat{\widehat{x}}(t)$ is the Hilbert transform of the Hilbert transform of $x(t)$ and corresponds to the output of two cascaded phase shifters. However, this output is known to be $-x(t)$, so $\widehat{\widehat{x}}(t) = -x(t)$, and we arrive at the *inverse Hilbert transform* relationship that

$$\boxed{x(t) = -\widehat{x}(t) * \frac{1}{\pi t} = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\widehat{x}(\tau)}{t-\tau}\, d\tau.} \tag{4.1.5}$$

Taken together, $x(t)$ and $\widehat{x}(t)$ are called a *Hilbert pair*. Hilbert pairs enjoy the unique property that $x(t) + i\widehat{x}(t)$ is an *analytic function*.[2]

[2] For the proof, see Titchmarsh, E. C., 1948: *Introduction to the Theory of Fourier Integrals*. Oxford University Press, p. 125.



Descended from a Prussian middle-class family, David Hilbert (1862–1943) would make significant contributions in the fields of algebraic form, algebraic number theory, foundations of geometry, analysis, mathematical physics, and the foundations of mathematics. Hilbert transforms arose during his study of integral equations (Hilbert, D., 1912: *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen.* Teubner, p. 75). (Portrait courtesy of Photo AKG, London, with permission.)

Because of the singularity at $\tau = t$, the integrals in Equation 4.1.4 and Equation 4.1.5 must be taken in the *Cauchy principal value* sense by approaching the singularity point from both sides, namely

$$\int_{-\infty}^{\infty} f(\tau)\,d\tau = \lim_{\epsilon\to 0}\left[\int_{-\infty}^{t-\epsilon} f(\tau)\,d\tau + \int_{t+\epsilon}^{\infty} f(\tau)\,d\tau\right], \qquad (\mathbf{4.1.6})$$

so that the infinities to the right and left of $\tau = t$ cancel each other. See Section 1.10. We also note that the Hilbert transform is basically a convolution and does not produce a change of domain; if x is a function of time, then $\widehat{x}$ is also a function of time. This is quite different from what we encountered with Laplace or Fourier transforms.

From its origin in phase shifting, Hilbert transforms of sinusoidal functions are trivial. Some examples are

$$\cos(\widehat{\omega t + \varphi}) = \cos\left(\omega t + \varphi - \tfrac{\pi}{2}\right) = \operatorname{sgn}(\omega)\sin(\omega t + \varphi). \qquad (\mathbf{4.1.7})$$

Similarly,

$$\sin(\widehat{\omega t + \varphi}) = -\operatorname{sgn}(\omega)\cos(\omega t + \varphi), \qquad (\mathbf{4.1.8})$$

and

$$\widehat{e^{i\omega t+i\varphi}} = -i\ \text{sgn}(\omega)e^{i\omega t+i\varphi}. \qquad (\mathbf{4.1.9})$$

Thus, Hilbert transformation does not change the amplitude of sine or cosine but does change their phase by $\pm\pi/2$.

• Example 4.1.1

Let us apply the integral definition of the Hilbert transform, Equation 4.1.4, to find the Hilbert transform of $\sin(\omega t)$, $\omega \neq 0$.

From the definition,

$$\mathcal{H}[\sin(\omega t)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin(\omega\tau)}{t-\tau}\,d\tau. \qquad (\mathbf{4.1.10})$$

If $x = t-\tau$, then

$$\mathcal{H}[\sin(\omega t)] = -\frac{\cos(\omega t)}{\pi}\int_{-\infty}^{\infty} \frac{\sin(\omega x)}{x}\,dx = -\cos(\omega t)\ \text{sgn}(\omega). \qquad (\mathbf{4.1.11})$$

□

• Example 4.1.2

Let us compute the Hilbert transform of $x(t) = \sin(t)/(t^2+1)$ from the definition of the Hilbert transform, Equation 4.1.4.

From the definition,

$$\widehat{x}(t) = \frac{1}{\pi}PV\int_{-\infty}^{\infty} \frac{\sin(\tau)}{(t-\tau)(\tau^2+1)}\,d\tau = \frac{1}{\pi}\Im\left[PV\int_{-\infty}^{\infty} \frac{e^{i\tau}}{(t-\tau)(\tau^2+1)}\,d\tau\right]. \qquad (\mathbf{4.1.12})$$

Because of the singularity on the real axis at $\tau = t$, we treat the integrals in Equation 4.1.12 in the sense of Cauchy principal value.

To evaluate Equation 4.1.12, we convert it into a closed contour integration by introducing a semicircle C_R of infinite radius in the upper half-plane. This yields a closed contour C, which consists of the real line plus this semicircle. Therefore, Equation 4.1.12 can be rewritten

$$PV\int_{-\infty}^{\infty} \frac{e^{i\tau}}{(t-\tau)(\tau^2+1)}\,d\tau = PV\oint_C \frac{e^{iz}}{(t-z)(z^2+1)}\,dz - \int_{C_R} \frac{e^{iz}}{(t-z)(z^2+1)}\,dz. \qquad (\mathbf{4.1.13})$$

The second integral on the right side of Equation 4.1.13 vanishes by Equation 1.9.7.

The evaluation of the closed integral in Equation 4.1.13 follows from the residue theorem. We have that

$$\text{Res}\left[\frac{e^{iz}}{(t-z)(z^2+1)};t\right] = \lim_{z\to t}\frac{(z-t)\,e^{iz}}{(t-z)(z^2+1)} = -\frac{e^{it}}{t^2+1}, \qquad (\mathbf{4.1.14})$$

and

$$\text{Res}\left[\frac{e^{iz}}{(t-z)(z^2+1)};i\right] = \lim_{z\to i}\frac{(z-i)\,e^{iz}}{(t-z)(z^2+1)} = \frac{e^{-1}}{2i(t-i)}. \qquad (\mathbf{4.1.15})$$

We do not have a contribution from $z=-i$ because it lies *outside* of the closed contour.

The Hilbert Transform of Some Common Functions

	function, $x(t)$	Hilbert transform, $\widehat{x}(t)$
1.	$\begin{cases} 1, & a < t < b \\ 0, & \text{otherwise} \end{cases}$	$\dfrac{1}{\pi} \ln\left\lvert \dfrac{t-a}{t-b} \right\rvert$
2.	$\sin(\omega t + \varphi)$	$-\operatorname{sgn}(\omega)\cos(\omega t + \varphi)$
3.	$\cos(\omega t + \varphi)$	$\operatorname{sgn}(\omega)\sin(\omega t + \varphi)$
4.	$e^{i\omega t + \varphi i}$	$-i \operatorname{sgn}(\omega) e^{i\omega t + \varphi i}$
5.	$\dfrac{1}{t}$	$-\pi\delta(t)$
6.	$\dfrac{1}{t^2+a^2}, \quad 0 < \Re(a)$	$\dfrac{t}{a(t^2+a^2)}$
7.	$\dfrac{\lambda t + \mu a}{t^2+a^2}, \quad 0 < \Re(a)$	$\dfrac{\mu t - \lambda a}{t^2+a^2}$
8.	$\dfrac{1}{1+t^4}$	$\dfrac{t(1+t^2)}{\sqrt{2}\,(1+t^4)}$
9.	$\dfrac{\sin(at)}{t}, \quad 0 < a$	$\dfrac{1-\cos(at)}{t}$
10.	$\dfrac{\sin(t)}{1+t^2}$	$\dfrac{e^{-1} - \cos(t)}{1+t^2}$
11.	$\sin(at)J_1(at), \quad 0 < a$	$-\cos(at)J_1(at)$
12.	$\sin(at)J_n(bt), \quad 0 < b < a$	$-\cos(at)J_n(bt)$
13.	$\cos(at)J_1(at), \quad 0 < a$	$\sin(at)J_1(at)$
14.	$\cos(at)J_n(bt), \quad 0 < b < a$	$\sin(at)J_n(at)$
15.	$\begin{cases} \sqrt{a^2-t^2}, & -a < t < a \\ 0, & \text{otherwise} \end{cases}$	$\begin{cases} t + \sqrt{t^2-a^2}, & -\infty < t < -a \\ t, & -a < t < a \\ t - \sqrt{t^2-a^2}, & a < t < \infty \end{cases}$
16.	$\sin\left(a\sqrt{t}\right) H(t), \quad 0 < a$	$\begin{cases} -e^{-a\sqrt{\lvert t\rvert}}, & -\infty < t < 0 \\ -\cos\left(a\sqrt{t}\right), & 0 < t < \infty \end{cases}$

Therefore,

$$PV \int_{-\infty}^{\infty} \frac{e^{i\tau}}{(t-\tau)(\tau^2+1)}\, d\tau = -\frac{\pi i\, e^{it}}{t^2+1} + \frac{\pi\, e^{-1}(t+i)}{t^2+1}. \tag{4.1.16}$$

Only one half of the value of the residue at $z = t$ was included; this reflects the semicircular indentation around the singularity there. Substituting Equation 4.1.16 into Equation 4.1.12, we obtain the final result that

$$\mathcal{H}\left[\frac{\sin(t)}{t^2+1}\right] = \frac{e^{-1} - \cos(t)}{t^2+1}. \tag{4.1.17}$$

□

• Example 4.1.3

Let us employ the relationship that the Fourier transform of $\widehat{x}(t)$ equals $-i\,\mathrm{sgn}(\omega)$ times the Fourier transform of $x(t)$ to find the Hilbert transform of $x(t) = e^{-t^2}$.

Because $\mathcal{F}(e^{-t^2}) = \sqrt{\pi}e^{-\omega^2/4}$,

$$\widehat{X}(\omega) = -i\sqrt{\pi}\,\mathrm{sgn}(\omega)e^{-\omega^2/4}. \tag{4.1.18}$$

Therefore,

$$\begin{aligned}
\widehat{x}(t) &= \frac{i}{2\sqrt{\pi}} \int_{-\infty}^{0} e^{it\omega-\omega^2/4}\, d\omega - \frac{i}{2\sqrt{\pi}} \int_{0}^{\infty} e^{it\omega-\omega^2/4}\, d\omega & \text{(4.1.19)}\\
&= \frac{i}{\sqrt{\pi}} \int_{-\infty}^{0} e^{2it\eta-\eta^2}\, d\eta - \frac{i}{\sqrt{\pi}} \int_{0}^{\infty} e^{2it\eta-\eta^2}\, d\eta & \text{(4.1.20)}\\
&= \frac{e^{-t^2}}{\sqrt{\pi}} \int_{-i\infty}^{t} e^{-s^2}\, ds - \frac{e^{-t^2}}{\sqrt{\pi}} \int_{t}^{i\infty} e^{-s^2}\, ds = \frac{2e^{-t^2}}{\sqrt{\pi}} \int_{0}^{t} e^{-s^2}\, ds, & \text{(4.1.21)}
\end{aligned}$$

where $s = t + \eta i$. The integral in Equation 4.1.21 is the well-known *Dawson's integral.*[3] See Gautschi and Waldvogel[4] for an alternative derivation. □

• Example 4.1.4: Numerical computation of the Hilbert transform

Recently André Weideman[5] devised a particularly efficient method for *numerically* computing the Hilbert transform when $x(t)$ is known exactly for any real t and enjoys the property that

$$\int_{-\infty}^{\infty} |x(t)|^2\, dt < \infty. \tag{4.1.22}$$

---

[3] Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in Fortran: The Art of Scientific Computing.* Cambridge University Press, Section 6.10.

[4] Gautschi, W., and J. Waldvogel, 2000: Computing the Hilbert transform of the generalized Laguerre and Hermite weight functions. *BIT*, **41**, 490–503.

[5] Weideman, J. A. C., 1995: Computing the Hilbert transform on the real line. *Math. Comput.*, **64**, 745–762.

Given Equation 4.1.22, the function $x(t)$ can be represented by the rational expansion

$$x(t) = \sum_{n=-\infty}^{\infty} a_n \rho_n(t), \tag{4.1.23}$$

where $\rho_n(t)$ is the set of rational functions

$$\rho_n(t) = \frac{(1+it)^n}{(1-it)^{n+1}}, \qquad n = 0, \pm 1, \pm 2, \cdots, \tag{4.1.24}$$

and

$$a_n = \frac{1}{\pi} \int_{-\infty}^{\infty} x(t) \rho_n^*(t)\, dt \tag{4.1.25}$$

or

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[1 - i \tan\left(\tfrac{1}{2}\theta\right)\right] x\left[\tan\left(\tfrac{1}{2}\theta\right)\right] e^{-in\theta}\, d\theta, \tag{4.1.26}$$

if we introduce the substitution $t = \tan(\theta/2)$.

Why is Equation 4.1.23 useful? Taking the Hilbert transform of both sides of this equation,

$$\widehat{x}(t) = \sum_{n=-\infty}^{\infty} a_n \widehat{\rho}_n(t). \tag{4.1.27}$$

Using contour integration, we find that

$$\widehat{\rho}_n(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{(1+i\tau)^n}{(1-i\tau)^{n+1}(t-\tau)}\, d\tau = -i \operatorname{sgn}(n) \rho_n(t), \tag{4.1.28}$$

where $\operatorname{sgn}(t)$ is the signum function with $\operatorname{sgn}(0) = 1$. Therefore,

$$\widehat{x}(t) = -i \sum_{n=-\infty}^{\infty} \operatorname{sgn}(n)\, a_n\, \rho_n(t). \tag{4.1.29}$$

We must now approximate Equation 4.1.29 so that we can evaluate it numerically. We do this by introducing the following truncated version:

$$\widehat{x}_N(t) = -i \sum_{n=-N}^{N-1} \operatorname{sgn}(n)\, A_n\, \rho_n(t). \tag{4.1.30}$$

This particular truncation was chosen because $\rho_n(t)$ and $\rho_{-n-1}(t)$ are a conjugate pair. The coefficient a_n has become A_n, which equals

$$A_n = \frac{1}{N} \sum_{j=-N+1}^{N-1} \left[1 - i \tan\left(\tfrac{1}{2}\theta_j\right)\right] x\left[\tan\left(\tfrac{1}{2}\theta_j\right)\right] e^{-in\theta_j}, \tag{4.1.31}$$

where $\theta_j = \pi j/N$. The terms corresponding to $j = \pm N$ have been set to zero because it is assumed that $x(t)$ vanishes rapidly with $t \to \pm\infty$. Finally, we substitute θ for t and transform Equation 4.1.30 into

$$\widehat{x}_N(t_j) = -\frac{i}{1 - i\tan(\theta_j)} \sum_{n=-N}^{N-1} \operatorname{sgn}(n) A_n e^{in\theta_j}. \tag{4.1.32}$$

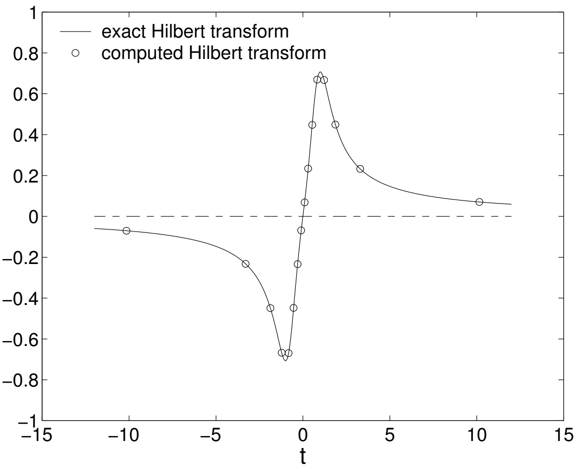


Figure 4.1.1: The Hilbert transform for $x(t) = 1/(1+t^4)$ computed from Weideman's algorithm.

The advantage of Equation 4.1.31 and Equation 4.1.32 is that they can be evaluated using fast Fourier transforms. For example, the following MATLAB script devised by Weideman illustrates his methods for $x(t) = 1/(1+t^4)$:

```
% initialize parameters used in computation
b = 1; N = 8; n = [-N:N-1]';
% set up collocation points and evaluate function there
t = b * tan(pi*(n+1/2)/(2*N)); F = 1./(1+t.^4);
% evaluate Equation 4.1.31
an = fftshift(fft(F.*(b-i*t)));
% compute Hilbert transform via Equation 4.1.32
hilbert = ifft(fftshift(i*(sign(n+1/2).*an)))./(b-i*t);
hilbert = -real(hilbert);
% find points at which we will compute exact answer
tt = [-12:0.02:12];
% compute exact answer
answer = tt.*(1+tt.^2)./(1+tt.^4)./sqrt(2);
fzero = zeros(size(tt));
% plot both computed Hilbert transform and exact answer
plot(tt,answer,'-',t,hilbert,'o',tt,fzero,'--')
xlabel('t','Fontsize',20)
legend('exact Hilbert transform','computed Hilbert transform')
legend boxoff
```

Figure 4.1.1 illustrates Weideman's algorithm for numerically computing the Hilbert transform of $1/(1+t^4)$.

There are two important points concerning Weideman's implementation of his algorithm. First, the collocation points originally given by $t_j = \tan[\pi j/(2N)]$, $j = -N, \ldots, N-1$ have changed to $t_j = \tan[(j+\frac{1}{2})\pi/(2N)]$, $j = -N, \ldots, N-1$. This change replaces the trapezoidal rule discretization for the Fourier coefficients with a midpoint rule. The advantages are twofold: First, it avoids the nuisance of dealing with a collocation point at infinity. Second, it actually yields more accurate results in many cases.

The discerning student will also notice that Weideman introduced a free parameter b, which we set to one. This rescaling parameter can have a major influence on the accuracy. The interested student is referred to the bottom of page 756 in Weideman's paper for further details. □

• Example 4.1.5: Discrete Hilbert transform

Quite often the function is given as discrete data points. How do we find the Hilbert transform in this case? We will now prove[6] that the equivalent *discrete* Hilbert transform is

$$\mathcal{H}(f_n) = \widehat{f}_k = \begin{cases} \dfrac{2}{\pi} \displaystyle\sum_{n \text{ odd}} \dfrac{f_n}{k-n}, & k \text{ even}, \\ \dfrac{2}{\pi} \displaystyle\sum_{n \text{ even}} \dfrac{f_n}{k-n}, & k \text{ odd}, \end{cases} \tag{4.1.33}$$

where f_n denotes a set of discrete data values that are sampled at $t = nT$ and both k and n run from $-\infty$ to ∞. The corresponding inverse is

$$f_n = \begin{cases} \dfrac{2}{\pi} \displaystyle\sum_{k \text{ odd}} \dfrac{\widehat{f}_k}{k-n}, & n \text{ even}, \\ \dfrac{2}{\pi} \displaystyle\sum_{k \text{ even}} \dfrac{\widehat{f}_k}{k-n}, & n \text{ odd}. \end{cases} \tag{4.1.34}$$

We begin our proof by inserting Equation 4.1.33 into Equation 4.1.34. For n even,

$$f_n = \frac{2}{\pi} \sum_{k \text{ odd}} \frac{1}{k-n} \left(\frac{2}{\pi} \sum_{p \text{ even}} \frac{f_p}{k-p} \right) = \frac{4}{\pi^2} \sum_{p \text{ even}} \sum_{k \text{ odd}} \frac{f_p}{(k-p)(k-n)} \tag{4.1.35}$$

$$= \frac{4}{\pi^2} \sum_{k \text{ odd}} \frac{f_n}{(k-n)^2} + \frac{4}{\pi^2} \sum_{p \text{ even}, p \neq n} \sum_{k \text{ odd}} (n-p) f_p \left\{ \frac{1}{k-n} - \frac{1}{k-p} \right\}. \tag{4.1.36}$$

The term within the curly brackets equals zero as k runs through all of its values. Therefore, Equation 4.1.36 reduces to

$$f_n = \frac{8}{\pi^2} f_n \left(1 + \frac{1}{3^2} + \frac{1}{5^2} + + \frac{1}{7^2} + \cdots \right). \tag{4.1.37}$$

However, the term in the brackets of Equation 4.1.37 equals $\pi^2/8$. Therefore, Equation 4.1.33 and Equation 4.1.34 is proved for n even. An identical proof follows for n odd.

A popular alternative[7] to Equation 4.1.33 involves the (fast) Fourier transform and the relationship that $\widehat{X}(\omega) = -i \operatorname{sgn}(\omega) X(\omega)$, where $X(\omega)$ and $\widehat{X}(\omega)$ denote the Fourier transform of $x(t)$ and $\widehat{x}(t)$, respectively. In this technique, a fast Fourier transform is taken of the data. This transformed dataset is then multiplied by $-i \operatorname{sgn}(\omega)$ and then back transformed to give the Hilbert transform.

---

[6] See Kak, S. C., 1970: The discrete Hilbert transform. *Proc. IEEE*, **58**, 585–586. For an alternative derivation, see Kress, R., and E. Martensen, 1970: Anwendung der Rechteckregel auf die reelle Hilbert-transformation mit unendlichem Intervall. *Z. Angew. Math. Mech.*, **50**, T61–T64.

[7] Čížek, V., 1970: Discrete Hilbert transform. *IEEE Trans. Audio Electroacoust.*, **AU-18**, 340–343.

Let $x(t)$ be a real, even function. Then $X(\omega)$, the Fourier transform of $x(t)$, is also an even function. Consequently,

$$\widehat{x}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{X}(\omega)e^{i\omega t}\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} -i\,\mathrm{sgn}(\omega)X(\omega)\left[\cos(\omega t) + i\sin(\omega t)\right]\,d\omega \qquad \mathbf{(4.1.38)}$$

$$= -\frac{i}{2\pi}\int_{-\infty}^{\infty} \mathrm{sgn}(\omega)X(\omega)\cos(\omega t)\,d\omega + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{sgn}(\omega)X(\omega)\sin(\omega t)\,d\omega \qquad \mathbf{(4.1.39)}$$

$$= \frac{1}{\pi}\int_0^{\infty} X(\omega)\sin(\omega t)\,d\omega. \qquad \mathbf{(4.1.40)}$$

Note that the Hilbert transform in this case is an odd function. Similarly, if $x(t)$ is a real, odd function,

$$\widehat{x}(t) = -\frac{i}{\pi}\int_0^{\infty} X(\omega)\cos(\omega t)\,d\omega, \qquad \mathbf{(4.1.41)}$$

and the Hilbert transform is an even function.

Problems

1. Show that the Hilbert transform of a constant function is zero.

2. Use Equation 4.1.4 to compute the Hilbert transform of $\cos(\omega t)$, $\omega \neq 0$.

3. Use Equation 4.1.4 to show that the Hilbert transform of the Dirac delta function $\delta(t)$ is $1/(\pi t)$.

4. Use Equation 4.1.4 to show that the Hilbert transform of $1/(t^2+1)$ is $t/(t^2+1)$.

5. The output $y(t)$ from an ideal lowpass filter can be expressed by the convolution integral

$$y(t) = x(t) * \frac{\sin(2\pi\omega t)}{\pi t},$$

where $x(t)$ is the input signal. Show that this expression can also be expressed in terms of Hilbert transforms as

$$y(t) = \mathcal{H}[x(t)\cos(2\pi\omega t)]\sin(2\pi\omega t) - \mathcal{H}[x(t)\sin(2\pi\omega t)]\cos(2\pi\omega t).$$

Following Example 4.1.3, find the Hilbert transforms of

6. $x(t) = \dfrac{1}{1+t^2}$

7. $x(t) = \begin{cases} 1, & -a < t < a \\ 0, & \text{otherwise} \end{cases}$

8. Using the commutative and associate properties of convolution, $f(t) * g(t) = g(t) * f(t)$ and $[f(t) * g(t)] * v(t) = f(t) * [g(t) * v(t)]$, respectively, and the definition of the Hilbert transform, Equation 4.1.4, show[8] that

$$\mathcal{H}[f(t) * g(t)] = \widehat{f}(t) * g(t) = f(t) * \widehat{g}(t).$$

---

[8] For an application, see Sakai, H., and G. A. Vanasse, 1966: Hilbert transform in Fourier spectroscopy. *J. Opt. Soc. Am.*, **56**, 131–132.

Using MATLAB, test Weideman's algorithm for the following cases. Why does the algorithm do well or not?

9. $\begin{cases} 1, & -1 < t < 1 \\ 0, & \text{otherwise} \end{cases}$ 10. $\sin(t)$ 11. $\dfrac{1}{t^2+1}$ 12. $\dfrac{\sin(t)}{1+t^4}$

For Problem 12, you will need

$$\mathcal{H}\left[\frac{\sin(t)}{t^4+1}\right] = \frac{e^{-1/\sqrt{2}}[\cos(1/\sqrt{2}\,) + \sin(1/\sqrt{2}\,)t^2] - \cos(t)}{t^4+1}.$$

4.2 SOME USEFUL PROPERTIES

In principle, we could construct any desired transform from the definition of the Hilbert transform. However, there are several general theorems that are much more effective in finding new transforms.

Linearity

From the definition of the Hilbert transform, it immediately follows that if $z(t) = c_1 x(t) + c_2 y(t)$, where c_1 and c_2 are arbitrary constants, then $\widehat{z}(t) = c_1\widehat{x}(t) + c_2\widehat{y}(t)$.

The energy in a signal and its Hilbert transform are the same.

Consider the energy spectral densities at input and output of a quadrature phase shifter. The output equals

$$|\widehat{X}(\omega)|^2 = \left|\mathcal{F}[\widehat{x}(t)]\right|^2 = |-i\ \mathrm{sgn}(\omega)|^2 |X(\omega)|^2 = |X(\omega)|^2. \tag{4.2.1}$$

Because the energy spectral density at input and output are the same, so are the total energies.

A signal and its Hilbert transform are orthogonal.

From Parseval's theorem,

$$\int_{-\infty}^{\infty} x(t)\widehat{x}(t)\,dt = \int_{-\infty}^{\infty} X(\omega)\widehat{X}^*(\omega)\,d\omega, \tag{4.2.2}$$

where $\widehat{X}(\omega) = \mathcal{F}[\widehat{x}(t)]$. Then,

$$\int_{-\infty}^{\infty} X(\omega)\widehat{X}^*(\omega)\,d\omega = \int_{-\infty}^{\infty} i\ \mathrm{sgn}(\omega)|X(\omega)|^2\,d\omega = 0, \tag{4.2.3}$$

because the integrand in the middle expression of Equation 4.2.3 is odd. Thus,

$$\int_{-\infty}^{\infty} x(t)\widehat{x}(t)\,dt = 0. \qquad \textbf{(4.2.4)}$$

The reason why a function and its Hilbert transform are orthogonal to each other follows from the fact that a Hilbert transformation of a function shifts the phase of each Fourier component of the function *forward* by $\pi/2$ for positive frequencies and *backward* for negative frequencies.

• Example 4.2.1

Let us verify the orthogonality condition for Hilbert transforms using $x(t) = 1/(1+t^2)$. Because $\widehat{x}(t) = t/(1+t^2)$,

$$\int_{-\infty}^{\infty} x(t)\widehat{x}(t)\,dt = \int_{-\infty}^{\infty} \frac{t}{(1+t^2)^2}\,dt = 0, \qquad \textbf{(4.2.5)}$$

since the integrand is an odd function. □

Shifting

Let us find the Hilbert transform of $x(t+a)$ if we know $\widehat{x}(t)$. From the definition of Hilbert transforms,

$$\mathcal{H}[x(t+a)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\eta+a)}{t-\eta}\,d\eta = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\tau)}{(t+a)-\tau}\,d\tau = \widehat{x}(t+a) \qquad \textbf{(4.2.6)}$$

or $\mathcal{H}[x(t+a)] = \widehat{x}(t+a)$.

Time scaling

Let $a > 0$. Then,

$$\mathcal{H}[x(at)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(a\eta)}{t-\eta}\,d\eta = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\tau)}{at-\tau}\,d\tau = \widehat{x}(at). \qquad \textbf{(4.2.7)}$$

On the other hand, if $a < 0$,

$$\mathcal{H}[x(at)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(a\eta)}{t-\eta}\,d\eta = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\tau)}{at-\tau}\,d\tau = -\widehat{x}(at). \qquad \textbf{(4.2.8)}$$

Thus, we have that $\mathcal{H}[x(at)] = \operatorname{sgn}(a)\,\widehat{x}(at)$.

Some General Properties of Hilbert Transforms

	function, x(t)	**Hilbert transform, $\widehat{\mathbf{x}}(\mathbf{t})$**
1.	$\widehat{x}(t)$	$-x(t)$
2.	$x(t) + y(t)$	$\widehat{x}(t) + \widehat{y}(t)$
3.	$x(t + a), \quad a$ real	$\widehat{x}(t + a)$
4.	$\dfrac{d^n x(t)}{dt^n}$	$\dfrac{d^n \widehat{x}(t)}{dt^n}$
5.	$x(at)$	$\operatorname{sgn}(a)\,\widehat{x}(at)$
6.	$tx(t)$	$t\widehat{x}(t) + \frac{1}{\pi}\int_{-\infty}^{\infty} x(\tau)\,d\tau$
7.	$(t + a)x(t)$	$(t + a)\widehat{x}(t) + \frac{1}{\pi}\int_{-\infty}^{\infty} x(\tau)\,d\tau$

Derivatives

Let us find the relationship between the nth derivative of $x(t)$ and its Hilbert transform. Using the derivative rule as it applies to Fourier transforms,

$$\mathcal{H}\left\{\mathcal{F}\left[\frac{d^n x}{dt^n}\right]\right\} = -i\operatorname{sgn}(\omega)(i\omega)^n X(\omega) = (i\omega)^n[-i\operatorname{sgn}(\omega)X(\omega)] = (i\omega)^n \widehat{X}(\omega) = \mathcal{F}\left[\frac{d^n \widehat{x}}{dt^n}\right]. \tag{4.2.9}$$

Taking the inverse Fourier transforms, we have that

$$\mathcal{H}\left(\frac{d^n x}{dt^n}\right) = \frac{d^n \widehat{x}}{dt^n}. \tag{4.2.10}$$

Convolution

Hilbert transforms enjoy a similar, but not identical, property with Fourier transforms with respect to convolution. If

$$w(t) = u(t) * v(t) = \int_{-\infty}^{\infty} u(\tau)v(t - \tau)\,d\tau = \int_{-\infty}^{\infty} u(t - \tau)v(\tau)\,d\tau, \tag{4.2.11}$$

then

$$\widehat{w}(t) = v(t) * \widehat{u}(t). \tag{4.2.12}$$

Proof: From the convolution theorem for Fourier transforms, $W(\omega) = V(\omega)U(\omega)$. Multiplying both sides of the equation by $-i\operatorname{sgn}(\omega)$,

$$\widehat{W}(\omega) = -i\operatorname{sgn}(\omega)W(\omega) = V(\omega)[-i\operatorname{sgn}(\omega)U(\omega)] = V(\omega)\widehat{U}(\omega). \tag{4.2.13}$$

Again, using the convolution theorem as it applies to Fourier transforms, we arrive at the final result. □

• Example 4.2.2

Given the functions $u(t) = \cos(t)$ and $v(t) = 1/(1+t^4)$, let us verify the convolution theorem as it applies to Hilbert transforms.

With $u(t) = \cos(t)$ and $v(t) = 1/(1+t^4)$,

$$w(t) = u(t) * v(t) = \int_{-\infty}^{\infty} \frac{\cos(t-x)}{1+x^4}\,dx \tag{4.2.14}$$

$$= \int_{-\infty}^{\infty} \frac{\cos(t)\cos(x)}{1+x^4}\,dx + \int_{-\infty}^{\infty} \frac{\sin(t)\sin(x)}{1+x^4}\,dx \tag{4.2.15}$$

$$= \frac{\pi}{\sqrt{2}} e^{-1/\sqrt{2}} \left[\cos\left(\frac{1}{\sqrt{2}}\right) + \sin\left(\frac{1}{\sqrt{2}}\right)\right] \cos(t) \tag{4.2.16}$$

so that

$$\widehat{w}(t) = \frac{\pi}{\sqrt{2}} e^{-1/\sqrt{2}} \left[\cos\left(\frac{1}{\sqrt{2}}\right) + \sin\left(\frac{1}{\sqrt{2}}\right)\right] \sin(t). \tag{4.2.17}$$

Because $\widehat{v}(t) = t(1+t^2)/[\sqrt{2}\,(1+t^4)]$,

$$u(t) * \widehat{v}(t) = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \cos(t-x) \frac{x(1+x^2)}{1+x^4}\,dx \tag{4.2.18}$$

$$= \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \frac{\cos(t)\cos(x)x(1+x^2)}{1+x^4}\,dx + \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \frac{\sin(t)\sin(x)x(1+x^2)}{1+x^4}\,dx \tag{4.2.19}$$

$$= \frac{1}{\sqrt{2}} \sin(t) \int_{-\infty}^{\infty} \frac{x(1+x^2)\sin(x)}{1+x^4}\,dx \tag{4.2.20}$$

$$= \frac{\pi}{\sqrt{2}} e^{-1/\sqrt{2}} \left[\cos\left(\frac{1}{\sqrt{2}}\right) + \sin\left(\frac{1}{\sqrt{2}}\right)\right] \sin(t), \tag{4.2.21}$$

and the convolution theorem for Hilbert transforms holds true in this case. □

Product theorem

Let $f(t)$ and $g(t)$ denote complex functions with Fourier transforms $F(\omega)$ and $G(\omega)$, respectively. If

1) $F(\omega)$ vanishes for $|\omega| > a$, and $G(\omega)$ vanishes for $|\omega| < a$, where $a > 0$,

or

2) $f(t)$ and $g(t)$ are analytic functions (their real and imaginary parts are Hilbert pairs),

then the Hilbert transform of the product of $f(t)$ and $g(t)$ is

$$\mathcal{H}[f(t)g(t)] = f(t)\widehat{g}(t). \tag{4.2.22}$$

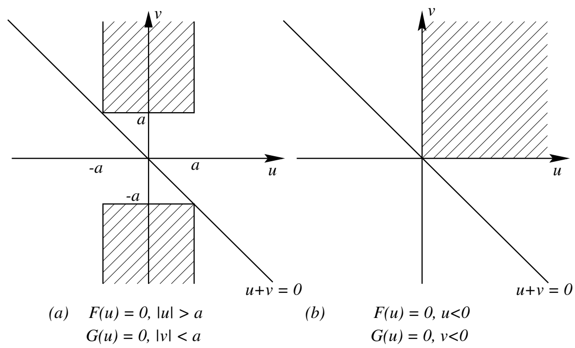


Figure 4.2.1: Region of integration in the proof of the product theorem.

Proof:[9] The product $f(t)g(t)$ can be expressed as

$$f(t)g(t) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u)G(v)e^{i(u+v)t}\,dv\,du. \tag{4.2.23}$$

Because $\mathcal{H}(e^{ibt}) = i\ \mathrm{sgn}(b)e^{ibt}$,

$$\mathcal{H}[f(t)g(t)] = \frac{i}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u)G(v)\ \mathrm{sgn}(u+v)e^{i(u+v)t}\,dv\,du. \tag{4.2.24}$$

The shaded regions of Figure 4.2.1 are those in which the product $F(u)G(v)$ is nonvanishing for the conditions of the theorem. In Figure 4.2.1(a) the nonoverlapping Fourier transforms yield two semi-infinite strips in which the product is nonvanishing. In Figure 4.2.1(b), for analytic functions, the Fourier transforms vanish for negative arguments[10] so that the product is nonvanishing only in the first quadrant. In both cases $\mathrm{sgn}(u+v) = \mathrm{sgn}(v)$ over the regions of integration in which the integrand is nonvanishing. Thus,

$$\mathcal{H}[f(t)g(t)] = \frac{i}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u)G(v)\ \mathrm{sgn}(v)e^{i(u+v)t}\,dv\,du \tag{4.2.25}$$

$$= f(t)\frac{i}{2\pi}\int_{-\infty}^{\infty} G(v)\ \mathrm{sgn}(v)e^{ivt}\,dv = f(t)\widehat{g}(t). \tag{4.2.26}$$

□

• Example 4.2.3: Hilbert transforms of band-pass functions

In communications, we have the double-sideband, amplitude-modulated signal given by $a(t)\cos(\omega t + \varphi)$, where φ is constant. From the product theorem, its Hilbert transform equals $a(t)\sin(\omega t + \varphi)$, $\omega > 0$, provided that the highest frequency component in $a(t)$ is less than ω. Paradoxically, the Hilbert transform of more general $a(t)\cos[\omega t + \varphi(t)]$, which equals $a(t)\sin[\omega t + \varphi(t)]$, has no such restriction.

---

[9] See Bedrosian, E., 1963: A product theorem for Hilbert transforms. *Proc. IEEE*, **51**, 868–869. This theorem has been extended to functions of n-dimensional real vectors by Stark, H., 1971: An extension of the Hilbert transform product theorem. *Proc. IEEE*, **59**, 1359–1360.

[10] Titchmarsh, E. C., 1948: *Introduction to the Theory of Fourier Integrals*. Oxford University Press, p. 128.

Problems

Verify the orthogonality property of Hilbert transforms using

1. $x(t) = 1/(1+t^4)$

2. $x(t) = \sin(t)/(1+t^2)$

3. $x(t) = \begin{cases} 1, & 0 < t < a \\ 0, & \text{otherwise} \end{cases}$

Verify the convolution theorem for Hilbert transforms using

4. $u(t) = \begin{cases} 1, & 0 < t < a, \\ 0, & \text{otherwise}, \end{cases} \quad v(t) = \sin(t)$

5. $u(t) = \cos(t), \quad v(t) = \dfrac{1}{1+t^2}$

6. Use the product theorem to show that

$$\mathcal{H}[\sin(at)J_n(bt)] = -\cos(at)J_n(bt), \qquad 0 < b < a,$$

if $n = 0, 1, 2, 3, \ldots$.
Hint:

$$\mathcal{F}[J_n(bt)] = \frac{2(-1)^m}{\sqrt{b^2-\omega^2}}\, T_n\!\left(\frac{|\omega|}{b}\right) H(b-|\omega|),$$

where $T_n(\cdot)$ is a Chebyshev polynomial of the first kind and $m = n/2$ or $(n-1)/2$, depending upon which definition gives an integer.

7. Given cosine and sine integrals:

$$Ci(x) = -\int_x^\infty \frac{\cos(t)}{t}\, dt, \qquad Si(x) = -\int_x^\infty \frac{\sin(t)}{t}\, dt,$$

and

$$\mathcal{H}[Ci(a|t|)] = -\text{sgn}(t) Si(a|t|), \quad 0 < a,$$

use the product rule to show that

$$\mathcal{H}[\sin(bt)Ci(a|t|)] = -\text{sgn}(t)\sin(bt)Si(a|t|), \qquad 0 < b < a.$$

Hint:

$$\mathcal{F}[Ci(a|t|)] = \begin{cases} 0, & 0 < |\omega| < a, \\ -\pi/|\omega|, & a < |\omega| < \infty, \end{cases} \qquad 0 < a.$$

8. Prove that

$$\mathcal{H}[tx(t)] = t\widehat{x}(t) - \frac{1}{\pi}\int_{-\infty}^\infty x(\tau)\, d\tau.$$

Hint:

$$\frac{\tau x(\tau)}{t-\tau} = \frac{t x(\tau)}{t-\tau} - x(\tau).$$

4.3 ANALYTIC SIGNALS

The monochromatic signal $A\cos(\omega_0 t + \varphi)$ appears in many physical and engineering applications. It is common to represent this signal by the complex representation $Ae^{i(\omega_0 t+\varphi)}$. These two representations are related to each other by

$$A\cos(\omega_0 t + \varphi) = \Re\left[Ae^{i(\omega_0 t+\varphi)}\right] = \tfrac{1}{2}\left[Ae^{i(\omega_0 t+\varphi)} + Ae^{-i(\omega_0 t+\varphi)}\right]. \tag{4.3.1}$$

Furthermore, the Fourier transform of $A\cos(\omega_0 t + \varphi)$ is

$$\mathcal{F}[A\cos(\omega_0 t + \varphi)] = \tfrac{1}{2}\left[Ae^{i\varphi}\delta(\omega-\omega_0) + Ae^{-i\varphi}\delta(\omega+\omega_0)\right], \tag{4.3.2}$$

while the Fourier transform of $Ae^{i(\omega_0 t+\varphi)}$ is

$$\mathcal{F}\left[Ae^{i(\omega_0 t+\varphi)}\right] = Ae^{i\varphi}\delta(\omega-\omega_0). \tag{4.3.3}$$

As Equation 4.3.2 and Equation 4.3.3 clearly show, in passing from the real signal to its complex representation, we double the strength of the positive frequencies and remove entirely the negative frequencies.

Let us generalize these concepts to nonmonochromatic signals. For the real signal $x(t)$ with Fourier transform $X(\omega)$ and the complex signal $z(t)$ with Fourier transform $Z(\omega)$, the previous paragraph shows that our generalization must have the property:

$$Z(\omega) = X(\omega) + \operatorname{sgn}(\omega)X(\omega) \tag{4.3.4}$$

or

$$Z(\omega) = \begin{cases} 2X(\omega), & \omega > 0, \\ X(\omega), & \omega = 0, \\ 0, & \omega < 0. \end{cases} \tag{4.3.5}$$

Taking the inverse of Equation 4.3.4, we have the definition of an *analytic signal* as

$$z(t) = x(t) + i\widehat{x}(t), \tag{4.3.6}$$

where $x(t)$ is a real signal and $\widehat{x}(t)$ is its Hilbert transform.

• Example 4.3.1

In Figure 4.3.1 the amplitude spectrum of the analytic signal is graphed when $x(t)$ is the rectangular pulse,

$$x(t) = \begin{cases} 1, & |t| < a, \\ 0, & |t| > a. \end{cases}$$

Note that the amplitude spectrum equals zero for $\omega < 0$ and twice the amplitude spectrum for $\omega > 0$. □

• Example 4.3.2

Let us find the energy of an analytic signal.

The energy of an analytic signal is

$$\int_{-\infty}^{\infty} |z(t)|^2\,dt = \int_{-\infty}^{\infty} x^2(t)\,dt + \int_{-\infty}^{\infty} \widehat{x}^2(t)\,dt = 2\int_{-\infty}^{\infty} x^2(t)\,dt = 2\int_{-\infty}^{\infty} |X(\omega)|^2\,d\omega \tag{4.3.7}$$

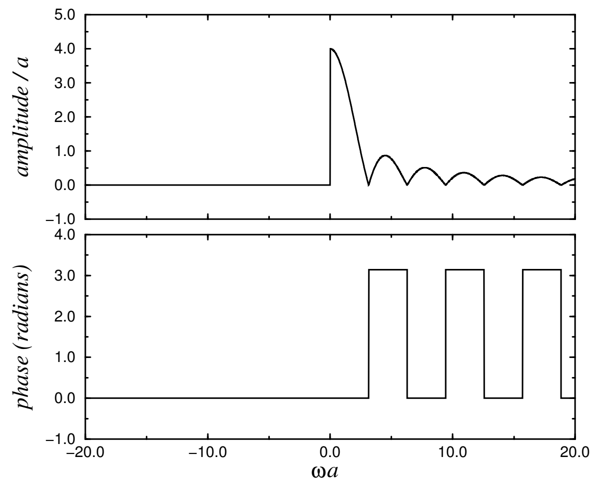


Figure 4.3.1: The spectrum of the analytic signal when $x(t)$ is the rectangular pulse given in Example 4.3.1.

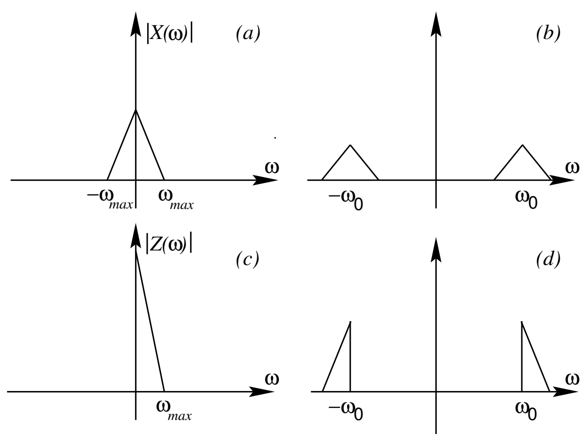


Figure 4.3.2: Given a function $x(t)$ with an amplitude spectrum shown in (a), frame (b) shows the amplitude spectrum of the amplitude-modulated signal $x(t)\cos(\omega_0 t)$ while frames (c) and (d) give the amplitude spectrum of the analytic signal $z(t)$ and $x(t)\cos(\omega_0 t) - \widehat{x}(t)\sin(\omega_0 t)$, respectively.

by Parseval's theorem. Thus, the analytic signal has twice the energy of the corresponding real signal. □

Consider the function $x(t)$ whose amplitude spectrum is shown in Figure 4.3.2(a). If we were to amplitude modulate $x(t)$ with $\cos(\omega_0 t)$, then the amplitude spectrum of this modulated signal would appear as pictured in Figure 4.3.2(b).

Consider now the signal

$$y(t) = x(t)\cos(\omega_0 t) - \widehat{x}(t)\sin(\omega_0 t) = \Re\big\{[x(t) + i\widehat{x}(t)]e^{i\omega_0 t}\big\} \tag{4.3.8}$$

$$= \Re\big\{z(t)e^{i\omega_0 t}\big\} = \tfrac{1}{2}\left[z(t)e^{i\omega_0 t} + z^*(t)e^{-i\omega_0 t}\right], \tag{4.3.9}$$

where $z(t)$ is the analytic signal of $x(t)$. We have plotted the amplitude spectrum $|Z(\omega)|$ in Figure 4.3.2(c). If we computed the amplitude spectrum of $y(t)$, we would find that

$$Y(\omega) = \tfrac{1}{2}Z(\omega - \omega_0) + \tfrac{1}{2}Z(-\omega - \omega_0) \tag{4.3.10}$$

$$Y(\omega) = \begin{cases} X(\omega - \omega_0), & \omega_0 \le \omega \le \omega_0 + \omega_{\max}, \\ X^*(-\omega - \omega_0), & -\omega_0 - \omega_{\max} \le \omega \le \omega_0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.3.11}$$

We have sketched this amplitude spectrum $|Y(\omega)|$ in Figure 4.3.2(d). Each triangular part is called the *single sideband signal* because it contains the upper frequencies ($|\omega| > \omega_0$) of the modulated signal $x(t)\cos(\omega_0 t)$. Similarly, if we had used $x(t)\cos(\omega_0 t) + \widehat{x}(t)\sin(\omega_0 t)$, we would only have obtained the lower sidebands. Consequently, a communication system using $x(t)\cos(\omega_0 t) - \widehat{x}(t)\sin(\omega_0 t)$ or $x(t)\cos(\omega_0 t) + \widehat{x}(t)\sin(\omega_0 t)$ would realize a 50% savings in its frequency bandwidth over one transmitting $x(t)\cos(\omega_0 t)$.

Problems

1. Find the analytic signal corresponding to $x(t) = \cos(\omega t)$, $\omega > 0$.

2. Show that the polar form of an analytic signal can be written

$$z(t) = |z(t)|e^{i\varphi(t)},$$

where

$$|z(t)|^2 = x^2(t) + \widehat{x}^2(t), \quad \varphi(t) = \tan^{-1}\left[\frac{\widehat{x}(t)}{x(t)}\right].$$

3. Analytic signals are often used with narrow-band waveforms with carrier frequency ω_0. If $\varphi(t) = \omega_0 t + \varphi'(t)$, show that the analytic signal can be written $z(t) = r(t)e^{i\omega_0 t}$, where $r(t) = |z(t)|e^{i\varphi'(t)}$. The function $r(t)$ is called the *complex envelope* or the *phasor amplitude*; this is a generalization of the phasor idea beyond pure alternating currents.

4.4 CAUSALITY: THE KRAMERS-KRONIG RELATIONSHIP

Causality is the physical principle which states that an event cannot proceed its cause. In this section we explore what effect this principle has on Hilbert transforms.

We begin by introducing the concept of causal functions. A *causal function* is a function that equals zero for all $t < 0$. As with all functions we can write it in terms of an even $x_e(t)$ and an odd $x_o(t)$ part as $x(t) = x_e(t) + x_o(t)$. Because $x(t)$ is causal, $x_o(t) = \operatorname{sgn}(t)x_e(t)$ and

$$x(t) = x_e(t) + \operatorname{sgn}(t)x_e(t). \tag{4.4.1}$$

Taking the Fourier transform of Equation 4.4.1, we find that the Fourier transform of *all* causal functions are of the form

$$X(\omega) = X_e(\omega) - i\widehat{X}_e(\omega), \tag{4.4.2}$$

where

$$\widehat{X}_e(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{X_e(\tau)}{\omega - \tau}\,d\tau, \qquad \text{and} \qquad X_e(\omega) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\widehat{X}_e(\tau)}{\omega - \tau}\,d\tau, \tag{4.4.3}$$

because

$$2\pi\mathcal{F}[x_e(t)\operatorname{sgn}(t)] = \frac{2}{i\omega} * X_e(\omega) = \frac{2}{i}\int_{-\infty}^{\infty} \frac{X_e(\tau)}{\omega - \tau}d\tau. \tag{4.4.4}$$

Equation 4.4.3 first arose in dielectric theory and, taken together, are called the *Kramers*[11] *and Kronig*[12] *relation* after their discoverers, who derived these relationships during their work on the dispersion of light by gaseous atoms or molecules.

• Example 4.4.1

Let us verify the Kramers-Kronig relation using the causal time function $x(t) = H(t)$.

Because $x_e(t) = \frac{1}{2}$ and $X_e(\omega) = \pi\delta(\omega)$,

$$\widehat{X}_e(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\pi\,\delta(\tau)}{\omega - \tau}\,d\tau = -\frac{1}{\omega}. \tag{4.4.5}$$

Consequently, by the Kramers-Kronig relation,

$$\mathcal{F}[H(t)] = X_e(\omega) - i\widehat{X}_e(\omega) = \pi\delta(\omega) + \frac{i}{\omega}. \tag{4.4.6}$$

□

• Example 4.4.2

A simple example of a causal function is the impulse response or Green's function introduced in earlier chapters. From Equation 4.4.2 we have the result that the transfer function $G(\omega)$, the Fourier transform of the impulse response, must yield the Hilbert transform pair $G_e(\omega) - i\widehat{G}_e(\omega)$.

For example, if $g(t) = e^{-t}H(t)$, then $G(\omega) = 1/(1 + i\omega)$. Because

$$\frac{1}{1 + i\omega} = \frac{1}{\omega^2 + 1} - i\frac{\omega}{\omega^2 + 1}, \tag{4.4.7}$$

we have the Hilbert transform pair of

$$x(t) = \frac{1}{t^2 + 1} \qquad \text{and} \qquad \widehat{x}(t) = \frac{t}{t^2 + 1}. \tag{4.4.8}$$

□

• Example 4.4.3

Let us verify the Kramers-Kronig relation for the Hilbert transform pair

$$x(t) = \frac{1}{t^4 + 1} \qquad \text{and} \qquad \widehat{x}(t) = \frac{t(t^2 + 1)}{\sqrt{2}(t^4 + 1)} \tag{4.4.9}$$

by direct integration.

---

[11] Kramers, H. A., 1929: Die Dispersion und Absorption von Röntgenstrahlen. *Phys. Z.*, **30**, 522–523.

[12] Kronig, R. de L., 1926: On the theory of dispersion of x-rays. *J. Opt. Soc. Am.*, **12**, 547–551.

From Equation 4.4.3, we have that

$$\frac{\omega(\omega^2+1)}{\sqrt{2}(\omega^4+1)} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{d\tau}{(\tau^4+1)(\omega-\tau)}. \tag{4.4.10}$$

Applying the residue theorem to the right side of Equation 4.4.10, we obtain

$$\begin{aligned}\frac{\omega(\omega^2+1)}{\sqrt{2}(\omega^4+1)} &= i\,\mathrm{Res}\left[\frac{1}{(z^4+1)(\omega-z)};\omega\right] + 2i\,\mathrm{Res}\left[\frac{1}{(z^4+1)(\omega-z)};e^{\pi i/4}\right] \\ &\quad + 2i\,\mathrm{Res}\left[\frac{1}{(z^4+1)(\omega-z)};e^{3\pi i/4}\right].\end{aligned} \tag{4.4.11}$$

We only include one half of the value of the residue at $\tau = \omega$ because the singularity lies on the path of integration and we must treat this integration along the lines of a Cauchy principal value. Evaluating the residues, we find

$$\mathrm{Res}\left[\frac{1}{(z^4+1)(\omega-z)};\omega\right] = -\frac{1}{\omega^4+1}, \tag{4.4.12}$$

$$\mathrm{Res}\left[\frac{1}{(z^4+1)(\omega-z)};e^{\pi i/4}\right] = \frac{\sqrt{2}-(1+i)\omega}{4\sqrt{2}\left[\left(\omega-\frac{1}{\sqrt{2}}\right)^2+\frac{1}{2}\right]}, \tag{4.4.13}$$

and

$$\mathrm{Res}\left[\frac{1}{(z^4+1)(\omega-z)};e^{3\pi i/4}\right] = \frac{\sqrt{2}+(1-i)\omega}{4\sqrt{2}\left[\left(\omega+\frac{1}{\sqrt{2}}\right)^2+\frac{1}{2}\right]}. \tag{4.4.14}$$

Substituting Equation 4.4.12 through Equation 4.4.14 into the right side of Equation 4.4.11, we obtain the left side.

Problems

1. For a causal function $x(t)$, prove that $x_o(t) = \mathrm{sgn}(t)x_e(t)$ and $x_e(t) = \mathrm{sgn}(t)x_o(t)$.

2. Redo our analysis if $x(t)$ is a negative time function, i.e., $x(t) = 0$ if $t > 0$. Verify your result using $x(t) = e^t H(-t)$.

3. Using $g(t) = te^{-t}H(t)$, find the corresponding Hilbert transform pairs.

4. Using $g(t) = e^{-t}\cos(\omega t)H(t)$, find the corresponding Hilbert transform pairs.

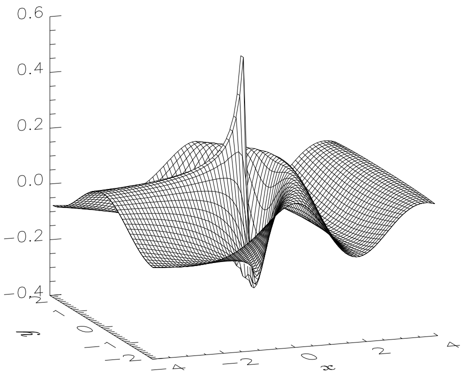
5. Verify the Kramers-Kronig relation for the Hilbert transform pair:

$$x(t) = \frac{1}{t^2+1} \qquad \text{and} \qquad \widehat{x}(t) = \frac{t}{t^2+1}$$

by direct integration.

Further Reading

Hahn, S. L., 1996: *Hilbert Transforms in Signal Processing.* Artech House, 442 pp. Covers the basic theory and gives some practical applications.



Chapter 5
Green's Functions

An important aspect of engineering mathematics is the solution of linear ordinary and partial differential equations. As an undergraduate you were probably introduced to the method of separation of variables, which leads to a solution in terms of an eigenfunction expansion. However, this method is not the only one; there is Duhamel's principle which uses the superposition integral. Here we expand upon this idea and illustrate how a solution, called a *Green's function*, to a differential equation forced by the Dirac delta function can be used in an integral representation of a solution when the forcing is arbitrary.

5.1 WHAT IS A GREEN'S FUNCTION?

The following examples taken from engineering show how Green's functions naturally appear during the solution of initial-value and boundary-value problems. We also show that the solution $u(x)$ can be expressed as an integral involving the Green's function and $f(x)$.

Circuit theory

In electrical engineering, one of the simplest electrical devices consists of a voltage source $v(t)$ connected to a resistor with resistance R and an inductor with inductance L. See Figure 5.1.1. Denoting the current by $i(t)$, the equation that governs this circuit is

$$L\frac{di}{dt} + Ri = v(t). \tag{5.1.1}$$

Consider now the following experiment: With the circuit initially dead, we allow the voltage to suddenly become $V_0/\Delta\tau$ during a very short duration $\Delta\tau$ starting at $t = \tau$.

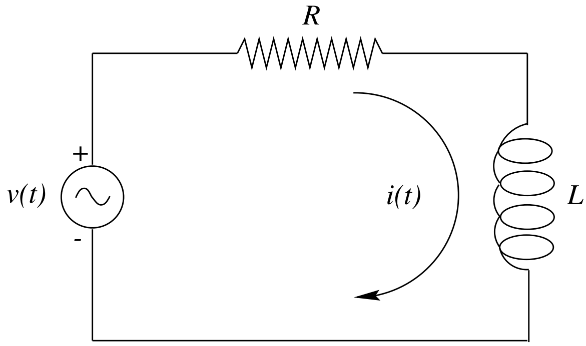


Figure 5.1.1: The RL electrical circuit driven by the voltage $v(t)$.

Then, at $t = \tau + \Delta\tau$, we again turn off the voltage supply. Mathematically, for $t > \tau + \Delta\tau$, the circuit's performance obeys the homogeneous differential equation:

$$L\frac{di}{dt} + Ri = 0, \qquad t > \tau + \Delta\tau, \tag{5.1.2}$$

whose solution is

$$i(t) = I_0 e^{-Rt/L}, \qquad t > \tau + \Delta\tau, \tag{5.1.3}$$

where I_0 is a constant and L/R is the *time constant* of the circuit. Because the voltage $v(t)$ during $\tau < t < \tau + \Delta\tau$ is $V_0/\Delta\tau$, then

$$\int_\tau^{\tau+\Delta\tau} v(t)\,dt = V_0. \tag{5.1.4}$$

Therefore, over the interval $\tau < t < \tau + \Delta\tau$, Equation 5.1.1 can be integrated to yield

$$L\int_\tau^{\tau+\Delta\tau} di + R\int_\tau^{\tau+\Delta\tau} i(t)\,dt = \int_\tau^{\tau+\Delta\tau} v(t)\,dt, \tag{5.1.5}$$

or

$$L\left[i(\tau+\Delta\tau) - i(\tau)\right] + R\int_\tau^{\tau+\Delta\tau} i(t)\,dt = V_0. \tag{5.1.6}$$

If $i(t)$ remains continuous as $\Delta\tau$ becomes small, then

$$R\int_\tau^{\tau+\Delta\tau} i(t)\,dt \approx 0. \tag{5.1.7}$$

Finally, because

$$i(\tau) = 0 \qquad \text{and} \qquad i(\tau+\Delta\tau) = I_0 e^{-R(\tau+\Delta\tau)/L} \approx I_0 e^{-R\tau/L}, \tag{5.1.8}$$

for small $\Delta\tau$, Equation 5.1.6 reduces to

$$LI_0 e^{-R\tau/L} = V_0, \qquad \text{or} \qquad I_0 = \frac{V_0}{L}e^{R\tau/L}. \tag{5.1.9}$$

Therefore, Equation 5.1.3 can be written as

$$i(t) = \begin{cases} 0, & t < \tau, \\ V_0 e^{-R(t-\tau)/L}/L, & \tau \le t, \end{cases} \tag{5.1.10}$$

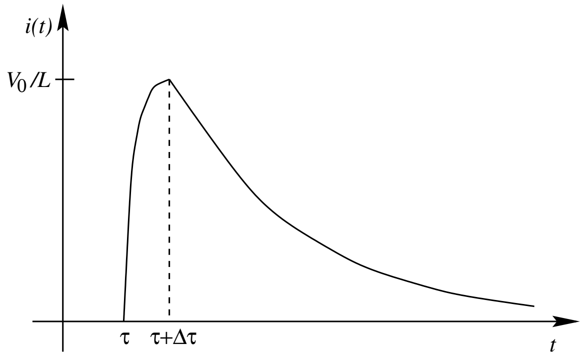


Figure 5.1.2: The current $i(t)$ within an RL circuit when the voltage $V_0/\Delta\tau$ is introduced between the times $\tau < t < \tau + \Delta\tau$.

after using Equation 5.1.9. Equation 5.1.10 is plotted in Figure 5.1.2.

Consider now a new experiment with the same circuit where we subject the circuit to N voltage impulses, each of duration $\Delta\tau$ and amplitude $V_i/\Delta\tau$ with $i = 0, 1, \ldots, N$, occurring at $t = \tau_i$. See Figure 5.1.3. The current response is then

$$i(t) = \begin{cases} 0, & t < \tau_0, \\ V_0 e^{-R(t-\tau_0)/L}/L, & \tau_0 < t < \tau_1, \\ V_0 e^{-R(t-\tau_0)/L}/L + V_1 e^{-R(t-\tau_1)/L}/L, & \tau_1 < t < \tau_2, \\ \vdots & \vdots \\ \sum_{i=0}^{N} V_i e^{-R(t-\tau_i)/L}/L, & \tau_N < t < \tau_{N+1}. \end{cases} \tag{5.1.11}$$

Finally, consider our circuit subjected to a continuous voltage source $v(t)$. Over each successive interval $d\tau$, the step change in voltage is $v(\tau)\,d\tau$. Consequently, from Equation 5.1.11 the response $i(t)$ is now given by the *superposition integral*

$$i(t) = \int_\tau^t \frac{v(\tau)}{L} e^{-R(t-\tau)/L}\,d\tau, \qquad \text{or} \qquad i(t) = \int_\tau^t v(\tau) g(t|\tau)\,d\tau, \tag{5.1.12}$$

where

$$g(t|\tau) = \frac{e^{-R(t-\tau)/L}}{L}, \qquad \tau < t. \tag{5.1.13}$$

Here we have assumed that $i(t) = v(t) = 0$ for $t < \tau$. In Equation 5.1.13, $g(t|\tau)$ is called the *Green's function.* As this equation shows, given the Green's function to Equation 5.1.1, the response $i(t)$ to any voltage source $v(t)$ can be obtained by convolving the voltage source with the Green's function.

We now show that we could have found the Green's function, Equation 5.1.13, by solving Equation 5.1.1 subject to an impulse- or delta-forcing function. Mathematically, this corresponds to solving the following initial-value problem:

$$L\frac{dg}{dt} + Rg = \delta(t-\tau), \qquad g(0|\tau) = 0. \tag{5.1.14}$$

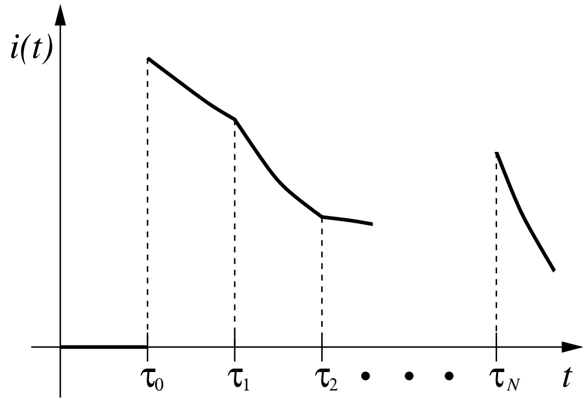


Figure 5.1.3: The current $i(t)$ within an RL circuit when the voltage is changed at $t = \tau_0$, $t = \tau_1$, and so forth.

Taking the Laplace transform of Equation 5.1.14, we find that

$$G(s|\tau) = \frac{e^{-s\tau}}{Ls + R}, \qquad \text{or} \qquad g(t|\tau) = \frac{e^{-R(t-\tau)/L}}{L} H(t-\tau), \tag{5.1.15}$$

where $H(\cdot)$ is the Heaviside step function. As our short derivation showed, the most direct route to finding a Green's function is solving the differential equation when its forcing equals the impulse or delta function. This is the technique that we will use throughout this chapter.

Statics

Consider a string of length L that is connected at both ends to supports and is subjected to a load (external force per unit length) of $f(x)$. We wish to find the displacement $u(x)$ of the string. If the load $f(x)$ acts downward (negative direction), the displacement $u(x)$ of the string is given by the differential equation:

$$T\frac{d^2u}{dx^2} = f(x), \tag{5.1.16}$$

where T denotes the uniform tensile force of the string. Because the string is stationary at both ends, the displacement $u(x)$ satisfies the boundary conditions $u(0) = u(L) = 0$.

Instead of directly solving for the displacement $u(x)$ of the string subject to the load $f(x)$, let us find the displacement that results from a load $\delta(x-\xi)$ concentrated at the point $x = \xi$. See Figure 5.1.4. For this load, the differential equation, Equation 5.1.16, becomes

$$T\frac{d^2g}{dx^2} = \delta(x-\xi), \tag{5.1.17}$$

subject to the boundary conditions $g(0|\xi) = g(L|\xi) = 0$.

In Equation 5.1.17, $g(x|\xi)$ denotes the displacement of the string when it is subjected to an impulse load at $x = \xi$. In line with our circuit theory example, it gives the *Green's function* for our statics problem. Once found, the displacement $u(x)$ of the string subject to

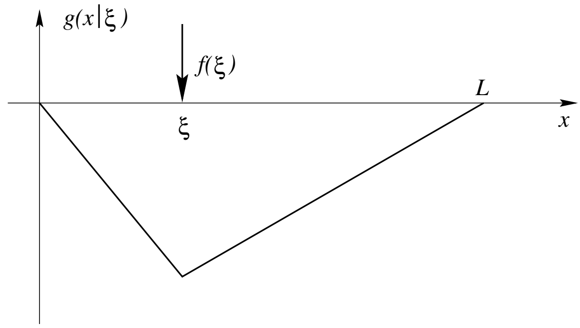


Figure 5.1.4: The response, commonly called a Green's function, of a string fixed at both ends to a point load at $x = \xi$.

any arbitrary load $f(x)$ can be found by convolving the load $f(x)$ with the Green's function $g(x|\xi)$ as we did earlier.

Let us now find this Green's function. At any point $x \neq \xi$, Equation 5.1.17 reduces to the homogeneous differential equation:

$$\frac{d^2 g}{dx^2} = 0, \tag{5.1.18}$$

which has the solution

$$g(x|\xi) = \begin{cases} ax + b, & 0 \le x < \xi, \\ cx + d, & \xi < x \le L. \end{cases} \tag{5.1.19}$$

Applying the boundary conditions, Equation 5.1.19, we find that

$$g(0|\xi) = a \cdot 0 + b = b = 0, \qquad \text{and} \qquad g(L|\xi) = cL + d = 0, \quad \text{or} \quad d = -cL. \tag{5.1.20}$$

Therefore, we can rewrite Equation 5.1.19 as

$$g(x|\xi) = \begin{cases} ax, & 0 \le x < \xi, \\ c(x - L), & \xi < x \le L, \end{cases} \tag{5.1.21}$$

where a and c are undetermined constants.

At $x = \xi$, the displacement $u(x)$ of the string must be continuous; otherwise, the string would be broken. Therefore, the Green's function given by Equation 5.1.21 must also be continuous there. Thus,

$$a\xi = c(\xi - L), \qquad \text{or} \qquad c = \frac{a\xi}{\xi - L}. \tag{5.1.22}$$

From Equation 5.1.13 the second derivative of $g(x|\xi)$ must equal the impulse function. Therefore, the first derivative of $g(x|\xi)$, obtained by integrating this equation, must be discontinuous by the amount $1/T$ or

$$\lim_{\epsilon \to 0} \left[\frac{dg(\xi + \epsilon|\xi)}{dx} - \frac{dg(\xi - \epsilon|\xi)}{dx} \right] = \frac{1}{T}, \tag{5.1.23}$$

in which case

$$\frac{dg(\xi^+|\xi)}{dx} - \frac{dg(\xi^-|\xi)}{dx} = \frac{1}{T}, \tag{5.1.24}$$

where ξ^+ and ξ^- denote points lying just above or below ξ, respectively. Using Equation 5.1.24, we find that

$$\frac{dg(\xi^-|\xi)}{dx} = a, \qquad \text{and} \qquad \frac{dg(\xi^+|\xi)}{dx} = c = \frac{a\xi}{\xi - L}. \tag{5.1.25}$$

Thus, Equation 5.1.25 leads to

$$\frac{a\xi}{\xi - L} - a = \frac{1}{T} = \frac{aL}{\xi - L}, \qquad \text{or} \qquad a = \frac{\xi - L}{LT}, \tag{5.1.26}$$

and the Green's function is

$$g(x|\xi) = \frac{1}{TL}(x_> - L)x_<, \tag{5.1.27}$$

where $x_< = \min(x, \xi)$ and $x_> = \max(x, \xi)$. To find the displacement $u(x)$ subject to the load $f(x)$, we proceed as we did in the previous example. The result of this analysis is

$$u(x) = \int_0^L f(\xi)g(x|\xi)\, d\xi = \frac{x - L}{TL}\int_0^x f(\xi)\,\xi\, d\xi + \frac{x}{TL}\int_x^L f(\xi)\,(\xi - L)\, d\xi, \tag{5.1.28}$$

since $\xi < x$ in the first integral and $x < \xi$ in the second integral of Equation 5.1.28.

Integral Equations

Consider the Sturm-Liouville problem

$$y'' + \lambda y = 0, \qquad y(0) = y(L) = 0. \tag{5.1.29}$$

From its general theory, nontrivial solutions exist only if

$$\lambda_n = \frac{n^2\pi^2}{L^2}, \qquad y_n(x) = \sin\left(\frac{n\pi x}{L}\right), \tag{5.1.30}$$

where $n = 1, 2, 3, \ldots$.

Consider now a new boundary-value problem:

$$\frac{d^2y}{dx^2} = -f(x), \qquad y(0) = y(L) = 0. \tag{5.1.31}$$

In the next section (Equation 5.2.76), we will show that we can write its solution by

$$y(x) = \int_0^L f(\xi)g(x|\xi)\, d\xi, \tag{5.1.32}$$

where the Green's function $g(x|\xi)$ is given by

$$\frac{d^2g}{dx^2} = -\delta(x - \xi), \qquad g(0|\xi) = g(L|\xi) = 0, \qquad \text{or} \qquad g(x|\xi) = (L - x_>)x_</L, \tag{5.1.33}$$

where $x_> = \max(x, \xi)$ and $x_< = \min(x, \xi)$.

We can now use Equation 5.1.29 to rewrite Equation 5.1.31 as $\lambda y(\xi) = f(\xi)$. Multiplying this equation by $g(x|\xi)$ and integrating from 0 to L, we find that

$$\int_0^L f(\xi)g(x|\xi)\,d\xi = \lambda \int_0^L y(\xi)g(x|\xi)\,d\xi, \tag{5.1.34}$$

or

$$y(x) - \lambda \int_0^L y(\xi)g(x|\xi)\,d\xi = 0. \tag{5.1.35}$$

Because of the equivalence of Equation 5.1.29 and Equation 5.1.35, the solutions to the *integral equation*, Equation 5.1.35, are $\lambda_n = n^2\pi^2/L^2$ with $y_n(x) = \sin(n\pi x/L)$. Direct substitution verifies this result. Thus, we can use Green's functions to construct integral equations that have known solutions. Indeed, it was the use of Green's functions to solve Fredholm integral equations that drew the attention of mathematicians at the turn of the twentieth century.[1]

5.2 ORDINARY DIFFERENTIAL EQUATIONS

Second-order differential equations are ubiquitous in engineering. In electrical engineering, many electrical circuits are governed by second-order, linear ordinary differential equations. In mechanical engineering they arise during the application of Newton's second law.

One of the drawbacks of solving ordinary differential equations with a forcing term is its lack of generality. Each new forcing function requires a repetition of the entire process. In this section we give some methods for finding the solution in a somewhat more general manner for stationary systems where the forcing, not any initially stored energy (i.e., nonzero initial conditions), produces the total output. Unfortunately, the solution must be written as an integral.

Consider the linear differential equation

$$y'' + 2y' + y = f(t), \tag{5.2.1}$$

subject to the initial conditions $y(0) = y'(0) = 0$. Solving this equation by Laplace transforms, we can write the Laplace transform of $y(t)$, $Y(s)$, as the product of two Laplace transforms:

$$Y(s) = \frac{1}{(s+1)^2}F(s). \tag{5.2.2}$$

One drawback in using Equation 5.2.2 is its dependence upon an unspecified Laplace transform $F(s)$. Is there a way to eliminate this dependence and yet retain the essence of the solution?

One way of obtaining a quantity that is independent of the forcing is to consider the ratio:

$$\frac{Y(s)}{F(s)} = G(s) = \frac{1}{(s+1)^2}. \tag{5.2.3}$$

This ratio is called the *transfer function* because we can transfer the input $F(s)$ into the output $Y(s)$ by multiplying $F(s)$ by $G(s)$. It depends only upon the properties of the system.

---

[1] See Section 36 in Kneser, A., 1911: *Integralgleichungen und ihre Anwendungen in der mathematischen Physik*. Braunschweig, 293 pp.

Let us now consider a problem related to Equation 5.2.1, namely

$$g'' + 2g' + g = \delta(t), \qquad t > 0, \tag{5.2.4}$$

with $g(0) = g'(0) = 0$. Because the forcing equals the Dirac delta function, $g(t)$ is called the *impulse response* or *Green's function*.[2] Computing $G(s)$,

$$G(s) = \frac{1}{(s+1)^2}. \tag{5.2.5}$$

From Equation 5.2.3 we see that $G(s)$ is also the transfer function. Thus, an alternative method for computing the transfer function is to subject the system to impulse forcing and the Laplace transform of the response is the transfer function.

From Equation 5.2.3,

$$Y(s) = G(s)F(s), \tag{5.2.6}$$

or

$$y(t) = g(t) * f(t). \tag{5.2.7}$$

That is, the convolution of the impulse response with the particular forcing gives the response of the system. Thus, we may describe a stationary system in one of two ways: (1) in the transform domain we have the transfer function, and (2) in the time domain there is the impulse response.

Despite the fundamental importance of the impulse response or Green's function for a given linear system, it is often quite difficult to determine, especially experimentally, and a more convenient practice is to deal with the response to the unit step $H(t)$. This response is called the *indicial admittance* or *step response*, which we shall denote by $a(t)$. Because $\mathcal{L}[H(t)] = 1/s$, we can determine the transfer function from the indicial admittance because $\mathcal{L}[a(t)] = G(s)\mathcal{L}[H(t)]$ or $sA(s) = G(s)$. Furthermore, because

$$\mathcal{L}[g(t)] = G(s) = \frac{\mathcal{L}[a(t)]}{\mathcal{L}[H(t)]}, \tag{5.2.8}$$

then

$$g(t) = \frac{da(t)}{dt}, \tag{5.2.9}$$

since $\mathcal{L}[f'(t)] = sF(s) - f(0^+)$.

• Example 5.2.1

Let us find the transfer function, impulse response, and step response for the system

$$y'' - 3y' + 2y = f(t), \tag{5.2.10}$$

with $y(0) = y'(0) = 0$. To find the impulse response, we solve

$$g'' - 3g' + 2g = \delta(t - \tau), \tag{5.2.11}$$

[2] For the origin of the Green's function, see Farina, J. E. G., 1976: The work and significance of George Green, the miller mathematician, 1793–1841. *Bull. Inst. Math. Appl.*, **12**, 98–105.

with $g(0) = g'(0) = 0$. We have generalized the problem to an arbitrary forcing at $t = \tau$ and now denote the Green's function by $g(t|\tau)$. We have done this so that our discussion will be consistent with the other sections in the chapter.

Taking the Laplace transform of Equation 5.2.11, we find that

$$G(s|\tau) = \frac{e^{-s\tau}}{s^2 - 3s + 2}, \tag{5.2.12}$$

which is the transfer function for this system when $\tau = 0$. The impulse response or Green's function equals the inverse of $G(s|\tau)$ or

$$g(t|\tau) = \left[e^{2(t-\tau)} - e^{t-\tau}\right] H(t-\tau). \tag{5.2.13}$$

To find the step response, we solve

$$a'' - 3a' + 2a = H(t), \tag{5.2.14}$$

with $a(0) = a'(0) = 0$. Taking the Laplace transform of Equation 5.2.14,

$$A(s) = \frac{1}{s(s-1)(s-2)}, \tag{5.2.15}$$

and the indicial admittance is given by the inverse of Equation 5.2.15, or

$$a(t) = \tfrac{1}{2} + \tfrac{1}{2}e^{2t} - e^t. \tag{5.2.16}$$

Note that $a'(t) = g(t|0)$. □

• Example 5.2.2

MATLAB's control toolbox contains several routines for the numerical computation of impulse and step responses if the transfer function can be written as the ratio of two polynomials. To illustrate this capacity, let us redo the previous example where the transfer function is given by Equation 5.2.12 with $\tau = 0$. The transfer function is introduced by loading in the polynomial in the numerator `num` and in the denominator `den` followed by calling `tf`. The MATLAB script

```
clear
% load in coefficients of the numerator and denominator
%     of the transfer function
num = [0 0 1]; den = [1 -3 2];
% create the transfer function
sys = tf(num,den);
% find the step response, a
[a,t] = step(sys);
% plot the indicial admittance
subplot(2,1,1), plot(t, a, 'o')
ylabel('indicial response','Fontsize',20)
% find the impulse response, g
[g,t] = impulse(sys);
% plot the impulse response
```

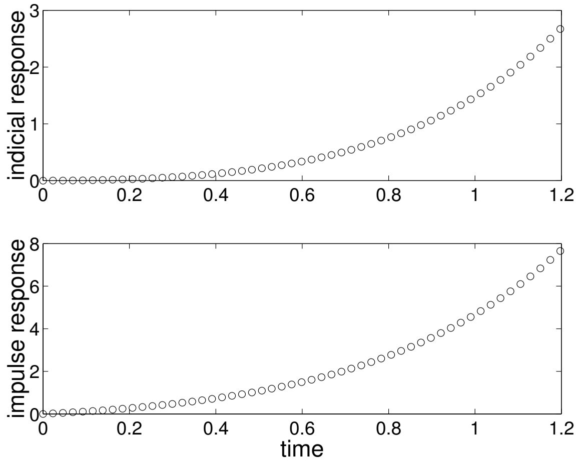


Figure 5.2.1: The impulse and step responses corresponding to the transfer function, Equation 5.2.12, with $\tau = 0$.

```
subplot(2,1,2), plot(t, g, 'o')
ylabel('impulse response','Fontsize',20)
xlabel('time','Fontsize',20)
```

shows how the impulse and step responses are found. Both of them are shown in Figure 5.2.1. □

• Example 5.2.3

There is an old joke about a man who took his car into a garage because of a terrible knocking sound. Upon his arrival the mechanic took one look at it and gave it a hefty kick.[3] Then, without a moment's hesitation he opened the hood, bent over, and tightened up a loose bolt. Turning to the owner, he said, "Your car is fine. That'll be $50." The owner felt that the charge was somewhat excessive, and demanded an itemized account. The mechanic said, "The kicking of the car and tightening one bolt, cost you a buck. The remaining $49 comes from knowing where to kick the car and finding the loose bolt."

Although the moral of the story may be about expertise as a marketable commodity, it also illustrates the concept of transfer function.[4] Let us model the car as a linear system where the equation

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_1 \frac{dy}{dt} + a_0 y = f(t) \qquad (\mathbf{5.2.17})$$

governs the response $y(t)$ to a forcing $f(t)$. Assuming that the car has been sitting still, the initial conditions are zero and the Laplace transform of Equation 5.2.17 is

$$K(s)Y(s) = F(s), \qquad (\mathbf{5.2.18})$$

where

$$K(s) = a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0. \qquad (\mathbf{5.2.19})$$

[3] This is obviously a very old joke.

[4] Originally suggested by Stern, M. D., 1987: Why the mechanic kicked the car - A teaching aid for transfer functions. *Math. Gaz.*, **71**, 62–64.

Hence,

$$Y(s) = \frac{F(s)}{K(s)} = G(s)F(s), \tag{5.2.20}$$

where the transfer function $G(s)$ clearly depends only on the internal workings of the car. So if we know the transfer function, we understand how the car vibrates because

$$y(t) = \int_0^t g(t-x)f(x)\,dx. \tag{5.2.21}$$

But what does this have to do with our mechanic? He realized that a short sharp kick mimics an impulse forcing with $f(t) = \delta(t)$ and $y(t) = g(t)$. Therefore, by observing the response of the car to his kick, he diagnosed the loose bolt and fixed the car. □

In the previous examples, we used Laplace transforms to solve for the Green's functions. However, there is a rich tradition of using Fourier transforms rather than Laplace transforms. In these particular cases, the Fourier transform of the Green's function is called *frequency response* or *steady-state transfer function* of our system when $\tau = 0$. Consider the following examples.

• Example 5.2.4: Spectrum of a damped harmonic oscillator

In mechanics the damped oscillations of a mass m attached to a spring with a spring constant k and damped with a velocity-dependent resistance are governed by the equation

$$my'' + cy' + ky = f(t), \tag{5.2.22}$$

where $y(t)$ denotes the displacement of the oscillator from its equilibrium position, c denotes the damping coefficient, and $f(t)$ denotes the forcing.

Assuming that both $f(t)$ and $y(t)$ have Fourier transforms, let us analyze this system by finding its frequency response. We begin by solving for the Green's function $g(t|\tau)$, which is given by

$$mg'' + cg' + kg = \delta(t-\tau), \tag{5.2.23}$$

because the Green's function is the response of a system to a delta function forcing. Taking the Fourier transform of both sides of Equation 5.2.23, the frequency response is

$$G(\omega|\tau) = \frac{e^{-i\omega\tau}}{k + ic\omega - m\omega^2} = \frac{e^{-i\omega\tau}/m}{\omega_0^2 + ic\omega/m - \omega^2}, \tag{5.2.24}$$

where $\omega_0^2 = k/m$ is the natural frequency of the system. The most useful quantity to plot is the frequency response or

$$|G(\omega|\tau)| = \frac{\omega_0^2}{k\sqrt{(\omega^2-\omega_0^2)^2 + \omega^2\omega_0^2(c^2/km)}} \tag{5.2.25}$$

$$= \frac{1}{k\sqrt{[(\omega/\omega_0)^2 - 1]^2 + (c^2/km)(\omega/\omega_0)^2}}. \tag{5.2.26}$$

In Figure 5.2.2 we plotted the frequency response as a function of $c^2/(km)$. Note that as the damping becomes larger, the sharp peak at $\omega = \omega_0$ essentially vanishes. As $c^2/(km) \to 0$,

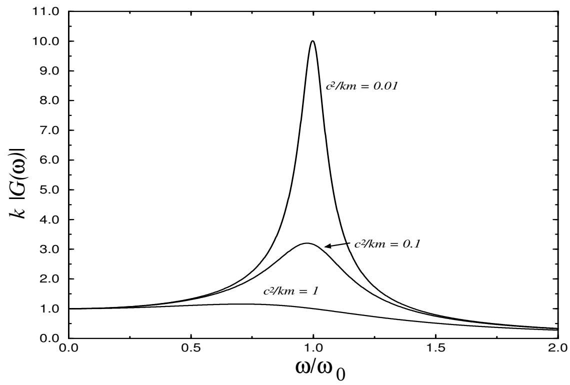


Figure 5.2.2: The variation of the frequency response for a damped harmonic oscillator as a function of driving frequency ω. See the text for the definition of the parameters.

we obtain a very finely tuned response curve. Let us now find the Green's function. From the definition of the inverse Fourier transform,

$$mg(t|\tau) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\omega t}}{\omega^2 - ic\omega/m - \omega_0^2}\,d\omega = -\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\omega t}}{(\omega-\omega_1)(\omega-\omega_2)}\,d\omega, \qquad \textbf{(5.2.27)}$$

where

$$\omega_{1,2} = \pm\sqrt{\omega_0^2 - \gamma^2} + \gamma i, \qquad \textbf{(5.2.28)}$$

and $\gamma = c/(2m) > 0$. We can evaluate Equation 5.2.27 by residues. Clearly the poles always lie in the upper half of the ω-plane. Thus, if $t < \tau$ in Equation 5.2.27 we can close the line integration along the real axis with a semicircle of infinite radius in the lower half of the ω-plane by Jordan's lemma. Because the integrand is analytic within the closed contour, $g(t|\tau) = 0$ for $t < \tau$. This is simply the causality condition,[5] the impulse forcing being the cause of the excitation. Clearly, causality is closely connected with the analyticity of the frequency response in the lower half of the ω-plane.

If $t > \tau$, we close the line integration along the real axis with a semicircle of infinite radius in the upper half of the ω-plane and obtain

$$mg(t|\tau) = 2\pi i\left(-\frac{1}{2\pi}\right)\left\{\operatorname{Res}\left[\frac{e^{iz(t-\tau)}}{(z-\omega_1)(z-\omega_2)};\omega_1\right] + \operatorname{Res}\left[\frac{e^{iz(t-\tau)}}{(z-\omega_1)(z-\omega_2)};\omega_2\right]\right\} \qquad \textbf{(5.2.29)}$$

$$= \frac{-i}{\omega_1-\omega_2}\left[e^{i\omega_1(t-\tau)} - e^{i\omega_2(t-\tau)}\right] = \frac{e^{-\gamma(t-\tau)}\sin\left[(t-\tau)\sqrt{\omega_0^2-\gamma^2}\right]}{\sqrt{\omega_0^2-\gamma^2}}H(t-\tau). \qquad \textbf{(5.2.30)}$$

Let us now examine the damped harmonic oscillator by describing the migration of the poles $\omega_{1,2}$ in the complex ω-plane as γ increases from 0 to ∞. See Figure 5.2.3. For $\gamma \ll \omega_0$ (weak damping), the poles $\omega_{1,2}$ are very near to the real axis, above the points $\pm\omega_0$, respectively. This corresponds to the narrow resonance band discussed earlier and we have an underdamped harmonic oscillator. As γ increases from 0 to ω_0, the poles

[5] The principle stating that an event cannot precede its cause.

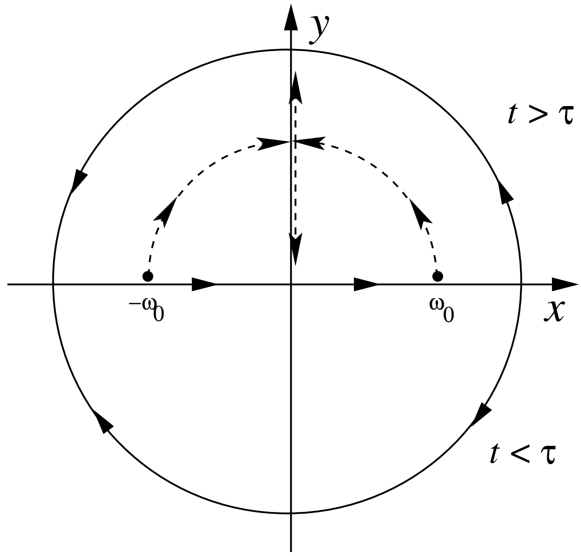


Figure 5.2.3: The migration of the poles of the frequency response of a damped harmonic oscillator as a function of γ.

approach the positive imaginary axis, moving along a semicircle of radius ω_0 centered at the origin. They coalesce at the point $i\omega_0$ for $\gamma = \omega_0$, yielding repeated roots, and we have a critically damped oscillator. For $\gamma > \omega_0$, the poles move in opposite directions along the positive imaginary axis; one of them approaches the origin, while the other tends to $i\infty$ as $\gamma \to \infty$. The solution then has two purely decaying, overdamped solutions. During the early 1950s, a similar diagram was invented by Evans[6] where the movement of closed-loop poles is plotted for all values of a system parameter, usually the gain. This *root-locus method* is very popular in system control theory for two reasons. First, the investigator can easily determine the contribution of a particular closed-loop pole to the transient response. Second, he can determine the manner in which open-loop poles or zeros should be introduced or their location modified so that he will achieve a desired performance characteristic for his system. □

• Example 5.2.5: Low-frequency filter

Consider the ordinary differential equation

$$Ry' + \frac{y}{C} = f(t), \tag{5.2.31}$$

where R and C are real, positive constants. If $y(t)$ denotes current, then Equation 5.2.31 would be the equation that gives the voltage across a capacitor in an RC circuit. Let us find the frequency response and Green's function for this system. We begin by writing Equation 5.2.31 as

$$Rg' + \frac{g}{C} = \delta(t - \tau), \tag{5.2.32}$$

where $g(t|\tau)$ denotes the Green's function. If the Fourier transform of $g(t|\tau)$ is $G(\omega|\tau)$, the frequency response $G(\omega|\tau)$ is given by

$$i\omega RG(\omega|\tau) + \frac{G(\omega|\tau)}{C} = e^{-i\omega\tau}, \tag{5.2.33}$$

---

[6] Evans, W. R., 1948: Graphical analysis of control systems. *Trans. AIEE*, **67**, 547–551; Evans, W. R., 1954: *Control-System Dynamics*. McGraw-Hill, 282 pp.

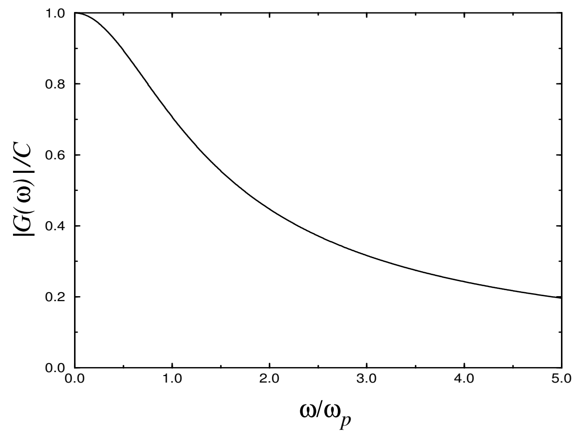


Figure 5.2.4: The variation of the frequency response, Equation 5.2.35, as a function of driving frequency ω. See the text for the definition of the parameters.

or

$$G(\omega|\tau) = \frac{e^{-i\omega\tau}}{i\omega R + 1/C} = \frac{Ce^{-i\omega\tau}}{1 + i\omega RC}, \tag{5.2.34}$$

and

$$|G(\omega|\tau)| = \frac{C}{\sqrt{1 + \omega^2 R^2 C^2}} = \frac{C}{\sqrt{1 + \omega^2/\omega_p^2}}, \tag{5.2.35}$$

where $\omega_p = 1/(RC)$ is an intrinsic constant of the system. In Figure 5.2.4 we plotted $|G(\omega|\tau)|$ as a function of ω. From this figure, we see that the response is largest for small ω and decreases as ω increases.

This is an example of a *low-frequency filter* because relatively more signal passes through at lower frequencies than at higher frequencies. To understand this, let us drive the system with a forcing function that has the Fourier transform $F(\omega)$. The response of the system will be $G(\omega, 0)F(\omega)$. Thus, that portion of the forcing function's spectrum at the lower frequencies is relatively unaffected because $|G(\omega, 0)|$ is near unity. However, at higher frequencies where $|G(\omega, 0)|$ is smaller, the magnitude of the output is greatly reduced. □

• **Example 5.2.6**

During his study of tumor growth, Adam[7] found the particular solution to an ordinary differential equation which, in its simplest form, is

$$y'' - \alpha^2 y = \begin{cases} |x|/L - 1, & |x| < L, \\ 0, & |x| > L, \end{cases} \tag{5.2.36}$$

by the method of Green's functions. Let us retrace his steps and see how he did it.

The first step is finding the Green's function. We do this by solving

$$g'' - \alpha^2 g = \delta(x), \tag{5.2.37}$$

subject to the boundary conditions $\lim_{|x|\to\infty} g(x) \to 0$. Taking the Fourier transform of Equation 5.2.37, we obtain

$$G(\omega) = -\frac{1}{\omega^2 + \alpha^2}. \tag{5.2.38}$$

[7] Adam, J. A., 1986: A simplified mathematical model of tumor growth. *Math. Biosci.*, **81**, 229–244.

The function $G(\omega)$ is the *frequency response* for our problem. Straightforward inversion yields the Green's function

$$g(x) = -\frac{e^{-\alpha|x|}}{2\alpha}. \tag{5.2.39}$$

Therefore, by the convolution integral, $y(x) = g(x) * f(x)$,

$$y(x) = \int_{-L}^{L} g(x-\xi)\,(|\xi|/L - 1)\,d\xi = \frac{1}{2\alpha}\int_{-L}^{L} (1 - |\xi|/L)\,e^{-\alpha|x-\xi|}\,d\xi. \tag{5.2.40}$$

To evaluate Equation 5.2.40 we must consider four separate cases: $-\infty < x < -L$, $-L < x < 0$, $0 < x < L$, and $L < x < \infty$. Turning to the $-\infty < x < -L$ case first, we have

$$y(x) = \frac{1}{2\alpha}\int_{-L}^{L} (1 - |\xi|/L)\,e^{\alpha(x-\xi)}\,d\xi \tag{5.2.41}$$

$$= \frac{e^{\alpha x}}{2\alpha}\int_{-L}^{0} (1 + \xi/L)\,e^{-\alpha\xi}\,d\xi + \frac{e^{\alpha x}}{2\alpha}\int_{0}^{L} (1 - \xi/L)\,e^{-\alpha\xi}\,d\xi \tag{5.2.42}$$

$$= \frac{e^{\alpha x}}{2\alpha^3 L}\left(e^{\alpha L} + e^{-\alpha L} - 2\right). \tag{5.2.43}$$

Similarly, for $x > L$,

$$y(x) = \frac{1}{2\alpha}\int_{-L}^{L} (1 - |\xi|/L)\,e^{-\alpha(x-\xi)}\,d\xi \tag{5.2.44}$$

$$= \frac{e^{-\alpha x}}{2\alpha}\int_{-L}^{0} (1 + \xi/L)\,e^{\alpha\xi}\,d\xi + \frac{e^{-\alpha x}}{2\alpha}\int_{0}^{L} (1 - \xi/L)\,e^{\alpha\xi}\,d\xi \tag{5.2.45}$$

$$= \frac{e^{-\alpha x}}{2\alpha^3 L}\left(e^{\alpha L} + e^{-\alpha L} - 2\right). \tag{5.2.46}$$

On the other hand, for $-L < x < 0$, we find that

$$y(x) = \frac{1}{2\alpha}\int_{-L}^{x} (1 - |\xi|/L)\,e^{-\alpha(x-\xi)}\,d\xi + \frac{1}{2\alpha}\int_{x}^{L} (1 - |\xi|/L)\,e^{\alpha(x-\xi)}\,d\xi \tag{5.2.47}$$

$$= \frac{e^{-\alpha x}}{2\alpha}\int_{-L}^{x} (1 + \xi/L)\,e^{\alpha\xi}\,d\xi + \frac{e^{\alpha x}}{2\alpha}\int_{x}^{0} (1 + \xi/L)\,e^{-\alpha\xi}\,d\xi + \frac{e^{\alpha x}}{2\alpha}\int_{0}^{L} (1 - \xi/L)\,e^{-\alpha\xi}\,d\xi \tag{5.2.48}$$

$$= \frac{1}{\alpha^3 L}\left[e^{-\alpha L}\cosh(\alpha x) + \alpha(x + L) - e^{\alpha x}\right]. \tag{5.2.49}$$

Finally, for $0 < x < L$, we have that

$$y(x) = \frac{1}{2\alpha}\int_{-L}^{x} (1 - |\xi|/L)\,e^{-\alpha(x-\xi)}\,d\xi + \frac{1}{2\alpha}\int_{x}^{L} (1 - |\xi|/L)\,e^{\alpha(x-\xi)}\,d\xi \tag{5.2.50}$$

$$= \frac{e^{-\alpha x}}{2\alpha}\int_{-L}^{0} (1 + \xi/L)\,e^{\alpha\xi}\,d\xi + \frac{e^{-\alpha x}}{2\alpha}\int_{0}^{x} (1 - \xi/L)\,e^{\alpha\xi}\,d\xi + \frac{e^{\alpha x}}{2\alpha}\int_{x}^{L} (1 - \xi/L)\,e^{-\alpha\xi}\,d\xi \tag{5.2.51}$$

$$= \frac{1}{\alpha^3 L}\left[e^{-\alpha L}\cosh(\alpha x) + \alpha(L - x) - e^{-\alpha x}\right]. \tag{5.2.52}$$

These results can be collapsed down into

$$y(x) = \frac{1}{\alpha^3 L}\left[e^{-\alpha L}\cosh(\alpha x) + \alpha(L - |x|) - e^{-\alpha|x|}\right] \tag{5.2.53}$$

if $|x| < L$, and

$$y(x) = \frac{e^{-\alpha|x|}}{2\alpha^3 L}\left(e^{\alpha L} + e^{-\alpha L} - 2\right) \tag{5.2.54}$$

if $|x| > L$. □

Superposition integral

So far we showed how the response of any system can be expressed in terms of its Green's function and the arbitrary forcing. Can we also determine the response using the indicial admittance $a(t)$?

Consider first a system that is dormant until a certain time $t = \tau_1$. At that instant we subject the system to a forcing $H(t - \tau_1)$. Then the response will be zero if $t < \tau_1$ and will equal the indicial admittance $a(t - \tau_1)$ when $t > \tau_1$ because the indicial admittance is the response of a system to the step function. Here $t - \tau_1$ is the time measured from the instant of change.

Next, suppose that we now force the system with the value $f(0)$ when $t = 0$ and hold that value until $t = \tau_1$. We then abruptly change the forcing by an amount $f(\tau_1) - f(0)$ to the value $f(\tau_1)$ at the time τ_1 and hold it at that value until $t = \tau_2$. Then we again abruptly change the forcing by an amount $f(\tau_2) - f(\tau_1)$ at the time τ_2, and so forth (see Figure 5.2.5). From the *linearity* of the problem, the response after the instant $t = \tau_n$ equals the sum

$$\begin{aligned} y(t) = f(0)a(t) &+ [f(\tau_1) - f(0)]a(t - \tau_1) + [f(\tau_2) - f(\tau_1)]a(t - \tau_2) \\ &+ \cdots + [f(\tau_n) - f(\tau_{n-1})]a(t - \tau_n). \end{aligned} \tag{5.2.55}$$

If we write $f(\tau_k) - f(\tau_{k-1}) = \Delta f_k$ and $\tau_k - \tau_{k-1} = \Delta\tau_k$, Equation 5.2.55 becomes

$$y(t) = f(0)a(t) + \sum_{k=1}^{n} a(t - \tau_k)\frac{\Delta f_k}{\Delta\tau_k}\,\Delta\tau_k. \tag{5.2.56}$$

Finally, proceeding to the limit as the number n of jumps becomes infinite, in such a manner that all jumps and intervals between successive jumps tend to zero, this sum has the limit

$$y(t) = f(0)a(t) + \int_0^t f'(\tau)a(t - \tau)\,d\tau. \tag{5.2.57}$$

Because the total response of the system equals the weighted sum (the weights being $a(t)$) of the forcing from the initial moment up to the time t, we refer to Equation 5.2.57 as the *superposition integral*, or *Duhamel's integral*,[8] named after the French mathematical

[8] Duhamel, J.-M.-C., 1833: Mémoire sur la méthode générale relative au mouvement de la chaleur dans les corps solides plongés dans des milieux dont la température varie avec le temps. *J. École Polytech.*, **22**, 20–77.

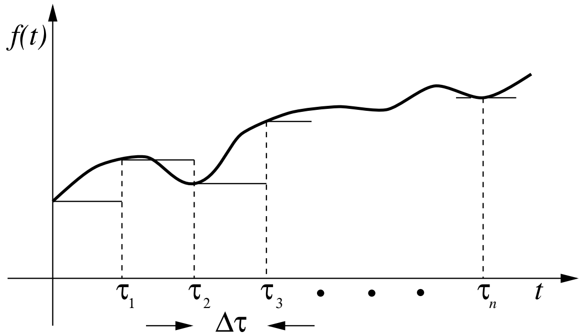


Figure 5.2.5: Diagram used in the derivation of Duhamel's integral.

physicist Jean-Marie-Constant Duhamel (1797–1872), who first derived it in conjunction with heat conduction.

We can also express Equation 5.2.57 in several different forms. Integration by parts yields

$$y(t) = f(t)a(0) + \int_0^t f(\tau)a'(t-\tau)\,d\tau = \frac{d}{dt}\left[\int_0^t f(\tau)a(t-\tau)\,d\tau\right]. \qquad \textbf{(5.2.58)}$$

• Example 5.2.7

Suppose that a system has the step response of $a(t) = A[1 - e^{-t/T}]$, where A and T are positive constants. Let us find the response if we force this system by $f(t) = kt$, where k is a constant.

From the superposition integral, Equation 5.2.57,

$$y(t) = 0 + \int_0^t kA[1 - e^{-(t-\tau)/T}]\,d\tau = kA[t - T(1 - e^{-t/T})]. \qquad \textbf{(5.2.59)}$$

□

Boundary-value problem

One of the purposes of this book is the solution of a wide class of nonhomogeneous ordinary differential equations of the form

$$\frac{d}{dx}\left[p(x)\frac{dy}{dx}\right] + s(x)y = -f(x), \qquad a \le x \le b, \qquad \textbf{(5.2.60)}$$

with

$$\alpha_1 y(a) + \alpha_2 y'(a) = 0, \qquad \beta_1 y(b) + \beta_2 y'(b) = 0. \qquad \textbf{(5.2.61)}$$

This is an example of a Sturm-Liouville-like equation

$$\frac{d}{dx}\left[p(x)\frac{dy}{dx}\right] + [q(x) + \lambda r(x)]y = -f(x), \qquad a \le x \le b, \qquad \textbf{(5.2.62)}$$

where λ is a parameter. Here we wish to develop the Green's function for this class of boundary-value problems.

We begin by determining the Green's function for the equation

$$\frac{d}{dx}\left[p(x)\frac{dg}{dx}\right] + s(x)g = -\delta(x-\xi), \tag{5.2.63}$$

subject to yet undetermined boundary conditions. We know that such a function exists for the special case $p(x) = 1$ and $s(x) = 0$, and we now show that this is *almost always* true in the general case. Presently we construct Green's functions by requiring that they satisfy the following conditions:

- $g(x|\xi)$ satisfies the *homogeneous* equation $f(x) = 0$ *except* at $x = \xi$,
- $g(x|\xi)$ satisfies certain *homogeneous* conditions, and
- $g(x|\xi)$ is continuous at $x = \xi$.

These homogeneous boundary conditions for a finite interval (a, b) will be

$$\alpha_1 g(a|\xi) + \alpha_2 g'(a|\xi) = 0, \qquad \beta_1 g(b|\xi) + \beta_2 g'(b|\xi) = 0, \tag{5.2.64}$$

where g' denotes the x derivative of $g(x|\xi)$ and neither a nor b equals ξ. The coefficients α_1 and α_2 cannot both be zero; this also holds for β_1 and β_2. These conditions include the commonly encountered Dirichlet, Neumann, and Robin boundary conditions.

What about the value of $g'(x|\xi)$ at $x = \xi$? Because $g(x|\xi)$ is a continuous function of x, Equation 5.2.63 dictates that there must be a discontinuity in $g'(x|\xi)$ at $x = \xi$. We now show that this discontinuity consists of a jump in the value $g'(x|\xi)$ at $x = \xi$. To prove this, we begin by integrating Equation 5.2.63 from $\xi - \epsilon$ to $\xi + \epsilon$, which yields

$$p(x)\frac{dg(x|\xi)}{dx}\bigg|_{\xi-\epsilon}^{\xi+\epsilon} + \int_{\xi-\epsilon}^{\xi+\epsilon} s(x)g(x|\xi)\,dx = -1. \tag{5.2.65}$$

Because $g(x|\xi)$ and $s(x)$ are both continuous at $x = \xi$,

$$\lim_{\epsilon\to 0}\int_{\xi-\epsilon}^{\xi+\epsilon} s(x)g(x|\xi)\,dx = 0. \tag{5.2.66}$$

Applying the limit $\epsilon \to 0$ to Equation 5.2.65, we have that

$$p(\xi)\left[\frac{dg(\xi^+|\xi)}{dx} - \frac{dg(\xi^-|\xi)}{dx}\right] = -1, \tag{5.2.67}$$

where ξ^+ and ξ^- denote points just above and below $x = \xi$, respectively. Consequently, our last requirement on $g(x|\xi)$ will be that

- dg/dx must have a jump discontinuity of magnitude $-1/p(\xi)$ at $x = \xi$.

Similar conditions hold for higher-order ordinary differential equations.[9]

Consider now the region $a \leq x < \xi$. Let $y_1(x)$ be a nontrivial solution of the *homogeneous* differential equation satisfying the boundary condition at $x = a$; then $\alpha_1 y_1(a) + \alpha_2 y_1'(a) = 0$. Because $g(x|\xi)$ must satisfy the same boundary condition, $\alpha_1 g(a|\xi) + \alpha_2 g'(a|\xi) = 0$. Since the set α_1, α_2 is nontrivial, then the Wronskian of y_1 and g must vanish at $x = a$ or $y_1(a)g'(a|\xi) - y_1'(a)g(a|\xi) = 0$. However, for $a \leq x < \xi$, both $y_1(x)$ and $g(x|\xi)$ satisfy the same differential equation, the homogeneous one. Therefore, their Wronskian is zero at all points and $g(x|\xi) = c_1 y_1(x)$ for $a \leq x < \xi$, where c_1 is an arbitrary constant. In a similar manner, if the nontrivial function $y_2(x)$ satisfies the homogeneous equation and the boundary conditions at $x = b$, then $g(x|\xi) = c_2 y_2(x)$ for $\xi < x \leq b$. The continuity condition of g and the jump discontinuity of g' at $x = \xi$ imply

$$c_1 y_1(\xi) - c_2 y_2(\xi) = 0, \qquad c_1 y_1'(\xi) - c_2 y_2'(\xi) = 1/p(\xi). \tag{5.2.68}$$

We can solve Equation 5.2.68 for c_1 and c_2 provided the Wronskian of y_1 and y_2 does not vanish at $x = \xi$, or

$$y_1(\xi)y_2'(\xi) - y_2(\xi)y_1'(\xi) \neq 0. \tag{5.2.69}$$

In other words, $y_1(x)$ must *not* be a multiple of $y_2(x)$. Is this always true? The answer is "generally yes." If the homogeneous equation admits no nontrivial solutions satisfying both boundary conditions at the same time,[10] then $y_1(x)$ and $y_2(x)$ must be linearly independent. On the other hand, if the homogeneous equation possesses a single solution, say $y_0(x)$, which also satisfies $\alpha_1 y_0(a) + \alpha_2 y_0'(a) = 0$ and $\beta_1 y_0(b) + \beta_2 y_0'(b) = 0$, then $y_1(x)$ will be a multiple of $y_0(x)$ and so is $y_2(x)$. Then they are multiples of each other and their Wronskian vanishes. This would occur, for example, if the differential equation is a Sturm-Liouville equation, λ equals the eigenvalue, and $y_0(x)$ is the corresponding eigenfunction. No Green's function exists in this case.

• Example 5.2.8

Consider the problem of finding the Green's function for $g'' + k^2 g = -\delta(x - \xi)$, $0 < x < L$, subject to the boundary conditions $g(0|\xi) = g(L|\xi) = 0$ with $k \neq 0$. The corresponding homogeneous equation is $y'' + k^2 y = 0$. Consequently, $g(x|\xi) = c_1 y_1(x) = c_1 \sin(kx)$ for $0 \leq x \leq \xi$, while $g(x|\xi) = c_2 y_2(x) = c_2 \sin[k(L - x)]$ for $\xi \leq x \leq L$.

Let us compute the Wronskian. For our particular problem,

$$W(x) = y_1(x)y_2'(x) - y_1'(x)y_2(x) \tag{5.2.70}$$

$$= -k\sin(kx)\cos[k(L-x)] - k\cos(kx)\sin[k(L-x)] \tag{5.2.71}$$

$$= -k\sin[k(x + L - x)] = -k\sin(kL), \tag{5.2.72}$$

and $W(\xi) = -k\sin(kL)$. Therefore, the Green's function will exist as long as $kL \neq n\pi$. If $kL = n\pi$, $y_1(x)$ and $y_2(x)$ are linearly *dependent* with $y_0(x) = c_3 \sin(n\pi x/L)$, the solution to the regular Sturm-Liouville problem $y'' + \lambda y = 0$, and $y(0) = y(L) = 0$. □

---

[9] Ince, E. L., 1956: *Ordinary Differential Equations*. Dover Publications, Inc. See Section 11.1.

[10] In the theory of differential equations, this system would be called *incompatible*: one that admits no solution, save $y = 0$, which is also continuous for all x in the interval (a, b) and satisfies the homogeneous boundary conditions.

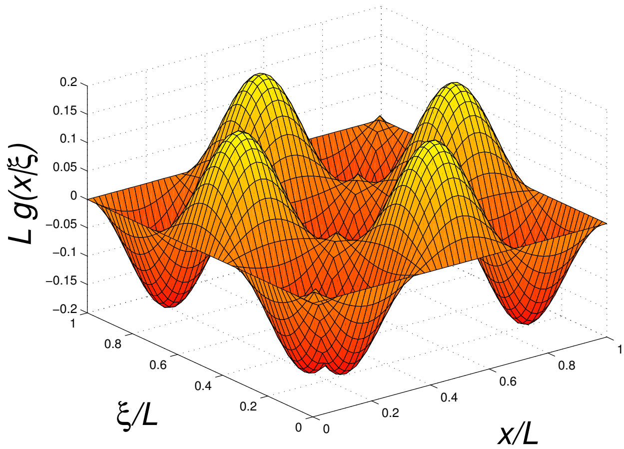


Figure 5.2.6: The Green's function, Equation 5.2.75, divided by L, as functions of x and ξ when $kL = 10$.

Let us now proceed to find $g(x|\xi)$ when it does exist. The system, Equation 5.2.68, has the unique solution

$$c_1 = -\frac{y_2(\xi)}{p(\xi)W(\xi)}, \quad \text{and} \quad c_2 = -\frac{y_1(\xi)}{p(\xi)W(\xi)}, \tag{5.2.73}$$

where $W(\xi)$ is the Wronskian of $y_1(x)$ and $y_2(x)$ at $x = \xi$. Therefore,

$$g(x|\xi) = -\frac{y_1(x_<)y_2(x_>)}{p(\xi)W(\xi)}. \tag{5.2.74}$$

Clearly $g(x|\xi)$ is symmetric in x and ξ. It is also unique. The proof of the uniqueness is as follows: We can always choose a different $y_1(x)$, but it will be a multiple of the "old" $y_1(x)$, and the Wronskian will be multiplied by the same factor, leaving $g(x|\xi)$ the same. This is also true if we modify $y_2(x)$ in a similar manner.

• **Example 5.2.9**

Let us find the Green's function for $g'' + k^2 g = -\delta(x - \xi)$, $0 < x < L$, subject to the boundary conditions $g(0|\xi) = g(L|\xi) = 0$. As we showed in the previous example, $y_1(x) = c_1 \sin(kx)$, $y_2(x) = c_2 \sin[k(L - x)]$, and $W(\xi) = -k\sin(kL)$. Substituting into Equation 5.2.74, we have that

$$g(x|\xi) = \frac{\sin(kx_<)\sin[k(L - x_>)]}{k\,\sin(kL)}, \tag{5.2.75}$$

where $x_< = \min(x, \xi)$ and $x_> = \max(x, \xi)$. Figure 5.2.6 illustrates Equation 5.2.75. □

So far, we showed that the Green's function for Equation 5.2.63 exists, is symmetric, and enjoys certain properties (see the material in the boxes after Equation 5.2.63 and Equation 5.2.67). But how does this help us solve Equation 5.2.63? We now prove that

$$y(x) = \int_a^b g(x|\xi)f(\xi)\,d\xi \tag{5.2.76}$$

is the solution to the nonhomogeneous differential equation, Equation 5.2.63, and the homogeneous boundary conditions, Equation 5.2.64.

We begin by noting that in Equation 5.2.76 x is a parameter while ξ is the dummy variable. As we perform the integration, we must switch from the form for $g(x|\xi)$ for $\xi \le x$ to the second form for $\xi \ge x$ when ξ equals x; thus,

$$y(x) = \int_a^x g(x|\xi) f(\xi)\, d\xi + \int_x^b g(x|\xi) f(\xi)\, d\xi. \tag{5.2.77}$$

Differentiation yields

$$\frac{d}{dx}\int_a^x g(x|\xi) f(\xi)\, d\xi = \int_a^x \frac{dg(x|\xi)}{dx} f(\xi)\, d\xi + g(x|x^-) f(x), \tag{5.2.78}$$

and

$$\frac{d}{dx}\int_x^b g(x|\xi) f(\xi)\, d\xi = \int_x^b \frac{dg(x|\xi)}{dx} f(\xi)\, d\xi - g(x|x^+) f(x). \tag{5.2.79}$$

Because $g(x|\xi)$ is continuous everywhere, we have that $g(x|x^+) = g(x|x^-)$ so that

$$\frac{dy}{dx} = \int_a^x \frac{dg(x|\xi)}{dx} f(\xi)\, d\xi + \int_x^b \frac{dg(x|\xi)}{dx} f(\xi)\, d\xi. \tag{5.2.80}$$

Differentiating once more gives

$$\frac{d^2y}{dx^2} = \int_a^x \frac{d^2g(x|\xi)}{dx^2} f(\xi)\, d\xi + \frac{dg(x|x^-)}{dx} f(x) + \int_x^b \frac{d^2g(x|\xi)}{dx^2} f(\xi)\, d\xi - \frac{dg(x|x^+)}{dx} f(x). \tag{5.2.81}$$

The second and fourth terms on the right side of Equation 5.2.81 will not cancel in this case; on the contrary,

$$\frac{dg(x|x^-)}{dx} - \frac{dg(x|x^+)}{dx} = -\frac{1}{p(x)}. \tag{5.2.82}$$

To show this, we note that the term $dg(x|x^-)/dx$ denotes a differentiation of $g(x|\xi)$ with respect to x using the $x > \xi$ form and then letting $\xi \to x$. Thus,

$$\frac{dg(x|x^-)}{dx} = -\lim_{\substack{\xi\to x\\ \xi<x}} \frac{y_2'(x) y_1(\xi)}{p(\xi)W(\xi)} = -\frac{y_2'(x) y_1(x)}{p(x)W(x)}, \tag{5.2.83}$$

while for $dg(x|x^+)/dx$ we use the $x < \xi$ form or

$$\frac{dg(x|x^+)}{dx} = -\lim_{\substack{\xi\to x\\ \xi>x}} \frac{y_1'(x) y_2(\xi)}{p(\xi)W(\xi)} = -\frac{y_1'(x) y_2(x)}{p(x)W(x)}. \tag{5.2.84}$$

Upon introducing these results into the differential equation

$$p(x)\frac{d^2y}{dx^2} + p'(x)\frac{dy}{dx} + s(x)y = -f(x), \tag{5.2.85}$$

we have

$$\begin{aligned} &\int_a^x [p(x)g''(x|\xi) + p'(x)g'(x|\xi) + s(x)g(x|\xi)] f(\xi)\, d\xi \\ &\quad + \int_x^b [p(x)g''(x|\xi) + p'(x)g'(x|\xi) + s(x)g(x|\xi)] f(\xi)\, d\xi - p(x)\frac{f(x)}{p(x)} = -f(x). \end{aligned} \tag{5.2.86}$$

Because

$$p(x)g''(x|\xi) + p'(x)g'(x|\xi) + s(x)g(x|\xi) = 0, \tag{5.2.87}$$

except for $x = \xi$, Equation 5.2.86, and thus Equation 5.2.63, is satisfied. Although Equation 5.2.87 does not hold at the point $x = \xi$, the results are still valid because that one point does not affect the values of the integrals. As for the boundary conditions,

$$y(a) = \int_a^b g(a|\xi)f(\xi)\,d\xi, \qquad y'(a) = \int_a^b \frac{dg(a|\xi)}{dx} f(\xi)\,d\xi, \tag{5.2.88}$$

and $\alpha_1 y(a) + \alpha_2 y'(a) = 0$ from Equation 5.2.64. A similar proof holds for $x = b$.

Finally, let us consider the solution for the nonhomogeneous boundary conditions $\alpha_1 y(a) + \alpha_2 y'(a) = \alpha$, and $\beta_1 y(b) + \beta_2 y'(b) = \beta$. The solution in this case is

$$\boxed{y(x) = \frac{\alpha y_2(x)}{\alpha_1 y_2(a) + \alpha_2 y_2'(a)} + \frac{\beta y_1(x)}{\beta_1 y_1(b) + \beta_2 y_1'(b)} + \int_a^b g(x|\xi)f(\xi)\,d\xi.} \tag{5.2.89}$$

A quick check shows that Equation 5.2.89 satisfies the differential equation and both nonhomogeneous boundary conditions.

Eigenfunction expansion

We just showed how Green's functions can be used to solve the nonhomogeneous linear differential equation. The next question is how do you find the Green's function? Here we present the most common method: *series expansion*. This is not surprising given its success in solving the Sturm-Liouville problem.

Consider the nonhomogeneous problem

$$y'' = -f(x), \qquad \text{with} \qquad y(0) = y(L) = 0. \tag{5.2.90}$$

The Green's function $g(x|\xi)$ must therefore satisfy

$$g'' = -\delta(x - \xi), \quad \text{with} \quad g(0|\xi) = g(L|\xi) = 0. \tag{5.2.91}$$

Because $g(x|\xi)$ vanishes at the ends of the interval $(0, L)$, this suggests that it can be expanded in a series of suitably chosen orthogonal functions such as, for instance, the Fourier sine series

$$g(x|\xi) = \sum_{n=1}^{\infty} G_n(\xi) \sin\left(\frac{n\pi x}{L}\right), \tag{5.2.92}$$

where the expansion coefficients G_n are dependent on the parameter ξ. Although we chose the orthogonal set of functions $\sin(n\pi x/L)$, we could have used other orthogonal functions as long as they vanish at the endpoints.

Because

$$g''(x|\xi) = \sum_{n=1}^{\infty} \left(-\frac{n^2\pi^2}{L^2}\right) G_n(\xi) \sin\left(\frac{n\pi x}{L}\right), \tag{5.2.93}$$

and

$$\delta(x-\xi) = \sum_{n=1}^{\infty} A_n(\xi) \sin\left(\frac{n\pi x}{L}\right), \tag{5.2.94}$$

where

$$A_n(\xi) = \frac{2}{L} \int_0^L \delta(x-\xi) \sin\left(\frac{n\pi x}{L}\right) dx = \frac{2}{L} \sin\left(\frac{n\pi\xi}{L}\right), \tag{5.2.95}$$

we have that

$$-\sum_{n=1}^{\infty} \left(\frac{n^2\pi^2}{L^2}\right) G_n(\xi) \sin\left(\frac{n\pi x}{L}\right) = -\frac{2}{L} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi\xi}{L}\right) \sin\left(\frac{n\pi x}{L}\right), \tag{5.2.96}$$

after substituting Equation 5.2.93 through Equation 5.2.95 into the differential equation, Equation 5.2.91. Since Equation 5.2.96 must hold for any arbitrary x,

$$\left(\frac{n^2\pi^2}{L^2}\right) G_n(\xi) = \frac{2}{L} \sin\left(\frac{n\pi\xi}{L}\right). \tag{5.2.97}$$

Thus, the Green's function is

$$g(x|\xi) = \frac{2L}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin\left(\frac{n\pi\xi}{L}\right) \sin\left(\frac{n\pi x}{L}\right). \tag{5.2.98}$$

How might we use Equation 5.2.98? We can use this series to construct the solution of the nonhomogeneous equation, Equation 5.2.90, via the formula

$$y(x) = \int_0^L g(x|\xi)\, f(\xi)\, d\xi. \tag{5.2.99}$$

This leads to

$$y(x) = \frac{2L}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin\left(\frac{n\pi x}{L}\right) \int_0^L f(\xi) \sin\left(\frac{n\pi\xi}{L}\right) d\xi, \tag{5.2.100}$$

or

$$y(x) = \frac{L^2}{\pi^2} \sum_{n=1}^{\infty} \frac{a_n}{n^2} \sin\left(\frac{n\pi x}{L}\right), \tag{5.2.101}$$

where a_n are the Fourier sine coefficients of $f(x)$.

• Example 5.2.10

Consider now the more complicated boundary-value problem

$$y'' + k^2 y = -f(x), \qquad \text{with} \qquad y(0) = y(L) = 0. \tag{5.2.102}$$

The Green's function $g(x|\xi)$ must now satisfy

$$g'' + k^2 g = -\delta(x-\xi), \quad \text{and} \quad g(0|\xi) = g(L|\xi) = 0. \tag{5.2.103}$$

Once again, we use the Fourier sine expansion

$$g(x|\xi) = \sum_{n=1}^{\infty} G_n(\xi) \sin\left(\frac{n\pi x}{L}\right). \qquad \textbf{(5.2.104)}$$

Direct substitution of Equation 5.2.104 and Equation 5.2.94 into Equation 5.2.103 and grouping by corresponding harmonics yields

$$-\frac{n^2\pi^2}{L^2} G_n(\xi) + k^2 G_n(\xi) = -\frac{2}{L}\sin\left(\frac{n\pi\xi}{L}\right), \qquad \textbf{(5.2.105)}$$

or

$$G_n(\xi) = \frac{2}{L}\,\frac{\sin(n\pi\xi/L)}{n^2\pi^2/L^2 - k^2}. \qquad \textbf{(5.2.106)}$$

Thus, the Green's function is

$$g(x|\xi) = \frac{2}{L}\sum_{n=1}^{\infty} \frac{\sin(n\pi\xi/L)\sin(n\pi x/L)}{n^2\pi^2/L^2 - k^2}. \qquad \textbf{(5.2.107)}$$

Examining Equation 5.2.107 more closely, we note that it enjoys the symmetry property that $g(x|\xi) = g(\xi|x)$. □

• **Example 5.2.11**

Let us find the series expansion for the Green's function for

$$xg'' + g' + \left(k^2 x - \frac{m^2}{x}\right) g = -\delta(x-\xi), \quad 0 < x < L, \qquad \textbf{(5.2.108)}$$

where $m \ge 0$ and is an integer. The boundary conditions are

$$\lim_{x\to 0} |g(x|\xi)| < \infty, \quad \text{and} \quad g(L|\xi) = 0. \qquad \textbf{(5.2.109)}$$

To find this series, consider the Fourier-Bessel series

$$g(x|\xi) = \sum_{n=1}^{\infty} G_n(\xi) J_m(k_{nm}x), \qquad \textbf{(5.2.110)}$$

where k_{nm} is the nth root of $J_m(k_{nm}L) = 0$. This series enjoys the advantage that it satisfies the boundary conditions and we will not have to introduce any homogeneous solutions so that $g(x|\xi)$ satisfies the boundary conditions.

Substituting Equation 5.2.110 into Equation 5.2.108 after we divide by x and using the Fourier-Bessel expansion for the delta function, we have that

$$(k^2 - k_{nm}^2) G_n(\xi) = -\frac{2k_{nm}^2 J_m(k_{nm}\xi)}{L^2[J_{m+1}(k_{nm}L)]^2} = -\frac{2J_m(k_{nm}\xi)}{L^2[J_m'(k_{nm}L)]^2}, \qquad \textbf{(5.2.111)}$$

so that

$$g(x|\xi) = \frac{2}{L^2}\sum_{n=1}^{\infty} \frac{J_m(k_{nm}\xi) J_m(k_{nm}x)}{(k_{nm}^2 - k^2)[J_m'(k_{nm}L)]^2}. \qquad \textbf{(5.2.112)}$$

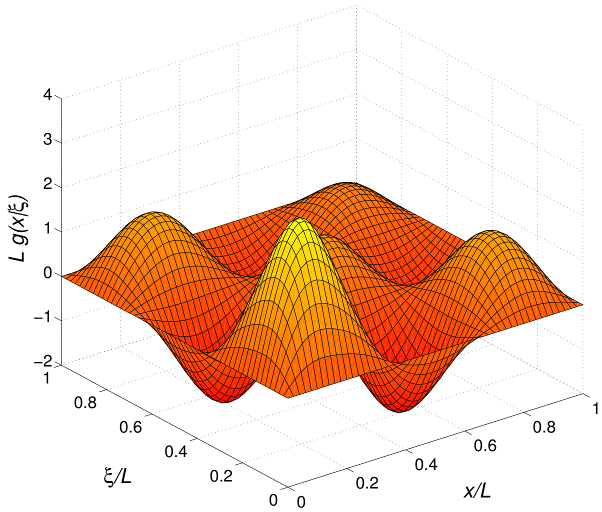


Figure 5.2.7: The Green's function, Equation 5.2.112, as functions of x/L and ξ/L when $kL = 10$ and $m = 1$.

Equation 5.2.112 is plotted in Figure 5.2.7. □

We summarize the expansion technique as follows: Suppose that we want to solve the differential equation

$$Ly(x) = -f(x), \tag{5.2.113}$$

with some condition $By(x) = 0$ along the boundary, where L now denotes the Sturm-Liouville differential operator

$$L = \frac{d}{dx}\left[p(x)\frac{d}{dx}\right] + [q(x) + \lambda r(x)], \tag{5.2.114}$$

and B is the boundary condition operator

$$B = \begin{cases} \alpha_1 + \alpha_2 \dfrac{d}{dx}, & \text{at } x = a, \\ \beta_1 + \beta_2 \dfrac{d}{dx}, & \text{at } x = b. \end{cases} \tag{5.2.115}$$

We begin by seeking a Green's function $g(x|\xi)$, which satisfies

$$Lg = -\delta(x - \xi), \qquad Bg = 0. \tag{5.2.116}$$

To find the Green's function, we utilize the set of eigenfunctions $\varphi_n(x)$ associated with the regular Sturm-Liouville problem

$$\frac{d}{dx}\left[p(x)\frac{d\varphi_n}{dx}\right] + [q(x) + \lambda_n r(x)]\varphi_n = 0, \tag{5.2.117}$$

where $\varphi_n(x)$ satisfies the same boundary conditions as $y(x)$. If g exists and if the set $\{\varphi_n\}$ is complete, then $g(x|\xi)$ can be represented by the series

$$g(x|\xi) = \sum_{n=1}^{\infty} G_n(\xi)\varphi_n(x). \tag{5.2.118}$$

Applying L to Equation 5.2.118,

$$Lg(x|\xi) = \sum_{n=1}^{\infty} G_n(\xi)L[\varphi_n(x)] = \sum_{n=1}^{\infty} G_n(\xi)(\lambda - \lambda_n)r(x)\varphi_n(x) = -\delta(x-\xi), \tag{5.2.119}$$

if λ does not equal any of the eigenvalues λ_n. Multiplying both sides of Equation 5.2.119 by $\varphi_m(x)$ and integrating over x,

$$\sum_{n=1}^{\infty} G_n(\xi)(\lambda - \lambda_n)\int_a^b r(x)\varphi_n(x)\varphi_m(x)\,dx = -\varphi_m(\xi). \tag{5.2.120}$$

If the eigenfunctions are *orthonormal*,

$$\int_a^b r(x)\varphi_n(x)\varphi_m(x)\,dx = \begin{cases} 1, & n = m, \\ 0, & n \neq m, \end{cases} \quad \text{and} \quad G_n(\xi) = \frac{\varphi_n(\xi)}{\lambda_n - \lambda}. \tag{5.2.121}$$

This leads directly to the *bilinear formula*:

$$\boxed{g(x|\xi) = \sum_{n=1}^{\infty} \frac{\varphi_n(\xi)\varphi_n(x)}{\lambda_n - \lambda},} \tag{5.2.122}$$

which permits us to write the Green's function at once if the eigenvalues and eigenfunctions of L are known.

Problems

For the following initial-value problems, find the transfer function, impulse response, Green's function, and step response. Assume that all of the necessary initial conditions are zero and $\tau > 0$. If you have MATLAB's control toolbox, use MATLAB to check your work.

1. $g' + kg = \delta(t-\tau)$
2. $g'' - 2g' - 3g = \delta(t-\tau)$
3. $g'' + 4g' + 3g = \delta(t-\tau)$
4. $g'' - 2g' + 5g = \delta(t-\tau)$
5. $g'' - 3g' + 2g = \delta(t-\tau)$
6. $g'' + 4g' + 4g = \delta(t-\tau)$
7. $g'' - 9g = \delta(t-\tau)$
8. $g'' + g = \delta(t-\tau)$
9. $g'' - g' = \delta(t-\tau)$

Find the Green's function and the corresponding bilinear expansion using eigenfunctions from the regular Sturm-Liouville problem $\varphi_n'' + k_n^2\varphi_n = 0$ for

$$g'' = -\delta(x-\xi), \qquad 0 < x, \xi < L,$$

which satisfy the following boundary conditions:

10. $g(0|\xi) - \alpha g'(0|\xi) = 0, \alpha \neq 0, -L,$ $\quad g(L|\xi) = 0,$

11. $g(0|\xi) - g'(0|\xi) = 0,$ $\quad g(L|\xi) - g'(L|\xi) = 0,$

12. $g(0|\xi) - g'(0|\xi) = 0,$ $\quad g(L|\xi) + g'(L|\xi) = 0.$

Find the Green's function[11] and the corresponding bilinear expansion using eigenfunctions from the regular Sturm-Liouville problem $\varphi_n'' + k_n^2\varphi_n = 0$ for

$$g'' - k^2 g = -\delta(x-\xi), \qquad 0 < x, \xi < L,$$

which satisfy the following boundary conditions:

13. $g(0|\xi) = 0,$ $\quad g(L|\xi) = 0,$

14. $g'(0|\xi) = 0,$ $\quad g'(L|\xi) = 0,$

15. $g(0|\xi) = 0,$ $\quad g(L|\xi) + g'(L|\xi) = 0,$

16. $g(0|\xi) = 0,$ $\quad g(L|\xi) - g'(L|\xi) = 0,$

17. $a\,g(0|\xi) + g'(0|\xi) = 0,$ $\quad g'(L|\xi) = 0,$

18. $g(0|\xi) + g'(0|\xi) = 0,$ $\quad g(L|\xi) - g'(L|\xi) = 0.$

5.3 JOINT TRANSFORM METHOD

In the previous section an important method for finding Green's function involved either Laplace or Fourier transforms. In the following sections we wish to find Green's functions for partial differential equations. Again, transform methods play an important role. We will always use the Laplace transform to eliminate the temporal dependence. However, for the spatial dimension we will use either a Fourier series or Fourier transform. Our choice will be dictated by the domain: If it reaches to infinity, then we will employ Fourier transforms. On the other hand, a domain of finite length calls for an eigenfunction expansion. The following two examples illustrate our solution technique for domains of infinite and finite extent.

• Example 5.3.1: One-dimensional Klein-Gordon equation

The Klein-Gordon equation is a form of the wave equation that arose in particle physics as the relativistic scalar wave equation describing particles with nonzero rest mass. In this example, we find its Green's function when there is only one spatial dimension:

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c^2}\left(\frac{\partial^2 g}{\partial t^2} + a^2 g\right) = -\delta(x-\xi)\delta(t-\tau), \tag{5.3.1}$$

---

[11] Problem 18 was used by Chakrabarti, A., and T. Sahoo, 1996: Reflection of water waves by a nearly vertical porous wall. *J. Austral. Math. Soc., Ser. B*, **37**, 417–429.

where $-\infty < x, \xi < \infty$, $0 < t, \tau$, c is a real, positive constant (the wave speed), and a is a real, nonnegative constant. The corresponding boundary conditions are

$$\lim_{|x|\to\infty} g(x,t|\xi,\tau) \to 0, \tag{5.3.2}$$

and the initial conditions are

$$g(x,0|\xi,\tau) = g_t(x,0|\xi,\tau) = 0. \tag{5.3.3}$$

We begin by taking the Laplace transform of Equation 5.3.1 and find that

$$\frac{d^2G}{dx^2} - \left(\frac{s^2+a^2}{c^2}\right) G = -\delta(x-\xi)e^{-s\tau}. \tag{5.3.4}$$

Applying Fourier transforms to Equation 5.3.4, we obtain

$$G(x,s|\xi,\tau) = \frac{c^2}{2\pi} e^{-s\tau} \int_{-\infty}^{\infty} \frac{e^{ik(x-\xi)}}{s^2+a^2+k^2c^2}\,dk = \frac{c^2}{\pi} e^{-s\tau} \int_0^{\infty} \frac{\cos[k(x-\xi)]}{s^2+a^2+k^2c^2}\,dk. \tag{5.3.5}$$

Inverting the Laplace transform and employing the second shifting theorem,

$$g(x,t|\xi,\tau) = \frac{c^2}{\pi} H(t-\tau) \int_0^{\infty} \frac{\sin\left[(t-\tau)\sqrt{a^2+k^2c^2}\,\right] \cos[k(x-\xi)]}{\sqrt{a^2+k^2c^2}}\,dk. \tag{5.3.6}$$

Equation 5.3.6 represents a superposition of homogeneous solutions (normal modes) to Equation 5.3.1. An intriguing aspect of Equation 5.3.6 is that this solution occurs everywhere after $t > \tau$. If $|x-\xi| > c(t-\tau)$, these wave solutions destructively interfere so that we have zero there while they constructively interfere at those times and places where the physical waves are present.

Applying integral tables to Equation 5.3.6, the final result is

$$\boxed{g(x,t|\xi,\tau) = \frac{c}{2} J_0\left[a\sqrt{(t-\tau)^2-(x-\xi)^2/c^2}\,\right] H[c(t-\tau)-|x-\xi|].} \tag{5.3.7}$$

Figure 5.3.1 illustrates this Green's function. Thus, the Green's function for the Klein-Gordon equation yields waves that propagate to the right and left from $x = 0$ with the wave front located at $x = \pm ct$. At a given point, after the passage of the wave front, the solution vibrates with an ever-decreasing amplitude and at a frequency that approaches a, the so-called *cutoff frequency*, at $t \to \infty$.

Why is a called a cutoff frequency? From Equation 5.3.5, we see that, although the spectral representation includes all of the wavenumbers k running from $-\infty$ to ∞, the frequency $\omega = \sqrt{c^2k^2+a^2}$ is restricted to the range $\omega \ge a$ from Equation 5.3.6. Thus, a is the lowest possible frequency that a wave solution to the Klein-Gordon equation may have for a real value of k. □

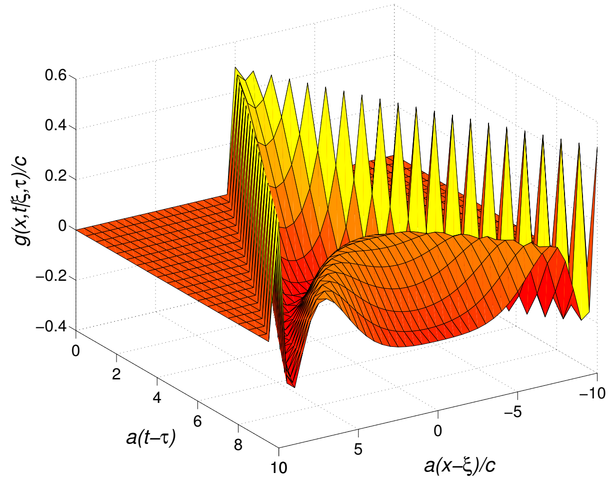


Figure 5.3.1: The free-space Green's function $g(x,t|\xi,\tau)/c$ for the one-dimensional Klein-Gordon equation at different distances $a(x-\xi)/c$ and times $a(t-\tau)$.

• **Example 5.3.2: One-dimensional wave equation on the interval** $0 < x < L$

One of the classic problems of mathematical physics involves finding the displacement of a taut string between two supports when an external force is applied. The governing equation is

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = f(x,t), \qquad 0 < x < L, \quad 0 < t, \tag{5.3.8}$$

where c is the constant phase speed.

In this example, we find the Green's function for this problem by considering the following problem:

$$\frac{\partial^2 g}{\partial t^2} - c^2 \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \quad 0 < x,\xi < L, \quad 0 < t,\tau, \tag{5.3.9}$$

with the boundary conditions

$$\alpha_1 g(0,t|\xi,\tau) + \beta_1 g_x(0,t|\xi,\tau) = 0, \quad 0 < t, \tag{5.3.10}$$

and

$$\alpha_2 g(L,t|\xi,\tau) + \beta_2 g_x(L,t|\xi,\tau) = 0, \quad 0 < t, \tag{5.3.11}$$

and the initial conditions

$$g(x,0|\xi,\tau) = g_t(x,0|\xi,\tau) = 0, \quad 0 < x < L. \tag{5.3.12}$$

We start by taking the Laplace transform of Equation 5.3.9 and find that

$$\frac{d^2 G}{dx^2} - \frac{s^2}{c^2} G = -\frac{\delta(x-\xi)}{c^2} e^{-s\tau}, \qquad 0 < x < L, \tag{5.3.13}$$

with

$$\alpha_1 G(0,s|\xi,\tau) + \beta_1 G'(0,s|\xi,\tau) = 0, \tag{5.3.14}$$

and

$$\alpha_2 G(L,s|\xi,\tau) + \beta_2 G'(L,s|\xi,\tau) = 0. \tag{5.3.15}$$

Problems similar to Equation 5.3.13 through Equation 5.3.15 were considered in the previous section. There, solutions were developed in terms of an eigenfunction expansion. Applying the same technique here,

$$G(x,s|\xi,\tau) = e^{-s\tau} \sum_{n=1}^{\infty} \frac{\varphi_n(\xi)\varphi_n(x)}{s^2 + c^2 k_n^2}, \tag{5.3.16}$$

where $\varphi_n(x)$ is the nth *orthonormal* eigenfunction to the regular Sturm-Liouville problem

$$\varphi''(x) + k^2\varphi(x) = 0, \qquad 0 < x < L, \tag{5.3.17}$$

subject to the boundary conditions

$$\alpha_1\varphi(0) + \beta_1\varphi'(0) = 0, \tag{5.3.18}$$

and

$$\alpha_2\varphi(L) + \beta_2\varphi'(L) = 0. \tag{5.3.19}$$

Taking the inverse of Equation 5.3.16, we have that the Green's function is

$$\boxed{g(x,t|\xi,\tau) = \left\{ \sum_{n=1}^{\infty} \varphi_n(\xi)\varphi_n(x) \frac{\sin[k_n c(t-\tau)]}{k_n c} \right\} H(t-\tau).} \tag{5.3.20}$$

Let us illustrate our results to find the Green's function for

$$\frac{\partial^2 g}{\partial t^2} - c^2 \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \tag{5.3.21}$$

with the boundary conditions

$$g(0,t|\xi,\tau) = g(L,t|\xi,\tau) = 0, \qquad 0 < t, \tag{5.3.22}$$

and the initial conditions

$$g(x,0|\xi,\tau) = g_t(x,0|\xi,\tau) = 0, \qquad 0 < x < L. \tag{5.3.23}$$

For this example, the Sturm-Liouville problem is

$$\varphi''(x) + k^2\varphi(x) = 0, \qquad 0 < x < L, \tag{5.3.24}$$

with the boundary conditions $\varphi(0) = \varphi(L) = 0$. The nth *orthonormal* eigenfunction for this problem is

$$\varphi_n(x) = \sqrt{\frac{2}{L}}\, \sin\left(\frac{n\pi x}{L}\right). \tag{5.3.25}$$

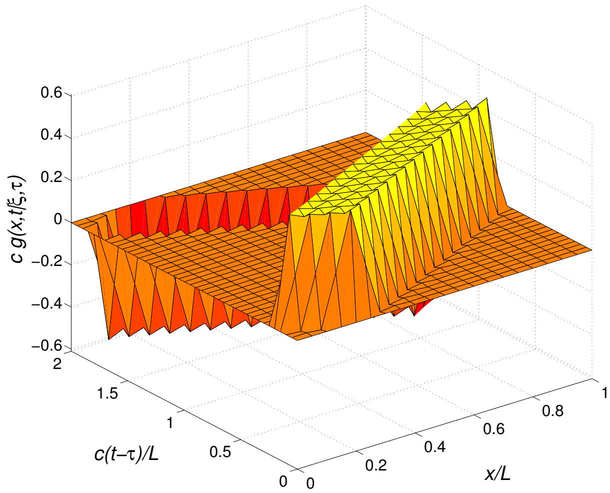


Figure 5.3.2: The Green's function $cg(x,t|\xi,\tau)$ given by Equation 5.3.26 for the one-dimensional wave equation over the interval $0 < x < L$ as a function of location x/L and time $c(t-\tau)/L$ with $\xi/L = 0.2$. The boundary conditions are $g(0,t|\xi,\tau) = g(L,t|\xi,\tau) = 0$.

Consequently, from Equation 5.3.20, the Green's function is

$$g(x,t|\xi,\tau) = \frac{2}{\pi c}\left\{\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi\xi}{L}\right)\sin\left(\frac{n\pi x}{L}\right)\sin\left[\frac{n\pi c(t-\tau)}{L}\right]\right\}H(t-\tau). \tag{5.3.26}$$

See Figure 5.3.2.

5.4 WAVE EQUATION

In Section 5.2, we showed how Green's functions could be used to solve initial- and boundary-value problems involving ordinary differential equations. When we approach partial differential equations, similar considerations hold, although the complexity increases. In the next three sections, we work through the classic groupings of the wave, heat, and Helmholtz's equations in one spatial dimension. All of these results can be generalized to three dimensions.

Of these three groups, we start with the wave equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = -q(x,t), \tag{5.4.1}$$

where t denotes time, x is the position, c is the phase velocity of the wave, and $q(x,t)$ is the source density. In addition to Equation 5.4.1 it is necessary to state boundary and initial conditions to obtain a unique solution. The condition on the boundary can be either Dirichlet or Neumann or a linear combination of both (Robin condition). The conditions in time must be Cauchy, that is, we must specify the value of $u(x,t)$ and its time derivative at $t = t_0$ for each point of the region under consideration.

We begin by proving that we can express the solution to Equation 5.4.1 in terms of boundary conditions, initial conditions, and the Green's function, which is found by solving

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 g}{\partial t^2} = -\delta(x-\xi)\delta(t-\tau), \tag{5.4.2}$$

where ξ denotes the position of a source that is excited at $t = \tau$. Equation 5.4.2 expresses the effect of an impulse as it propagates from $x = \xi$ as time increases from $t = \tau$. For $t < \tau$, causality requires that $g(x,t|\xi,\tau) = g_t(x,t|\xi,\tau) = 0$ if the impulse is the sole source of the disturbance. We also require that g satisfies the homogeneous form of the boundary condition satisfied by u.

Our derivation starts with the equations

$$\frac{\partial^2 u(\xi,\tau)}{\partial \xi^2} - \frac{1}{c^2}\frac{\partial^2 u(\xi,\tau)}{\partial \tau^2} = -q(\xi,\tau), \tag{5.4.3}$$

and

$$\frac{\partial^2 g(x,t|\xi,\tau)}{\partial \xi^2} - \frac{1}{c^2}\frac{\partial^2 g(x,t|\xi,\tau)}{\partial \tau^2} = -\delta(x-\xi)\delta(t-\tau), \tag{5.4.4}$$

where we obtain Equation 5.4.4 from a combination of Equation 5.4.2 plus reciprocity, namely $g(x,t|\xi,\tau) = g(\xi,-\tau|x,-t)$. Next we multiply Equation 5.4.3 by $g(x,t|\xi,\tau)$ and Equation 5.4.4 by $u(\xi,\tau)$ and subtract. Integrating over ξ from a to b, where a and b are the endpoints of the spatial domain, and over τ from 0 to t^+, where t^+ denotes a time slightly later than t so that we avoid ending the integration exactly at the peak of the delta function, we obtain

$$\begin{aligned}\int_0^{t^+}\int_a^b \biggl\{ g(x,t|\xi,\tau)\frac{\partial^2 u(\xi,\tau)}{\partial \xi^2} - u(\xi,\tau)\frac{\partial^2 g(x,t|\xi,\tau)}{\partial \xi^2} \\ + \frac{1}{c^2}\left[u(\xi,\tau)\frac{\partial^2 g(x,t|\xi,\tau)}{\partial \tau^2} - g(x,t|\xi,\tau)\frac{\partial^2 u(\xi,\tau)}{\partial \tau^2}\right]\biggr\}\,d\xi\,d\tau \\ = u(x,t) - \int_0^{t^+}\int_a^b q(\xi,\tau)\; g(x,t|\xi,\tau)\,d\xi\,d\tau.\end{aligned} \tag{5.4.5}$$

Because

$$\begin{aligned} g(x,t|\xi,\tau)\frac{\partial^2 u(\xi,\tau)}{\partial \xi^2} - u(\xi,\tau)\frac{\partial^2 g(x,t|\xi,\tau)}{\partial \xi^2} \\ = \frac{\partial}{\partial \xi}\left[g(x,t|\xi,\tau)\frac{\partial u(\xi,\tau)}{\partial \xi}\right] - \frac{\partial}{\partial \xi}\left[u(\xi,\tau)\frac{\partial g(x,t|\xi,\tau)}{\partial \xi}\right],\end{aligned} \tag{5.4.6}$$

and

$$\begin{aligned} g(x,t|\xi,\tau)\frac{\partial^2 u(\xi,\tau)}{\partial \tau^2} - u(\xi,\tau)\frac{\partial^2 g(x,t|\xi,\tau)}{\partial \tau^2} \\ = \frac{\partial}{\partial \tau}\left[g(x,t|\xi,\tau)\frac{\partial u(\xi,\tau)}{\partial \tau}\right] - \frac{\partial}{\partial \tau}\left[u(\xi,\tau)\frac{\partial g(x,t|\xi,\tau)}{\partial \tau}\right],\end{aligned} \tag{5.4.7}$$

we find that

$$\begin{aligned}\int_0^{t^+}\left[g(x,t|\xi,\tau)\frac{\partial u(\xi,\tau)}{\partial \xi} - u(\xi,\tau)\frac{\partial g(x,t|\xi,\tau)}{\partial \xi}\right]_{\xi=a}^{\xi=b} d\tau \\ + \frac{1}{c^2}\int_a^b \left[u(\xi,\tau)\frac{\partial g(x,t|\xi,\tau)}{\partial \tau} - g(x,t|\xi,\tau)\frac{\partial u(\xi,\tau)}{\partial \tau}\right]_{\tau=0}^{\tau=t^+} d\xi \\ + \int_0^{t^+}\int_a^b q(\xi,\tau)\; g(x,t|\xi,\tau)\,d\xi\,d\tau = u(x,t).\end{aligned} \tag{5.4.8}$$

The integrand in the first integral is specified by the boundary conditions. In the second integral, the integrand vanishes at $t = t^+$ from the initial conditions on $g(x,t|\xi,\tau)$. The limit at $t = 0$ is determined by the initial conditions. Hence,

$$\begin{aligned} u(x,t) &= \int_0^{t^+} \int_a^b q(\xi,\tau) g(x,t|\xi,\tau)\, d\xi\, d\tau \\ &+ \int_0^{t^+} \left[g(x,t|\xi,\tau) \frac{\partial u(\xi,\tau)}{\partial \xi} - u(\xi,\tau) \frac{\partial g(x,t|\xi,\tau)}{\partial \xi} \right]_{\xi=a}^{\xi=b} d\tau \\ &- \frac{1}{c^2} \int_a^b \left[u(\xi,0) \frac{\partial g(x,t|\xi,0)}{\partial \tau} - g(x,t|\xi,0) \frac{\partial u(\xi,0)}{\partial \tau} \right] d\xi. \end{aligned} \tag{5.4.9}$$

Equation 5.4.9 gives the complete solution of the nonhomogeneous problem. The first two integrals on the right side of this equation represent the effect of the source and the boundary conditions, respectively. The last term involves the initial conditions; it can be interpreted as asking what sort of source is needed so that the function $u(x,t)$ starts in the desired manner.

• Example 5.4.1

Let us apply the Green's function technique to solve

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad 0 < t, \tag{5.4.10}$$

subject to the boundary conditions $u(0,t) = 0$ and $u(1,t) = t$, $0 < t$, and the initial conditions $u(x,0) = x$ and $u_t(x,0) = 0$, $0 < x < 1$.

Because there is no source term and $c = 1$, Equation 5.4.9 becomes

$$\begin{aligned} u(x,t) &= \int_0^t \left[g(x,t|1,\tau) u_\xi(1,\tau) - u(1,\tau) g_\xi(x,t|1,\tau) \right] d\tau \\ &- \int_0^t \left[g(x,t|0,\tau) u_\xi(0,\tau) - u(0,\tau) g_\xi(x,t|0,\tau) \right] d\tau \\ &- \int_0^1 \left[u(\xi,0) g_\tau(x,t|\xi,0) - g(x,t|\xi,0) u_\xi(\xi,0) \right] d\xi. \end{aligned} \tag{5.4.11}$$

Therefore we must first compute the Green's function for this problem. However, we have already done this in Example 5.3.2 and it is given by Equation 5.3.26 with $c = L = 1$. Next, we note that $g(x,t|1,\tau) = g(x,t|0,\tau) = 0$ and $u(0,\tau) = u_\tau(\xi,0) = 0$. Consequently, Equation 5.4.11 reduces to only two nonvanishing integrals:

$$u(x,t) = -\int_0^t u(1,\tau) g_\xi(x,t|1,\tau)\, d\tau - \int_0^1 u(\xi,0) g_\tau(x,t|\xi,0)\, d\xi. \tag{5.4.12}$$

If we now substitute for $g(x,t|\xi,\tau)$ and reverse the order of integration and summation,

$$\begin{aligned} \int_0^t u(1,\tau) g_\xi(x,t|1,\tau)\, d\tau &= 2 \sum_{n=1}^\infty (-1)^n \sin(n\pi x) \int_0^t \tau \sin[n\pi(t-\tau)]\, d\tau \\ &= 2t \sum_{n=1}^\infty (-1)^n \sin(n\pi x) \int_0^t \sin[n\pi(t-\tau)]\, d(t-\tau) \end{aligned} \tag{5.4.13}$$

$$-2\sum_{n=1}^{\infty}(-1)^n \sin(n\pi x)\int_0^t (t-\tau)\sin[n\pi(t-\tau)]\,d(t-\tau) \qquad \textbf{(5.4.14)}$$

$$\int_0^t u(1,\tau)g_\xi(x,t|1,\tau)\,d\tau = -2t\sum_{n=1}^{\infty}(-1)^n \sin(n\pi x)\left.\frac{\cos[n\pi(t-\tau)]}{n\pi}\right|_0^t \qquad \textbf{(5.4.15)}$$

$$-2\sum_{n=1}^{\infty}(-1)^n \sin(n\pi x)\left.\left\{\frac{\sin[n\pi(t-\tau)]}{n^2\pi^2}-(t-\tau)\frac{\cos[n\pi(t-\tau)]}{n\pi}\right\}\right|_0^t$$

$$= -\frac{2t}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(n\pi x)+\frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\sin(n\pi x)\sin(n\pi t), \qquad \textbf{(5.4.16)}$$

and

$$\int_0^1 u(\xi,0)g_\tau(x,t|\xi,0)\,d\xi = -2\sum_{n=1}^{\infty}\sin(n\pi x)\cos(n\pi t)\int_0^1 \xi\sin(n\pi\xi)\,d\xi \qquad \textbf{(5.4.17)}$$

$$= -2\sum_{n=1}^{\infty}\sin(n\pi x)\cos(n\pi t)\left.\left[\frac{\sin(n\pi\xi)}{n^2\pi^2}-\frac{\xi\,\cos(n\pi\xi)}{n\pi}\right]\right|_0^1 \qquad \textbf{(5.4.18)}$$

$$= \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(n\pi x)\cos(n\pi t). \qquad \textbf{(5.4.19)}$$

Substituting Equation 5.4.16 and Equation 5.4.19 into Equation 5.4.12, we finally obtain

$$u(x,t) = -\frac{2t}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(n\pi x)-\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin(n\pi x)\cos(n\pi t)$$
$$+\frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\sin(n\pi x)\sin(n\pi t). \qquad \textbf{(5.4.20)}$$

The first summation in Equation 5.4.20 is the Fourier sine expansion for $f(x)=x$ over the interval $0<x<1$. Indeed, a quick check shows that the particular solution $u_p(x,t)=xt$ satisfies the partial differential equation and boundary conditions. The remaining two summations are necessary so that $u(x,0)=x$ and $u_t(x,0)=0$. □

To apply Equation 5.4.9 to other problems, we must now find the Green's function for a specific domain. In the following examples we illustrate how this is done using the joint transform method introduced in the previous section. Note that both examples given there were for the wave equation.

• Example 5.4.2: One-dimensional wave equation in an unlimited domain

The simplest possible example of Green's functions for the wave equation is the one-dimensional vibrating string problem.[12] In this problem the Green's function is given by the equation

$$\frac{\partial^2 g}{\partial t^2}-c^2\frac{\partial^2 g}{\partial x^2}=c^2\delta(x-\xi)\delta(t-\tau), \qquad \textbf{(5.4.21)}$$

[12] See also Graff, K. F., 1991: *Wave Motion in Elastic Solids*. Dover Publications, Inc., Section 1.1.8.

where $-\infty < x, \xi < \infty$, and $0 < t, \tau$. If the initial conditions equal zero, the Laplace transform of Equation 5.4.21 is

$$\frac{d^2G}{dx^2} - \frac{s^2}{c^2}G = -\delta(x-\xi)e^{-s\tau}, \tag{5.4.22}$$

where $G(x,s|\xi,\tau)$ is the Laplace transform of $g(x,t|\xi,\tau)$. To solve Equation 5.4.22 we take its Fourier transform and obtain the algebraic equation

$$\overline{G}(k,s|\xi,\tau) = \frac{\exp(-ik\xi - s\tau)}{k^2 + s^2/c^2}. \tag{5.4.23}$$

Having found the joint Laplace-Fourier transform of $g(x,t|\xi,\tau)$, we must work our way back to the Green's function. From the definition of the Fourier transform, we have that

$$G(x,s|\xi,\tau) = \frac{e^{-s\tau}}{2\pi}\int_{-\infty}^{\infty} \frac{e^{ik(x-\xi)}}{k^2 + s^2/c^2}\,dk. \tag{5.4.24}$$

To evaluate the Fourier-type integral, Equation 5.4.24, we apply the residue theorem. See Section 2.1. Performing the calculation,

$$G(x,s|\xi,\tau) = \frac{c\ \exp(-s\tau - s|x-\xi|/c)}{2s}. \tag{5.4.25}$$

Finally, taking the inverse Laplace transform of Equation 5.4.25,

$$g(x,t|\xi,\tau) = \frac{c}{2}H(t-\tau-|x-\xi|/c)\,, \tag{5.4.26}$$

or

$$\boxed{g(x,t|\xi,\tau) = \frac{c}{2}H[c(t-\tau)+(x-\xi)]\,H[c(t-\tau)-(x-\xi)]\,.} \tag{5.4.27}$$

We can use Equation 5.4.26 and the *method of images* to obtain the Green's function for

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 g}{\partial t^2} = \delta(x-\xi)\delta(t-\tau), \qquad 0 < x, t, \xi, \tau, \tag{5.4.28}$$

subject to the boundary condition $g(0,t|\xi,\tau) = 0$.

We begin by noting that the free-space Green's function,[13] Equation 5.4.26, is the particular solution to Equation 5.4.28. Therefore, we need only find a homogeneous solution $f(x,t|\xi,\tau)$ so that

$$g(x,t|\xi,\tau) = \frac{c}{2}H(t-\tau-|x-\xi|/c) + f(x,t|\xi,\tau) \tag{5.4.29}$$

[13] In electromagnetic theory, a free-space Green's function is the particular solution of the differential equation valid over a domain of infinite extent, where the Green's function remains bounded as we approach infinity, or satisfies a radiation condition there.

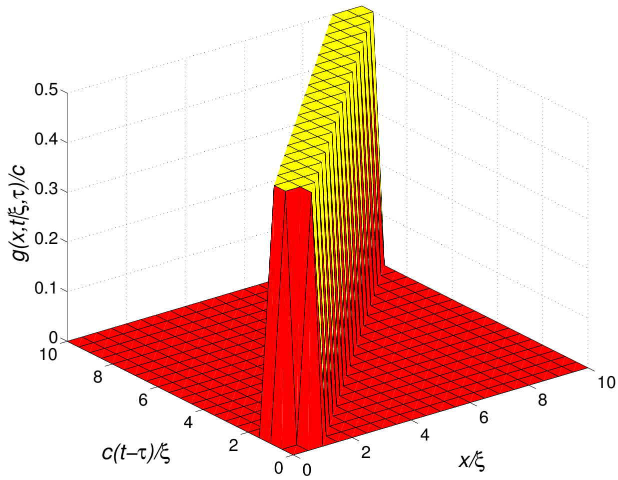


Figure 5.4.1: The Green's function $g(x,t|\xi,\tau)/c$ given by Equation 5.4.30 for the one-dimensional wave equation for $x > 0$ at different distances x/ξ and times $c(t-\tau)$ subject to the boundary condition $g(0,t|\xi,\tau) = 0$.

satisfies the boundary condition at $x = 0$.

To find $f(x,t|\xi,\tau)$, let us introduce a source at $x = -\xi$ at $t = \tau$. The corresponding free-space Green's function is $H(t - \tau - |x + \xi|/c)$. If, along the boundary $x = 0$ for any time t, this Green's function destructively interferes with the free-space Green's function associated with the source at $x = \xi$, then we have our solution. This will occur if our new source has a negative sign, resulting in the combined Green's function

$$g(x,t|\xi,\tau) = \frac{c}{2}\left[H(t-\tau-|x-\xi|/c) - H(t-\tau-|x+\xi|/c)\right]. \tag{5.4.30}$$

See Figure 5.4.1. Because Equation 5.4.30 satisfies the boundary condition, we need no further sources.

In a similar manner, we can use Equation 5.4.26 and the method of images to find the Green's function for

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 g}{\partial t^2} = \delta(x-\xi)\delta(t-\tau), \quad 0 < x, t, \xi, \tau, \tag{5.4.31}$$

subject to the boundary condition $g_x(0,t|\xi,\tau) = 0$.

We begin by examining the related problem

$$\frac{\partial^2 g}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 g}{\partial t^2} = \delta(x-\xi)\delta(t-\tau) + \delta(x+\xi)\delta(t-\tau), \tag{5.4.32}$$

where $-\infty < x, \xi < \infty$, and $0 < t, \tau$. In this particular case, we have chosen an image that is the mirror reflection of $\delta(x-\xi)$. This was dictated by the fact that the Green's function must be an even function of x along $x = 0$ for any time t. In line with this argument,

$$g(x,t|\xi,\tau) = \frac{c}{2}\left[H(t-\tau-|x-\xi|/c) + H(t-\tau-|x+\xi|/c)\right]. \tag{5.4.33}$$

$\square$

- **Example 5.4.3: Equation of telegraphy**

When the vibrating string problem includes the effect of air resistance, Equation 5.4.21 becomes

$$\frac{\partial^2 g}{\partial t^2} + 2\gamma\frac{\partial g}{\partial t} - c^2\frac{\partial^2 g}{\partial x^2} = c^2\delta(x-\xi)\delta(t-\tau), \tag{5.4.34}$$

where $-\infty < x, \xi < \infty$, and $0 < t, \tau$, with the boundary conditions

$$\lim_{|x|\to\infty} g(x,t|\xi,\tau) \to 0 \tag{5.4.35}$$

and the initial conditions

$$g(x,0|\xi,\tau) = g_t(x,0|\xi,\tau) = 0. \tag{5.4.36}$$

Let us find the Green's function.

Our analysis begins by introducing an intermediate dependent variable $w(x,t|\xi,\tau)$, where $g(x,t|\xi,\tau) = e^{-\gamma t}w(x,t|\xi,\tau)$. Substituting for $g(x,t|\xi,\tau)$, we now have

$$\frac{\partial^2 w}{\partial t^2} - \gamma^2 w - c^2\frac{\partial^2 w}{\partial x^2} = c^2\delta(x-\xi)\delta(t-\tau)e^{\gamma\tau}. \tag{5.4.37}$$

Taking the Laplace transform of Equation 5.4.37, we obtain

$$\frac{d^2W}{dx^2} - \left(\frac{s^2-\gamma^2}{c^2}\right)W = -\delta(x-\xi)e^{\gamma\tau - s\tau}. \tag{5.4.38}$$

Using Fourier transforms as in Example 5.3.1, the solution to Equation 5.4.38 is

$$W(x,s|\xi,\tau) = \frac{\exp[-|x-\xi|\sqrt{(s^2-\gamma^2)/c^2} + \gamma\tau - s\tau]}{2\sqrt{(s^2-\gamma^2)/c^2}}. \tag{5.4.39}$$

Employing tables to invert the Laplace transform and the second shifting theorem, we have that

$$w(x,t|\xi,\tau) = \frac{c}{2}e^{\gamma\tau}I_0\left[\gamma\sqrt{(t-\tau)^2 - (x-\xi)^2/c^2}\,\right]H[c(t-\tau) - |x-\xi|], \tag{5.4.40}$$

or

$$\boxed{g(x,t|\xi,\tau) = \frac{c}{2}e^{-\gamma(t-\tau)}I_0\left[\gamma\sqrt{(t-\tau)^2 - (x-\xi)^2/c^2}\,\right]H[c(t-\tau) - |x-\xi|].} \tag{5.4.41}$$

Figure 5.4.2 illustrates Equation 5.4.41 when $\gamma = 1$. □

- **Example 5.4.4**

Let us solve[14] the one-dimensional wave equation on an infinite domain:

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = \cos(\omega t)\delta[x - X(t)], \tag{5.4.42}$$

---

[14] See Knowles, J. K., 1968: Propagation of one-dimensional waves from a source in random motion. *J. Acoust. Soc. Am.*, **43**, 948–957.

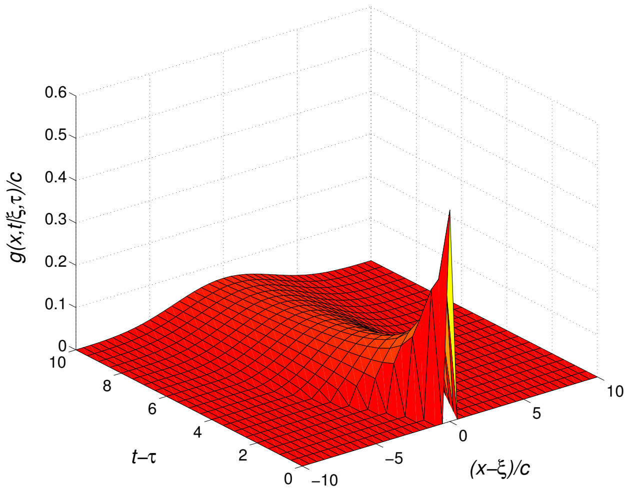


Figure 5.4.2: The free-space Green's function $g(x,t|\xi,\tau)/c$ for the one-dimensional equation of telegraphy with $\gamma = 1$ at different distances $(x-\xi)/c$ and times $t-\tau$.

subject to the boundary conditions

$$\lim_{|x|\to\infty} u(x,t) \to 0, \qquad 0 < t, \tag{5.4.43}$$

and initial conditions

$$u(x,0) = u_t(x,0) = 0, \qquad -\infty < x < \infty. \tag{5.4.44}$$

Here ω is a constant and $X(t)$ is some function of time.

With the given boundary and initial conditions, only the first integral in Equation 5.4.9 does not vanish. Substituting the source term $q(x,t) = \cos(\omega t)\delta[x - X(t)]$ and the Green's function given by Equation 5.4.26, we have that

$$u(x,t) = \int_0^t \int_{-\infty}^{\infty} q(\xi,\tau) g(x,t|\xi,\tau)\, d\xi\, d\tau \tag{5.4.45}$$

$$= \frac{1}{2}\int_0^t \int_{-\infty}^{\infty} \cos(\omega\tau)\delta[\xi - X(\tau)] H(t-\tau) - |x-\xi|)\, d\xi\, d\tau \tag{5.4.46}$$

$$= \frac{1}{2}\int_0^t H[t - \tau - |X(\tau) - x|] \cos(\omega\tau)\, d\tau, \tag{5.4.47}$$

since $c = 1$.

Problems

1. By direct substitution, show[15] that

$$g(x,t|0,0) = J_0(\sqrt{xt}\,)H(x)H(t)$$

[15] First proven by Picard, É., 1894: Sur une équation aux dérivées partielles de la théorie de la propagation de l'électricité. *Bull. Soc. Math.*, **22**, 2–8.

is the free-space Green's function governed by

$$\frac{\partial^2 g}{\partial x \partial t} + \tfrac{1}{4} g = \delta(x)\delta(t), \qquad -\infty < x, t < \infty.$$

2. Use Equation 5.3.20 to construct the Green's function for the one-dimensional wave equation

$$\frac{\partial^2 g}{\partial t^2} - \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \quad 0 < x, \xi < L, \quad 0 < t, \tau,$$

subject to the boundary conditions $g(0,t|\xi,\tau) = g_x(L,t|\xi,\tau) = 0$, $0 < t$, and the initial conditions that $g(x,0|\xi,\tau) = g_t(x,0|\xi,\tau) = 0$, $0 < x < L$.

3. Use Equation 5.3.20 to construct the Green's function for the one-dimensional wave equation

$$\frac{\partial^2 g}{\partial t^2} - \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \quad 0 < x, \xi < L, \quad 0 < t, \tau,$$

subject to the boundary conditions $g_x(0,t|\xi,\tau) = g_x(L,t|\xi,\tau) = 0$, $0 < t$, and the initial conditions that $g(x,0|\xi,\tau) = g_t(x,0|\xi,\tau) = 0$, $0 < x < L$.

4. Use the Green's function given by Equation 5.3.26 to write down the solution to the wave equation $u_{tt} = u_{xx}$ on the interval $0 < x < L$ with the boundary conditions $u(0,t) = u(L,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = \cos(\pi x/L)$ and $u_t(x,0) = 0$, $0 < x < L$.

5. Use the Green's function given by Equation 5.3.26 to write down the solution to the wave equation $u_{tt} = u_{xx}$ on the interval $0 < x < L$ with the boundary conditions $u(0,t) = e^{-t}$ and $u(L,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = \sin(\pi x/L)$ and $u_t(x,0) = 1$, $0 < x < L$.

6. Use the Green's function that you found in Problem 2 to write down the solution to the wave equation $u_{tt} = u_{xx}$ on the interval $0 < x < L$ with the boundary conditions $u(0,t) = 0$ and $u_x(L,t) = 1$, $0 < t$, and the initial conditions $u(x,0) = x$ and $u_t(x,0) = 1$, $0 < x < L$.

7. Use the Green's function that you found in Problem 3 to write down the solution to the wave equation $u_{tt} = u_{xx}$ on the interval $0 < x < L$ with the boundary conditions $u_x(0,t) = 1$ and $u_x(L,t) = 0$, $0 < t$, and the initial conditions $u(x,0) = 1$ and $u_t(x,0) = 0$, $0 < x < L$.

8. Find the Green's function[16] governed by

$$\frac{\partial^2 g}{\partial t^2} + 2\frac{\partial g}{\partial t} - \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \quad 0 < x, \xi < L, \quad 0 < t, \tau,$$

subject to the boundary conditions

$$g_x(0,t|\xi,\tau) = g_x(L,t|\xi,\tau) = 0, \qquad 0 < t,$$

---

[16] Özişik, M. N., and B. Vick, 1984: Propagation and reflection of thermal waves in a finite medium. *Int. J. Heat Mass Transfer*, **27**, 1845–1854; Tang, D.-W., and N. Araki, 1996: Propagation of non-Fourier temperature wave in finite medium under laser-pulse heating (in Japanese). *Nihon Kikai Gakkai Rombumshu (Trans. Japan Soc. Mech. Engrs.), Ser. B*, **62**, 1136–1141.

and the initial conditions

$$g(x, 0|\xi, \tau) = g_t(x, 0|\xi, \tau) = 0, \qquad 0 < x < L.$$

Step 1: If the Green's function can be written as the Fourier half-range cosine series

$$g(x, t|\xi, \tau) = \frac{1}{L} G_0(t|\tau) + \frac{2}{L} \sum_{n=1}^{\infty} G_n(t|\tau) \cos\left(\frac{n\pi x}{L}\right),$$

so that it satisfies the boundary conditions, show that $G_n(t|\tau)$ is governed by

$$G_n'' + 2G_n' + \frac{n^2\pi^2}{L^2} G_n = \cos\left(\frac{n\pi\xi}{L}\right) \delta(t-\tau), \qquad 0 \le n.$$

Step 2: Show that

$$G_0(t|\tau) = e^{-(t-\tau)} \sinh(t-\tau) H(t-\tau),$$

and

$$G_n(t|\tau) = \cos\left(\frac{n\pi\xi}{L}\right) e^{-(t-\tau)} \frac{\sin[\beta_n(t-\tau)]}{\beta_n} H(t-\tau), \quad 1 \le n,$$

where $\beta_n = \sqrt{(n\pi/L)^2 - 1}$.

Step 3: Combine the results from Steps 1 and 2 and show that

$$\begin{aligned} g(x, t|\xi, \tau) = {} & e^{-(t-\tau)} \sinh(t-\tau) H(t-\tau)/L \\ & + 2e^{-(t-\tau)} H(t-\tau)/L \\ & \times \sum_{n=1}^{\infty} \frac{\sin[\beta_n(t-\tau)]}{\beta_n} \cos\left(\frac{n\pi\xi}{L}\right) \cos\left(\frac{n\pi x}{L}\right). \end{aligned}$$

5.5 HEAT EQUATION

In this section we present the Green's function[17] for the heat equation

$$\frac{\partial u}{\partial t} - a^2 \frac{\partial^2 u}{\partial x^2} = q(x, t), \tag{5.5.1}$$

where t denotes time, x is the position, a^2 is the diffusivity, and $q(x, t)$ is the source density. In addition to Equation 5.5.1, boundary conditions must be specified to ensure the uniqueness of solution; the most common ones are Dirichlet, Neumann, and Robin (a linear combination of the first two). An initial condition $u(x, t = t_0)$ is also needed.

The heat equation differs in many ways from the wave equation and the Green's function must, of course, manifest these differences. The most notable one is the asymmetry of the

[17] See also Carslaw, H. S., and J. C. Jaeger, 1959: *Conduction of Heat in Solids.* Clarendon Press, Chapter 14; Beck, J. V., K. D. Cole, A. Haji-Sheikh, and B. Litkouhi, 1992: *Heat Conduction Using Green's Functions.* Hemisphere Publishing Corp., 523 pp.; Özişik, M. N., 1993: *Heat Conduction.* John Wiley & Sons, Inc., Chapter 6.

heat equation with respect to time. This merely reflects the fact that the heat equation differentiates between past and future as entropy continually increases.

We begin by proving that we can express the solution to Equation 5.5.1 in terms of boundary conditions, the initial condition, and the Green's function, which is found by solving

$$\frac{\partial g}{\partial t} - a^2 \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \tag{5.5.2}$$

where ξ denotes the position of the source. From causality[18] we know that $g(x,t|\xi,\tau) = 0$ if $t < \tau$. We again require that the Green's function $g(x,t|\xi,\tau)$ satisfies the homogeneous form of the boundary condition on $u(x,t)$. For example, if u satisfies a homogeneous or nonhomogeneous Dirichlet condition, then the Green's function will satisfy the corresponding *homogeneous* Dirichlet condition. Although we will focus on the mathematical aspects of the problem, Equation 5.5.2 can be given the physical interpretation of the temperature distribution within a medium when a unit of heat is introduced at ξ at time τ.

We now establish that the solution to the nonhomogeneous heat equation can be expressed in terms of the Green's function, boundary conditions, and the initial condition. We begin with the equations

$$a^2 \frac{\partial^2 u(\xi,\tau)}{\partial \xi^2} - \frac{\partial u(\xi,\tau)}{\partial \tau} = -q(\xi,\tau), \tag{5.5.3}$$

and

$$a^2 \frac{\partial^2 g(x,t|\xi,\tau)}{\partial \xi^2} + \frac{\partial g(x,t|\xi,\tau)}{\partial \tau} = -\delta(x-\xi)\delta(t-\tau). \tag{5.5.4}$$

As we did in the previous section, we multiply Equation 5.5.3 by $g(x,t|\xi,\tau)$ and Equation 5.5.4 by $u(\xi,\tau)$ and subtract. Integrating over ξ from a to b, where a and b are the endpoints of the spatial domain, and over τ from 0 to t^+, where t^+ denotes a time slightly later than t so that we avoid ending the integration exactly at the peak of the delta function, we find

$$\begin{aligned} a^2 \int_0^{t^+} \int_a^b & \left[u(\xi,\tau) \frac{\partial^2 g(x,t|\xi,\tau)}{\partial \xi^2} - g(x,t|\xi,\tau) \frac{\partial^2 u(\xi,\tau)}{\partial \xi^2} \right] d\xi\, d\tau \\ & + \int_0^{t^+} \int_a^b \left[u(\xi,\tau) \frac{\partial g(x,t|\xi,\tau)}{\partial \tau} + g(x,t|\xi,\tau) \frac{\partial u(\xi,\tau)}{\partial \tau} \right] d\xi\, d\tau \\ & = \int_0^{t^+} \int_a^b q(\xi,\tau) g(x,t|\xi,\tau)\, d\xi\, d\tau - u(x,t). \end{aligned} \tag{5.5.5}$$

Applying Equation 5.4.6 and performing the time integration in the second integral, we finally obtain

$$\begin{aligned} u(x,t) = & \int_0^{t^+} \int_a^b q(\xi,\tau) g(x,t|\xi,\tau)\, d\xi\, d\tau \\ & + a^2 \int_0^{t^+} \left[g(x,t|\xi,\tau) \frac{\partial u(\xi,\tau)}{\partial \xi} - u(\xi,\tau) \frac{\partial g(x,t|\xi,\tau)}{\partial \xi} \right]_{\xi=a}^{\xi=b} d\tau \\ & + \int_a^b u(\xi,0) g(x,t|\xi,0)\, d\xi, \end{aligned} \tag{5.5.6}$$

---

[18] The principle stating that an event cannot precede its cause.

where we used $g(x,t|\xi,t^+)=0$. The first two terms in Equation 5.5.6 represent the familiar effects of volume sources and boundary conditions, while the third term includes the effects of the initial data.

• Example 5.5.1: One-dimensional heat equation in an unlimited domain

The Green's function for the one-dimensional heat equation is governed by

$$\frac{\partial g}{\partial t}-a^2\frac{\partial^2 g}{\partial x^2}=\delta(x-\xi)\delta(t-\tau), \quad -\infty<x,\xi<\infty, \quad 0<t,\tau, \tag{5.5.7}$$

subject to the boundary conditions $\lim_{|x|\to\infty} g(x,t|\xi,\tau)\to 0$, and the initial condition $g(x,0|\xi,\tau)=0$. Let us find $g(x,t|\xi,\tau)$.

We begin by taking the Laplace transform of Equation 5.5.7 and find that

$$\frac{d^2G}{dx^2}-\frac{s}{a^2}G=-\frac{\delta(x-\xi)}{a^2}e^{-s\tau}. \tag{5.5.8}$$

Next, we take the Fourier transform of Equation 5.5.8 so that

$$(k^2+b^2)\overline{G}(k,s|\xi,\tau)=\frac{e^{-ik\xi}e^{-s\tau}}{a^2}, \tag{5.5.9}$$

where $\overline{G}(k,s|\xi,\tau)$ is the Fourier transform of $G(x,s|\xi,\tau)$ and $b^2=s/a^2$.

To find $G(x,s|\xi,\tau)$, we use the inversion integral

$$G(x,s|\xi,\tau)=\frac{e^{-s\tau}}{2\pi a^2}\int_{-\infty}^{\infty}\frac{e^{i(x-\xi)k}}{k^2+b^2}\,dk. \tag{5.5.10}$$

Transforming Equation 5.5.10 into a closed contour via Jordan's lemma, we evaluate it by the residue theorem and find that

$$G(x,s|\xi,\tau)=\frac{e^{-|x-\xi|\sqrt{s}/a-s\tau}}{2a\sqrt{s}}. \tag{5.5.11}$$

From a table of Laplace transforms we finally obtain

$$\boxed{g(x,t|\xi,\tau)=\frac{H(t-\tau)}{\sqrt{4\pi a^2(t-\tau)}}\exp\left[-\frac{(x-\xi)^2}{4a^2(t-\tau)}\right],} \tag{5.5.12}$$

after applying the second shifting theorem. □

The primary use of the fundamental or free-space Green's function[19] is as a *particular* solution to the Green's function problem. For this reason, it is often called the *fundamental*

[19] In electromagnetic theory, a free-space Green's function is the particular solution of the differential equation valid over a domain of infinite extent, where the Green's function remains bounded as we approach infinity, or satisfies a radiation condition there.

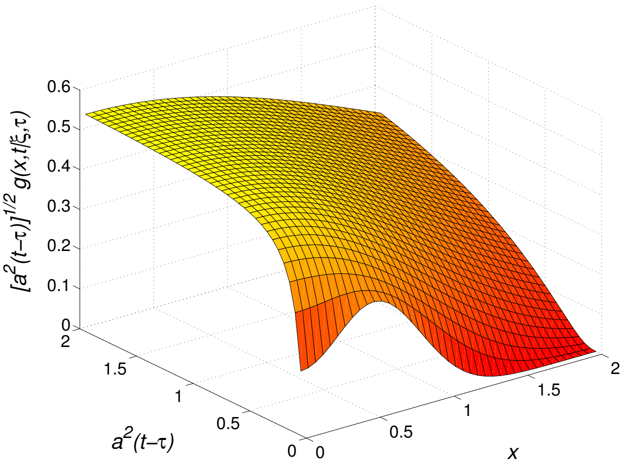


Figure 5.5.1: The Green's function, Equation 5.5.17, for the one-dimensional heat equation on the semi-infinite domain $0 < x < \infty$, and $0 \le t - \tau$, when the left boundary condition is $g_x(0, t|\xi, \tau) = 0$ and $\xi = 0.5$.

heat conduction solution. Consequently, we usually must find a homogeneous solution so that the sum of the free-space Green's function plus the homogeneous solution satisfies any boundary conditions. The following examples show some commonly employed techniques.

• Example 5.5.2

Let us find the Green's function for the following problem:

$$\frac{\partial g}{\partial t} - a^2 \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \qquad 0 < x, \xi < \infty, \quad 0 < t, \tau, \tag{5.5.13}$$

subject to the boundary conditions $g(0, t|\xi, \tau) = 0$, $\lim_{x\to\infty} g(x, t|\xi, \tau) \to 0$, and the initial condition $g(x, 0|\xi, \tau) = 0$. From the boundary condition $g(0, t|\xi, \tau) = 0$, we deduce that $g(x, t|\xi, \tau)$ must be an odd function in x over the open interval $(-\infty, \infty)$. We find this Green's function by introducing an image source of $-\delta(x + \xi)$ and resolving Equation 5.5.7 with the source $\delta(x-\xi)\delta(t-\tau) - \delta(x+\xi)\delta(t-\tau)$. Because Equation 5.5.7 is linear, Equation 5.5.12 gives the solution for each delta function and the Green's function for Equation 5.5.13 is

$$g(x, t|\xi, \tau) = \frac{H(t-\tau)}{\sqrt{4\pi a^2(t-\tau)}} \left\{ \exp\left[-\frac{(x-\xi)^2}{4a^2(t-\tau)} \right] - \exp\left[-\frac{(x+\xi)^2}{4a^2(t-\tau)} \right] \right\} \tag{5.5.14}$$

$$= \frac{H(t-\tau)}{\sqrt{\pi a^2(t-\tau)}} \exp\left[-\frac{x^2+\xi^2}{4a^2(t-\tau)} \right] \sinh\left[\frac{x\xi}{2a^2(t-\tau)} \right]. \tag{5.5.15}$$

In a similar manner, if the boundary condition at $x = 0$ changes to $g_x(0, t|\xi, \tau) = 0$, then Equation 5.5.14 through Equation 5.5.15 become

$$g(x, t|\xi, \tau) = \frac{H(t-\tau)}{\sqrt{4\pi a^2(t-\tau)}} \left\{ \exp\left[-\frac{(x-\xi)^2}{4a^2(t-\tau)} \right] + \exp\left[-\frac{(x+\xi)^2}{4a^2(t-\tau)} \right] \right\} \tag{5.5.16}$$

$$= \frac{H(t-\tau)}{\sqrt{\pi a^2(t-\tau)}} \exp\left[-\frac{x^2+\xi^2}{4a^2(t-\tau)} \right] \cosh\left[\frac{x\xi}{2a^2(t-\tau)} \right]. \tag{5.5.17}$$

Figure 5.5.1 illustrates Equation 5.5.17 for the special case when $\xi = 0.5$. □

• Example 5.5.3: One-dimensional heat equation on the interval $0 < x < L$

Here we find the Green's function for the one-dimensional heat equation over the interval $0 < x < L$ associated with the problem

$$\frac{\partial u}{\partial t} - a^2 \frac{\partial^2 u}{\partial x^2} = f(x,t), \qquad 0 < x < L, \quad 0 < t, \tag{5.5.18}$$

where a^2 is the diffusivity constant.

To find the Green's function for this problem, consider the following problem:

$$\frac{\partial g}{\partial t} - a^2 \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \qquad 0 < x, \xi < L, \quad 0 < t, \tau, \tag{5.5.19}$$

with the boundary conditions

$$\alpha_1 g(0,t|\xi,\tau) + \beta_1 g_x(0,t|\xi,\tau) = 0, \quad 0 < t, \tag{5.5.20}$$

and

$$\alpha_2 g(L,t|\xi,\tau) + \beta_2 g_x(L,t|\xi,\tau) = 0, \quad 0 < t, \tag{5.5.21}$$

and the initial condition

$$g(x,0|\xi,\tau) = 0, \quad 0 < x < L. \tag{5.5.22}$$

We begin by taking the Laplace transform of Equation 5.5.19 and find that

$$\frac{d^2 G}{dx^2} - \frac{s}{a^2} G = -\frac{\delta(x-\xi)}{a^2} e^{-s\tau}, \qquad 0 < x < L, \tag{5.5.23}$$

with

$$\alpha_1 G(0,s|\xi,\tau) + \beta_1 G'(0,s|\xi,\tau) = 0, \tag{5.5.24}$$

and

$$\alpha_2 G(L,s|\xi,\tau) + \beta_2 G'(L,s|\xi,\tau) = 0. \tag{5.5.25}$$

Problems similar to Equation 5.5.23 through Equation 5.5.25 were considered in Section 5.2. Applying this technique of eigenfunction expansions, we have that

$$G(x,s|\xi,\tau) = e^{-s\tau} \sum_{n=1}^{\infty} \frac{\varphi_n(\xi)\varphi_n(x)}{s + a^2 k_n^2}, \tag{5.5.26}$$

where $\varphi_n(x)$ is the nth *orthonormal* eigenfunction to the regular Sturm-Liouville problem

$$\varphi''(x) + k^2 \varphi(x) = 0, \qquad 0 < x < L, \tag{5.5.27}$$

subject to the boundary conditions

$$\alpha_1 \varphi(0) + \beta_1 \varphi'(0) = 0, \tag{5.5.28}$$

and

$$\alpha_2 \varphi(L) + \beta_2 \varphi'(L) = 0. \tag{5.5.29}$$

Taking the inverse of Equation 5.5.26, we have that

$$\boxed{g(x,t|\xi,\tau) = \left[\sum_{n=1}^{\infty} \varphi_n(\xi)\varphi_n(x)e^{-k_n^2 a^2 (t-\tau)}\right] H(t-\tau).} \tag{5.5.30}$$

For example, let us find the Green's function for the heat equation on a finite domain

$$\frac{\partial g}{\partial t} - a^2 \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \quad 0 < x,\xi < L, \quad 0 < t,\tau, \tag{5.5.31}$$

with the boundary conditions $g(0,t|\xi,\tau) = g(L,t|\xi,\tau) = 0$, $0 < t$, and the initial condition $g(x,0|\xi,\tau) = 0$, $0 < x < L$.

The Sturm-Liouville problem is

$$\varphi''(x) + k^2\varphi(x) = 0, \qquad 0 < x < L, \tag{5.5.32}$$

with the boundary conditions $\varphi(0) = \varphi(L) = 0$. The nth *orthonormal* eigenfunction to Equation 5.5.32 is

$$\varphi_n(x) = \sqrt{\frac{2}{L}}\, \sin\left(\frac{n\pi x}{L}\right). \tag{5.5.33}$$

Substituting Equation 5.5.33 into Equation 5.5.30, we find that

$$g(x,t|\xi,\tau) = \frac{2}{L}\left\{\sum_{n=1}^{\infty} \sin\left(\frac{n\pi\xi}{L}\right) \sin\left(\frac{n\pi x}{L}\right) e^{-a^2 n^2 \pi^2 (t-\tau)/L^2}\right\} H(t-\tau). \tag{5.5.34}$$

On the other hand, the Green's function for the heat equation on a finite domain governed by

$$\frac{\partial g}{\partial t} - a^2 \frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \qquad 0 < x,\xi < L, \quad 0 < t,\tau, \tag{5.5.35}$$

with the boundary conditions

$$g_x(0,t|\xi,\tau) = 0, \quad g_x(L,t|\xi,\tau) + hg(L,t|\xi,\tau) = 0, \quad 0 < t, \tag{5.5.36}$$

and the initial condition $g(x,0|\xi,\tau) = 0$, $0 < x < L$, yields the Sturm-Liouville problem that we must now solve:

$$\varphi''(x) + \lambda\varphi(x) = 0, \quad \varphi'(0) = 0, \quad \varphi'(L) + h\varphi(L) = 0. \tag{5.5.37}$$

The nth *orthonormal* eigenfunction for Equation 5.5.37 is

$$\varphi_n(x) = \sqrt{\frac{2(k_n^2 + h^2)}{L(k_n^2 + h^2) + h}}\, \cos(k_n x), \tag{5.5.38}$$

where k_n is the nth root of $k\tan(kL) = h$. We also used the identity that $(k_n^2 + h^2)\sin^2(k_n h) = h^2$. Substituting Equation 5.5.38 into Equation 5.5.30, we finally obtain

$$g(x,t|\xi,\tau) = \frac{2}{L}\left\{\sum_{n=1}^{\infty} \frac{[(k_nL)^2 + (hL)^2]\cos(k_n\xi)\cos(k_nx)}{(k_nL)^2 + (hL)^2 + hL} e^{-a^2k_n^2(t-\tau)}\right\} H(t-\tau). \quad \mathbf{(5.5.39)}$$

□

• **Example 5.5.4**

Let us use Green's functions to solve the heat equation

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < L, \quad 0 < t, \quad \mathbf{(5.5.40)}$$

subject to the boundary conditions

$$u(0,t) = 0, \quad u(L,t) = t, \qquad 0 < t, \quad \mathbf{(5.5.41)}$$

and the initial condition

$$u(x,0) = 0, \qquad 0 < x < L. \quad \mathbf{(5.5.42)}$$

Because there is no source term, Equation 5.5.6 simplifies to

$$\begin{aligned} u(x,t) = a^2 & \int_0^t [g(x,t|L,\tau)u_\xi(L,\tau) - u(L,\tau)g_\xi(x,t|L,\tau)]\, d\tau \qquad \mathbf{(5.5.43)} \\ & - a^2\int_0^t [g(x,t|0,\tau)u_\xi(0,\tau) - u(0,\tau)g_\xi(x,t|0,\tau)]\, d\tau + \int_0^L u(\xi,0)g(x,t|\xi,0)\, d\xi. \end{aligned}$$

The Green's function that should be used here is the one given by Equation 5.5.34. Further simplification occurs by noting that $g(x,t|0,\tau) = g(x,t|L,\tau) = 0$ as well as $u(0,\tau) = u(\xi,0) = 0$. Therefore we are left with the single integral

$$u(x,t) = -a^2\int_0^t u(L,\tau)g_\xi(x,t|L,\tau)\, d\tau. \quad \mathbf{(5.5.44)}$$

Upon substituting for $g(x,t|L,\tau)$ and reversing the order of integration and summation,

$$\begin{aligned} u(x,t) &= -\frac{2\pi a^2}{L^2}\sum_{n=1}^{\infty}(-1)^n n\sin\left(\frac{n\pi x}{L}\right)\int_0^t \tau\exp\left[\frac{a^2n^2\pi^2}{L^2}(\tau - t)\right] d\tau && \mathbf{(5.5.45)} \\ &= -\frac{2L^2}{a^2\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3}\sin\left(\frac{n\pi x}{L}\right)\exp\left[\frac{a^2n^2\pi^2}{L^2}(\tau - t)\right]\left(\frac{a^2n^2\pi^2\tau}{L^2} - 1\right)\Bigg|_0^t && \mathbf{(5.5.46)} \\ &= -\frac{2L^2}{a^2\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3}\sin\left(\frac{n\pi x}{L}\right)\left[\frac{a^2n^2\pi^2t}{L^2} - 1 + \exp\left(-\frac{a^2n^2\pi^2t}{L^2}\right)\right]. && \mathbf{(5.5.47)} \end{aligned}$$

Figure 5.5.2 illustrates Equation 5.5.47. This solution could also have been found using Duhamel's integral. □

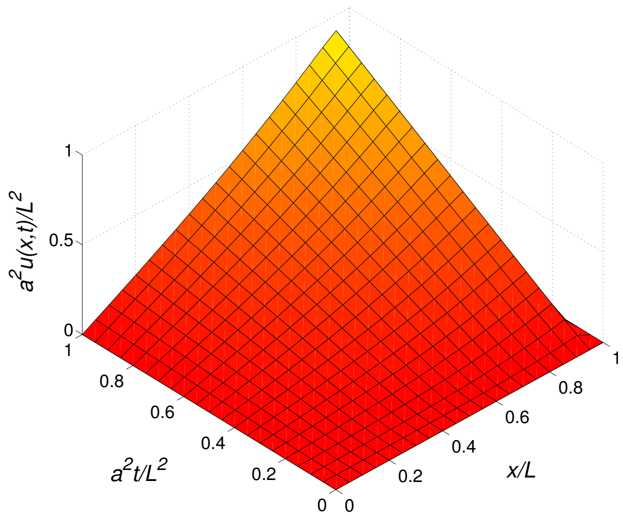


Figure 5.5.2: The temperature distribution within a bar when the temperature is initially at zero and then the ends are held at zero at $x = 0$ and t at $x = L$.

• Example 5.5.5: Heat equation within a cylinder

In this example, we find the Green's function for the heat equation in cylindrical coordinates

$$\frac{\partial g}{\partial t} - \frac{a^2}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g}{\partial r}\right) = \frac{\delta(r-\rho)\delta(t-\tau)}{2\pi r}, \qquad 0 < r, \rho < b, \quad 0 < t, \tau, \tag{5.5.48}$$

subject to the boundary conditions $\lim_{r\to 0} |g(r, t|\rho, \tau)| < \infty$, $g(b, t|\rho, \tau) = 0$, and the initial condition $g(r, 0|\rho, \tau) = 0$.

As usual, we begin by taking the Laplace transform of Equation 5.5.48, or

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dG}{dr}\right) - \frac{s}{a^2}G = -\frac{e^{-s\tau}}{2\pi a^2 r}\delta(r-\rho). \tag{5.5.49}$$

Next we re-express $\delta(r-\rho)/r$ as the Fourier-Bessel expansion

$$\frac{\delta(r-\rho)}{2\pi r} = \sum_{n=1}^{\infty} A_n J_0(k_n r/b), \tag{5.5.50}$$

where k_n is the nth root of $J_0(k) = 0$, and

$$A_n = \frac{2}{b^2 J_1^2(k_n)} \int_0^b \frac{\delta(r-\rho)}{2\pi r} J_0(k_n r/b)\, r\, dr = \frac{J_0(k_n\rho/b)}{\pi b^2 J_1^2(k_n)} \tag{5.5.51}$$

so that

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dG}{dr}\right) - \frac{s}{a^2}G = -\frac{e^{-s\tau}}{\pi a^2 b^2} \sum_{n=1}^{\infty} \frac{J_0(k_n\rho/b) J_0(k_n r/b)}{J_1^2(k_n)}. \tag{5.5.52}$$

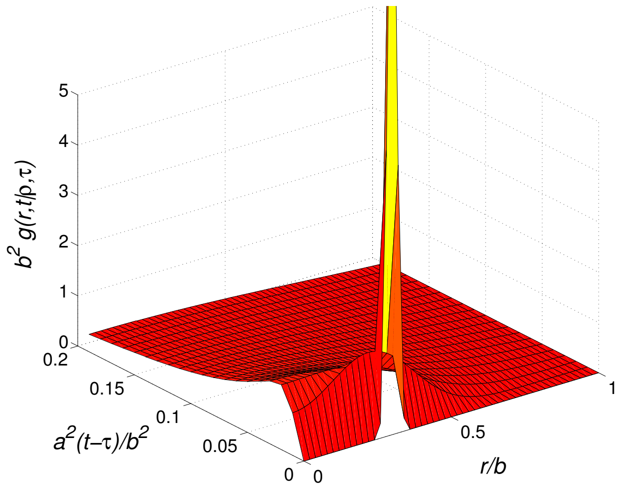


Figure 5.5.3: The Green's function, Equation 5.5.54, for the axisymmetric heat equation, Equation 5.5.48, with a Dirichlet boundary condition at $r = b$. Here $\rho/b = 0.3$ and the graph starts at $a^2(t-\tau)/b^2 = 0.001$ to avoid the delta function at $t - \tau = 0$.

The solution to Equation 5.5.52 is

$$G(r,s|\rho,\tau) = \frac{e^{-s\tau}}{\pi} \sum_{n=1}^{\infty} \frac{J_0(k_n\rho/b)J_0(k_n r/b)}{(sb^2 + a^2k_n^2)J_1^2(k_n)}. \tag{5.5.53}$$

Taking the inverse of Equation 5.5.53 and applying the second shifting theorem,

$$g(r,t|\rho,\tau) = \frac{H(t-\tau)}{\pi b^2} \sum_{n=1}^{\infty} \frac{J_0(k_n\rho/b)J_0(k_n r/b)}{J_1^2(k_n)} e^{-a^2k_n^2(t-\tau)/b^2}. \tag{5.5.54}$$

See Figure 5.5.3.

If we modify the boundary condition at $r = b$ so that it now reads

$$g_r(b,t|\rho,\tau) + hg(b,t|\rho,\tau) = 0, \tag{5.5.55}$$

where $h \geq 0$, our analysis now leads to

$$g(r,t|\rho,\tau) = \frac{H(t-\tau)}{\pi b^2} \sum_{n=1}^{\infty} \frac{J_0(k_n\rho/b)J_0(k_n r/b)}{J_0^2(k_n) + J_1^2(k_n)} e^{-a^2k_n^2(t-\tau)/b^2}, \tag{5.5.56}$$

where k_n are the positive roots of $k\,J_1(k) - hb\,J_0(k) = 0$. If $h = 0$, we must add $1/(\pi b^2)$ to Equation 5.5.56.

Problems

1. Find the free-space Green's function for the linearized Ginzburg-Landau equation[20]

$$\frac{\partial g}{\partial t} + v\frac{\partial g}{\partial x} - ag - b\frac{\partial^2 g}{\partial x^2} = \delta(x-\xi)\delta(t-\tau), \; -\infty < x, \xi < \infty, \; 0 < t, \tau,$$

[20] See Deissler, R. J., 1985: Noise-sustained structure, intermittency, and the Ginzburg-Landau equation. *J. Stat. Phys.*, **40**, 371–395.

with $b > 0$.

Step 1: Taking the Laplace transform of the partial differential equation, show that it reduces to the ordinary differential equation

$$b\frac{d^2G}{dx^2} - v\frac{dG}{dx} + aG - sG = -\delta(x-\xi)e^{-s\tau}.$$

Step 2: Using Fourier transforms, show that

$$G(x,s|\xi,\tau) = \frac{e^{-s\tau}}{2\pi}\int_{-\infty}^{\infty} \frac{e^{ik(x-\xi)}}{s+ikv+bk^2-a}\,dk,$$

or

$$g(x,t|\xi,\tau) = \frac{e^{a(t-\tau)}}{\pi}H(t-\tau)\int_0^\infty e^{-b(t-\tau)k^2}\cos\{k[x-\xi-v(t-\tau)]\}\,dk.$$

Step 3: Evaluate the second integral and show that

$$g(x,t|\xi,\tau) = \frac{e^{a(t-\tau)}H(t-\tau)}{2\sqrt{\pi b(t-\tau)}}\exp\left\{-\frac{[x-\xi-v(t-\tau)]^2}{4b(t-\tau)}\right\}.$$

2. Use Green's functions to show that the solution[21] to

$$\frac{\partial u}{\partial t} = a^2\frac{\partial^2 u}{\partial x^2}, \qquad 0 < x, t,$$

subject to the boundary conditions

$$u(0,t) = 0, \quad \lim_{x\to\infty} u(x,t) \to 0, \qquad 0 < t,$$

and the initial condition

$$u(x,0) = f(x), \qquad 0 < x < \infty,$$

is

$$u(x,t) = \frac{e^{-x^2/(4a^2t)}}{a\sqrt{\pi t}}\int_0^\infty f(\tau)\sinh\left(\frac{x\tau}{2a^2t}\right)e^{-\tau^2/(4a^2t)}\,d\tau.$$

3. Use Equation 5.5.30 to construct the Green's function for the one-dimensional heat equation $g_t - g_{xx} = \delta(x-\xi)\delta(t-\tau)$ for $0 < x < L$, $0 < t$, with the initial condition that $g(x,0|\xi,\tau) = 0$, $0 < x < L$, and the boundary conditions that $g(0,t|\xi,\tau) = g_x(L,t|\xi,\tau) = 0$ for $0 < t$. Assume that $L \neq \pi$.

4. Use Equation 5.5.30 to construct the Green's function for the one-dimensional heat equation $g_t - g_{xx} = \delta(x-\xi)\delta(t-\tau)$ for $0 < x < L$, $0 < t$, with the initial condition that

[21] See Gilev, S. D., and T. Yu. Mikhaĭlova, 1996: Current wave in shock compression of matter in a magnetic field. *Tech. Phys.*, **41**, 407–411.

$g(x,0|\xi,\tau) = 0$, $0 < x < L$, and the boundary conditions that $g_x(0,t|\xi,\tau) = g_x(L,t|\xi,\tau) = 0$ for $0 < t$.

5. Use Equation 5.5.43 and the Green's function given by Equation 5.5.34 to find the solution to the heat equation $u_t = u_{xx}$ for $0 < x < L$, $0 < t$, with the initial data $u(x,0) = 1$, $0 < x < L$, and the boundary conditions $u(0,t) = e^{-t}$ and $u(L,t) = 0$ when $0 < t$.

6. Use Equation 5.5.43 and the Green's function that you found in Problem 3 to find the solution to the heat equation $u_t = u_{xx}$ for $0 < x < L$, $0 < t$, with the initial data $u(x,0) = 1$, $0 < x < L$, and the boundary conditions $u(0,t) = \sin(t)$ and $u_x(L,t) = 0$ when $0 < t$.

7. Use Equation 5.5.43 and the Green's function that you found in Problem 4 to find the solution to the heat equation $u_t = u_{xx}$ for $0 < x < L$, $0 < t$, with the initial data $u(x,0) = 1$, $0 < x < L$, and the boundary conditions $u_x(0,t) = 1$ and $u_x(L,t) = 0$ when $0 < t$.

8. Find the Green's function for

$$\frac{\partial g}{\partial t} - a^2 \frac{\partial^2 g}{\partial x^2} + a^2 k^2 g = \delta(x-\xi)\delta(t-\tau), \quad 0 < x, \xi < L, \quad 0 < t, \tau,$$

subject to the boundary conditions

$$g(0,t|\xi,\tau) = g_x(L,t|\xi,\tau) = 0, \qquad 0 < t,$$

and the initial condition

$$g(x,0|\xi,\tau) = 0, \qquad 0 < x < L,$$

where a and k are real constants.

5.6 HELMHOLTZ'S EQUATION

In the previous sections, we sought solutions to the heat and wave equations via Green's functions. In this section, we turn to the reduced wave equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \lambda u = -f(x,y). \tag{5.6.1}$$

Equation 5.6.1, generally known as *Helmholtz's equation*, includes the special case of *Poisson's equation* when $\lambda = 0$. Poisson's equation has a special place in the theory of Green's functions because George Green (1793–1841) invented his technique for its solution.

The reduced wave equation arises during the solution of the harmonically forced wave equation[22] by separation of variables. In one spatial dimension, the problem is

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = -f(x)e^{-i\omega t}. \tag{5.6.2}$$

Equation 5.6.2 occurs, for example, in the mathematical analysis of a stretched string over some interval subject to an external, harmonic forcing. Assuming that $u(x,t)$ is bounded

[22] See, for example, Graff, K. F., 1991: *Wave Motion in Elastic Solids*. Dover Publications, Inc., Section 1.4.

everywhere, we seek solutions of the form $u(x,t) = y(x)e^{-i\omega t}$. Upon substituting this solution into Equation 5.6.2 we obtain the ordinary differential equation

$$y'' + k_0^2 y = -f(x), \tag{5.6.3}$$

where $k_0^2 = \omega^2/c^2$. This is an example of the one-dimensional Helmholtz equation.

Let us now use Green's functions to solve the Helmholtz equation, Equation 5.6.1, where the Green's function is given by the Helmholtz equation

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} + \lambda g = -\delta(x-\xi)\delta(y-\eta). \tag{5.6.4}$$

The most commonly encountered boundary conditions are

- the *Dirichlet boundary condition*, where g vanishes on the boundary,
- the *Neumann boundary condition*, where the normal gradient of g vanishes on the boundary, and
- the *Robin boundary condition*, which is the linear combination of the Dirichlet and Neumann conditions.

We begin by multiplying Equation 5.6.1 by $g(x,y|\xi,\eta)$ and Equation 5.6.4 by $u(x,y)$, subtract and integrate over the region $a < x < b, c < y < d$. We find that

$$\begin{aligned} u(\xi,\eta) = \int_c^d \int_a^b \biggl\{ g(x,y|\xi,\eta) \left[\frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2} \right] \\ - u(x,y) \left[\frac{\partial^2 g(x,y|\xi,\eta)}{\partial x^2} + \frac{\partial^2 g(x,y|\xi,\eta)}{\partial y^2} \right] \biggr\} \, dx \, dy \\ + \int_c^d \int_a^b f(x,y) g(x,y|\xi,\eta) \, dx \, dy \end{aligned} \tag{5.6.5}$$

$$\begin{aligned} = \int_c^d \int_a^b \left\{ \frac{\partial}{\partial x} \left[g(x,y|\xi,\eta) \frac{\partial u(x,y)}{\partial x} \right] - \frac{\partial}{\partial x} \left[u(x,y) \frac{\partial g(x,y|\xi,\eta)}{\partial x} \right] \right\} \, dx \, dy \\ + \int_c^d \int_a^b \left\{ \frac{\partial}{\partial y} \left[g(x,y|\xi,\eta) \frac{\partial u(x,y)}{\partial y} \right] - \frac{\partial}{\partial y} \left[u(x,y) \frac{\partial g(x,y|\xi,\eta)}{\partial y} \right] \right\} \, dx \, dy \\ + \int_c^d \int_a^b f(x,y) g(x,y|\xi,\eta) \, dx \, dy \end{aligned} \tag{5.6.6}$$

$$\begin{aligned} = \int_c^d \left[g(x,y|\xi,\eta) \frac{\partial u(x,y)}{\partial x} - u(x,y) \frac{\partial g(x,y|\xi,\eta)}{\partial x} \right]_{x=a}^{x=b} \, dy \\ + \int_a^b \left[g(x,y|\xi,\eta) \frac{\partial u(x,y)}{\partial y} - u(x,y) \frac{\partial g(x,y|\xi,\eta)}{\partial y} \right]_{y=c}^{y=d} \, dx \\ + \int_c^d \int_a^b f(x,y) g(x,y|\xi,\eta) \, dx \, dy. \end{aligned} \tag{5.6.7}$$

Because (ξ,η) is an arbitrary point inside the rectangle, we denote it in general by (x,y). Furthermore, the variable (x,y) is now merely a dummy integration variable that we now denote by (ξ,η). Upon making these substitutions and using the symmetry condition

$g(x,y|\xi,\eta) = g(\xi,\eta|x,y)$, we have that

$$u(x,y) = \int_c^d \left[g(x,y|\xi,\eta) \frac{\partial u(\xi,\eta)}{\partial \xi} - u(\xi,\eta) \frac{\partial g(x,y|\xi,\eta)}{\partial \xi} \right]_{\xi=a}^{\xi=b} d\eta$$
$$+ \int_a^b \left[g(x,y|\xi,\eta) \frac{\partial u(\xi,\eta)}{\partial \eta} - u(\xi,\eta) \frac{\partial g(x,y|\xi,\eta)}{\partial \eta} \right]_{\eta=c}^{\eta=d} d\xi$$
$$+ \int_c^d \int_a^b f(\xi,\eta) g(x,y|\xi,\eta) \, d\xi \, d\eta. \qquad (\mathbf{5.6.8})$$

Equation 5.6.8 shows that the solution of Helmholtz's equation depends upon the sources inside the rectangle and values of $u(x,y)$ and $(\partial u/\partial x, \partial u/\partial y)$ along the boundary. On the other hand, we must still find the particular Green's function for a given problem; this Green's function depends directly upon the boundary conditions. At this point, we work out several special cases.

1. *Nonhomogeneous Helmholtz equation and homogeneous Dirichlet boundary conditions*

In this case, let us assume that we can find a Green's function that also satisfies the same Dirichlet boundary conditions as $u(x,y)$. Once the Green's function is found, then Equation 5.6.8 reduces to

$$u(x,y) = \int_c^d \int_a^b f(\xi,\eta) g(x,y|\xi,\eta) \, d\xi d\eta. \qquad (\mathbf{5.6.9})$$

A possible source of difficulty would be the nonexistence of the Green's function. From our experience in Section 5.2, we know that this will occur if λ equals one of the eigenvalues of the corresponding homogeneous problem. An example of this occurs in acoustics when the Green's function for the Helmholtz equation does not exist at *resonance*.

2. *Homogeneous Helmholtz equation and nonhomogeneous Dirichlet boundary conditions*

In this particular case, $f(x,y) = 0$. For convenience, let us use the Green's function from the previous example so that $g(x,y|\xi,\eta) = 0$ along all of the boundaries. Under these conditions, Equation 5.6.8 becomes

$$u(x,y) = -\int_a^b u(\xi,\eta) \frac{\partial g(x,y|\xi,\eta)}{\partial \eta} \bigg|_{\eta=c}^{\eta=d} d\xi - \int_c^d u(\xi,\eta) \frac{\partial g(x,y|\xi,\eta)}{\partial \xi} \bigg|_{\xi=a}^{\xi=b} d\eta. \qquad (\mathbf{5.6.10})$$

Consequently, the solution is determined once we compute the normal gradient of the Green's function along the boundary.

3. *Nonhomogeneous Helmholtz equation and homogeneous Neumann boundary conditions*

If we require that $u(x,y)$ satisfies the nonhomogeneous Helmholtz equation with homogeneous Neumann boundary conditions, then the governing equations are Equation 5.6.1

and the boundary conditions $u_x = 0$ along $x = a$ and $x = b$, and $u_y = 0$ along $y = c$ and $y = d$. Integrating Equation 5.6.1, we have that

$$\int_c^d \left[\frac{\partial u(b,y)}{\partial x} - \frac{\partial u(a,y)}{\partial x}\right] dy + \int_a^b \left[\frac{\partial u(x,d)}{\partial y} - \frac{\partial u(x,c)}{\partial y}\right] dx \\ + \lambda \int_c^d \int_a^b u(x,y)\, dx\, dy = -\int_c^d \int_a^b f(x,y)\, dx\, dy. \qquad \textbf{(5.6.11)}$$

Because the first two integrals in Equation 5.6.11 must vanish in the case of homogeneous Neumann boundary conditions, this equation cannot be satisfied if $\lambda = 0$ unless

$$\int_c^d \int_a^b f(x,y)\, dx\, dy = 0. \qquad \textbf{(5.6.12)}$$

A physical interpretation of Equation 5.6.12 is as follows: Consider the physical process of steady-state heat conduction within a rectangular region. The temperature $u(x,y)$ is given by Poisson's equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -f(x,y), \qquad \textbf{(5.6.13)}$$

where $f(x,y)$ is proportional to the density of the heat sources and sinks. The boundary conditions $u_x(a,y) = u_x(b,y) = 0$ and $u_y(x,c) = u_y(x,d) = 0$ imply that there is no heat exchange across the boundary. Consequently, no steady-state temperature distribution can exist unless the heat sources are balanced by heat sinks. This balance of heat sources and sinks is given by Equation 5.6.12.

Having provided an overview of how Green's functions can be used to solve Poisson and Helmholtz equations, let us now determine several of them for commonly encountered domains.

- **Example 5.6.1: Free-space Green's function for the one-dimensional Helmholtz equation**

Let us find the Green's function for the one-dimensional Helmholtz equation

$$g'' + k_0^2 g = -\delta(x-\xi), \qquad -\infty < x, \xi < \infty. \qquad \textbf{(5.6.14)}$$

If we solve Equation 5.6.14 by piecing together homogeneous solutions, then

$$g(x|\xi) = Ae^{-ik_0(x-\xi)} + Be^{ik_0(x-\xi)}, \qquad \textbf{(5.6.15)}$$

for $x < \xi$, while

$$g(x|\xi) = Ce^{-ik_0(x-\xi)} + De^{ik_0(x-\xi)}, \qquad \textbf{(5.6.16)}$$

for $\xi < x$.

Let us examine Equation 5.6.15 more closely. The solution represents two propagating waves. Because $x < \xi$, the first term is a wave propagating out to infinity, while the second term gives a wave propagating in from infinity. This is seen most clearly by including the $e^{-i\omega t}$ term into Equation 5.6.15, or

$$g(x|\xi)e^{-i\omega t} = Ae^{-ik_0(x-\xi)-i\omega t} + Be^{ik_0(x-\xi)-i\omega t}. \qquad \textbf{(5.6.17)}$$

Because we have a source only at $x = \xi$, solutions that represent waves originating at infinity are nonphysical and we must discard them. This requirement that there are only outwardly

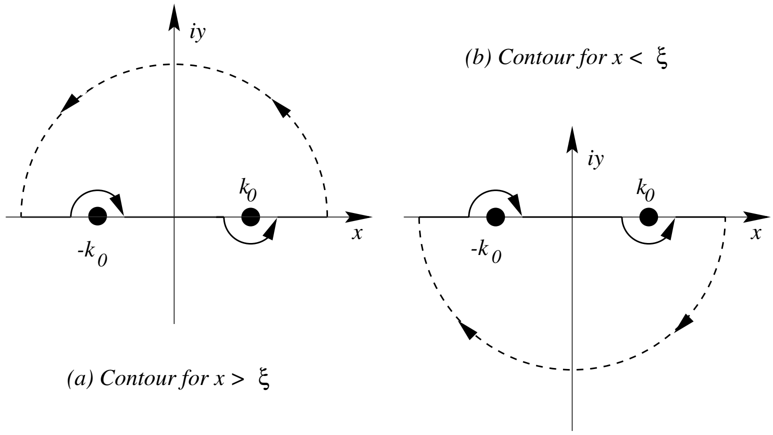


Figure 5.6.1: Contour used to evaluate Equation 5.6.21.

propagating wave solutions is commonly called *Sommerfeld's radiation condition*.[23] Similar considerations hold for Equation 5.6.16 and we must take $C = 0$.

To evaluate A and D, we use the continuity conditions on the Green's function:

$$g(\xi^+|\xi) = g(\xi^-|\xi), \quad \text{and} \quad g'(\xi^+|\xi) - g'(\xi^-|\xi) = -1, \tag{5.6.18}$$

or

$$A = D, \qquad \text{and} \qquad ik_0 D + ik_0 A = -1. \tag{5.6.19}$$

Therefore,

$$\boxed{g(x|\xi) = \frac{i}{2k_0} e^{ik_0|x-\xi|}.} \tag{5.6.20}$$

We can also solve Equation 5.6.14 by Fourier transforms. Assuming that the Fourier transform of $g(x|\xi)$ exists and denoting it by $G(k|\xi)$, we find that

$$G(k|\xi) = \frac{e^{-ik\xi}}{k^2 - k_0^2}, \qquad \text{and} \qquad g(x|\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{ik(x-\xi)}}{k^2 - k_0^2}\, dk. \tag{5.6.21}$$

Immediately we see that there is a problem with the singularities lying on the path of integration at $k = \pm k_0$. How do we avoid them?

There are four possible ways that we might circumvent the singularities. One of them is shown in Figure 5.6.1. Applying Jordan's lemma to close the line integral along the real axis (as shown in Figure 5.6.1),

$$\boxed{g(x|\xi) = \frac{1}{2\pi}\oint_C \frac{e^{iz(x-\xi)}}{z^2 - k_0^2}\, dz.} \tag{5.6.22}$$

[23] Sommerfeld, A., 1912: Die Greensche Funktion der Schwingungsgleichung. *Jahresber. Deutschen Math.- Vereinung*, **21**, 309–353.

Free-Space Green's Function for the Poisson and Helmholtz Equations

Dimension	Poisson Equation	Helmholtz Equation
One	no solution	$g(x\|\xi) = \dfrac{i}{2k_0} e^{ik_0\|x-\xi\|}$
Two	$g(x,y\|\xi,\eta) = -\dfrac{\ln(r)}{2\pi}$	$g(x,y\|\xi,\eta) = \dfrac{i}{4} H_0^{(1)}(k_0 r)$
$r = \sqrt{(x-\xi)^2 + (y-\eta)^2}$		

Note: For the Helmholtz equation, we have taken the temporal forcing to be $e^{-i\omega t}$ and $k_0 = \omega/c$.

For $x < \xi$,

$$g(x|\xi) = -i \operatorname{Res}\left[\frac{e^{iz(x-\xi)}}{z^2 - k_0^2}; -k_0\right] = \frac{i}{2k_0} e^{-ik_0(x-\xi)}, \tag{5.6.23}$$

while

$$g(x|\xi) = i \operatorname{Res}\left[\frac{e^{iz(x-\xi)}}{z^2 - k_0^2}; k_0\right] = \frac{i}{2k_0} e^{ik_0(x-\xi)}, \tag{5.6.24}$$

for $x > \xi$. A quick check shows that these solutions agree with Equation 5.6.20. If we try the three other possible paths around the singularities, we obtain incorrect solutions. □

• **Example 5.6.2: Free-space Green's function for the two-dimensional Helmholtz equation**

At this point, we have found two forms of the free-space Green's function for the one-dimensional Helmholtz equation. The first form is the analytic solution, Equation 5.6.20, while the second is the integral representation, Equation 5.6.21, where the line integration along the real axis is shown in Figure 5.6.1.

In the case of two dimensions, the Green's function[24] for the Helmholtz equation symmetric about the point (ξ, η) is the solution of the equation

$$\frac{d^2 g}{dr^2} + \frac{1}{r}\frac{dg}{dr} + k_0^2 g = -\frac{\delta(r)}{2\pi r}, \tag{5.6.25}$$

where $r = \sqrt{(x-\xi)^2 + (y-\eta)^2}$. The homogeneous form of Equation 5.6.25 is Bessel's differential equation of order zero. Consequently, the general solution in terms of Hankel functions is

$$g(\mathbf{r}|\mathbf{r}_0) = A\, H_0^{(1)}(k_0 r) + B\, H_0^{(2)}(k_0 r). \tag{5.6.26}$$

Why have we chosen to use Hankel functions rather than $J_0(\cdot)$ and $Y_0(\cdot)$? As we argued earlier, solutions to the Helmholtz equation must represent *outwardly* propagating waves (the Sommerfeld radiation condition). If we again assume that the temporal behavior is $e^{-i\omega t}$ and use the asymptotic expressions for Hankel functions, we see that $H_0^{(1)}(k_0 r)$ represents outwardly propagating waves and $B = 0$.

[24] For an alternative derivation, see Graff, K. F., 1991: *Wave Motion in Elastic Solids*. Dover Publications, Inc., pp. 284–285.

What is the value of A? Integrating Equation 5.6.26 over a small circle around the point $r = 0$ and taking the limit as the radius of the circle vanishes, $A = i/4$ and

$$\boxed{g(\mathbf{r}|\mathbf{r}_0) = \frac{i}{4} H_0^{(1)}(k_0 r).} \tag{5.6.27}$$

If a real function is needed, then the free-space Green's function equals the Neumann function $Y_0(k_0 r)$ divided by -4. □

• Example 5.6.3: Free-space Green's function for the two-dimensional Laplace equation

In this subsection, we find the free-space Green's function for Poisson's equation in two dimensions. This Green's function is governed by

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 g}{\partial \theta^2} = -\frac{\delta(r-\rho)\delta(\theta-\theta')}{r}. \tag{5.6.28}$$

If we now choose our coordinate system so that the origin is located at the point source, $r = \sqrt{(x-\xi)^2 + (y-\eta)^2}$ and $\rho = 0$. Multiplying both sides of this simplified Equation 5.6.28 by $r\,dr\,d\theta$ and integrating over a circle of radius ϵ, we obtain -1 on the right side from the surface integration over the delta functions. On the left side,

$$\int_0^{2\pi} r\frac{\partial g}{\partial r}\bigg|_{r=\epsilon} d\theta = -1. \tag{5.6.29}$$

The Green's function $g(r,\theta|0,\theta') = -\ln(r)/(2\pi)$ satisfies Equation 5.6.29.

To find an alternative form of the free-space Green's function when the point of excitation and the origin of the coordinate system do not coincide, we first note that

$$\delta(\theta-\theta') = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{in(\theta-\theta')}. \tag{5.6.30}$$

This suggests that the Green's function should be of the form

$$g(r,\theta|\rho,\theta') = \sum_{n=-\infty}^{\infty} g_n(r|\rho)e^{in(\theta-\theta')}. \tag{5.6.31}$$

Substituting Equation 5.6.30 and Equation 5.6.31 into Equation 5.6.29, we obtain the ordinary differential equation

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dg_n}{dr}\right) - \frac{n^2}{r^2}g_n = -\frac{\delta(r-\rho)}{2\pi r}. \tag{5.6.32}$$

The homogeneous solution to Equation 5.6.32 is

$$g_0(r|\rho) = \begin{cases} a, & 0 \le r \le \rho, \\ b\,\ln(r), & \rho \le r < \infty, \end{cases} \tag{5.6.33}$$

and

$$g_n(r|\rho) = \begin{cases} c\,(r/\rho)^n, & 0 \le r \le \rho\,, \\ d\,(\rho/r)^n, & \rho \le r < \infty, \end{cases} \tag{5.6.34}$$

if $n \neq 0$.

At $r = \rho$, the g_n's must be continuous, in which case,

$$a = b\ln(\rho), \qquad \text{and} \qquad c = d. \tag{5.6.35}$$

On the other hand,

$$\rho \frac{dg_n}{dr}\bigg|_{r=\rho^-}^{r=\rho^+} = -\frac{1}{2\pi}, \tag{5.6.36}$$

or

$$a = -\frac{\ln(\rho)}{2\pi}, \quad b = -\frac{1}{2\pi}, \quad \text{and} \quad c = d = \frac{1}{4\pi n}. \tag{5.6.37}$$

Therefore,

$$g(r,\theta|\rho,\theta') = -\frac{\ln(r_>)}{2\pi} + \frac{1}{2\pi}\sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{r_<}{r_>}\right)^n \cos[n(\theta-\theta')], \tag{5.6.38}$$

where $r_> = \max(r,\rho)$ and $r_< = \min(r,\rho)$.

We can simplify Equation 5.6.38 by noting that

$$\ln\bigl[1+\rho^2 - 2\rho\cos(\theta-\theta')\bigr] = -2\sum_{n=1}^{\infty} \frac{\rho^n \cos[n(\theta-\theta')]}{n}, \tag{5.6.39}$$

if $|\rho| < 1$. Applying this relationship to Equation 5.6.38, we find that

$$g(r,\theta|\rho,\theta') = -\frac{1}{4\pi}\ln\bigl[r^2+\rho^2-2r\rho\cos(\theta-\theta')\bigr]\,. \tag{5.6.40}$$

Note that when $\rho = 0$ we recover $g(r,\theta|0,\theta') = -\ln(r)/(2\pi)$. □

• Example 5.6.4: Two-dimensional Poisson equation over a rectangular domain

Consider the two-dimensional Poisson equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -f(x,y). \tag{5.6.41}$$

This equation arises in equilibrium problems, such as the static deflection of a rectangular membrane. In that case, $f(x,y)$ represents the external load per unit area, divided by the tension in the membrane. The solution $u(x,y)$ must satisfy certain boundary conditions. For the present, let us choose $u(0,y) = u(a,y) = 0$, and $u(x,0) = u(x,b) = 0$.

To find the Green's function for Equation 5.6.41 we must solve the partial differential equation

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = -\delta(x-\xi)\delta(y-\eta), \quad 0 < x,\xi < a, \quad 0 < y,\eta < b, \tag{5.6.42}$$

subject to the boundary conditions

$$g(0,y|\xi,\eta) = g(a,y|\xi,\eta) = g(x,0|\xi,\eta) = g(x,b|\xi,\eta) = 0. \tag{5.6.43}$$

From Equation 5.6.9,

$$u(x, y) = \int_0^a \int_0^b g(x, y|\xi, \eta) f(\xi, \eta)\, d\eta\, d\xi. \tag{5.6.44}$$

One approach to finding the Green's function is to expand it in terms of the eigenfunctions $\varphi(x, y)$ of the differential equation

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = -\lambda \varphi, \tag{5.6.45}$$

and the boundary conditions, Equation 5.6.43. The eigenvalues are

$$\lambda_{nm} = \frac{n^2\pi^2}{a^2} + \frac{m^2\pi^2}{b^2}, \tag{5.6.46}$$

where $n = 1, 2, 3, \ldots$, $m = 1, 2, 3, \ldots$, and the corresponding eigenfunctions are

$$\varphi_{nm}(x, y) = \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right). \tag{5.6.47}$$

Therefore, we seek $g(x, y|\xi, \eta)$ in the form

$$g(x, y|\xi, \eta) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} A_{nm} \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right). \tag{5.6.48}$$

Because the delta functions can be written

$$\delta(x-\xi)\delta(y-\eta) = \frac{4}{ab} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi\eta}{b}\right) \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right), \tag{5.6.49}$$

we find that

$$\left(\frac{n^2\pi^2}{a^2} + \frac{m^2\pi^2}{b^2}\right) A_{nm} = \frac{4}{ab} \sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi\eta}{b}\right), \tag{5.6.50}$$

after substituting Equations 5.6.48 and 5.6.49 into Equation 5.6.42, and setting the corresponding harmonics equal to each other. Therefore, the *bilinear formula* for the Green's function of Poisson's equation is

$$\boxed{g(x, y|\xi, \eta) = \frac{4}{ab} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{n\pi\xi}{a}\right) \sin\left(\frac{m\pi y}{b}\right) \sin\left(\frac{m\pi\eta}{b}\right)}{n^2\pi^2/a^2 + m^2\pi^2/b^2}.} \tag{5.6.51}$$

Thus, solutions to Poisson's equation can now be written as

$$u(x, y) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{a_{nm}}{n^2\pi^2/a^2 + m^2\pi^2/b^2} \sin\left(\frac{n\pi x}{a}\right) \sin\left(\frac{m\pi y}{b}\right), \tag{5.6.52}$$

where a_{nm} are the Fourier coefficients for the function $f(x,y)$:

$$a_{nm} = \frac{4}{ab}\int_0^a \int_0^b f(x,y)\sin\left(\frac{n\pi x}{a}\right)\sin\left(\frac{m\pi y}{b}\right)\,dy\,dx. \tag{5.6.53}$$

Another form of the Green's function can be obtained by considering each direction separately. To satisfy the boundary conditions along the edges $y = 0$ and $y = b$, we write the Green's function as the Fourier series

$$g(x,y|\xi,\eta) = \sum_{m=1}^{\infty} G_m(x|\xi)\sin\left(\frac{m\pi y}{b}\right), \tag{5.6.54}$$

where the coefficients $G_m(x|\xi)$ are left as undetermined functions of x, ξ, and m. Substituting this series into the partial differential equation for g, multiplying by $2\sin(n\pi y/b)/b$, and integrating over y, we find that

$$\frac{d^2G_n}{dx^2} - \frac{n^2\pi^2}{b^2}G_n = -\frac{2}{b}\sin\left(\frac{n\pi\eta}{b}\right)\delta(x-\xi). \tag{5.6.55}$$

This differential equation shows that the expansion coefficients $G_n(x|\xi)$ are one-dimensional Green's functions; we can find them, as we did in Section 5.2, by piecing together homogeneous solutions to Equation 5.6.55 that are valid over various intervals. For the region $0 \le x \le \xi$, the solution to this equation that vanishes at $x = 0$ is

$$G_n(x|\xi) = A_n\sinh\left(\frac{n\pi x}{b}\right), \tag{5.6.56}$$

where A_n is presently arbitrary. The corresponding solution for $\xi \le x \le a$ is

$$G_n(x|\xi) = B_n\sinh\left[\frac{n\pi(a-x)}{b}\right]. \tag{5.6.57}$$

Note that this solution vanishes at $x = a$. Because the Green's function must be continuous at $x = \xi$,

$$A_n\sinh\left(\frac{n\pi\xi}{b}\right) = B_n\sinh\left[\frac{n\pi(a-\xi)}{b}\right]. \tag{5.6.58}$$

On the other hand, the appropriate jump discontinuity of $G_n'(x|\xi)$ yields

$$-\frac{n\pi}{b}B_n\cosh\left[\frac{n\pi(a-\xi)}{b}\right] - \frac{n\pi}{b}A_n\cosh\left(\frac{n\pi\xi}{b}\right) = -\frac{2}{b}\sin\left(\frac{n\pi\eta}{b}\right). \tag{5.6.59}$$

Solving for A_n and B_n,

$$A_n = \frac{2}{n\pi}\sin\left(\frac{n\pi\eta}{b}\right)\frac{\sinh[n\pi(a-\xi)/b]}{\sinh(n\pi a/b)}, \tag{5.6.60}$$

and

$$B_n = \frac{2}{n\pi}\sin\left(\frac{n\pi\eta}{b}\right)\frac{\sinh(n\pi\xi/b)}{\sinh(n\pi a/b)}. \tag{5.6.61}$$

This yields the Green's function

$$g(x,y|\xi,\eta) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sinh[n\pi(a-x_>)/b]\sinh(n\pi x_</b)}{n\ \sinh(n\pi a/b)}\sin\left(\frac{n\pi\eta}{b}\right)\sin\left(\frac{n\pi y}{b}\right), \tag{5.6.62}$$

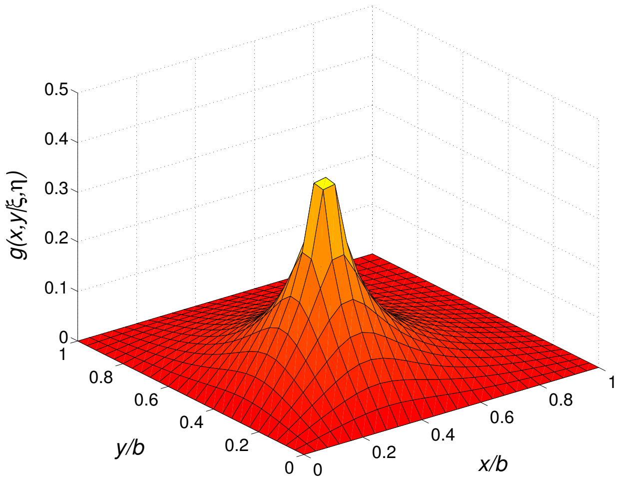


Figure 5.6.2: The Green's function, Equation 5.6.62 or Equation 5.6.63, for the planar Poisson equation over a rectangular area with Dirichlet boundary conditions on all sides when $a = b$ and $\xi/b = \eta/b = 0.3$.

where $x_> = \max(x, \xi)$ and $x_< = \min(x, \xi)$. Figure 5.6.2 illustrates Equation 5.6.62 in the case of a square domain with $\xi/b = \eta/b = 0.3$.

If we began with a Fourier expansion in the y-direction, we would have obtained

$$g(x, y|\xi, \eta) = \frac{2}{\pi} \sum_{m=1}^{\infty} \frac{\sinh[m\pi(b - y_>)/a] \sinh(m\pi y_</a)}{m \sinh(m\pi b/a)} \sin\left(\frac{m\pi\xi}{a}\right) \sin\left(\frac{m\pi x}{a}\right), \quad \textbf{(5.6.63)}$$

where $y_> = \max(y, \eta)$ and $y_< = \min(y, \eta)$. □

• Example 5.6.5: Two-dimensional Helmholtz equation over a rectangular domain

The problem to be solved is

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} + k_0^2 g = -\delta(x - \xi)\delta(y - \eta), \quad \textbf{(5.6.64)}$$

where $0 < x, \xi < a$, and $0 < y, \eta < b$, subject to the boundary conditions that

$$g(0, y|\xi, \eta) = g(a, y|\xi, \eta) = g(x, 0|\xi, \eta) = g(x, b|\xi, \eta) = 0. \quad \textbf{(5.6.65)}$$

We use the same technique to solve Equation 5.6.64 as we did in the previous example by assuming that the Green's function has the form

$$g(x, y|\xi, \eta) = \sum_{m=1}^{\infty} G_m(x|\xi) \sin\left(\frac{m\pi y}{b}\right), \quad \textbf{(5.6.66)}$$

where the coefficients $G_m(x|\xi)$ are undetermined functions of x, ξ, and η. Substituting this series into Equation 5.6.64, multiplying by $2\sin(n\pi y/b)/b$, and integrating over y, we find that

$$\frac{d^2 G_n}{dx^2} - \left(\frac{n^2\pi^2}{b^2} - k_0^2\right) G_n = -\frac{2}{b} \sin\left(\frac{n\pi\eta}{b}\right) \delta(x - \xi). \quad \textbf{(5.6.67)}$$

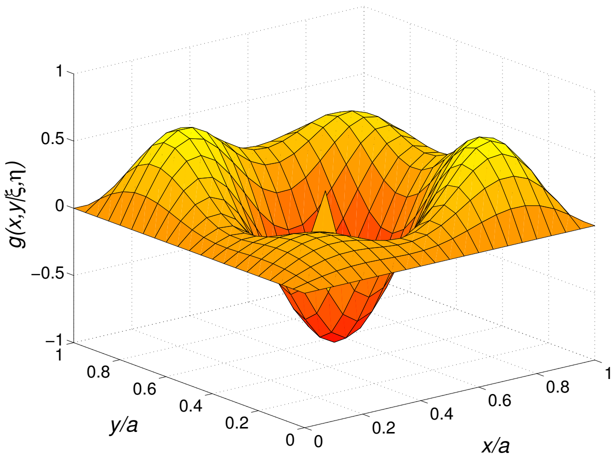


Figure 5.6.3: The Green's function, Equation 5.6.72, for Helmholtz's equation over a rectangular region with a Dirichlet boundary condition on the sides when $a = b$, $k_0 a = 10$, and $\xi/a = \eta/a = 0.35$.

The first method for solving Equation 5.6.67 involves writing

$$\delta(x-\xi) = \frac{2}{a}\sum_{m=1}^{\infty} \sin\left(\frac{m\pi\xi}{a}\right)\sin\left(\frac{m\pi x}{a}\right), \tag{5.6.68}$$

and

$$G_n(x|\xi) = \frac{2}{a}\sum_{m=1}^{\infty} a_{nm}\sin\left(\frac{m\pi x}{a}\right). \tag{5.6.69}$$

Upon substituting Equations 5.6.68 and 5.6.69 into Equation 5.6.67, we obtain

$$\begin{aligned}\sum_{m=1}^{\infty}\left(k_0^2 - \frac{m^2\pi^2}{a^2} - \frac{n^2\pi^2}{b^2}\right) a_{nm}\sin\left(\frac{m\pi x}{a}\right) \\ = -\frac{4}{ab}\sum_{m=1}^{\infty} \sin\left(\frac{n\pi\eta}{b}\right)\sin\left(\frac{m\pi\xi}{a}\right)\sin\left(\frac{m\pi x}{a}\right).\end{aligned} \tag{5.6.70}$$

Matching similar harmonics,

$$a_{nm} = \frac{4\sin(m\pi\xi/a)\sin(n\pi\eta/b)}{ab(m^2\pi^2/a^2 + n^2\pi^2/b^2 - k_0^2)}, \tag{5.6.71}$$

and the *bilinear form of the Green's function* is

$$\boxed{g(x,y|\xi,\eta) = \frac{4}{ab}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \frac{\sin(m\pi\xi/a)\sin(n\pi\eta/b)\sin(m\pi x/a)\sin(n\pi y/b)}{m^2\pi^2/a^2 + n^2\pi^2/b^2 - k_0^2}.} \tag{5.6.72}$$

See Figure 5.6.3. The bilinear form of the Green's function for the two-dimensional Helmholtz equation with Neumann boundary conditions is left as Problem 8.

As in the previous example, we can construct an alternative to the bilinear form of the Green's function, Equation 5.6.72, by writing Equation 5.6.67 as

$$\frac{d^2G_n}{dx^2} - k_n^2 G_n = -\frac{2}{b}\sin\left(\frac{n\pi\eta}{b}\right)\delta(x-\xi), \tag{5.6.73}$$

where $k_n^2 = n^2\pi^2/b^2 - k_0^2$. The homogeneous solution to Equation 5.6.73 is now

$$G_n(x|\xi) = \begin{cases} A_n \sinh(k_n x), & 0 \le x \le \xi, \\ B_n \sinh[k_n(a-x)], & \xi \le x \le a. \end{cases} \tag{5.6.74}$$

This solution satisfies the boundary conditions at both endpoints.

Because $G_n(x|\xi)$ must be continuous at $x = \xi$,

$$A_n \sinh(k_n\xi) = B_n \sinh[k_n(a-\xi)]. \tag{5.6.75}$$

On the other hand, the jump discontinuity involving $G_n'(x|\xi)$ yields

$$-k_n B_n \cosh[k_n(a-\xi)] - k_n A_n \cosh(k_n\xi) = -\frac{2}{b}\sin\left(\frac{n\pi\eta}{b}\right). \tag{5.6.76}$$

Solving for A_n and B_n,

$$A_n = \frac{2}{bk_n}\sin\left(\frac{n\pi\eta}{b}\right)\frac{\sinh[k_n(a-\xi)]}{\sinh(k_n a)}, \tag{5.6.77}$$

and

$$B_n = \frac{2}{bk_n}\sin\left(\frac{n\pi\eta}{b}\right)\frac{\sinh(k_n\xi)}{\sinh(k_n a)}. \tag{5.6.78}$$

This yields the Green's function

$$\begin{aligned} g(x,y|\xi,\eta) = {} & \frac{2}{b}\sum_{n=1}^{N}\frac{\sin[\kappa_n(a-x_>)]\sin(\kappa_n x_<)}{\kappa_n\,\sin(\kappa_n a)}\sin\left(\frac{n\pi\eta}{b}\right)\sin\left(\frac{n\pi y}{b}\right) \\ & + \frac{2}{b}\sum_{n=N+1}^{\infty}\frac{\sinh[k_n(a-x_>)]\sinh(k_n x_<)}{k_n\,\sinh(k_n a)}\sin\left(\frac{n\pi\eta}{b}\right)\sin\left(\frac{n\pi y}{b}\right), \end{aligned} \tag{5.6.79}$$

where $x_> = \max(x,\xi)$ and $x_< = \min(x,\xi)$. Here N denotes the largest value of n such that $k_n^2 < 0$, and $\kappa_n^2 = k_0^2 - n^2\pi^2/b^2$. If we began with a Fourier expansion in the y direction, we would have obtained

$$\begin{aligned} g(x,y|\xi,\eta) = {} & \frac{2}{a}\sum_{m=1}^{M}\frac{\sin[\kappa_m(b-y_>)]\sin(\kappa_m y_<)}{\kappa_m\,\sin(\kappa_m b)}\sin\left(\frac{m\pi\xi}{a}\right)\sin\left(\frac{m\pi x}{a}\right) \\ & + \frac{2}{a}\sum_{m=M+1}^{\infty}\frac{\sinh[k_m(b-y_>)]\sinh(k_m y_<)}{k_m\,\sinh(k_m b)}\sin\left(\frac{m\pi\xi}{a}\right)\sin\left(\frac{m\pi x}{a}\right), \end{aligned} \tag{5.6.80}$$

where M denotes the largest value of m such that $k_m^2 < 0$, $k_m^2 = m^2\pi^2/a^2 - k_0^2$, $\kappa_m^2 = k_0^2 - m^2\pi^2/a^2$, $y_< = \min(y,\eta)$, and $y_> = \max(y,\eta)$. □

• **Example 5.6.6: Two-dimensional Helmholtz equation over a circular disk**

In this example, we find the Green's function for the Helmholtz equation when the domain consists of the circular region $0 < r < a$. The Green's function is governed by the partial differential equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 g}{\partial \theta^2} + k_0^2 g = -\frac{\delta(r-\rho)\delta(\theta-\theta')}{r}, \tag{5.6.81}$$

where $0 < r, \rho < a$, and $0 \le \theta, \theta' \le 2\pi$, with the boundary conditions

$$\lim_{r\to 0} |g(r,\theta|\rho,\theta')| < \infty, \qquad g(a,\theta|\rho,\theta') = 0, \qquad 0 \le \theta, \theta' \le 2\pi. \tag{5.6.82}$$

The Green's function must be periodic in θ.

We begin by noting that

$$\delta(\theta-\theta') = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}\cos[n(\theta-\theta')] = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\cos[n(\theta-\theta')]. \tag{5.6.83}$$

Therefore, the solution has the form

$$g(r,\theta|\rho,\theta') = \sum_{n=-\infty}^{\infty} g_n(r|\rho)\cos[n(\theta-\theta')]. \tag{5.6.84}$$

Substituting Equation 5.6.83 and Equation 5.6.84 into Equation 5.6.81 and simplifying, we find that

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dg_n}{dr}\right) - \frac{n^2}{r^2}g_n + k_0^2 g_n = -\frac{\delta(r-\rho)}{2\pi r}. \tag{5.6.85}$$

The solution to Equation 5.6.85 is the Fourier-Bessel series

$$g_n(r|\rho) = \sum_{m=1}^{\infty} A_{nm} J_n\left(\frac{k_{nm} r}{a}\right), \tag{5.6.86}$$

where k_{nm} is the mth root of $J_n(k) = 0$. Upon substituting Equation 5.6.86 into Equation 5.6.85 and solving for A_{nm}, we have that

$$(k_0^2 - k_{nm}^2/a^2)A_{nm} = -\frac{1}{\pi a^2 J'^2_n(k_{nm})}\int_0^a \delta(r-\rho) J_n\left(\frac{k_{nm} r}{a}\right)\,dr, \tag{5.6.87}$$

or

$$A_{nm} = \frac{J_n(k_{nm}\rho/a)}{\pi(k_{nm}^2 - k_0^2 a^2) J'^2_n(k_{nm})}. \tag{5.6.88}$$

Thus, the Green's function[25] is

$$\boxed{g(r,\theta|\rho,\theta') = \frac{1}{\pi}\sum_{n=-\infty}^{\infty}\sum_{m=1}^{\infty}\frac{J_n(k_{nm}\rho/a)J_n(k_{nm}r/a)}{(k_{nm}^2 - k_0^2 a^2)J'^2_n(k_{nm})}\cos[n(\theta-\theta')].} \tag{5.6.89}$$

---

[25] For an example of its use, see Zhang, D. R., and C. F. Foo, 1999: Fields analysis in a solid magnetic toroidal core with circular cross section based on Green's function. *IEEE Trans. Magnetics*, **35**, 3760–3762.

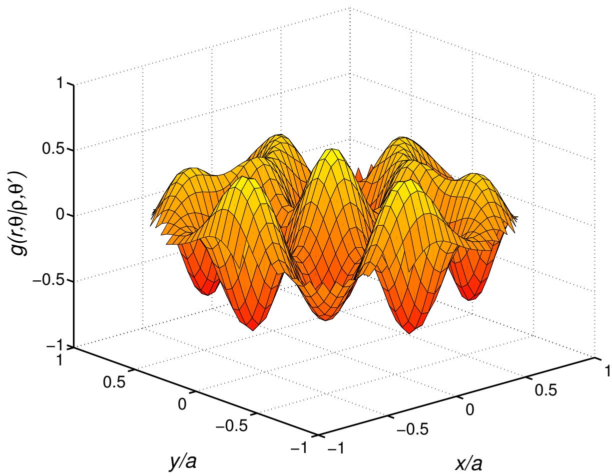


Figure 5.6.4: The Green's function, Equation 5.6.89, for Helmholtz's equation within a circular disk with a Dirichlet boundary condition on the rim when $k_0a = 10$, $\rho/a = 0.35\sqrt{2}$, and $\theta' = \pi/4$.

See Figure 5.6.4.

Problems

1. Using a Fourier sine expansion in the x-direction, construct the Green's function governed by the planar Poisson equation

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = -\delta(x-\xi)\delta(y-\eta), \quad 0 < x, \xi < a, \quad -\infty < y, \eta < \infty,$$

with the Dirichlet boundary conditions

$$g(0, y|\xi, \eta) = g(a, y|\xi, \eta) = 0, \quad -\infty < y < \infty,$$

and the conditions at infinity that

$$\lim_{|y|\to\infty} g(x, y|\xi, \eta) \to 0, \qquad 0 < x < a.$$

2. Using a generalized Fourier expansion in the x-direction, construct the Green's function governed by the planar Poisson equation

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = -\delta(x-\xi)\delta(y-\eta), \quad 0 < x, \xi < a, \quad -\infty < y, \eta < \infty,$$

with the Neumann and Dirichlet boundary conditions

$$g_x(0, y|\xi, \eta) = g(a, y|\xi, \eta) = 0, \quad -\infty < y < \infty,$$

and the conditions at infinity that

$$\lim_{|y|\to\infty} g(x, y|\xi, \eta) \to 0, \qquad 0 < x < a.$$

3. Using a Fourier sine expansion in the y-direction, show that the Green's function governed by the planar Poisson equation

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = -\delta(x-\xi)\delta(y-\eta), \quad 0 < x, \xi < a, \quad 0 < y, \eta < b,$$

with the boundary conditions

$$g(x, 0|\xi, \eta) = g(x, b|\xi, \eta) = 0,$$

and

$$g(0, y|\xi, \eta) = g_x(a, y|\xi, \eta) + \beta\, g(a, y|\xi, \eta) = 0, \qquad \beta \geq 0,$$

is

$$g(x, y|\xi, \eta) = \sum_{n=1}^{\infty} \frac{\sinh(\nu x_<)\,\{\nu\cosh[\nu(a - x_>)] + \beta\sinh[\nu(a - x_>)]\}}{\nu^2\cosh(\nu a) + \beta\nu\sinh(\nu a)} \sin\left(\frac{n\pi\eta}{b}\right)\sin\left(\frac{n\pi y}{b}\right),$$

where $\nu = n\pi/b$, $x_> = \max(x, \xi)$, and $x_< = \min(x, \xi)$.

4. Using the Fourier series representation of the delta function in a circular domain:

$$\delta(\theta - \theta') = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}\cos[n(\theta - \theta')], \qquad 0 \leq \theta, \theta' \leq 2\pi,$$

construct the Green's function governed by the planar Poisson equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 g}{\partial \theta^2} = -\frac{\delta(r-\rho)\delta(\theta-\theta')}{r},$$

where $a < r, \rho < b$, and $0 \leq \theta, \theta' \leq 2\pi$, subject to the boundary conditions $g(a, \theta|\rho, \theta') = g(b, \theta|\rho, \theta') = 0$ and periodicity in θ.

5. Construct the Green's function governed by the planar Poisson equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 g}{\partial \theta^2} = -\frac{\delta(r-\rho)\delta(\theta-\theta')}{r},$$

where $0 < r, \rho < \infty$, and $0 < \theta, \theta' < \beta$, subject to the boundary conditions that $g(r, 0|\rho, \theta')$ $= g(r, \beta|\rho, \theta') = 0$ for all r. Hint:

$$\delta(\theta - \theta') = \frac{2}{\beta}\sum_{n=1}^{\infty}\sin\left(\frac{n\pi\theta'}{\beta}\right)\sin\left(\frac{n\pi\theta}{\beta}\right).$$

6. Construct the Green's function governed by the planar Poisson equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial g}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 g}{\partial \theta^2} = -\frac{\delta(r-\rho)\delta(\theta-\theta')}{r},$$

where $0 < r, \rho < a$, and $0 < \theta, \theta' < \beta$, subject to the boundary conditions $g(r, 0|\rho, \theta') = g(r, \beta|\rho, \theta') = g(a, \theta|\rho, \theta') = 0$. Hint:

$$\delta(\theta - \theta') = \frac{2}{\beta} \sum_{n=1}^{\infty} \sin\left(\frac{n\pi\theta'}{\beta}\right) \sin\left(\frac{n\pi\theta}{\beta}\right).$$

7. Using a Fourier sine series in the z-direction and the fact that

$$\delta(x - b) = \frac{2b}{a^2} \sum_{k=1}^{\infty} \frac{J_0(\mu_k b/a) J_0(\mu_k x/a)}{J_1^2(\mu_k)}, \qquad 0 < x, b < a,$$

where μ_k is the kth positive root of $J_0(\mu) = 0$, find the Green's function governed by the axisymmetric Poisson equation

$$\frac{\partial^2 g}{\partial r^2} + \frac{1}{r}\frac{\partial g}{\partial r} + \frac{\partial^2 g}{\partial z^2} = -\frac{\delta(r-\rho)\delta(z-\zeta)}{2\pi r},$$

where $0 < r, \rho < a$, and $0 < z, \zeta < L$, subject to the boundary conditions

$$g(r, 0|\rho, \zeta) = g(r, L|\rho, \zeta) = 0, \quad 0 < r < a,$$

and

$$\lim_{r \to 0} |g(r, z|\rho, \zeta)| < \infty, \quad g(a, z|\rho, \zeta) = 0, \quad 0 < z < L.$$

8. Following Example 5.6.5 except for using Fourier cosine series, construct the Green's function[26] governed by the planar Helmholtz equation

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} + k_0^2 g = -\delta(x-\xi)\delta(y-\eta), \quad 0 < x, \xi < a, \quad 0 < y, \eta < b,$$

subject to the Neumann boundary conditions

$$g_x(0, y|\xi, \eta) = g_x(a, y|\xi, \eta) = 0, \quad 0 < y < b,$$

and

$$g_y(x, 0|\xi, \eta) = g_y(x, b|\xi, \eta) = 0, \quad 0 < x < a.$$

9. Using Fourier sine transforms,

$$g(x, y|\xi, \eta) = \frac{2}{\pi} \int_0^{\infty} G(k, y|\xi, \eta) \sin(kx)\, dk,$$

where

$$G(k, y|\xi, \eta) = \int_0^{\infty} g(x, y|\xi, \eta) \sin(kx)\, dx,$$

[26] Kulkarni et al. (Kulkarni, S., F. G. Leppington, and E. G. Broadbent, 2001: Vibrations in several interconnected regions: A comparison of SEA, ray theory and numerical results. *Wave Motion*, **33**, 79–96) solved this problem when the domain has two different, constant k_0^2's.

find the Green's function governed by

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} = -\delta(x-\xi)\delta(y-\eta),$$

for the quarter space $0 < x, y$, with the boundary conditions

$$g(0, y|\xi, \eta) = g(x, 0|\xi, \eta) = 0,$$

and

$$\lim_{x,y\to\infty} g(x, y|\xi, \eta) \to 0.$$

Step 1: Taking the Fourier sine transform in the x direction, show that the partial differential equation reduces to the ordinary differential equation

$$\frac{d^2 G}{dy^2} - k^2 G = -\sin(k\xi)\delta(y-\eta),$$

with the boundary conditions

$$G(k, 0|\xi, \eta) = 0, \qquad \text{and} \qquad \lim_{y\to\infty} G(k, y|\xi, \eta) \to 0.$$

Step 2: Show that the particular solution to the ordinary differential equation in Step 1 is

$$G_p(k, y|\xi, \eta) = \frac{\sin(k\xi)}{2k} e^{-k|y-\eta|}.$$

You may want to review Example 5.2.8.

Step 3: Find the homogeneous solution to the ordinary differential equation in Step 1 so that the general solution satisfies the boundary conditions. Show that the general solution is

$$G(k, y|\xi, \eta) = \frac{\sin(k\xi)}{2k} \left[e^{-k|y-\eta|} - e^{-k(y+\eta)}\right].$$

Step 4: Taking the inverse, show that

$$g(x, y|\xi, \eta) = \frac{1}{\pi} \int_0^\infty \left[e^{-k|y-\eta|} - e^{-k(y+\eta)}\right] \sin(k\xi) \sin(kx) \frac{dk}{k}.$$

Step 5: Performing the integration,[27] show that

$$g(x, y|\xi, \eta) = -\frac{1}{4\pi} \ln\left\{ \frac{[(x-\xi)^2 + (y-\eta)^2][(x+\xi)^2 + (y+\eta)^2]}{[(x-\xi)^2 + (y+\eta)^2][(x+\xi)^2 + (y-\eta)^2]} \right\}.$$

---

[27] Gradshteyn, I. S., and I. M. Ryzhik, 1965: *Table of Integrals, Series, and Products*. Academic Press, Section 3.947, Formula 1.

10. Find the free-space Green's function[28] governed by

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} - g = -\delta(x-\xi)\delta(y-\eta), \qquad -\infty < x, y, \xi, \eta < \infty.$$

Step 1: Introducing the Fourier transform

$$g(x, y|\xi, \eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(k, y|\xi, \eta)e^{ikx}\, dk,$$

where

$$G(k, y|\xi, \eta) = \int_{-\infty}^{\infty} g(x, y|\xi, \eta)e^{-ikx}\, dx,$$

show that the governing partial differential equation can be transformed into the ordinary differential equation

$$\frac{d^2G}{dy^2} - \left(k^2+1\right)G = -e^{-ik\xi}\delta(y-\eta).$$

Step 2: Introducing the Fourier transform in the y-direction,

$$G(k, y|\xi, \eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{G}(k, \ell|\xi, \eta)e^{i\ell y}\, d\ell,$$

where

$$\overline{G}(k, \ell|\xi, \eta) = \int_{-\infty}^{\infty} G(k, y|\xi, \eta)e^{-i\ell y}\, dy,$$

solve the ordinary differential equation in Step 1 and show that

$$G(k, y|\xi, \eta) = \frac{e^{-ik\xi}}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\ell(y-\eta)}}{k^2+\ell^2+1}\, d\ell.$$

Step 3: Complete the problem by showing that

$$\begin{aligned} g(x, y|\xi, \eta) &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{ik(x-\xi)}e^{i\ell(y-\eta)}}{k^2+\ell^2+1}\, d\ell\, dk \\ &= \frac{1}{4\pi^2}\int_0^{\infty}\int_0^{2\pi} \frac{e^{ir\kappa\cos(\theta-\varphi)}}{\kappa^2+1}\kappa\, d\theta\, d\kappa \\ &= \frac{1}{2\pi}\int_0^{\infty} \frac{J_0(r\kappa)}{\kappa^2+1}\kappa\, d\kappa = \frac{K_0(r)}{2\pi}, \end{aligned}$$

where $r = \sqrt{(x-\xi)^2+(y-\eta)^2}$, $k = \kappa\cos(\theta)$, $\ell = \kappa\sin(\theta)$, $x - \xi = r\cos(\varphi)$, and $y - \eta = r\sin(\varphi)$. You will need to use integral tables[29] to obtain the final result.

---

[28] For its use, see Geisler, J. E., 1970: Linear theory of the response of a two layer ocean to a moving hurricane. *Geophys. Fluid Dyn.*, **1**, 249–272.

[29] Gradshteyn and Ryzhik, op. cit., Section 6.532, Formula 6.

11. Find the free-space Green's function governed by

$$\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2} - \frac{\partial g}{\partial x} = -\delta(x-\xi)\delta(y-\eta), \qquad -\infty < x, y, \xi, \eta < \infty.$$

Step 1: By introducing $\varphi(x, y|\xi, \eta)$ such that

$$g(x, y|\xi, \eta) = e^{x/2}\varphi(x, y|\xi, \eta),$$

show that the partial differential equation for $g(x, y|\xi, \eta)$ becomes

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} - \frac{\varphi}{4} = -e^{-\xi/2}\delta(x-\xi)\delta(y-\eta).$$

Step 2: After taking the Fourier transform with respect to x of the partial differential equation in Step 1, show that it becomes the ordinary differential equation

$$\frac{d^2\Phi}{dy^2} - \left(k^2 + \tfrac{1}{4}\right)\Phi = -e^{-\xi/2 - ik\xi}\delta(y-\eta).$$

Step 3: Introducing the same transformation as in Step 3 of the previous problem, show that

$$\Phi(k, y|\xi, \eta) = \frac{e^{-\xi/2-ik\xi}}{2\pi}\int_{-\infty}^{\infty} \frac{e^{i\ell(y-\eta)}}{k^2 + \ell^2 + \frac{1}{4}}\, d\ell,$$

and

$$\varphi(x, y|\xi, \eta) = \frac{e^{-\xi/2}}{2\pi} K_0(\tfrac{1}{2}r),$$

where $r = \sqrt{(x-\xi)^2 + (y-\eta)^2}$.

Step 4: Using the transformation introduced in Step 1, show that

$$g(x, y|\xi, \eta) = \frac{e^{(x-\xi)/2}}{2\pi} K_0(\tfrac{1}{2}r).$$

5.7 GALERKIN METHOD

In the previous sections we developed various analytic expressions for Green's functions. We close this chapter by showing how to construct a numerical approximation.

Finite elements can be used to solve differential equations by introducing subdomains known as *finite elements* rather than a grid of nodal points. The solution is then represented within each element by an interpolating polynomial. Unlike finite difference schemes that are constructed from Taylor expansions, the theory behind finite elements introduces concepts from functional analysis and variational methods to formulate the algebraic equations.

There are several paths that lead to the same finite element formulation. The two most common techniques are the Galerkin and collocation methods. In this section we focus on the *Galerkin method*. This method employs a rational polynomial, called a *basis function*, that satisfies the boundary conditions.

We begin by considering the Sturm-Liouville problem governed by

$$\frac{d^2\psi_n}{dx^2} + \lambda_n \psi_n = 0, \qquad 0 < x < L, \tag{5.7.1}$$

subject to the boundary conditions $\psi_n(0) = \psi_n(L) = 0$. Although we could solve this problem exactly, we will pretend that we cannot. Rather, we will assume that we can express it by

$$\psi_n(x) = \sum_{j=1}^{N} \alpha_{nj} f_j(x), \tag{5.7.2}$$

where $f_j(x)$ is our *basis function*. Clearly, it is desirable that $f_j(0) = f_j(L) = 0$.

How do we compute α_{nj}? We begin by multiplying Equation 5.7.1 by $f_i(x)$ and integrating the resulting equation from 0 and L. This yields

$$\int_0^L f_i(x) \frac{d^2\psi_n}{dx^2}\, dx + \lambda_n \int_0^L f_i(x)\psi_n(x)\, dx = 0, \tag{5.7.3}$$

where $i = 1, 2, 3, \ldots, N$. Next, we substitute Equation 5.7.2 and find that

$$\sum_{j=1}^{N} \left[\int_0^L f_i(x) f_j''(x)\, dx + \lambda_n \int_0^L f_i(x) f_j(x)\, dx \right] \alpha_{nj} = 0. \tag{5.7.4}$$

We can write Equation 5.7.4 as

$$(A + \lambda_n B)\mathbf{d} = \mathbf{0}, \tag{5.7.5}$$

where

$$a_{ij} = \int_0^L f_i(x) f_j''(x)\, dx = -\int_0^L f_i'(x) f_j'(x)\, dx, \tag{5.7.6}$$

$$b_{ij} = \int_0^L f_i(x) f_j(x)\, dx, \tag{5.7.7}$$

and the vector $\mathbf{d}$ contains the elements α_{nj}.

There are several obvious choices for $f_j(x)$:

• Example 5.7.1

The simplest choice for $f_j(x) = \sin(j\pi x/L)$. If we select $N = 2$, Equation 5.7.2 becomes

$$\psi_n(x) = \alpha_{n1} \sin\left(\frac{\pi x}{L}\right) + \alpha_{n2} \sin\left(\frac{2\pi x}{L}\right). \tag{5.7.8}$$

From Equation 5.7.6 and Equation 5.7.7,

$$a_{ij} = -\left(\frac{j\pi}{L}\right)^2 \int_0^L \sin\left(\frac{i\pi x}{L}\right) \sin\left(\frac{j\pi x}{L}\right) dx, \qquad i = 1, 2,\ j = 1, 2; \tag{5.7.9}$$

and

$$b_{ij} = \int_0^L \sin\left(\frac{i\pi x}{L}\right) \sin\left(\frac{j\pi x}{L}\right) dx, \qquad i = 1, 2,\ j = 1, 2. \tag{5.7.10}$$

Performing the integrations, $a_{12} = a_{21} = b_{12} = b_{21} = 0$, $a_{11} = -\pi^2/(2L)$, $a_{22} = -2\pi^2/L$, and $b_{11} = b_{22} = L/2$.

Returning to Equation 5.7.5, it becomes

$$\begin{pmatrix} -\pi^2/2 + \lambda_n L^2/2 & 0 \\ 0 & -2\pi^2 + \lambda_n L^2/2 \end{pmatrix} \begin{pmatrix} \alpha_{n1} \\ \alpha_{n2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.7.11}$$

In order for Equation 5.7.11 to have a unique solution,

$$\begin{vmatrix} -\pi^2/2 + \lambda_n L^2/2 & 0 \\ 0 & -2\pi^2 + \lambda_n L^2/2 \end{vmatrix} = 0. \tag{5.7.12}$$

Equation 5.7.12 yields $4\lambda_1 = \lambda_2 = 4\pi^2/L^2$.

In summary,

$$\psi_1(x) = \sin(\pi x/L), \qquad \lambda_1 = \pi^2/L^2; \tag{5.7.13}$$

and

$$\psi_2(x) = \sin(2\pi x/L), \qquad \lambda_2 = 4\pi^2/L^2, \tag{5.7.14}$$

with $\alpha_{12} = \alpha_{21} = 0$. Here we have chosen that $\alpha_{11} = \alpha_{22} = 1$. □

• Example 5.7.2

Another possible choice for $f_j(x)$ involves polynomials of the form $(1 - x/L)(x/L)^j$ with $j = 1, 2$. Unlike the previous example, we have *nonorthogonal* basis functions here. Note that $f_j(0) = f_j(L) = 0$. Therefore, Equation 5.7.2 becomes

$$\psi_n(x) = \alpha_{n1}(1 - x/L)(x/L) + \alpha_{n2}(1 - x/L)(x/L)^2. \tag{5.7.15}$$

From Equation 5.7.6 and Equation 5.7.7,

$$a_{ij} = -\frac{1}{L^2}\int_0^L \left(1 - \frac{x}{L}\right)\left(\frac{x}{L}\right)^i \left[j(j-1)\left(\frac{x}{L}\right)^{j-2} - j(j+1)\left(\frac{x}{L}\right)^{j-1} \right] dx \tag{5.7.16}$$

$$= \frac{1}{L}\left[\frac{j(j-1)}{i+j-1} - \frac{j(j-1)}{i+j} - \frac{j(j+1)}{i+j} + \frac{j(j+1)}{i+j+1}\right], \tag{5.7.17}$$

with $i = 1, 2$ and $j = 1, 2$. Similarly,

$$b_{ij} = \int_0^L \left(1 - \frac{x}{L}\right)\left(\frac{x}{L}\right)^i \left(1 - \frac{x}{L}\right)\left(\frac{x}{L}\right)^j dx \tag{5.7.18}$$

$$= L\left[\frac{1}{i+j+1} - \frac{2}{i+j+2} + \frac{1}{i+j+3}\right]. \tag{5.7.19}$$

Performing the computations, $a_{11} = -1/(3L)$, $a_{12} = a_{21} = -1/(6L)$, $a_{22} = -2/(15L)$, $b_{11} = L/30$, $b_{12} = b_{21} = L/60$, and $b_{22} = L/105$.

Returning to Equation 5.7.5, it becomes

$$\begin{pmatrix} -1/3 + \lambda_n L^2/30 & -1/6 + \lambda_n L^2/60 \\ -1/6 + \lambda_n L^2/60 & -2/15 + \lambda_n L^2/105 \end{pmatrix} \begin{pmatrix} \alpha_{n1} \\ \alpha_{n2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.7.20}$$

In order for Equation 5.7.20 to have a unique solution,

$$\begin{vmatrix} -1/3 + \lambda_n L^2/30 & -1/6 + \lambda_n L^2/60 \\ -1/6 + \lambda_n L^2/60 & -2/15 + \lambda_n L^2/105 \end{vmatrix} = 0. \qquad (\mathbf{5.7.21})$$

Equation 5.7.21 yields $\lambda_1 L^2 = 10$ and $\lambda_2 L^2 = 42$. Note how close these values of λ are to those found in the previous example. Returning to Equation 5.7.20, we find that $\alpha_{11} = 1$, $\alpha_{12} = 0$, $\alpha_{22} = 1$, and $\alpha_{21} = -1/2$.

In summary,

$$\psi_1(x) = \left(1 - \frac{x}{L}\right)\frac{x}{L}, \quad \lambda_1 = \frac{10}{L^2}; \qquad (\mathbf{5.7.22})$$

and

$$\psi_2(x) = -\frac{1}{2}\left(1 - \frac{x}{L}\right)\frac{x}{L} + \left(1 - \frac{x}{L}\right)\left(\frac{x}{L}\right)^2, \quad \lambda_2 = \frac{42}{L^2}. \qquad (\mathbf{5.7.23})$$

Because $f_j(x)$ are linearly independent, their use in the Galerkin expansion is quite acceptable. However, because these functions are not particularly orthogonal, their usefulness will become more difficult as N increases. Consequently, the choice of orthogonal functions is often best. □

How do we employ the Galerkin technique to approximate Green's functions? We begin by considering the inhomogeneous heat conduction problem:

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = F(x,t), \qquad 0 < x < L, \quad 0 < t, \qquad (\mathbf{5.7.24})$$

with the boundary conditions

$$u(0,t) = u(L,t) = 0, \qquad 0 < t, \qquad (\mathbf{5.7.25})$$

and the initial condition $u(x,0) = 0$, $0 < x < L$.

Let us write the solution to this problem as

$$u(x,t) = \sum_{n=1}^{N} c_n(t)\psi_n(x)e^{-\lambda_n t}. \qquad (\mathbf{5.7.26})$$

Direct substitution of Equation 5.7.26 into Equation 5.7.24, followed by multiplying the resulting equation by $f_i(x)$ and integrating from 0 to L, gives

$$\sum_{n=1}^{N} c_n(t)e^{-\lambda_n t}\int_0^L f_i(x)\frac{d^2\psi_n}{dx^2}\,dx - \sum_{n=1}^{N}\left(\frac{dc_n}{dt} - \lambda_n c_n\right)e^{-\lambda_n t}\int_0^L f_i(x)\psi_n(x)\,dx$$
$$= -\int_0^L f_i(x)F(x,t)\,dx. \qquad (\mathbf{5.7.27})$$

Because

$$\frac{d^2\psi_n}{dx^2} + \lambda_n\psi_n = 0, \qquad (\mathbf{5.7.28})$$

Equation 5.7.27 simplifies to

$$\sum_{n=1}^{N}\frac{dc_n}{dt}e^{-\lambda_n t}\int_0^L f_i(x)\psi_n(x)\,dx = \int_0^L f_i(x)F(x,t)\,dx = F_i^*(t), \qquad (\mathbf{5.7.29})$$

where $i = 1, 2, \ldots, N$.

We must now find c_n. We can write Equation 5.7.29 as

$$\sum_{n=1}^{N} e_{in} e^{-\lambda_n t} \frac{dc_n}{dt}\, dx = F_i^*(t), \tag{5.7.30}$$

where

$$e_{in} = \sum_{j=1}^{N} \alpha_{nj} b_{ji}. \tag{5.7.31}$$

Using linear algebra, we find that

$$e^{-\lambda_n t} \frac{dc_n}{dt} = \sum_{i=1}^{N} p_{ni} F_i^*(t), \tag{5.7.32}$$

where p_{ni} are the elements of an array $P = E^{-1}$ and $E = (DB)^T$. The arrays D and B consist of elements α_{ij} and b_{ij}, respectively. Solving Equation 5.7.32, we find that

$$c_n(t) = A_n + \sum_{i=1}^{N} p_{ni} \int_0^t F_i^*(\eta) e^{\lambda_n \eta}\, d\eta. \tag{5.7.33}$$

Because $u(x, 0) = 0$, $c_n(0) = 0$ and $A_n = 0$.

We are now ready to find our Green's function. Let us set $F(x, t) = \delta(x - \xi)\delta(t - \tau)$. Then $F_i^*(t) = f_i(\xi)\delta(t - \tau)$ and

$$c_n(t) = H(t - \tau) \sum_{i=1}^{N} p_{ni} f_i(\xi) e^{\lambda_n \tau}. \tag{5.7.34}$$

From Equation 5.7.2, Equation 5.7.26, and Equation 5.7.34, we obtain the final result that

$$g(x, t|\xi, \tau) = H(t - \tau) \sum_{n=1}^{N} \sum_{j=1}^{N} \sum_{i=1}^{N} \alpha_{nj} p_{ni} f_i(\xi) f_j(x) e^{-\lambda_n (t-\tau)}. \tag{5.7.35}$$

□

• Example 5.7.3

In Example 5.5.2, we solved the Green's function problem:

$$\frac{\partial g}{\partial t} - \frac{\partial^2 g}{\partial x^2} = \delta(x - \xi)\delta(t - \tau), \tag{5.7.36}$$

with the boundary condition

$$g(0, t|\xi, \tau) = g(L, t|\xi, \tau) = 0, \tag{5.7.37}$$

and the initial condition $g(x, 0|\xi, \tau) = 0$. There we found the solution (Equation 5.5.34):

$$g(x, t|\xi, \tau) = H(t - \tau) \sum_{n=1}^{\infty} \psi_n(\xi) \psi_n(x) e^{-k_n^2 (t-\tau)}, \tag{5.7.38}$$

where we have the orthonormal eigenfunctions

$$\psi_n(x) = \sqrt{2/L}\,\sin(k_n x), \qquad k_n = n\pi/L. \tag{5.7.39}$$

Let us use the basis function $f_j(x) = (1 - x/L)(x/L)^j$ to find the approximate Green's function to Equation 5.7.36. Here $j = 1, 2, 3, \ldots, N$,

For $N = 2$, we showed in Example 5.7.2 that

$$B = L\begin{pmatrix} 1/30 & 1/60 \\ 1/60 & 1/105 \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ -1/2 & 1 \end{pmatrix}. \tag{5.7.40}$$

Consequently,

$$BD = \frac{L}{840}\begin{pmatrix} 28 & 14 \\ 0 & 1 \end{pmatrix}, \qquad E = \frac{L}{840}\begin{pmatrix} 28 & 0 \\ 14 & 1 \end{pmatrix}. \tag{5.7.41}$$

Using Gaussian elimination,

$$P = E^{-1} = \frac{1}{L}\begin{pmatrix} 30 & 0 \\ -420 & 840 \end{pmatrix}. \tag{5.7.42}$$

Therefore, the two-term approximation to the Green's function, Equation 5.7.38, is

$$\begin{aligned} g(x,t|\xi,\tau) = {} & \frac{30}{L}\frac{x}{L}\left(1-\frac{x}{L}\right)\frac{\xi}{L}\left(1-\frac{\xi}{L}\right)\exp\left[-\frac{10(t-\tau)}{L^2}\right]H(t-\tau) \\ & + \left[\frac{210}{L}\frac{x}{L}\left(1-\frac{x}{L}\right)\frac{\xi}{L}\left(1-\frac{\xi}{L}\right) - \frac{420}{L}\frac{x}{L}\left(1-\frac{x}{L}\right)\left(\frac{\xi}{L}\right)^2\left(1-\frac{\xi}{L}\right)\right. \\ & \left. - \frac{420}{L}\left(\frac{x}{L}\right)^2\left(1-\frac{x}{L}\right)\frac{\xi}{L}\left(1-\frac{\xi}{L}\right) + \frac{840}{L}\left(\frac{x}{L}\right)^2\left(1-\frac{x}{L}\right)\left(\frac{\xi}{L}\right)^2\left(1-\frac{\xi}{L}\right)\right] \\ & \times \exp\left[-\frac{42(t-\tau)}{L^2}\right]H(t-\tau). \end{aligned} \tag{5.7.43}$$

For $N > 2$, hand computations are very cumbersome and numerical computations are necessary. For a specific N, we first compute the arrays A and B via Equation 5.7.17 and Equation 5.7.19.

```
for j = 1:N
for i = 1:N
  A(i,j) = j*(j-1)/(i+j-1) - j*(j-1)/(i+j) ...
         + j*(j+1)/(i+j+1) - j*(j+1)/(i+j) ;
  B(i,j) = 1/(i+j+1) - 2/(i+j+2) + 1/(i+j+3);
end; end
```

Next we compute the λ_n's and corresponding eigenfunctions: `[v,d] = eig(A,-B)`. Table 5.7.1 gives $L^2\lambda_n$ for several values of N.

Once we found the eigenvalues and eigenvectors, we now compute the matrices D, E, and P. For convenience we have reordered the eigenvalues so that their numerical value increases with n. Furthermore, we have set α_{nn} equal to one for $n = 1, 2, \ldots, N$.

```
[lambda,ix] = sort(temp);
for i = 1:N
for j = 1:N
  D(i,j) = v(j,ix(i));
```

Table 5.7.1: The Value of $L^2\lambda_n$ for $n = 1, 2, \ldots, N$ as a Function of N.

n	Exact	$N = 2$	$N = 3$	$N = 4$	$N = 6$	$N = 8$	$N = 10$
1	9.87	10.00	9.87	9.87	9.87	9.87	9.87
2	39.48	42.00	42.00	39.50	39.48	39.48	39.48
3	88.83		102.13	102.13	89.17	88.83	88.83
4	157.91			200.50	159.99	157.96	157.91
5	246.74				350.96	254.42	247.04
6	355.31				570.53	376.47	356.65
7	483.61					878.88	531.55
8	631.65					1298.03	725.34
9	799.44						1850.98
10	986.96						2548.73

```
end; end
for i = 1:N
  denom = D(i,i);
  for j = 1:N
    D(i,j) = D(i,j) / denom;
end; end
E = transpose(D*B);
P = inv(E);
```

Having computed the matrices D and P, our final task is the computation of the Green's function using Equation 5.7.35. The MATLAB code is:

```
phi_i(1) = (1-xi)*xi;
for i = 2:N
  phi_i(i) = xi*phi_i(i-1);
end

for ii = 1:idim

  x = (ii-1)*dx;
  phi_j(1) = (1-x)*x;
  for j = 2:N
    phi_j(j) = x*phi_j(j-1);
  end

  for n = 1:N
  for j = 1:N
  for i = 1:N
    g(ii) = g(ii) + D(n,j).*P(n,i).*phi_j(j).*phi_i(i) ...
              .*exp(-lambda(n)*time);
end; end; end

end
```

In this code the parameter `time` denotes the quantity $(t-\tau)/L^2$. Figure 5.7.1 compares this approximate Green's function for various N against the exact solution. One of the

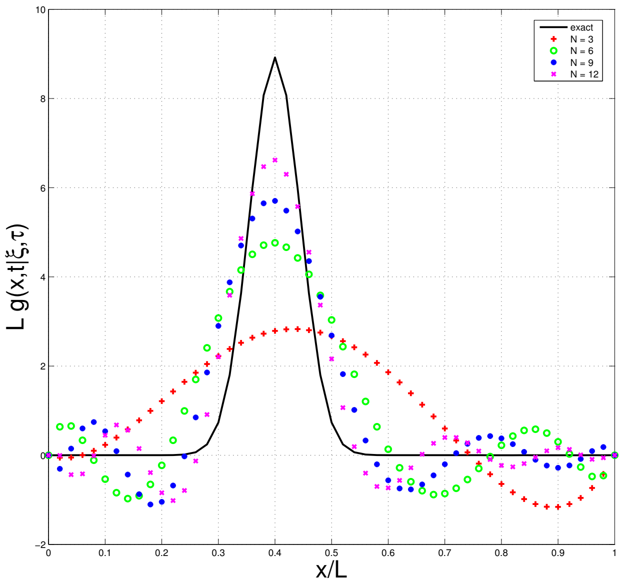


Figure 5.7.1: Comparison of the exact Green's function $Lg(x,t|\xi,\tau)$ for a one-dimensional heat equation given by Equation 5.7.38 (solid line) and approximate Green's functions found by the Galerkin method for different values of N. Here $(t-\tau)/L^2 = 0.001$ and $\xi = 0.4$.

problems with this method is finding the inverse of the array E. As N increases, the accuracy of the inverse becomes poorer.

Further Readings

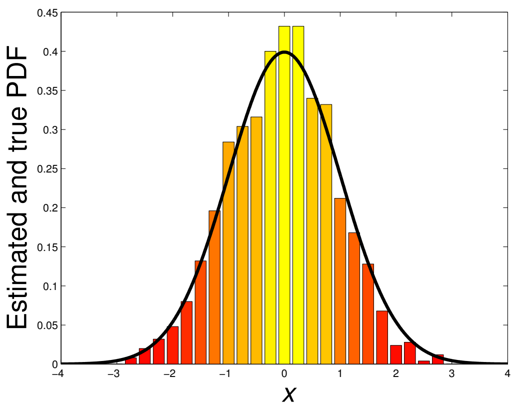
Beck, J. V., K. D. Cole, A. Haji-Sheikh, and B. Litkouhi, 1992: *Heat Conduction Using Green's Functions.* Hemisphere Publishing Corp., 523 pp. Detailed study of solving heat conduction problems via Green's functions.

Carslaw, H. S., and J. C. Jaeger, 1959: *Conduction of Heat in Solids.* Oxford University Press, Chapter 14. An early classic for finding the Green's function for the heat equation.

Duffy, D. G., 2015: *Green's Functions with Applications.* Chapman & Hall/ CRC, 464 pp. A source book.

Özişik, M. N., 1993: *Heat Conduction.* John Wiley & Sons, Chapter 6. A book of how to solve partial differential equations of heat conduction.

Stakgold, I., 1997: *Green's Functions and Boundary Value Problems.* Wiley-Interscience, 720 pp. A systematic approach to boundary-value problems.



Chapter 6

Probability

So far in this book we presented mathematical techniques that are used to solve deterministic problems—problems in which the underlying physical processes are known exactly. In this and the next chapter we turn to problems in which uncertainty is key.

Although probability theory was first developed to explain the behavior of games of chance,[1] its usefulness in the physical sciences and engineering became apparent by the late nineteenth century. Consider, for example, the biological process of birth and death. If b denotes the birth rate and d is the death rate, the size of the population $P(t)$ at time t is

$$P(t) = P(0)e^{(b-d)t}. \tag{6.0.1}$$

Let us examine the situation when $P(0) = 1$ and $b = 2d$ so that a birth is twice as likely to occur as a death. Then, Equation 6.0.1 predicts exponential growth with $P(t) = e^{dt}$. But the first event may be a death, a one-in-three chance since $d/(b + d) = 1/3$, and this would result in the population immediately becoming extinct. Consequently we see that for small populations, chance fluctuations become important and a deterministic model is inadequate.

The purpose of this and the next chapter is to introduce mathematical techniques that will lead to realistic models where chance plays an important role, and show under what conditions deterministic models will work. In this chapter we present those concepts that we will need in the next chapter to explain random processes.

---

[1] Todhunter, I., 1949: *A History of the Mathematical Theory of Probability from the Time of Pascal to That of Laplace*. Chelsea, 624 pp.; Hald, A., 1990: *A History of Probability and Statistics and Their Applications before 1750*. John Wiley & Sons, 586 pp.

6.1 REVIEW OF SET THEORY

Often we must count various objects in order to compute a probability. Sets provide a formal method to aid in these computations. Here we review important concepts from set theory.

Sets are collections of objects, such as the number of undergraduate students at a college. We define a set A either by naming the objects or describing the objects. For example, the set of natural numbers can be either enumerated:

$$A = \{1, 2, 3, 4, \ldots\}, \tag{6.1.1}$$

or described:

$$A = \{I : I \text{ is an integer and } I \geq 1\}. \tag{6.1.2}$$

Each object in set A is called an *element* and each element is *distinct*. Furthermore, the *ordering* of the elements within the set is not important.

Two sets are said to be equal if they contain the same elements and are written $A = B$. An element x of a set A is denoted by $x \in A$. A set with no elements is called a *empty* or *null* set and denoted by $\emptyset$. On the other hand, a *universal set* is the set of all elements under consideration.

A set B is *subset* of a set A, written $B \subset A$, if every element in B is also an element of A. For example, if $A = \{x : 0 \leq x < \infty\}$ and $S = \{x : -\infty < x < \infty\}$, then $A \subset S$. We can also use this concept to define the equality of sets A and B. Equality occurs when $A \subset B$ and $B \subset A$.

The *complement* of A, written $\overline{A}$, is the set of elements in S but not in A. For example, if $A = \{x : 0 \leq x < \infty\}$ and $S = \{x : -\infty < x < \infty\}$, then $\overline{A} = \{x : -\infty < x < 0\}$.

Two sets can be combined together to form a new set. This *union* of A and B, written $A \cup B$, creates a new set that contains elements that belong to A and/or B. This definition can be extended to multiple sets $A_1, A_2, \ldots, A_N$ so that the union is the set of elements for which each element belongs to at least one of these sets. It is written

$$A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_N = \bigcup_{i=1}^{N} A_i. \tag{6.1.3}$$

The *intersection* of sets A and B, written $A \cap B$, is defined as the set of elements that belong to both A and B. This definition can also be extended to multiple sets $A_1, A_2, \ldots, A_N$ so that the intersection is the set of elements for which each element belongs to all of these sets. It is written

$$A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_N = \bigcap_{i=1}^{N} A_i. \tag{6.1.4}$$

If two sets A and B have no elements in common, they are said to be *disjoint* and $A \cap B = \emptyset$.

A popular tool for visualizing set operations is the *Venn diagram*.[2] For sets A and B Figure 6.1.1 pictorially illustrates $A \cup B$, $A \cap B$, $\overline{A}$, and $A \cap \overline{B}$.

With these definitions a number of results follow: $\overline{\overline{A}} = A$, $A \cup \overline{A} = S$, $A \cap \overline{A} = \emptyset$, $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$, $A \cup S = S$, $A \cap S = A$, $\overline{S} = \emptyset$, and $\overline{\emptyset} = S$. Here S denotes the universal set.

Sets obey various rules similar to those encountered in algebra. They include:

[2] Venn, J., 2008: *Symbolic Logic*. Kessinger, 492 pp.

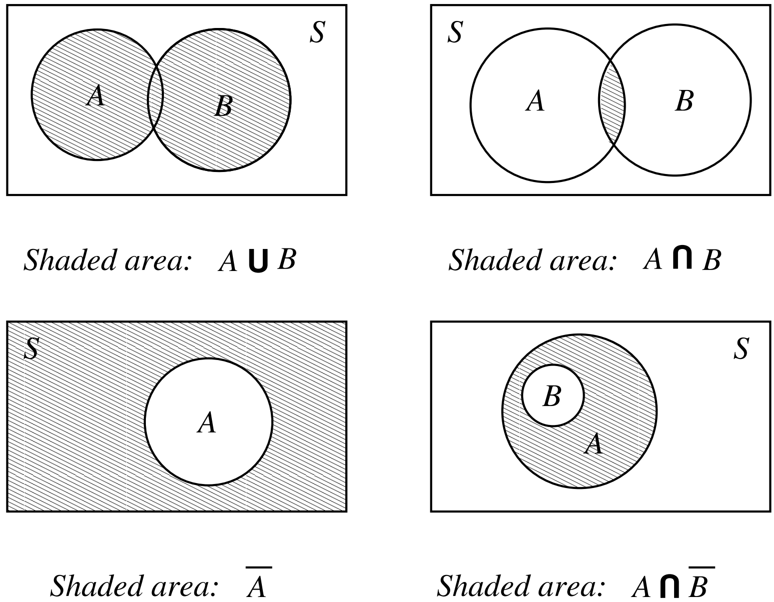


Figure 6.1.1: Examples of Venn diagrams for various configurations of sets A and B. Note that in the case of the lower right diagram, $B \subset A$.

1. Commutative properties: $A \cup B = B \cup A$, $\quad A \cap B = B \cap A$.
2. Associate properties: $A \cup (B \cup C) = (A \cup B) \cup C$, $\quad A \cap (B \cap C) = (A \cap B) \cap C$.
3. Distributive properties: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $\quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
4. De Morgan's law: $A \cup B = \overline{\overline{A} \cap \overline{B}}$.

Finally we define the *size* of a set. *Discrete* sets may have a finite number of elements or *countably infinite* number of elements. By countably infinite we mean that we could in theory count the number of elements in the sets. Two simple examples are $A = \{1, 2, 3, 4, 5, 6\}$ and $A = \{1, 4, 16, 64, \ldots\}$. Discrete sets lie in opposition to *continuous* sets where the elements are infinite in number and cannot be counted. A simple example is $A = \{x : 0 \leq x \leq 2\}$.

Problems

1. If $B \subset A$, use Venn diagrams to show that $A = B \cup (\overline{B} \cap A)$ and $B \cap (\overline{B} \cap A) = \emptyset$. Hint: Use the Venn diagram in the lower right frame of Figure 6.1.1.

2. Using Venn diagrams, show that $A \cup B = A \cup (\overline{A} \cap B)$ and $B = (A \cap B) \cup (\overline{A} \cap B)$. Hint: For $A \cap B$, use the upper right frame from Figure 6.1.1.

6.2 CLASSIC PROBABILITY

All questions of probability begin with the concept of an *experiment* where the governing principle is chance. The set of all possible outcomes of a random experiment is called the *sample space* (or universal set); we shall denote it by S. An element of S is called a *sample point*. The number of elements in S can be finite as in the flipping of a coin twice, infinite but countable such as repeatedly tossing a coin and counting the number of heads, or infinite and uncountable, as measuring the lifetime of a light bulb.

Any subset of the sample set S is called an *event*. If this event contains a single point, then the event is *elementary* or *simple*.

• Example 6.2.1

Consider an experiment that consists of two steps. In the first step, a die is tossed. If the number of dots on the top of the die is even, a coin is flipped; if the number of dots on the die is odd, a ball is selected from a box containing blue and green balls. The sample space is $S = \{1B, 1G, 2H, 2T, 3B, 3G, 4H, 4T, 5B, 5G, 6H, 6T\}$. The event A of obtaining a blue ball is $A = \{1B, 3B, 5B\}$, of obtaining a green ball is $B = \{1G, 3G, 5G\}$, and obtaining an even number of dots when the die is tossed is $C = \{2H, 2T, 4H, 4T, 6H, 6T\}$. The simple event of obtaining a 1 on the die and a blue ball is $D = \{1B\}$. □

Equally likely outcomes

An important class of probability problems consists of those whose outcomes are equally likely. The expression "equally likely" is essentially an intuitive one. For example, if a coin is tossed it seems reasonable that the coin is just as likely to fall "heads" as to fall "tails." Probability seeks to quantify this common sense.

Consider a sample space S of an experiment that consists of finitely many outcomes that are equally likely. Then the probability of an event A is

$$P(A) = \frac{\text{Number of points in } A}{\text{Number of points in } S}. \qquad \textbf{(6.2.1)}$$

With this simple definition we are ready to do some simple problems. An important aid in our counting is whether we can count a particular sample only once (sampling without replacement) or repeatedly (sampling with replacement). The following examples illustrate both cases.

• Example 6.2.2: Balls drawn from urns with replacement

Imagine the situation where we have an urn that has k red balls and $N - k$ black balls. A classic problem asks: What is the chance of two balls being drawn, one after another with replacement, where the first ball is red and the second one is black?

We begin by labeling the k red balls with $1, 2, 3, \ldots, k$ and black balls are numbered $k + 1, k + 2, \ldots, N$. The possible outcomes of the experiment can be written as a 2-tuple (z_1, z_2), where $z_1 \in 1, 2, 3, \ldots, N$ and $z_2 \in 1, 2, 3, \ldots, N$. A successful outcome is a red ball followed by a black one; we can express this case by $E = \{(z_1, z_2) : z_1 = 1, 2, \ldots, k; z_2 = k + 1, k + 2, \ldots, N\}$. Now the total number of 2-tuples in the sample space is N^2 while the total number of 2-tuples in E is $k(N - k)$. Therefore, the probability is

$$P(E) = \frac{k(N-k)}{N^2} = p(1-p), \qquad \textbf{(6.2.2)}$$

where $p = k/N$. □

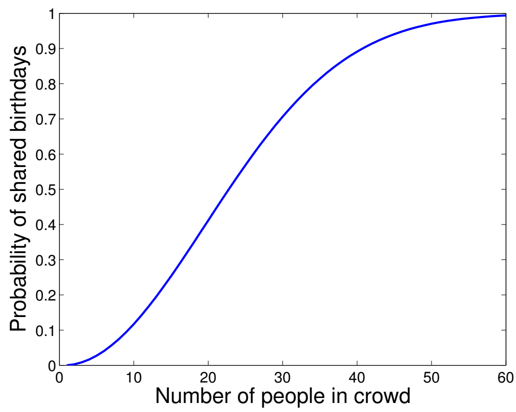


Figure 6.2.1: The probability that a pair of individuals in a crowd of n people share the same birthday.

• Example 6.2.3: Balls drawn from urns without replacement

Let us redo the previous example where the same ball now cannot be chosen twice. We can express this mathematically by the condition $z_1 \neq z_2$. The sample space has $N(N-1)$ balls and the number of successful 2-tuples is again $k(N-k)$. The probability is therefore given by

$$P(E) = \frac{k(N-k)}{N(N-1)} = \frac{k}{N}\frac{N-k}{N}\frac{N}{N-1} = p(1-p)\frac{N}{N-1}. \qquad \textbf{(6.2.3)}$$

The restriction that we cannot replace the original ball has resulted in a higher probability. Why? We have reduced the number of red balls and thereby reduced the chance that we again selected another red ball while the situation with the black balls remains unchanged. □

• Example 6.2.4: The birthday problem[3]

A classic problem in probability is: What is the chance that at least two individuals share the same birthday in a crowd of n people? Actually it is easier to solve the complementary problem: What is the chance that no one in a crowd of n individuals shares the same birthday?

For simplicity let us assume that there are only 365 days in the year. Each individual then has a birthday on one of these 365 days. Therefore, there are a total of $(365)^n$ possible outcomes in a given crowd.

Consider now each individual separately. The first person has a birthday on one of 365 days. The second person, who cannot have the same birthday, has one of the remaining 364 days. Therefore, if A denotes the event that no two people have the same birthday and each outcome is equally likely, then

$$P(A) = \frac{n(A)}{n(S)} = \frac{(365)(364)\cdots(365-n+1)}{(365)^n}. \qquad \textbf{(6.2.4)}$$

To solve the original question, we note that $P(\overline{A}) = 1 - P(A)$ where $P(\overline{A})$ denotes the probability that at least two individuals share the same birthday.

---

[3] First posed by von Mises, R., 1939: Über Aufteilungs- und Besetzungswahrscheinlichkeiten. *Rev. Fac. Sci. Istambul*, **4**, 145–163.

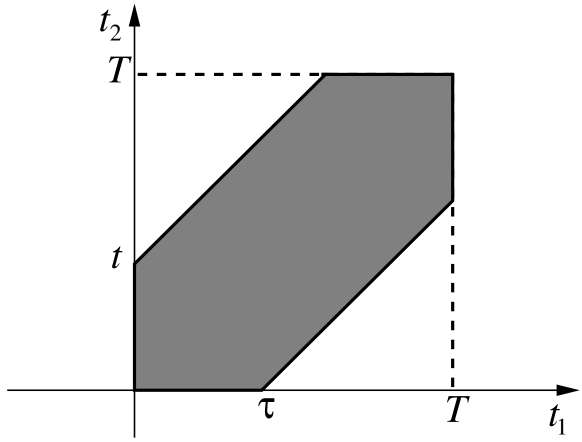


Figure 6.2.2: The graphical solution of whether two fellows can chat online between noon and T minutes after noon. The shaded area denotes the cases when the two will both be online whereas the rectangle gives the sample space.

If $n = 50$, $P(A) \approx 0.03$ and $P(\overline{A}) \approx 0.97$. On the other hand, if $n = 23$, $P(A) \approx 0.493$ and $P(\overline{A}) \approx 0.507$. Figure 6.2.1 illustrates $P(\overline{A})$ as a function of n. Nymann[4] computed the probability that in a group of n people, at least one pair will have the same birthday with at least one such pair among the first k people. □

In the previous examples we counted the objects in sets A and S. Sometimes we can define these sets only as areas on a graph. This graphical definition of probability is

$$P(A) = \frac{\text{Area covered by set } A}{\text{Area covered by set } S}. \tag{6.2.5}$$

The following example illustrates this definition.

• Example 6.2.5

Two friends, Joe and Dave, want to chat online but they will log on independently between noon and T minutes after noon. Because of their schedules, Joe can only wait t minutes after his log-on while Dave can only spare τ minutes. Neither fellow can stay beyond T minutes after noon. What is the chance that they will chat?

Let us denote Joe's log-on time by t_1 and Dave's log-on time by t_2. Joe and Dave will chat if $0 < t_2 - t_1 < t$ and $0 < t_1 - t_2 < \tau$. In Figure 6.2.2 we show the situation where these inequalities are both satisfied. The area of the sample space is T^2. Therefore, from the geometrical definition of probability, the probability $P(A)$ that they will chat is

$$P(A) = \frac{T^2 - (T - t)^2/2 - (T - \tau)^2/2}{T^2}. \tag{6.2.6}$$

□

So far there has been a single event that interests us and we have only had to compute $P(A)$. Suppose we now have two events that we wish to follow. How are the probabilities $P(A)$ and $P(B)$ related?

---

[4] Nymann, J. E., 1975: Another generalization of the birthday problem. *Math. Mag.*, **53**, 111–125.

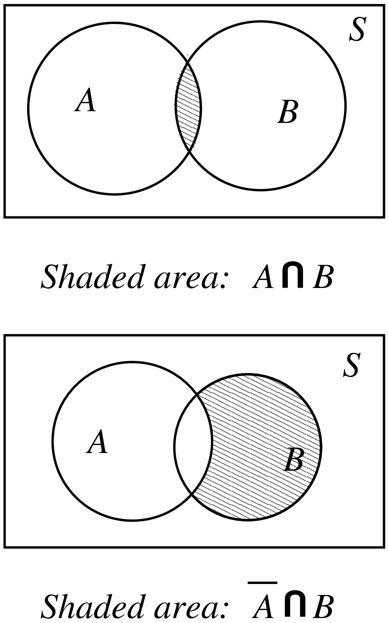


Figure 6.2.3: The Venn diagram used in the derivation of Property 5.

Consider the case of flipping a coin. We could define event A as obtaining a head, $A = \{\text{head}\}$. Event B could be obtaining a tail, $B = \{\text{tail}\}$. Clearly $A \cup B = \{\text{head}, \text{tail}\} = S$, the sample space. Furthermore, $A \cap B = \emptyset$ and A and B are *mutually exclusive*. We already know that $P(A) = P(B) = \frac{1}{2}$. But what happens if $A \cap B$ is *not* an empty set?

From our definition of probability and previous examples, we introduce the following three basic axions:

Axion 1: $P(A), P(B) \geq 0$,
Axion 2: $P(S) = 1$,
Axion 3: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.

The first two axions are clearly true from the definition of probability and sample space. It is the third axion that needs some attention. Here we have two mutually exclusive events A and B in the sample space S. Because the number of points in $A \cup B$ equals the number of points in A plus the number of points in B, $n(A \cup B) = n(A) + n(B)$. Dividing both sides of this equation by the number of sample points and applying Equation 6.2.1, we obtain Axion 3 when $A \cap B = \emptyset$.

From these three axioms, the following properties can be written down:

1. $P(\overline{A}) = 1 - P(A)$
2. $P(\emptyset) = 0$
3. $P(A) \leq P(B)$ if $A \subset B$
4. $P(A) \leq 1$
5. $P(A \cup B) + P(A \cap B) = P(A) + P(B)$.

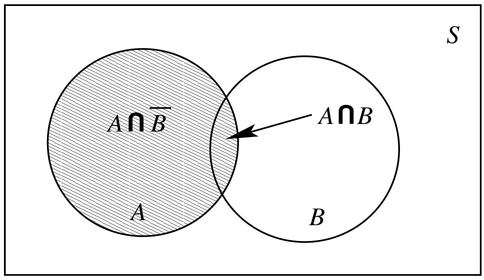


Figure 6.2.4: The Venn diagram that shows that $A = (A \cap \overline{B}) \cup (A \cap B)$.

All of these properties follow readily from our definition of probability except for Property 5 and this is an important one. To prove this property from Axion 3, consider the Venn diagram shown in Figure 6.2.3. From this figure we see that

$$A \cup B = A \cup (\overline{A} \cap B) \qquad \text{and} \qquad B = (A \cap B) \cup (\overline{A} \cap B). \tag{6.2.7}$$

From Axion 3, we have that

$$P(A \cup B) = P(A) + P(\overline{A} \cap B), \tag{6.2.8}$$

and

$$P(B) = P(A \cap B) + P(\overline{A} \cap B). \tag{6.2.9}$$

Eliminating $P(\overline{A} \cap B)$ between Equation 6.2.8 and Equation 6.2.9, we obtain Property 5.

The following example illustrates a probability problem with two events A and B.

• **Example 6.2.6**

Consider Figure 6.2.4. From this figure, we see that $A = (A \cap \overline{B}) \cup (A \cap B)$. Because $A \cap \overline{B}$ and $A \cap B$ are mutually exclusive, then from Axion 3 we have that

$$P(A) = P(A \cap \overline{B}) + P(A \cap B). \tag{6.2.10}$$

□

• **Example 6.2.7**

A company has 400 employees. Every quarter, 100 of them are tested for drugs. The company's policy is to test everyone at random, whether they have been previously tested or not. What is the chance that someone is *not* tested?

The chance that someone *will* be tested is 1/4. Therefore, the chance that someone will *not* be tested is $1 - 1/4 = 3/4$. □

Permutations and combinations

By now it should be evident that your success at computing probabilities lies in correctly counting the objects in a given set. Here we examine two important concepts for systemic counting: permutations and combinations.

A *permutation* consists of ordering n objects *without any regard to their order*. For example, the six permutations of the three letters a, b, and c are abc, acb, bac, bca, cab, and cba. The number of permutations equals $n!$.

In a combination of given objects, we select one or more objects without regard to their order. There are two types of combinations: (1) n different objects, taken k at a time, without repetition, and (2) n different objects, taken k at a time, with repetitions. In the first case, the number of sets that can be made up from n objects, each set containing k different objects and no two sets containing exactly the same k things, equals

$$\text{number of different combinations} = \binom{n}{k} \equiv \frac{n!}{k!(n-k)!}. \tag{6.2.11}$$

Using the three letters a, b, and c, there are three combinations, taken two letters at a time, without repetition: ab, ac, and bc.

In the second case, the number of sets, consisting of k objects chosen from the n objects and each being used as often as desired, is

$$\text{number of different combinations} = \binom{n+k-1}{k}. \tag{6.2.12}$$

Returning to our example using three letters, there are six combinations with repetitions: ab, ac, bc, aa, bb, and cc.

• Example 6.2.8

An urn contains r red balls and b blue balls. If a random sample of size m is chosen, what is the probability that it contains exactly k red balls?

If we choose a random sample of size m, we obtain $\binom{r+b}{m}$ possible outcomes. The number of samples that includes k red balls and $m-k$ blue balls is $\binom{r}{k}\binom{b}{m-k}$. Therefore, the probability that a sample of size m contains exactly k red balls is

$$\frac{\binom{r}{k}\binom{b}{m-k}}{\binom{r+b}{m}}. \qquad \square$$

• Example 6.2.9

A dog kennel has 50 dogs, including 5 German shepherds. (a) What is the probability of choosing 3 German shepherds if 10 dogs are randomly selected? (b) What is the probability of choosing all of the German shepherds in a group of 10 dogs that is chosen at random?

Let S denote the sample space of groups of 10 dogs. The number of those groups is $n(S) = 50!/(10!40!)$. Let A_i denote the set of 10 dogs that contain i German shepherds. Then the number of groups of 10 dogs that contain i German shepherds is $n(A_i) = 10!/[i!(10-i)!]$. Therefore, the probability that out of 50 dogs, we can select at random 10 dogs that include i German shepherds is

$$P(A_i) = \frac{n(A_i)}{n(S)} = \frac{10!10!40!}{i!(10-i)!50!}. \tag{6.2.13}$$

Thus, $P(A_3) = 1.1682 \times 10^{-8}$ and $P(A_5) = 2.453 \times 10^{-8}$. □

• **Example 6.2.10**

Consider an urn with n red balls and n blue balls inside. Let $R = \{r_1, r_2, \ldots, r_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$. Then the number of subsets of $R \cup B$ with n elements is $\binom{2n}{n}$. On the other hand, any subset of $R \cup B$ with n elements can be written as the union of a subset of R with i elements and a subset of B with $n - i$ elements for some $0 \leq i \leq n$. Because, for each i, there are $\binom{n}{i}\binom{n}{n-i}$ such subsets, the total number of subsets of red and blue balls with n elements equals $\sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i}$. Since both approaches must be equivalent,

$$\binom{2n}{n} = \sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i} = \sum_{i=0}^{n}\binom{n}{i}^2 \tag{6.2.14}$$

because $\binom{n}{n-i} = \binom{n}{i}$. □

Conditional probability

Often we are interested in the probability of an event A provided event B occurs. Denoting this *conditional probability* by $P(A|B)$, its probability is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0, \tag{6.2.15}$$

where $P(A \cap B)$ is the joint probability of A and B. Similarly,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \qquad P(A) > 0. \tag{6.2.16}$$

Therefore,

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \tag{6.2.17}$$

and we obtain the famous *Bayes' rule*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{6.2.18}$$

• **Example 6.2.11**

Consider a box containing 10 pencils. Three of the pencils are defective with broken lead. If we draw 2 pencils out at random, what is the chance that we will have selected nondefective pencils?

There are two possible ways of selecting our two pencils: with and without replacement. Let Event A be that the first pencil is not defective and Event B be that the second pencil is not defective. Regardless of whether we replace the first pencil or not, $P(A) = \frac{7}{10}$ because each pencil is equally likely to be picked. If we then replace the first pencil, we have the same situation before any selection was made and $P(B|A) = P(A) = \frac{7}{10}$. Therefore,

$$P(A \cap B) = P(A)P(B|A) = 0.49. \qquad \textbf{(6.2.19)}$$

On the other hand, if we do not replace the first selected pencil, $P(B|A) = \frac{6}{9}$ because there is one fewer nondefective pencils. Consequently,

$$P(A \cap B) = P(A)P(B|A) = \frac{7}{10} \times \frac{6}{9} = \frac{14}{30} < 0.49. \qquad \textbf{(6.2.20)}$$

Why do we have a better chance of obtaining defective pencils if we don't replace the first one? Our removal of that first, nondefective pencil has reduced the uncertainty because we know that there are relatively more defective pencils in the remaining 9 pencils. This reduction in uncertainty must be reflected in a reduction in the chances that both selected pencils will be nondefective. □

Law of total probability

Conditional probabilities are useful because they allow us to simplify probability calculations. Suppose we have n mutually exclusive events $A_1, A_2, \ldots, A_n$ whose probabilities sum to unity, then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n), \qquad \textbf{(6.2.21)}$$

where B is an arbitrary event, and $P(B|A_i)$ is the conditional probability of B assuming A_i. In other words, the law (or formula) of total probability expresses the total probability of an outcome that can be realized via several distinct events.

• Example 6.2.12

There are three boxes, each containing a different number of light bulbs. The first box has 10 bulbs, of which 4 are dead. The second has 6 bulbs, of which one is dead. Finally, there is a third box of eight bulbs, of which 3 bulbs are dead. What is the probability of choosing a dead bulb if a bulb is randomly chosen from one of the three boxes?

The probability of choosing a dead bulb is

$$\begin{aligned} P(D) &= P(D|B_1)P(B_1) + P(D|B_2)P(B_2) + P(D|B_3)P(B_3) && \textbf{(6.2.22)} \\ &= \left(\frac{1}{3}\right)\left(\frac{4}{10}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{3}\right)\left(\frac{3}{8}\right) = \frac{113}{360}. && \textbf{(6.2.23)} \end{aligned}$$

If we had only one box with a total 24 bulbs, of which 8 were dead, then our chance of choosing a dead bulb would be $1/3 > 113/360$. □

Independent events

If events A and B satisfy the equation

$$P(A \cap B) = P(A)P(B), \qquad \textbf{(6.2.24)}$$

they are called *independent events*. From Equation 6.2.15 and Equation 6.2.16, we see that if Equation 6.2.24 holds, then

$$P(A|B) = P(A), \qquad P(B|A) = P(B), \tag{6.2.25}$$

assuming that $P(A) \neq 0$ and $P(B) \neq 0$. Therefore, the term "independent" refers to the fact that the probability of A does *not* depend on the occurrence or non-occurrence of B, and vice versa.

• Example 6.2.13

Imagine some activity where you get two chances to be successful (for example, jumping for fruit still on a tree or shooting basketballs). If each attempt is independent and the probability of success 0.6 is the same for each trial, what is the probability of success after (at most) two tries?

There are two ways of achieving success. We can be successful in the first attempt with $P(S_1) = 0.6$ or we can fail and then be successful on the second attempt: $P(F_1 \cap S_2) = P(F_1)P(S_2) = (0.4)(0.6) = 0.24$, since each attempt is independent. Therefore, the probability of achieving success in two tries is $0.6 + 0.24 = 0.84$. Alternatively, we can compute the probability of failure in two attempts: $P(F_1 \cap F_2) = 0.16$. Then the probability of success with two tries would be the complement of the probability of two failures: $1 - 0.16 = 0.84$. □

• Example 6.2.14

Consider the tossing of a fair die. Let event A denote the tossing of a 2 or 3. Then $P(A) = P(\{2, 3\}) = \frac{1}{3}$. Let event B denote tossing an odd number, $B = \{1, 3, 5\}$. Then $P(B) = \frac{1}{2}$.

Now $A \cap B = \{3\}$ and $P(A \cap B) = \frac{1}{6}$. Because $P(A \cap B) = P(A)P(B)$, events A and B are independent. □

Often we can characterize each outcome of an experiment consisting of n experiments as either a "success" or a "failure." If the probability of each individual success is p, then the probability of k successes and $n - k$ failures is $p^k(1 - p)^{n-k}$. Because there are $n!/[k!(n - k)!]$ ways of achieving these k successes, the probability of an event having k successes in n independent trials is

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \tag{6.2.26}$$

where p is the probability of a success during one of the independent trials.

• Example 6.2.15

What is the probability of having two boys in a four-child family?

Let us assume that the probability of having a male is 0.5. Taking the birth of one child as a single trial,

$$P_4(2) = \frac{4!}{2!2!} \left(\frac{1}{2}\right)^4 = \frac{3}{8}. \tag{6.2.27}$$

Note that this is *not* 0.5, as one might initially guess.

Problems

1. For the following experiments, describe the sample space:

(a) flipping a coin twice

(b) selecting two items out of three items $\{a, b, c\}$ without replacement

(c) selecting two items out of three items $\{a, b, c\}$ with replacement

(d) selecting three balls, one by one, from a box that contains four blue balls and five green balls without replacement

(e) selecting three balls, one by one, from a box that contains four blue balls and five green balls with replacement.

2. Consider two fair dice. What is the probability of throwing them so that the dots sum to seven?

3. In throwing a fair die, what is the probability of obtaining a one *or* two on the top side of the cube?

4. What is the probability of getting heads exactly (a) twice or (b) thrice if you flip a fair coin 6 times?

5. An urn contains six red balls, three blue balls, and two green balls. Two balls are randomly selected. What is the sample space for this experiment? Let X denote the number of green balls selected. What are the possible values of X? Calculate $P(X = 1)$.

6. Consider an urn with 30 blue balls and 50 red balls in it. These balls are identical except for their color. If they are well mixed and you draw 3 balls without replacement, what is the probability that the balls are all of the same color?

7. A deck of cards has 52 cards, including 4 jacks and 4 ten's. What is the probability of selecting a jack *or* ten?

8. Two boys and two girls take their place on a stage to receive an award. What is the probability that the boys take the two end seats?

9. A lottery consists of posting a 3-digit number given by selecting 3 balls from 10 balls, each ball having the number from 1 to 10. The balls are not replaced after they are drawn. What are your chances of winning the lottery if the order does not matter? What are your chances of winning the lottery if the order does matter? Write a short MATLAB code and verify your results. You may want to read about the MATLAB intrinsic function `randperm`.

10. A circle of radius 1 is inscribed in a square with sides of length 2. A point is selected at random in the square in such a manner that all the subsets of equal area of the square are equally likely to contain the point. What is the probability that it is inside the circle?

11. In a rural high school, 20% of the students play football and 10% of them play football and wrestle. If Ed, a randomly selected student of this high school, played football, what is the probability that he also wrestles for his high school?

12. You have a well-shuffled card deck. What is the probability the second card in the deck is an ace?

13. We have two urns: One has 4 red balls and 6 green balls, the other has 6 red and 4 green. We toss a fair coin. If heads, we pick a random ball from the first urn, if tails from the second. What is the probability of getting a red ball? How do your results compare with the probability of getting a red ball if all of the red and green balls had been placed into a single urn?

14. A customer decides between two dinners: a "cheap" one and an "expensive" one. The probability that the customer chooses the expensive meal is $P(E) = 0.2$. A customer who chooses the expensive meal likes it with a 80% probability $P(L|E) = 0.8$. A customer who chooses the cheap meal dislikes it with 70% probability $P(D|C) = 0.7$.

(a) Compute the probability that a customer (1) will choose a cheap meal, (2) will be disappointed with an expensive meal, and (3) will like the cheap meal.

(b) Use the law of total probability to compute the probability that a customer will be disappointed.

(c) If a customer found his dinner to his liking, what is the probability that he or she chose the expensive meal? Hint: Use Bayes' theorem.

15. Suppose that two points are *randomly* and *independently* selected from the interval $(0, 1)$. What is the probability the first one is greater than 1/4, and the second one is less than 3/4? Check your result using `rand` in MATLAB.

16. A certain brand of electronics chip is found to fail prematurely in 1% of all cases. If three of these chips are used in three independent sets of equipment, what is the probability that (a) all three will fail prematurely, (b) that two will fail prematurely, (c) that one will fail prematurely, and (d) that none will fail?

Project: Experimenting with MATLAB's Intrinsic Function rand

The MATLAB function `rand` can be used in simulations where sampling occurs with replacement. If we write `X = rand(1,100)`, the vector `X` contains 100 elements whose values vary between 0 and 1. Therefore, if you wish to simulate a fair die, then we can set up the following table:

`0 < X < 1/6`	die with one dot showing
`1/6 < X < 1/3`	die with two dots showing
`1/3 < X < 1/2`	die with three dots showing
`1/2 < X < 2/3`	die with four dots showing
`2/3 < X < 5/6`	die with five dots showing
`5/6 < X < 1`	die with six dots showing.

We can then write MATLAB code that counts the number of times that we obtain a one or two. Call this number `n`. Then the probability that we would obtain one or two dots on a fair die is `n/100`. Carry out this experiment and compare your answer with the result from Problem 2. What occurs as you do more and more experiments?

Table 6.2.1: The Probability of a Male (Female) Freshman Having Always Had New Male (Female) Roommates from a Pool of m Other Male (Female) Freshmen after n Random Reassignments during His (Her) Freshman Year. The Numerator Is the Probability for a Two-Person Room; the Denominator Is the Probability for a Three-Person Room.

	Total Number of Freshmen							
n	6	12	18	24	30	36	42	48
2	$\frac{0.8000}{0.3000}$	$\frac{0.9091}{0.6545}$	$\frac{0.9412}{0.7721}$	$\frac{0.9565}{0.8300}$	$\frac{0.9655}{0.8645}$	$\frac{0.9714}{0.8874}$	$\frac{0.9756}{0.9037}$	$\frac{0.9787}{0.9158}$
4	$\frac{0.1920}{0.0000}$	$\frac{0.5409}{0.4550}$	$\frac{0.6839}{0.1792}$	$\frac{0.7594}{0.3015}$	$\frac{0.8059}{0.3981}$	$\frac{0.8374}{0.4729}$	$\frac{0.8601}{0.5325}$	$\frac{0.8773}{0.5801}$
6	$\frac{0.0000}{0.0000}$	$\frac{0.1878}{0.0000}$	$\frac{0.3692}{0.0073}$	$\frac{0.4910}{0.0385}$	$\frac{0.5750}{0.0867}$	$\frac{0.6357}{0.1407}$	$\frac{0.6815}{0.1943}$	$\frac{0.7172}{0.2450}$
8	$\frac{0.0000}{0.0000}$	$\frac{0.0310}{0.0000}$	$\frac{0.1405}{0.0000}$	$\frac{0.2524}{0.0012}$	$\frac{0.3459}{0.0075}$	$\frac{0.4214}{0.0212}$	$\frac{0.4825}{0.0411}$	$\frac{0.5325}{0.0658}$

Project: Experimenting with MATLAB's Intrinsic Function `randperm`

MATLAB's intrinsic function `randperm(m)` creates a random ordering of the numbers from `1` to `m`. If you execute `perm = randperm(365)`, this would produce a vector of length 365 and each element has a value lying between 1 and 365. If you repeat the process, you would obtain another list of 365 numbers but they would be in a different order.

Let us simulate the birthday problem. Invoking the `randperm` command, use the first element to simulate the birthday of student 1 in a class of `N` students. Repeatedly invoking this command, create vector `birthdays` that contains the birthdays of the `N` students. Then find out if any of the days are duplicates of another. (Hint: You might want to explore the MATLAB command `unique`.) Repeating this experiment many times, compute the chance that a class of size `N` has at least two students that have the same birthday. Compare your results with Equation 6.2.4. What occurs as the number of experiments increases?

Project: The Roommate Problem

You are a freshman at a small all-male (all-female) college with m other freshmen. For *esprit de corps*, the administration requires that n times during your freshman year, you are randomly (with equal probability) assigned new roommates. The administration does *not*, however, require that you have never roomed with any of them previously.

(a) Assuming that there are 2 freshmen per room (so that $m + 1$ is even), what is the probability that all of your roommates during the year have never roomed with you before? Verify your answer by writing a MATLAB script that simulates this housing practice. I used the MATLAB intrinsic functions `randi(m,1,n)`, `unique` and `length` and ran the simulation 10 million times.

(b) Assuming that there are 3 freshmen per room (so that $m + 1$ is a multiple of 3), what is the probability that all of your roommates during the year have never roomed with you before? Verify your answer by writing a MATLAB script that simulates this housing practice. I used the MATLAB intrinsic functions `randperm`, `unique` and `length` and ran the simulation 10 million times.

6.3 DISCRETE RANDOM VARIABLES

In the previous section we presented the basic concepts of probability. In high school algebra you were introduced to the concept of a variable—a quantity that could vary unlike constants and parameters. Here we extend this idea to situations where the variations are due to randomness.

A *random variable* is a single-valued real function that assigns a real number, the *value*, to each sample point t of S. The variable can be discrete, such as the flipping of a coin, or continuous, such as the lifetime of a light bulb. The sample space S is the *domain* of the *random variable* $X(t)$, and the collection of all numbers $X(t)$ is the *range*. Two or more sample points can give the same value of $X(t)$, but we will never allow two different numbers in the range of $X(t)$ for a given t.

The term "random variable" is probably a poor one. Consider the simple example of tossing a coin. A random variable that describes this experiment is

$$X[s_i] = \begin{cases} 1, & s_1 = \text{head}, \\ 0, & s_2 = \text{tail}. \end{cases} \tag{6.3.1}$$

An obvious question is: What is random about Equation 6.3.1? If a head is tossed, we obtain the answer one; if a tail is tossed, we obtain a zero. Everything is well defined; there is no element of chance here. The randomness arises from the tossing of the coin. Until the experiment (tossing of the coin) is performed, we do not know the outcome of the experiment and the value of the random variable. Therefore, *a random variable is a variable that may take different values if a random experiment is conducted and its value is not known in advance.*

We begin our study of random variables by focusing on those arising from discrete events. If X is discrete, X assumes only finitely many or countably many values: x_1, x_2, x_3, ... For each possible value of x_i, there is a corresponding positive probability $p_X[x_1] = P(X = x_1), p_X[x_2] = P(X = x_2), \ldots$ given by the *probability mass function.* For values of x different from x_i, say $x_1 < x < x_2$, the probability mass function equals zero. Therefore, we have that

$$p_X[x_i] = \begin{cases} p_i, & x = x_i, \\ 0, & \text{otherwise}, \end{cases} \tag{6.3.2}$$

where $i = 1, 2, 3, \ldots$.

At this point it is convenient to introduce several special classes or types of random variables. First we have *independent random variables* where the realization of one does not affect the probability distribution of the other. Of equal importance are *identically distributed random variables* where the random variables have the same probability distribution. Finally we can combine both properties into *independent identically distributed* (i.i.d.) random variables. This last class occurs repeatedly in common applications.

• Example 6.3.1

Consider a fair die. We can describe the results from rolling this fair die via the discrete random variable X, which has the possible values $x_i = 1, 2, 3, 4, 5, 6$ with the probability $p_X[x_i] = \frac{1}{6}$ each. Note that $0 \le p_X[x_i] < 1$ here. Furthermore,

$$\sum_{i=1}^{6} p_X[x_i] = 1. \tag{6.3.3}$$

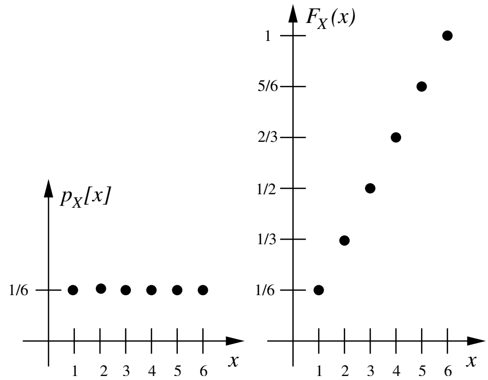


Figure 6.3.1: The probability mass function for a fair die.

Figure 6.3.1 illustrates the probability mass function. □

• Example 6.3.2

Let us now modify Example 6.3.1 so that

$$X[s_i] = \begin{cases} 1, & s_i = 1, 2, \\ 2, & s_i = 3, 4, \\ 3, & s_i = 5, 6. \end{cases} \tag{6.3.4}$$

The probability mass function becomes

$$p_X[1] = p_X[2] = p_X[3] = \tfrac{1}{3}. \tag{6.3.5}$$

□

• Example 6.3.3

Consider the probability mass function:

$$p_X[x_n] = \begin{cases} k(1/2)^n, & n = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases} \tag{6.3.6}$$

Let us (a) find the value of k, (b) find $P(X = 2)$, (c) find $P(X \leq 2)$, and (d) $P(X \geq 1)$.

From the properties of probability mass function,

$$k \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = k \frac{1}{1 - \frac{1}{2}} = 2k = 1. \tag{6.3.7}$$

Therefore, $k = \frac{1}{2}$. Note that $0 \leq p_X[x_n] \leq 1$.

Having found k, we immediately have

$$P(X = 2) = p_X[x_2] = \tfrac{1}{8}, \tag{6.3.8}$$

$$P(X \leq 2) = p_X[x_0] + p_X[x_1] + p_X[x_2] = \tfrac{7}{8}, \tag{6.3.9}$$

and

$$P(X \geq 1) = 1 - P(X = 0) = \tfrac{1}{2}. \tag{6.3.10}$$

□

Some Properties of the Probability Mass Function $p_X[x_i]$

$$0 \le p_X[x_k] < 1, \qquad p_X[x] = 0 \quad \text{if} \quad x \ne x_k, \qquad k = 1, 2, \ldots$$

$$\sum_n p_X[x_n] = 1$$

$$F_X(x) = P(X \le x) = \sum_{x_k \le x} p_X[x_k]$$

$$P(a < x \le b) = \sum_{a < x_k \le b} p_X[x_k]$$

Having introduced the probability mass function, an alternative means of describing the probabilities of a discrete random variable is the *cumulative distribution function.* It is defined as

$$F_X(x) = P(X \le x), \qquad -\infty < x < \infty. \tag{6.3.11}$$

It is computed via

$$F_X(x) = \sum_{x_i \le x} p_X[x_i] = \sum_{x_i \le x} p_i. \tag{6.3.12}$$

Consequently, combining Equation 6.3.11 and Equation 6.3.12, we obtain

$$P(a < x \le b) = \sum_{a < x_i \le b} p_i. \tag{6.3.13}$$

Equation 6.3.13 gives the probability over the interval $(a, b]$.

• Example 6.3.4

A Bernoulli experiment is a random experiment, the outcome of which is a success or failure. Consider now a sequence of independent Bernoulli trials with probability p of success from trial to trial. This sequence is observed until the first success occurs. Let X denote a random variable that equals the trial number on which the first success occurs. The probability mass function is then

$$p_X[x_n] = (1-p)^{n-1}p, \qquad n = 1, 2, 3, \ldots. \tag{6.3.14}$$

Let us compute the cumulative distribution function.

For geometric series, we begin by noting that

$$\sum_{n=0}^{\infty} ar^n = \sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}, \qquad |r| < 1. \tag{6.3.15}$$

Next we check Equation 6.3.14 and determine whether it is a valid probability mass function. It is because

$$\sum_{n=1}^{\infty} p_X[x_n] = \sum_{n=1}^{\infty} (1-p)^{n-1}p = \frac{p}{1-(1-p)} = 1, \tag{6.3.16}$$

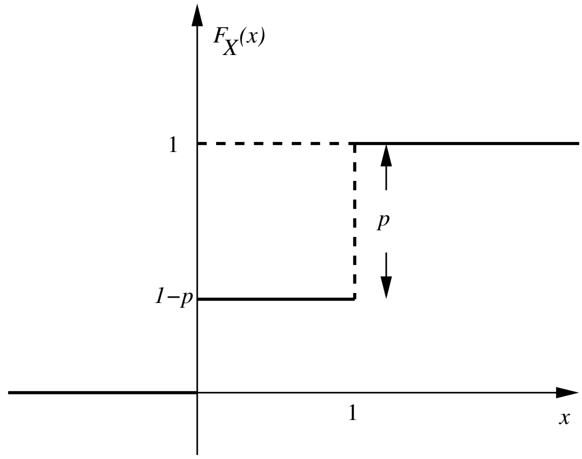


Figure 6.3.2: The cumulative distribution function for a Bernoulli random variable.

where we used Equation 6.3.15. Next, we note that

$$P(X > m) = \sum_{n=m+1}^{\infty} (1-p)^{n-1} p = \frac{(1-p)^m p}{1-(1-p)} = (1-p)^m. \qquad \textbf{(6.3.17)}$$

Therefore,

$$F_X(x) = P(X \le m) = 1 - P(X > m) = 1 - (1-p)^m, \qquad \textbf{(6.3.18)}$$

where $m = 1, 2, 3, \ldots$. □

• Example 6.3.5: Generating discrete random variables via MATLAB

In this example we show how to generate a discrete random variable using MATLAB's intrinsic function `rand`. This MATLAB command produces random, uniformly distributed (equally probable) reals over the interval $(0, 1)$. How can we use this function, when in the case of discrete random variables, we have only integer values, such as $k = 1, 2, 3, 4, 5, 6$, in the case of tossing a die?[5]

Consider the Bernoulli random variable $X = k$, $k = 0, 1$. As you will show in your homework, it has the cumulative distribution function of

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1-p, & 0 \le x < 1, \\ 1, & 1 \le x. \end{cases} \qquad \textbf{(6.3.19)}$$

See Figure 6.3.2.

Imagine now a program that includes the MATLAB function `rand`, which yields the value `t`. Then, if $0 < t \le 1-p$, Figure 6.3.2 gives us that $X = 0$. On the other hand, if $1 - p < t < 1$, then $X = 1$. Thus, to obtain `M` realizations of the Bernoulli random variable `X`, the MATLAB code would read for a given `p`:

```
clear;
```

[5] This technique is known as the inverse transform sampling method. See pages 85–102 in Devroye, L., 1986: *Non-Uniform Random Variable Generation*. Springer-Verlag, 843 pp.

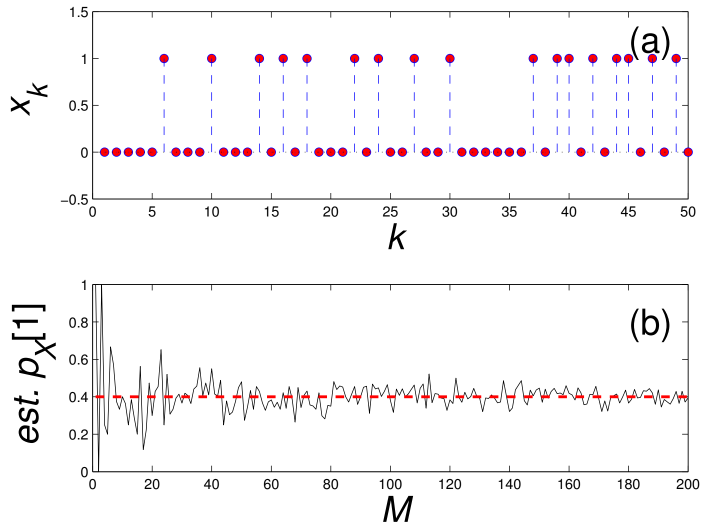


Figure 6.3.3: (a) Outcomes of the Bernoulli random variable generated by the MATLAB function `rand`. (b) The computed value of the probability mass function $p_X[1]$ as a function of M realization of the Bernoulli random variable. The dashed line is the line for the exact answer $p = 0.4$.

```
for i = 1:M
  t = rand(1,1);
    if (t <= 1-p) X(i,1) = 0;
    else
    X(i,1) = 1;
end; end
```

The end product of this code creates a vector `X` of length `M` consisting of a random variable with either zeros or ones. This is shown in Figure 6.3.3(a) when $p = 0.4$.

Once we have generated this random variable, we can use its relative frequency to compute its probability mass function and cumulative distribution function from

$$\hat{p}_X[x_k] = \frac{\text{Number of outcomes equal to } k}{M}, \tag{6.3.20}$$

and

$$\hat{F}_X(x) = \frac{\text{Number of outcomes } \leq x}{M}. \tag{6.3.21}$$

In Figure 6.3.3(b) we have computed the value of $\hat{p}_X[1]$. Clearly it should equal p. As this figure shows, we obtain poor results when M is small, with $\hat{p}_X[1]$ moving randomly above and below the correct answer. As M becomes larger, our estimate improves.

Problems

1. The Bernoulli distribution has the probability mass function

$$p_X[x_k] = P(X = k) = p^k(1-p)^{1-k}, \qquad k = 0, 1,$$

where $0 \leq p \leq 1$. (a) Show that this distribution is a valid probability mass function. (b) Find its cumulative distribution function.

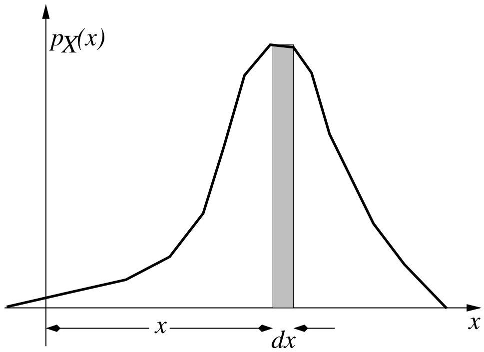


Figure 6.4.1: A probability density function.

2. An experiment is performed where a digit, ranging from 0 to 9, is repeatedly and randomly chosen. If X denotes the times that this experiment must be repeated until the digit 0 is selected, find $P(X)$.

3. A scientific company needs a programmer who knows an unusual programming language. If only 5% of programmers know this language, how many programmers should the company interview to have a 75% chance of finding such a programmer?

6.4 CONTINUOUS RANDOM VARIABLES

In the previous section we examined random variables that can assume only certain discrete values. Here we extend the concept of random variables so that they can take on values over a continuous interval. Typical examples of continuous random variables include the noisy portion of the voltage within an amplifier, the phase of a propagating wave, and the amount of precipitation.

An important quantity that we introduced in the previous section was the probability mass function. What is the corresponding function for continuous random variables? From the fundamental concepts of probability, we know that the probability of a continuous variable assuming one specific value out of its possible range values equals zero; it is merely one point out of an infinite number of points in the sample space. On the other hand, there is a finite probability that the value assumed by the random variable X will lie within an arbitrarily small interval dx and this probability will depend on the length of the interval.

Another factor that should influence the probability is the value of x. There is no reason why the probability of X should be independent of x. Consequently, an equation for probability in the interval $x < X \le x + dx$ requires a function $p_X(x)$, which acts as a weighting function and models the relative frequency behavior of X. For these reasons, the probability that a continuous random variable X will assume a value lying between x and $x + dx$ is given by

$$P(x < X \le x + dx) = p_X(x)\, dx. \qquad \mathbf{(6.4.1)}$$

Figure 6.4.1 illustrates a possible example of $p_X(x)$ where the shaded area equals the probability $P(x < X \le x + dx)$. Clearly the function $p_X(x) = P(x < X \le x + dx)/dx$ has the dimension of probability per infinitesimal interval dx and is called, for that reason, the *probability density.* Furthermore, although $p_X(x)\, dx \le 1$, this does *not* mean that $p_X(x) \le 1$. A family of random variables having the same probability density is *identically distributed.*

The function $p_X(x)$ must also satisfy several additional conditions. Because probability cannot be negative, $p_X(x) \ge 0$ of all x. Furthermore, as Figure 6.4.1 suggests, if we add

Some Properties of the Probability Density Function $p_X(x)$

$$p_X(x) \geq 0, \qquad \int_{-\infty}^{\infty} p_X(x)\,dx = 1$$

$$P(a < X \leq b) = \int_a^b p_X(x)\,dx$$

up all of the possible values of x, then we have a certain event. We can express this mathematically by

$$\int_{-\infty}^{\infty} p_X(x)\,dx = 1. \tag{6.4.2}$$

Thus, a probability density has the properties given by Equation 6.4.1 and Equation 6.4.2. It must also be a single-valued function of x. Note that these conditions do not require that $p_X(x)$ is a continuous function of x.

Let us now consider the probability $P(a < X \leq b)$ where a and b are constants. If we subdivide the range of x between a and b into infinitesimal intervals $(x, x + dx)$, the probability that the random variable will assume a value from one such interval is given by Equation 6.4.1. The probability that the variable will assume a value in the interval (a, b) equals the sum of the probabilities from each subinterval between a and b and is given by the area under the curve $p(x)$ between $x = a$ and $x = b$. Therefore,

$$P(a < X \leq b) = \int_a^b p_X(x)\,dx. \tag{6.4.3}$$

If $a = -\infty$, we have that

$$P(X \leq b) = \int_{-\infty}^b p_X(x)\,dx. \tag{6.4.4}$$

Alternatively, setting $b = \infty$,

$$P(a < X) = \int_a^{\infty} p_X(x)\,dx. \tag{6.4.5}$$

From Equation 6.4.3 we also have

$$P(X > a) = 1 - P(X < a) = 1 - \int_{-\infty}^a p_X(x)\,dx = \int_a^{\infty} p_X(x)\,dx. \tag{6.4.6}$$

From Equation 6.4.4 we now define

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x p_X(\xi)\,d\xi. \tag{6.4.7}$$

This function $F_X(x)$ is called the *cumulative distribution function*, or simply the distribution function, of the random variable X. Clearly,

$$p_X(x) = F_X'(x). \tag{6.4.8}$$

Therefore, from the properties of $p_X(x)$, we have that (1) $F_X(x)$ is a nondecreasing function of x, (2) $F_X(-\infty) = 0$, (3) $F_X(\infty) = 1$, and (4) $P(a < X \leq b) = F_X(b) - F_X(a)$.

• Example 6.4.1

The continuous random variable X has the probability density function

$$p_X(x) = \begin{cases} k\left(x - x^2\right), & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases} \qquad \textbf{(6.4.9)}$$

What must be the value of k? What is the cumulative distribution function? What is $P(X < 1/2)$?

From Equation 6.4.2, we have that

$$\int_{-\infty}^{\infty} p_X(x)\,dx = k\int_0^1 \left(x - x^2\right)\,dx = k\left.\frac{x^2}{2} - \frac{x^3}{3}\right|_0^1 = \frac{k}{6}. \qquad \textbf{(6.4.10)}$$

Therefore, k must equal 6.

Next, we note that

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} p_X(\xi)\,d\xi. \qquad \textbf{(6.4.11)}$$

If $x < 0$, $F_X(x) = 0$. For $0 < x < 1$, then

$$F_X(x) = 6\int_0^x \left(\xi - \xi^2\right)\,d\xi = 6\left.\left(\frac{\xi^2}{2} - \frac{\xi^3}{3}\right)\right|_0^x = 3x^2 - 2x^3. \qquad \textbf{(6.4.12)}$$

Finally, if $x > 1$,

$$F_X(x) = 6\int_0^1 \left(\xi - \xi^2\right)\,d\xi = 1. \qquad \textbf{(6.4.13)}$$

In summary,

$$F_X(x) = \begin{cases} 0, & 0 \le x, \\ 3x^2 - 2x^3, & 0 < x \le 1, \\ 1, & 1 < x. \end{cases} \qquad \textbf{(6.4.14)}$$

Because $P(X \le x) = F_X(x)$, we have that $P(X < \frac{1}{2}) = \frac{1}{2}$ and $P(X > \frac{1}{2}) = 1 - P(X < \frac{1}{2}) = \frac{1}{2}$. □

• Example 6.4.2: Generating continuous random variables via MATLAB[6]

In the previous section we showed how the MATLAB function `rand` can be used to generate outcomes for a discrete random variable. Similar considerations hold for a continuous random variable.

Consider the exponential random variable X. Its probability density function is

$$p_X(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & 0 < x, \end{cases} \qquad \textbf{(6.4.15)}$$

where $\lambda > 0$. For homework you will show that the corresponding cumulative distribution function is

$$F_X(x) = \begin{cases} 0, & x \le 0, \\ 1 - e^{-\lambda x}, & 0 < x. \end{cases} \qquad \textbf{(6.4.16)}$$

---

[6] This technique is known as the inverse transform sampling method. See pages 27–39 in Devroye, L., 1986: *Non-Uniform Random Variable Generation*. Springer-Verlag, 843 pp.

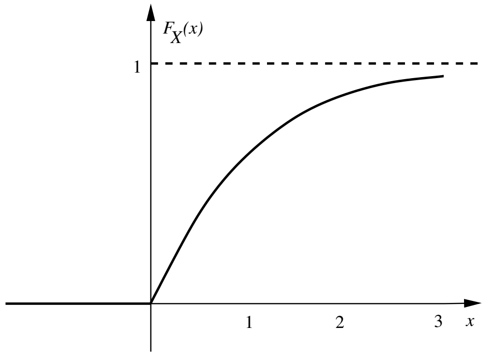


Figure 6.4.2: The cumulative distribution function for an exponential random variable.

Figure 6.4.2 illustrates this cumulative density function when $\lambda = 1$. How can we use these results to generate a MATLAB code that produces an exponential random variable?

Recall that both MATLAB function `rand` and the cumulative distribution function produce values that vary between 0 and 1. Given a value from `rand`, we can compute the corresponding $X = x$, which would give the same value from the cumulative distribution function. In short, we are creating random values for the cumulative distribution function and using those values to give the exponential random variable via

$$X = x = -\ln(1 - \texttt{rand})\,/\lambda, \tag{6.4.17}$$

where we have set $F_X(x) = \texttt{rand}$. Therefore, the MATLAB code to generate exponential random variables for a particular `lambda` is

```
clear;
for i = 1:M
  t = rand(1,1);
  X(i,1) = -log(1-t) / lambda;
end
```

where `M` is the number of experiments that we run. In Figure 6.4.3(a) we illustrate the first 200 outcomes from our numerical experiment to generate an exponential random variable.

To compute the probability density function we use the finite difference approximation of Equation 6.4.1, or

$$\hat{p}(x_0) = \frac{\text{Number of outcomes in } [x_0 - \Delta x/2, x_0 + \Delta x/2]}{M\Delta x}, \tag{6.4.18}$$

where Δx is the size of the bins into which we collect the various outcomes. Figure 6.4.3(b) illustrates this numerical estimation of the probability density function in the case of an exponential random variable. The function $\hat{p}_X(x)$ was created from the MATLAB code:

```
clear;
delta_x = 0.2; lambda = 1; M = 1000; % Initialize Δx, λ and M
% sample M outcomes from the uniformly distributed distribution
t = rand(M,1);
% generate the exponential random variable
x = - log(1-t)/lambda;
```

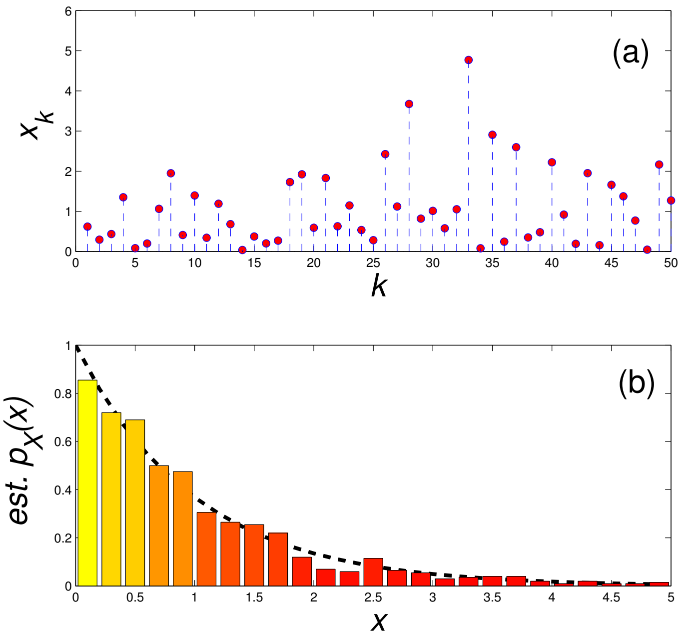


Figure 6.4.3: (a) Outcomes of a numerical experiment to generate an exponential random variable using the MATLAB function `rand`. (b) The function $\hat{p}_X(x)$ given by Equation 6.4.18 as a function of x for an exponential random variable with $M = 1000$. The dashed black line is the exact probability density function.

```
% create the various bins [x_0 - Δx/2, x_0 + Δx/2]
bincenters=[delta_x/2:delta_x:5];
bins=length(bincenters); % count the number of bins
% now bin the M outcomes into the various bins
[n,x_out] = hist(x,bincenters);
n = n / (delta_x*M); % compute the probability per bin
bar_h = bar(x_out,n); % create the bar graph
bar_child = get(bar_h,'Children');
set(bar_child,'CData',n);
colormap(Autumn);
```

Problems

1. The probability density function for the exponential random variable is

$$p_X(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & 0 < x, \end{cases}$$

with $\lambda > 0$. Find its cumulative distribution function.

2. Given the probability density function

$$p_X(x) = \begin{cases} kx, & 0 < x < 2, \\ 0, & \text{otherwise}, \end{cases}$$

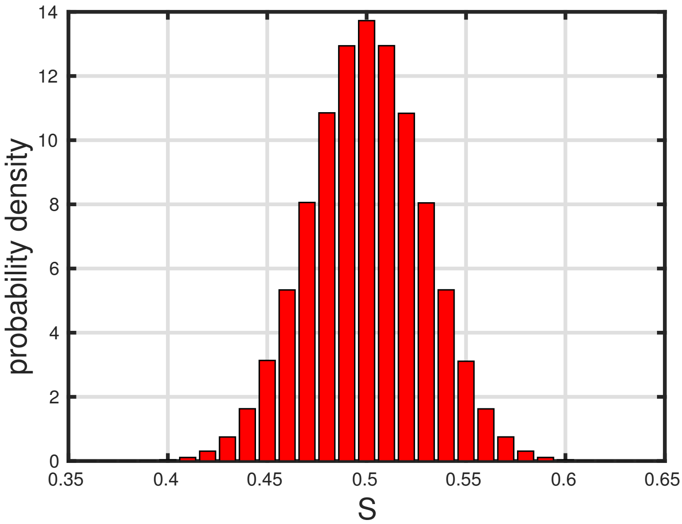


Figure 6.4.4: Computed probability density function for the sum $S = (X_1 + X_2 + X_3 + \cdots + X_{100})/100$, where X_i is the ith sample from a uniform distribution.

where k is a constant, (a) compute the value of k, (b) find the cumulative density function $F_X(x)$, and (c) find the $P(1 < X \leq 2)$.

3. Given the probability density function

$$p_X(x) = \begin{cases} k\,(1 - |x|)\,, & |x| < 1, \\ 0, & |x| > 1, \end{cases}$$

where k is a constant, (a) compute the value of k and (b) find the cumulative density function $F_X(x)$.

Project: Central Limit Theorem

Consider the sum $S = (X_1 + X_2 + X_3 + \cdots + X_{100})/100$, where X_i is the ith sample from a uniform distribution.

Step 1: Write a MATLAB program to compute the probability density function of S. See Figure 6.4.4.

Step 2: The *central limit theorem* states the distribution of the sum (or average) of a large number of independent, identically distributed random variables will be approximately normal, regardless of the underlying distribution. Do your numerical results agree with this theorem?

6.5 MEAN AND VARIANCE

In the previous two sections we explored the concepts of the random variable and distribution. Here we introduce two parameters, *mean* and *variance*, that are useful in characterizing a distribution.

The mean μ_X is defined by

$$\mu_X = E(X) = \begin{cases} \displaystyle\sum_k x_k \, p_X[x_k], & X \text{ discrete,} \\ \displaystyle\int_{-\infty}^{\infty} x \, p_X(x)\, dx, & X \text{ continuous.} \end{cases} \tag{6.5.1}$$

The mean provides the position of the center of the distribution. The operator $E(X)$, which is called the *expectation* of X, gives the average value of X that one should *expect* after many trials.

Two important properties involve the expectation of the sum and product of two random variables X and Y. The first one is

$$E(X + Y) = E(X) + E(Y). \tag{6.5.2}$$

Second, if X and Y are *independent* random variables, then

$$E(XY) = E(X)E(Y). \tag{6.5.3}$$

The proofs can be found elsewhere.[7]

The variance provides the spread of a distribution. It is computed via

$$\sigma_X^2 = \text{Var}(X) = E\{[X - E(X)]^2\}, \tag{6.5.4}$$

or

$$\sigma_X^2 = \begin{cases} \displaystyle\sum_k (x_k - \mu_X)^2 p_X[x_k], & X \text{ discrete,} \\ \displaystyle\int_{-\infty}^{\infty} (x - \mu_X)^2 p_X(x)\, dx, & X \text{ continuous.} \end{cases} \tag{6.5.5}$$

If we expand the right side of Equation 6.5.4, an alternative method for finding the variance is

$$\sigma_X^2 = \text{Var}(X) = E(X^2) - [E(X)]^2, \tag{6.5.6}$$

where

$$E(X^n) = \begin{cases} \displaystyle\sum_k x_k^n \, p_X[x_k], & X \text{ discrete,} \\ \displaystyle\int_{-\infty}^{\infty} x^n p_X(x)\, dx, & X \text{ continuous.} \end{cases} \tag{6.5.7}$$

• Example 6.5.1: Mean and variance of M equally likely outcomes

Consider the random variable $X = k$ where $k = 1, 2, \ldots, M$. If each event has an equally likely outcome, $p_X[x_k] = 1/M$. Then the expected or average or mean value is

$$\mu_X = \frac{1}{M} \sum_{k=1}^{M} x_k = \frac{M(M+1)}{2M} = \frac{M+1}{2}. \tag{6.5.8}$$

---

[7] For example, Kay, S. M., 2006: *Intuitive Probability and Random Processes Using MATLAB*. Springer, 833 pp. See Sections 7.7 and 12.7.

Note that the mean does *not* equal any of the possible values of X. Therefore, the expected value need not equal a value that will be actually observed.

Turning to the variance,

$$\begin{aligned}\operatorname{Var}(X) &= (M+1)\left[(2M+1)/6 - (M+1)/4\right] &\quad (\mathbf{6.5.9})\\ &= (M+1)\left[4M+2-3M-3\right]/12 &\quad (\mathbf{6.5.10})\\ &= (M+1)(M-1)/12 = (M^2-1)/12, &\quad (\mathbf{6.5.11})\end{aligned}$$

because

$$E(X^2) = \frac{1}{M}\sum_{k=1}^{M} x_k^2 = \frac{M(M+1)(2M+1)}{6M} = \frac{(M+1)(2M+1)}{6}. \quad (\mathbf{6.5.12})$$

We used Equation 6.5.6 to compute the variance. □

• Example 6.5.2

Let us find the mean and variance of the random variable X whose probability density function is

$$p_X(x) = \begin{cases} kx, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases} \quad (\mathbf{6.5.13})$$

From Equation 6.5.1, we have that

$$\mu_X = E(X) = \int_0^1 x(kx)\,dx = \left.\frac{kx^3}{3}\right|_0^1 = \frac{k}{3}. \quad (\mathbf{6.5.14})$$

From Equation 6.5.6, the variance of X is

$$\sigma_X^2 = \operatorname{Var}(X) = E(X^2) - [E(X)]^2 = \int_0^1 x^2(kx)\,dx - \frac{k^2}{9} = \left.\frac{kx^4}{4}\right|_0^1 - \frac{k^2}{9} = \frac{k}{4} - \frac{k^2}{9}. \quad (\mathbf{6.5.15})$$

□

• Example 6.5.3: Characteristic functions

The *characteristic function* of a random variable is defined by

$$\phi_X(\omega) = E[\exp(i\omega X)]. \quad (\mathbf{6.5.16})$$

If X is a discrete random variable, then

$$\phi_X(\omega) = \sum_{k=-\infty}^{\infty} p_X[x_k]e^{ik\omega}. \quad (\mathbf{6.5.17})$$

On the other hand, if X is a continuous random variable,

$$\phi_X(\omega) = \int_{-\infty}^{\infty} p_X(x)e^{i\omega x}\,dx, \quad (\mathbf{6.5.18})$$

the inverse Fourier transform (times 2π) of the Fourier transform, $p_X(x)$.

Characteristic functions are useful for computing various moments of a random variable via

$$E(X^n) = \frac{1}{i^n} \left. \frac{d^n \varphi_X(\omega)}{d\omega^n} \right|_{\omega=0}. \tag{6.5.19}$$

This follows by taking repeated differentiation of Equation 6.5.16 and then evaluating the differentiation at $\omega = 0$.

Consider, for example, the exponential probability density function $p_X(x) = \lambda e^{-\lambda x}$ with $x, \lambda > 0$. A straightforward calculation gives

$$\phi_X(\omega) = \frac{\lambda}{\lambda - \omega i}. \tag{6.5.20}$$

Substituting Equation 6.5.20 into Equation 6.5.19 yields

$$E(X^n) = \frac{n!}{\lambda^n}. \tag{6.5.21}$$

In particular,

$$E(X) = \frac{1}{\lambda} \qquad \text{and} \qquad E(X^2) = \frac{2}{\lambda^2}. \tag{6.5.22}$$

Consequently, $\mu_X = 1/\lambda$ and $\text{Var}(X) = E(X^2) - \mu_X^2 = 1/\lambda^2$. □

• Example 6.5.4: Characteristic function for a Gaussian distribution

Let us find the characteristic function for the Gaussian distribution and then use that characteristic function to compute the mean and variance.

Because

$$p_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \tag{6.5.23}$$

the characteristic function equals

$$\phi_X(\omega) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)+i\omega x}\, dx \tag{6.5.24}$$

$$= e^{i\omega\mu-\sigma^2\omega^2/2} \left\{ \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu-i\omega\sigma^2)^2}{2\sigma^2} \right] dx \right\} \tag{6.5.25}$$

$$= e^{i\omega\mu-\sigma^2\omega^2/2} \tag{6.5.26}$$

because the quantity within the wavy brackets equals one.

Given this characteristic function, Equation 6.5.26, we have that

$$\phi_X'(\omega) = (i\mu - \sigma^2\omega) e^{i\omega\mu-\sigma^2\omega^2/2}. \tag{6.5.27}$$

Therefore, $\phi_X'(0) = i\mu$ and from Equation 6.5.19, $\mu_X = E(X) = \mu$. Furthermore,

$$\phi_X''(\omega) = (i\mu - \sigma^2\omega)^2 e^{i\omega\mu-\sigma^2\omega^2/2} - \sigma^2 e^{i\omega\mu-\sigma^2\omega^2/2}. \tag{6.5.28}$$

Consequently, $\phi_X''(0) = -\mu^2 - \sigma^2$ and $\text{Var}(X) = E(X^2) - \mu_X^2 = \sigma^2$.

• **Example 6.5.5: (Weak) law of large numbers**

One of the reasons why independent identically distributed (i.i.d.) random variables play such a large role in probability and statistics lies in the (weak) *law of large numbers*. If $X_1, X_2, X_3, \ldots, X_n$ denote i.i.d. random variables and $A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then $P(|A_n - \mu| \geq \epsilon) \to 0$, as $n \to \infty$ for any $\epsilon > 0$. How is this law useful in daily life? Let us observe some random variable many times and take the average of these observations. The law of large numbers predicts that this average will converges to a *single value*, namely the mean.

Problems

1. Let $X(s)$ denote a discrete random variable associated with a fair coin toss. Then

$$X(s) = \begin{cases} 0, & s = \text{tail}, \\ 1, & s = \text{head}. \end{cases}$$

Find the expected value and variance of this random variable.

2. The geometric random variable X has the probability mass function:

$$p_X[x_k] = P(X = k) = p(1-p)^{k-1}, \qquad k = 1, 2, 3, \ldots.$$

Find its mean and variance. Hint:

$$\sum_{k=1}^{\infty} kr^{k-1} = \frac{1}{(1-r)^2}, \qquad \sum_{k=2}^{\infty} k(k-1)r^{k-2} = \frac{2}{(1-r)^3}, \quad |r| < 1,$$

and $E(X^2) = E[X(X-1)] + E(X)$.

3. Given

$$p_X(x) = \begin{cases} kx(2-x) & 0 < x < 2, \\ \quad 0, & \text{otherwise,} \end{cases}$$

(a) find k and (b) its mean and variance.

4. Given the probability density

$$p_X(x) = (a^2 - x^2)^{\nu - \frac{1}{2}}, \qquad \nu > -\tfrac{1}{2},$$

find its characteristic function using integral tables.

For the following distributions, first find their characteristic functions. Then compute the mean and variance using Equation 6.5.19.

5. Binomial distribution:

$$p_X[x_k] = \binom{n}{k} p^k q^{n-k}, \qquad 0 < p < 1,$$

where $q = 1 - p$. Hint: Use the binomial theorem to simplify Equation 6.5.17.

6. Poisson distribution:

$$p_X[x_k] = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad 0 < \lambda.$$

7. Geometric distribution:

$$p_X[x_k] = q^k p, \qquad 0 < p < 1,$$

where $q = 1 - p$.

8. Uniform distribution:

$$p_X(x) = \frac{H(x-a) - H(x-b)}{b-a}, \qquad b > a > 0.$$

Project: MATLAB's Intrinsic Function `mean` and `var`

MATLAB has the special commands `mean` and `var` to compute the mean and variance, respectively, of the random variable X. Use the MATLAB command `randn` to create a random variable `X(n)` of length `N`. Then, find the mean and variance of `X(n)`. How do these parameters vary with `N`?

Project: Monte Carlo Integration and Importance Sampling

Consider the integral $I = \int_0^1 \sqrt{1-x^2}\,dx = \pi/4$. If we were to compute it numerically by the conventional midpoint rule, the approximate value is given by

$$I_N = \frac{1}{N}\sum_{n=1}^{N} f(x_n), \tag{1}$$

where $f(x) = \sqrt{1-x^2}$ and $x_n = (n-1/2)/N$. For $N = 10$, 50, 100, and 500, the absolute value of the relative error is 2.7×10^{-3}, 2.4×10^{-4}, 8.6×10^{-5}, and 7.7×10^{-6}, respectively.

Monte Carlo integration is a simple alternative method for doing the numerical integration using random sampling. It is a particularly powerful technique for approximating complicated integrals. Here you will explore a simple one-dimensional version of this scheme.

Consider the random variable:

$$I_M = \frac{1}{M}\sum_{m=1}^{M} f(x_m), \tag{2}$$

where x_m is the mth sample point taken from the uniform distribution. I_M is a random variable because it is a function of the random variable x_m. Therefore,

$$\begin{aligned} E(I_M) &= \frac{1}{M}\sum_{m=1}^{M} E[f(x_m)] = \frac{1}{M}\sum_{m=1}^{M}\int_0^1 f(x)p(x)\,dx \\ &= \frac{1}{M}\sum_{m=1}^{M}\int_0^1 f(x)\,dx = \int_0^1 f(x)\,dx = I, \end{aligned}$$

because $p(x)$, the probability of the uniform distribution, equals 1. Furthermore, as we increase the number of samples M, I_M approaches I. By the *strong law of large numbers*,

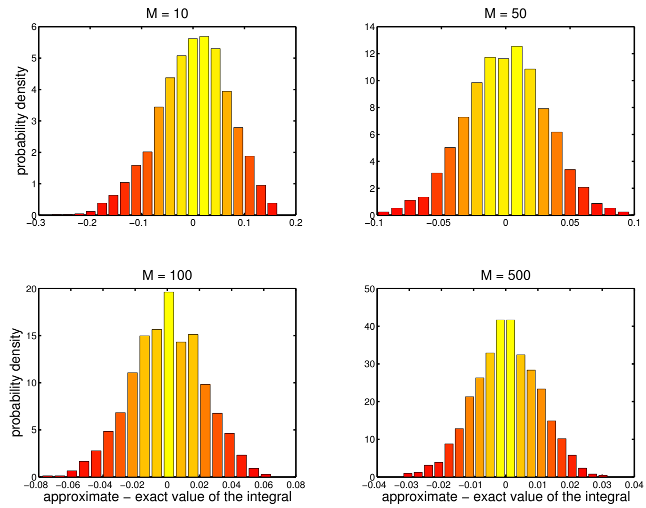


Figure 6.5.1: The probability density function arising from using Monte Carlo integration to compute $\int_0^1 \sqrt{1-x^2}\,dx$ for various values of M.

this limit is guaranteed to converge to the exact solution: $P(\lim_{M\to\infty} I_M - I) = 1$. Equation (2) is *not* the midpoint rule because the uniform grid x_n has been replaced by randomly spaced grid points.

Step 1: Write a MATLAB program that computes I_M for various values of M when x_m is selected from a uniform distribution. By running your code thousands of times, find the probability density as a function of the difference between I_M and I. Compute the mean and variance of I_M. How does the variance vary with M? See Figure 6.5.1.

The reason why standard Monte Carlo integration is not particularly good is the fact that we used a uniform distribution. A better idea would be to sample from regions where the integrand is larger. This is the essence of the concept of *importance sampling*: That certain values of the input random variable x_m in a simulation have more impact on the parameters being estimated than others.

We begin by noting that

$$I = \int_0^1 f(x)\,dx = \int_0^1 \frac{f(x)}{p_1(x)} p_1(x)\,dx,$$

where $p_1(x)$ is a new probability density function that replaces the uniform probability distribution and is relatively larger when $f(x)$ is larger and relatively smaller when $f(x)$ is smaller.

The question now becomes how to compute $p_1(x)$. We shall use the VEGAS algorithm, which constructs $p_1(x)$ by sampling $f(x)$ K times, where $K < M$. Within each kth subinterval we assume that there are M/K uniformly distributed points. Therefore,

$$p_1(x_m) = \frac{K\,f(s_m)}{\sum_{k=1}^{K} f(s_k)},$$

where s_k is the center point of the kth subinterval within which the mth point is located. For each m, we must find x_m. This is done in two steps: First we randomly choose the kth

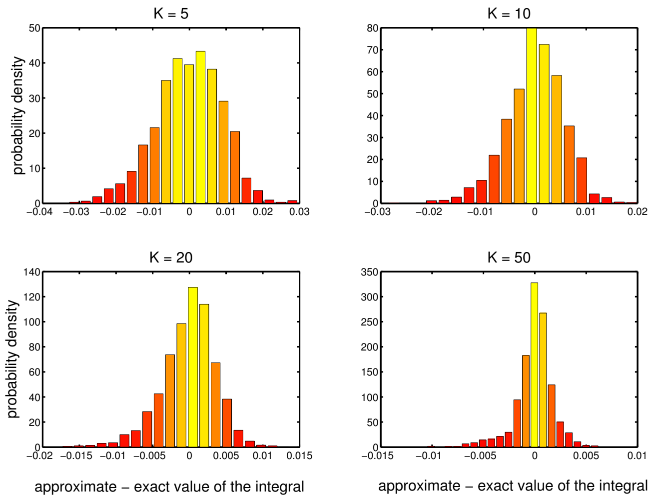


Figure 6.5.2: The probability density function arising from using importance sampling with Monte Carlo integration to compute $\int_0^1 \sqrt{1-x^2}\,dx$ for various values of K and $M = 100$.

subinterval using a uniform distribution. Then we randomly choose the point x_m within that subinterval using a uniform distribution. Therefore, our modified integration scheme becomes

$$I_M = \frac{1}{M}\sum_{m=1}^{M}\frac{f(x_m)}{p_1(x_m)}. \tag{3}$$

Now,

$$\begin{aligned} E(I_M) &= \frac{1}{M}\sum_{m=1}^{M} E\left[\frac{f(x_m)}{p_1(x_m)}\right] = \frac{1}{M}\sum_{m=1}^{M}\int_0^1 \frac{f(x)}{p_1(x)} p_1(x)p_2(x)\,dx \\ &= \frac{1}{M}\sum_{m=1}^{M}\int_0^1 f(x)\,dx = \int_0^1 f(x)\,dx = I, \end{aligned}$$

because $p_2(x) = 1$.

Step 2: Write a MATLAB program that computes I_M for various values of K for a fixed value of M. Recall that you must first select the subdivision using the MATLAB function `rand` and then the value of x_m within the subdivision using a uniform distribution. By running your code thousands of times, find the probability density as a function of the difference between I_M and I. Compute the mean and variance of I_M. How does the variance vary with M? See Figure 6.5.2.

6.6 SOME COMMONLY USED DISTRIBUTIONS

In the previous sections we introduced the concept of probability distributions and their description via mean and variance. In this section we focus on some special distributions, both discrete and continuous, that appear often in engineering.

Bernoulli distribution

Consider an experiment where the outcome can be classified as either a success or failure. The probability of a success is p and the probability of a failure is $1-p$. Then these "Bernoulli trials" have a random variable X associated with them where the probability mass function is given by

$$p_X[x_k] = P(X = k) = p^k(1-p)^{1-k}, \qquad k = 0, 1, \tag{6.6.1}$$

where $0 \le p \le 1$. From Equation 6.3.12 the cumulative density function of the Bernoulli random variable X is

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1-p, & 0 \le x < 1, \\ 1, & 1 \le x. \end{cases} \tag{6.6.2}$$

The mean and variance of the Bernoulli random variable X are

$$\mu_X = E(X) = p, \qquad \text{and} \qquad \sigma_X^2 = \mathrm{Var}(X) = p(1-p). \tag{6.6.3}$$

• Example 6.6.1

A simple pass and fail process is taking a final exam, which can be modeled by a Bernoulli distribution. Suppose a class passed a final exam with the probability of 0.75. If X denotes the random variable that someone passed the exam, then

$$E(X) = p = 0.75, \qquad \text{and} \qquad \mathrm{Var}(X) = p(1-p) = (0.75)(0.25) = 0.1875. \tag{6.6.4}$$

□

Geometric distribution

Consider again an experiment where we either have success with probability p or failure with probability $1-p$. This experiment is repeated until the first success occurs. Let random variable X denote the trial number on which this first success occurs. Its probability mass function is

$$p_X[x_k] = P(X = k) = p(1-p)^{k-1}, \qquad k = 1, 2, 3, \ldots. \tag{6.6.5}$$

From Equation 6.3.12 the cumulative density function of this geometric random variable X is

$$F_X(x) = P(X \le x) = 1 - (1-p)^k. \tag{6.6.6}$$

The mean and variance of the geometric random variable X are

$$\mu_X = E(X) = \frac{1}{p}, \qquad \text{and} \qquad \sigma_X^2 = \mathrm{Var}(X) = \frac{1-p}{p^2}. \tag{6.6.7}$$

- **Example 6.6.2**

A particle within an accelerator has the probability 0.01 of hitting a target material. (a) What is the probability that the first particle to hit the target is the 50th? (b) What is the probability that the target will be hit by *any* particle?

$$P(\text{first particle to hit is the 50th}) = 0.01(0.99)^{49} = 0.0061. \tag{6.6.8}$$

$$P(\text{target hit by any of first 50th particles}) = \sum_{n=1}^{50} 0.01(0.99)^{n-1} = 0.3950. \tag{6.6.9}$$

□

- **Example 6.6.3**

The police ticket 5% of parked cars. Assuming that the cars are ticketed independently, find the probability of 1 ticket on a block with 7 parked cars.

Each car is a Bernoulli trial with $P(\text{ticket}) = 0.05$. Therefore,

$$P(1 \text{ ticket on block}) = P(1 \text{ ticket in 7 trials}) = \binom{7}{1}(0.95)^6(0.05) = 0.2573. \tag{6.6.10}$$

□

Binomial distribution

Consider now an experiment in which n independent Bernoulli trials are performed and X represents the number of successes that occur in the n trials. In this case the random variable X is called *binomial* with parameters (n, p) with a probability mass function given by

$$p_X[x_k] = P(X = k) = \binom{n}{k} p^k(1-p)^{n-k}, \qquad k = 0, 1, \ldots, n, \tag{6.6.11}$$

where $0 \le p \le 1$, and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \tag{6.6.12}$$

the *binomial coefficient*. The term p^k arises from the k successes while $(1-p)^{n-k}$ is due to the failures. The binomial coefficient gives the number of ways that we pick those k successes from the n trials.

The corresponding cumulative density function of X is

$$F_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}, \qquad n \le x < n+1. \tag{6.6.13}$$

The mean and variance of the binomial random variable X are

$$\mu_X = E(X) = np, \qquad \text{and} \qquad \sigma_X^2 = \text{Var}(X) = np(1-p). \tag{6.6.14}$$

A Bernoulli random variable is the same as a binomial random variable when the parameters are $(1, n)$.

• Example 6.6.4

Let us find the probability of rolling the same side of a die (say, the side with N dots on it) at least 3 times when a fair die is rolled 4 times.

During our 4 tosses, we could obtain no rolls with N dots on the side ($k = 0$), one roll with N dots ($k = 1$), two rolls with N dots ($k = 2$), three rolls with N dots ($k = 3$), or four rolls with N dots ($k = 4$). If we define A as the event of rolling a die so that the side with N dots appears *at least* three times, then we must add the probabilities for $k = 3$ and $k = 4$. Therefore,

$$P(A) = p_X[x_3] + p_X[x_4] = \binom{4}{3} p^3(1-p)^1 + \binom{4}{4} p^4(1-p)^0 \tag{6.6.15}$$

$$= \frac{4!}{3!1!} p^3(1-p)^1 + \frac{4!}{4!0!} p^4(1-p)^0 = 0.0162 \tag{6.6.16}$$

because $p = \frac{1}{6}$. □

• Example 6.6.5

If 10 random binary digits are transmitted, what is the probability that *more* than seven 1's are included among them?

Let X denote the number of 1's among the 10 digits. Then

$$P(X > 7) = P(X = 8) + P(X = 9) + P(X = 10) = p_X[x_8] + p_X[x_9] + p_X[x_{10}] \tag{6.6.17}$$

$$= \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 + \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^1 + \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0 \tag{6.6.18}$$

$$= (45 + 10 + 1)\left(\frac{1}{2}\right)^{10} = \frac{56}{1024}. \tag{6.6.19}$$

□

Poisson distribution

The Poisson probability distribution arises as an approximation for the binomial distribution as $n \to \infty$ and $p \to 0$ such that np remains finite. To see this, let us rewrite the binomial distribution as follows:

$$P(X = k) = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\frac{\lambda^n}{n!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}, \tag{6.6.20}$$

if $\lambda = np$. For finite λ,

$$\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^k \to 1, \qquad \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}, \tag{6.6.21}$$

and

$$\lim_{n\to\infty}\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \to 1. \tag{6.6.22}$$

Therefore, for large n, small p and moderate λ, we can approximate the binomial distribution by the Poisson distribution:

$$p_X[x_k] = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots. \tag{6.6.23}$$

The corresponding cumulative density function of X is

$$F_X(x) = e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!}, \qquad n \le x < n+1. \tag{6.6.24}$$

The mean and variance of the Poisson random variable X are

$$\mu_X = E(X) = \lambda, \qquad \text{and} \qquad \sigma_X^2 = \text{Var}(X) = \lambda. \tag{6.6.25}$$

In addition to this approximation, the Poisson distribution is the probability distribution for a Poisson process. But that has to wait for the next chapter.

• Example 6.6.6

Consider a student union on a campus. On average 3 persons enter the union per minute. What is the probability that, during any given minute, 3 or more persons will enter the union?

To make use of Poisson's distribution to solve this problem, we must have both a large n and a small p with the average $\lambda = np = 3$. Therefore, we divide time into a large number of small intervals so that n is large while the probability that someone will enter the union is small. Assuming independence of events, we have a binomial distribution with large n. Let A denote the event that 3 or more persons will enter the union, then

$$P(\overline{A}) = p_X[0] + p_X[1] + p_X[2] = e^{-3}\left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right) = 0.423. \tag{6.6.26}$$

Therefore, $P(A) = 1 - P(\overline{A}) = 0.577.$ □

Uniform distribution

The continuous random variable X is called *uniform* if its probability density function is

$$p_X(x) = \begin{cases} 1/(b-a), & a < x < b, \\ 0, & \text{otherwise.} \end{cases} \tag{6.6.27}$$

The corresponding cumulative density function of X is

$$F_X(x) = \begin{cases} 0, & x \le a, \\ (x-a)/(b-a), & a < x < b, \\ 1, & b \le x. \end{cases} \tag{6.6.28}$$

The mean and variance of a uniform random variable X are

$$\mu_X = E(X) = \tfrac{1}{2}(a+b), \qquad \text{and} \qquad \sigma_X^2 = \text{Var}(X) = \frac{(b-a)^2}{12}. \tag{6.6.29}$$

Uniform distributions are used when we have no prior knowledge of the actual probability density function and all continuous values in some range appear equally likely.

Exponential distribution

The continuous random variable X is called *exponential* with parameter $\lambda > 0$ if its probability density function is

$$p_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x < 0. \end{cases} \tag{6.6.30}$$

The corresponding cumulative density function of X is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{6.6.31}$$

The mean and variance of an exponential random variable X are

$$\mu_X = E(X) = 1/\lambda, \qquad \text{and} \qquad \sigma_X^2 = \text{Var}(X) = 1/\lambda^2. \tag{6.6.32}$$

This distribution has the interesting property that is "memoryless." By memoryless, we mean that for a nonnegative random variable X, then

$$P(X > s + t | X > t) = P(X > s), \tag{6.6.33}$$

where $x, t \geq 0$. For example, if the lifetime of a light bulb is exponentially distributed, then the light bulb that has been in use for some hours is as good as a new light bulb with regard to the amount of time remaining until it fails.

To prove this, from Equation 6.2.4, Equation 6.6.33 becomes

$$\frac{P(X > s + t \text{ and } X > t)}{P(X > t)} = P(X > s), \tag{6.6.34}$$

or

$$P(X > s + t \text{ and } X > t) = P(X > t)P(X > s), \tag{6.6.35}$$

since $P(X > s + t \text{ and } X > t) = P(X > s + t)$. Now, because

$$P(X > s + t) = 1 - \left[1 - e^{-\lambda(s+t)}\right] = e^{-\lambda(s+t)}, \tag{6.6.36}$$

$$P(X > s) = 1 - \left(1 - e^{-\lambda s}\right) = e^{-\lambda s}, \tag{6.6.37}$$

and

$$P(X > t) = 1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t}. \tag{6.6.38}$$

Therefore, Equation 6.6.35 is satisfied and X is memoryless.

• Example 6.6.7

A component in an electrical circuit has an exponentially distributed failure time with a mean of 1000 hours. Calculate the time so that the probability of the time to failure is less than 10^{-3}.

Let the exponential random variable $X = k$ have the units of hours. Then $\lambda = 10^{-3}$. From the definition of the cumulative density function,

$$F_X(x_t) = P(X \leq x_t) = 0.001, \qquad \text{and} \qquad 1 - \exp(-\lambda x_t) = 0.001. \tag{6.6.39}$$

Solving for x_t,

$$x_t = -\ln(0.999)/\lambda = 1. \tag{6.6.40}$$

□

• Example 6.6.8

A computer contains a certain component whose time (in years) to failure is given by the random variable T distributed exponentially with $\lambda = 1/5$. If 5 of these components are installed in different computers, what is the probability that at least 2 of them will still work at the end of 8 years?

The probability that a component will last 8 years or longer is

$$P(T > 8) = e^{-8/5} = 0.2019, \tag{6.6.41}$$

because $\lambda = 1/5$.

Let X denote the number of components functioning after 8 years. Then,

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) \tag{6.6.42}$$

$$= 1 - \binom{5}{0} (0.2019)^0 (0.7981)^5 - \binom{5}{1} (0.2019)^1 (0.7981)^4 \tag{6.6.43}$$

$$= 0.2666. \tag{6.6.44}$$

□

Normal (or Gaussian) distribution

The normal distribution is the most important continuous distribution. It occurs in many applications and plays a key role in the study of random phenomena in nature.

A random variable X is called a *normal* random variable if its probability density function is

$$p_X(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\,\sigma}, \tag{6.6.45}$$

where the mean and variance of a normal random variable X are

$$\mu_X = E(X) = \mu, \qquad \text{and} \qquad \sigma_X^2 = \mathrm{Var}(X) = \sigma^2. \tag{6.6.46}$$

The distribution is symmetric with respect to $x = \mu$ and its shape is sometimes called "bell shaped." For small σ^2 we obtain a high peak and steep slope while with increasing σ^2 the curve becomes flatter and flatter.

The corresponding cumulative density function of X is

$$F_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-(\xi-\mu)^2/(2\sigma^2)}\, d\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\xi^2/2}\, d\xi. \tag{6.6.47}$$

The integral in Equation 6.6.46 must be evaluated numerically. It is convenient to introduce the *probability integral*:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\xi^2/2}\, d\xi. \tag{6.6.48}$$

Note that $\Phi(-z) = 1 - \Phi(z)$. Therefore,

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{6.6.49}$$

and

$$P(a < X \le b) = F_X(b) - F_X(a). \tag{6.6.50}$$

Consider now the intervals consisting of one σ, two σ, and three σ around the mean μ. Then, from Equation 6.6.50,

$$P(\mu - \sigma < X \le \mu + \sigma) = 0.68, \tag{6.6.51}$$

$$P(\mu - 2\sigma < X \le \mu + 2\sigma) = 0.955, \tag{6.6.52}$$

and

$$P(\mu - 3\sigma < X \le \mu + 3\sigma) = 0.997. \tag{6.6.53}$$

Therefore, approximately $\frac{2}{3}$ of the values will be distributed between $\mu - \sigma$ and $\mu + \sigma$, approximately 95% of the values will be distributed between $\mu - 2\sigma$ and $\mu + 2\sigma$, and almost all values will be distributed between $\mu - 3\sigma$ and $\mu + 3\sigma$. For most uses, then, all values will lie between $\mu - 3\sigma$ and $\mu + 3\sigma$, the so-called "three-sigma limits."

As stated earlier, the mean and variance of a normal random variable X are

$$\mu_X = E(X) = \mu, \qquad \text{and} \qquad \sigma_X^2 = \operatorname{Var}(X) = \sigma^2. \tag{6.6.54}$$

The notation $N(\mu;\sigma)$ commonly denotes that X is normal with mean μ and variance σ^2. The special case of a normal random variable Z with zero mean and unit variance, $N(0,1)$, is called a *standard normal random variable*.

Problems

1. Four coins are tossed simultaneously. Find the probability function for the random variable X that gives the number of heads. Then compute the probabilities of (a) obtaining no heads, (b) exactly one head, (c) at least one head, and (d) not less than four heads.

2. A binary source generates the digits 1 and 0 randomly with equal probability. (a) What is the probability that three 1's and three 0's will occur in a six-digit sequence? (b) What is the probability that *at least* three 1's will occur in a six-digit sequence?

3. Show that the probability of exactly n heads in $2n$ tosses of a fair coin is

$$p_X[x_n] = \frac{1 \cdot 3 \cdot 5 \cdots 2n-1}{2 \cdot 4 \cdot 6 \cdots 2n}.$$

4. If your cell phone rings, on average, 3 times between noon and 3 P.M., what is the probability that during that time period you will receive (a) no calls, (b) 6 or more calls, and (c) not more than 2 calls? Assume that the probability is given by a Poisson distribution.

5. A company sells blank DVDs in packages of 10. If the probability of a defective DVD is 0.001, (a) what is the probability that a package contains a defective DVD? (b) what is the probability that a package has two or more defective DVDs?

6. A school plans to offer a course on probability in a classroom that contains 20 seats. From experience they know that 95% of the students who enroll actually show up. If the school allows 22 students to enroll before the session is closed, what is the probability of the class being oversubscribed?

7. The lifetime (in hours) of a certain electronic device is a random variable T having a probability density function $p_T(t) = 100H(t-100)/t^2$. What is the probability that exactly 3 of 5 such devices must be replaced within the first 150 hours of operation? Assume that the events that the ith device must be replaced within this time are independent.

6.7 JOINT DISTRIBUTIONS

In the previous sections we introduced distributions that depended upon a single random variable. Here we generalize these techniques for two random variables. The range of the two-dimensional random variable (X, Y) is $R_{XY} = \{(x, y); \xi \in S \text{ and } X(\xi) = x, Y(\xi) = y\}$.

Discrete joint distribution

Let X and Y denote two *discrete* random variables defined on the same sample space (jointly distributed). The function $p_{XY}[x_i, y_j] = P[X = x_i, Y = y_j]$ is the *joint probability mass function* of X and Y. As one might expect, $p_{XY}[x_i, y_j] \ge 0$.

Let the sets of possible values of X and Y be A and B. If $x_i \notin A$ or $y_j \notin B$, then $p_{XY}[x_i, y_j] = 0$. Furthermore,

$$\sum_{x_i \in A, y_j \in B} p_{XY}[x_i, y_j] = 1. \qquad (6.7.1)$$

The *marginal probability functions* of X and Y are defined by

$$p_X[x_i] = \sum_{y_j \in B} p_{XY}[x_i, y_j], \qquad (6.7.2)$$

and

$$p_Y[y_j] = \sum_{x_i \in A} p_{XY}[x_i, y_j]. \qquad (6.7.3)$$

If X and Y are independent random variables, then $p_{XY}[x_i, y_j] = p_X[x_i] \cdot p_Y[y_j]$.

• Example 6.7.1

A joint probability mass function is given by

$$p_{XY}[x_i, y_j] = \begin{cases} k(x_i + 2y_j), & x_i = 1, 2, 3, y_j = 1, 2; \\ 0, & \text{otherwise.} \end{cases} \qquad (6.7.4)$$

Let us find the value of k, $p_X[x_i]$, and $p_Y[y_j]$.

From Equation 6.7.1, we have that

$$k \sum_{x_i=1}^{3} \sum_{y_j=1}^{2} (x_i + 2y_j) = 1, \tag{6.7.5}$$

or

$$k\big[(1+2)+(1+4)+(2+2)+(2+4)+(3+2)+(3+4)\big] = 1. \tag{6.7.6}$$

Therefore, $k = 1/30$.

Turning to $p_X[x_i]$ and $p_Y[y_j]$,

$$p_X[x_i] = k\sum_{y_j=1}^{2}(x_i+2y_j) = k(x_i+2)+k(x_i+4) = k(2x_i+6) = (x_i+3)/15, \tag{6.7.7}$$

where $x_i = 1, 2, 3$, and

$$p_Y[y_j] = k\sum_{x_i=1}^{3}(x_i+2y_j) = k(1+2y_j)+k(2+2y_j)+k(3+2y_j) = k(6+6y_j) = (1+y_j)/5, \tag{6.7.8}$$

where $y_j = 1, 2$. □

• Example 6.7.2

Consider an urn that contains 1 red ball, 2 blue balls, and 2 green balls. Let (X, Y) be a bivariate random variable where X and Y denote the number of red and blue balls, respectively, chosen from the urn. There are 18 possible ways that three balls can be drawn from the urn: *rbb*, *rbg*, *rgb*, *rgg*, *brb*, *brg*, *bbr*, *bbg*, *bgr*, *bgb*, *bgg*, *grb*, *grg*, *gbr*, *gbb*, *gbg*, *ggr*, and *ggb*.

The range of X and Y in the present problem is $R_{XY}=\{(0,1),(0,2),(1,0),(1,1),(1,2)\}$. The joint probability mass function of (X, Y) is given by $p_{XY}[x_i, y_j] = P(X = i, Y = j)$, where $x_i = 0, 1$ and $y_j = 0, 1, 2$. From our list of possible drawings, we find that $p_{XY}[0,0] = 0$, $p_{XY}[0,1] = 1/6$, $p_{XY}[0,2] = 1/6$, $p_{XY}[1,0] = 1/6$, $p_{XY}[1,1] = 1/3$, and $p_{XY}[1,2] = 1/6$. Note that all of these probabilities sum to one.

Given these probabilities, the marginal probabilities are $p_X[0] = 1/3$, $p_X[1] = 2/3$, $p_Y[0] = 1/3$, $p_Y[1] = 1/2$, and $p_Y[2] = 1/3$. Because $p_{XY}[0,0] \neq p_X[0]p_Y[0]$, X and Y are not independent variables. □

• Example 6.7.3

Consider a community where 50% of the families have a pet. Of these families, 60% have one pet, 30% have 2 pets, and 10% have 3 pets. Furthermore, each pet is equally likely (and independently) to be a male or female. If a family is chosen at random from the community, then we want to compute the joint probability that his family has M male pets and F female pets.

These probabilities are as follows:

$$P\{F = 0, M = 0\} = P\{\text{no pets}\} = 0.5, \tag{6.7.9}$$

$$P\{F=1, M=0\} = P\{1 \text{ female and total of 1 pet}\} \quad (\mathbf{6.7.10})$$
$$= P\{1 \text{ pet}\}P\{1 \text{ female}|1 \text{ pet}\} \quad (\mathbf{6.7.11})$$
$$= (0.5)(0.6) \times \frac{1}{2} = 0.15, \quad (\mathbf{6.7.12})$$

$$P\{F=2, M=0\} = P\{2 \text{ females and total of 2 pets}\} \quad (\mathbf{6.7.13})$$
$$= P\{2 \text{ pets}\}P\{2 \text{ females}|2 \text{ pets}\} \quad (\mathbf{6.7.14})$$
$$= (0.5)(0.3) \times \left(\frac{1}{2}\right)^2 = 0.0375, \quad (\mathbf{6.7.15})$$

and

$$P\{F=3, M=0\} = P\{3 \text{ females and total of 3 pets}\} \quad (\mathbf{6.7.16})$$
$$= P\{3 \text{ pets}\}P\{3 \text{ females}|3 \text{ pets}\} \quad (\mathbf{6.7.17})$$
$$= (0.5)(0.1) \times \left(\frac{1}{2}\right)^3 = 0.00625. \quad (\mathbf{6.7.18})$$

The remaining probabilities can be obtained in a similar manner. □

Continuous joint distribution

Let us now turn to the case when we have two continuous random variables. In analog with the definition given in Section 6.4, we define the two-dimensional probability density $p_{XY}(x, y)$ by

$$P(x < X \le x + dx, y < Y \le y + dy) = p_{XY}(x, y)\, dx\, dy. \quad (\mathbf{6.7.19})$$

Here, the comma in the probability parentheses means "and also."

Repeating the same analysis as in Section 6.4, we find that $p_{XY}(x, y)$ must be a single-valued function with $p_{XY}(x, y) \ge 0$, and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{XY}(x, y)\, dx\, dy = 1. \quad (\mathbf{6.7.20})$$

The joint distribution function of X and Y is

$$F_{XY}(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_{XY}(\xi, \eta)\, d\xi\, d\eta. \quad (\mathbf{6.7.21})$$

Therefore,

$$P(a < X \le b, c < Y \le d) = \int_{a}^{b} \int_{c}^{d} p_{XY}(\xi, \eta)\, d\xi\, d\eta. \quad (\mathbf{6.7.22})$$

The *marginal probability density functions* are defined by

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x, y)\, dy, \quad \text{and} \quad p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x, y)\, dx. \quad (\mathbf{6.7.23})$$

An important distinction exists upon whether the random variables are independent or not. Two variables X and Y are *independent* if and only if

$$p_{XY}(x,y) = p_X(x)p_Y(y), \tag{6.7.24}$$

and conversely.

• Example 6.7.4

The joint probability density function of bivariate random variables (X, Y) is

$$p_{XY}(x,y) = \begin{cases} kxy, & 0 < y < x < 1, \\ 0, & \text{otherwise,} \end{cases} \tag{6.7.25}$$

where k is a constant. (a) Find the value of k. (b) Are X and Y independent?

The range R_{XY} for this problem is a right triangle with its sides given by $x = 1$, $y = 0$, and $y = x$. From Equation 6.7.20,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{XY}(x,y)\,dx\,dy = k\int_0^1 x\left[\int_0^x y\,dy\right]dx = k\int_0^1 x\,\frac{y^2}{2}\bigg|_0^x\,dx \tag{6.7.26}$$

$$= \frac{k}{2}\int_0^1 x^3\,dx = \frac{k}{8}\,x^4\big|_0^1 = \frac{k}{8}. \tag{6.7.27}$$

Therefore, $k = 8$.

To check for independence we must first compute $p_X(x)$ and $p_Y(y)$. From Equation 6.7.23 and holding x constant,

$$p_X(x) = 8x\int_0^x y\,dy = 4x^3, \qquad 0 < x < 1; \tag{6.7.28}$$

$p_X(x) = 0$ otherwise. From Equation 6.7.23 and holding y constant,

$$p_Y(y) = 8y\int_y^1 x\,dx = 4y(1-y^2), \qquad 0 < y < 1. \tag{6.7.29}$$

Because $p_{XY}(x,y) \neq p_X(x)p_Y(y)$, X and Y are not independent. □

• Example 6.7.5: Buffon's needle problem

A classic application of joint probability distributions is the solution of Buffon's needle problem:[8] Consider an infinite plane with an infinite series of parallel lines spaced a unit distance apart. A needle of length $L < 1$ is thrown upward and we want to compute the probability that the stick will land so that it intersects one of these lines. See Figure 6.7.1.

There are two random variables that determine the needle's orientation: X, the distance from the lower end O of the needle to the nearest line above and Θ, the angle from the vertical to the needle. Of course, we assume that the position where the needle lands is random; otherwise, it would not be a probability problem.

---

[8] First posed in 1733, its solution is given on pages 100–104 of Buffon, G., 1777: Essai d'arithmétique morale. *Histoire naturelle, générale et particulière, Supplément*, **4**, 46–123.

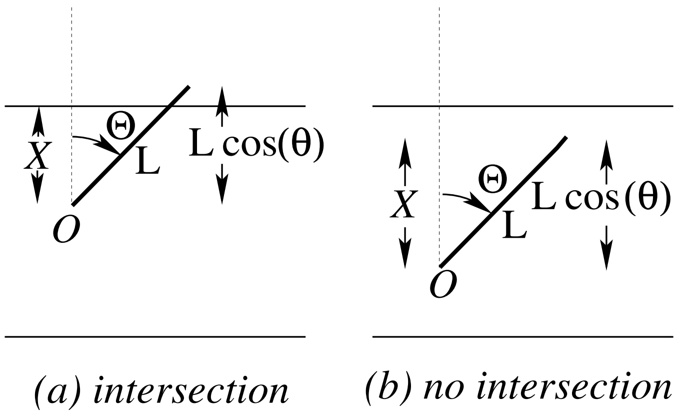


Figure 6.7.1: Schematic of Buffon's needle problem showing the random variables X and Θ.

Let us define X first. Its possible values lie between 0 and 1. Second, X is uniformly distributed on $(0, 1)$ with the probability density

$$p_X(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases} \tag{6.7.30}$$

Turning to Θ, its value lies between $-\pi/2$ to $\pi/2$ and is uniformly distributed between these values. Therefore, the probability density is

$$p_\Theta(\theta) = \begin{cases} 1/\pi, & -\pi/2 < \theta < \pi/2, \\ 0, & \text{otherwise.} \end{cases} \tag{6.7.31}$$

The probability p that we seek is

$$p = P\{\text{needle intersects line}\} = P\{X < L\cos(\Theta)\}. \tag{6.7.32}$$

Because X and Θ are independent, their joint density equals the product of the densities for X and Θ: $p_{X\Theta}(x, \theta) = p_X(x)p_\Theta(\theta)$.

The final challenge is to use $p_{X\Theta}(x, \theta)$ to compute p. In Section 6.2 we gave a geometric definition of probability. The area of the sample space is π because it consists of a rectangle in (X, Θ) space with $0 < x < 1$ and $-\pi/2 < \theta < \pi/2$. The values of X and Θ that lead to the intersection with a parallel line is $0 < x < L\cos(\theta)$ where $-\pi/2 < \theta < \pi/2$. Consequently, from Equation 6.2.5,

$$p = \int_{-\pi/2}^{\pi/2} \int_0^{L\cos(\theta)} p_{X\Theta}(x, \theta)\, dx\, d\theta = \int_{-\pi/2}^{\pi/2} \int_0^{L\cos(\theta)} p_X(x) p_\Theta(\theta)\, dx\, d\theta \tag{6.7.33}$$

$$= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \int_0^{L\cos(\theta)} dx\, d\theta = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} L\cos(\theta)\, d\theta = \frac{2L}{\pi}. \tag{6.7.34}$$

Consequently, given L, we can perform the tossing either physically or numerically, measure p, and compute the value of π. □

Convolution

It is often important to calculate the distribution of $X+Y$ from the distribution of X and Y when X and Y are independent. We shall derive the relationship for continuous random variables and then state the result for X and Y discrete.

Let X have a probability density function $p_X(x)$ and Y has the probability density $p_Y(y)$. Then the cumulative distribution function of $X+Y$ is

$$G_{X+Y}(a) = P(x+y \leq a) = \iint_{x+y\leq a} p_X(x)p_Y(y)\,dx\,dy \tag{6.7.35}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} p_X(x)p_Y(y)\,dx\,dy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{a-y} p_X(x)\,dx\right] p_Y(y)\,dy \tag{6.7.36}$$

$$= \int_{-\infty}^{\infty} F_X(a-y)p_Y(y)\,dx. \tag{6.7.37}$$

Therefore,

$$p_{X+Y}(a) = \frac{d}{da}\left[\int_{-\infty}^{\infty} F_X(a-y)p_Y(y)\,dy\right] = \int_{-\infty}^{\infty} p_X(a-y)p_Y(y)\,dy. \tag{6.7.38}$$

In the case when X and Y are discrete,

$$p_{X+Y}[a_k] = \sum_{i=-\infty}^{\infty} p_X[x_i]p_Y[a_k-x_i]. \tag{6.7.39}$$

Covariance

In Section 6.5 we introduced the concept of variance of a random variable X. There we showed that this quantity measures the dispersion, or spread, of the distribution of X about its expectation. What about the case of two jointly distributed random numbers?

Our first attempt might be to look at $\text{Var}(X)$ and $\text{Var}(Y)$. But this would simply display the dispersions of X and Y independently rather than jointly. Indeed, $\text{Var}(X)$ would give the spread along the x-direction while $\text{Var}(Y)$ would measure the dispersion along the y-direction.

Consider now $\text{Var}(aX+bY)$, the joint spread of X and Y along the $(ax+by)$-direction for two arbitrary real numbers a and b. Then

$$\text{Var}(aX+bY) = E[(aX+bY) - E(aX+bY)]^2 \tag{6.7.40}$$

$$= E[(aX+bY) - E(aX) - E(bY)]^2 \tag{6.7.41}$$

$$= E\{a[X-E(X)] + b[Y-E(Y)]\}^2 \tag{6.7.42}$$

$$= E\{a^2[X-E(X)]^2 + b^2[Y-E(Y)]^2 + 2ab[X-E(X)][Y-E(Y)]\} \tag{6.7.43}$$

$$= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2abE\{[X-E(X)][Y-E(Y)]\}. \tag{6.7.44}$$

Thus, the joint spread or dispersion of X and Y in any arbitrary direction $ax + by$ depends upon three parameters: $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, and $E\{[X - E(X)][Y - E(Y)]\}$. Because $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ give the dispersion of X and Y separately, it is the quantity $E\{[X-E(X)][Y-E(Y)]\}$ that measures the joint spread of X and Y. This last quantity,

$$\mathrm{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\}, \tag{6.7.45}$$

is called the *covariance* and is usually denoted by $\mathrm{Cov}(X, Y)$ because it determines how X and Y covary jointly. It only makes sense when we have two different random variables because in the case of a single random variable, $\mathrm{Cov}(X, X) = \sigma_X^2 = \mathrm{Var}(X)$. Furthermore, $\mathrm{Cov}(X, Y) \leq \sigma_X \sigma_Y$. In summary,

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y). \tag{6.7.46}$$

An alternative method for computing the covariance occurs if we recall that $\mu_X = E(X)$ and $\mu_Y = E(Y)$. Then

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] = E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) && (6.7.47)\\
&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y && (6.7.48)\\
&= E(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y && (6.7.49)\\
&= E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y), && (6.7.50)
\end{aligned}$$

where

$$E(XY) = \begin{cases} \displaystyle\sum_{x_i \in A, y_j \in B} x_i y_j \, p_{XY}[x_i, y_j], & X \text{ discrete}, \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, p_{XY}(x, y)\, dx\, dy, & X \text{ continuous}. \end{cases} \tag{6.7.51}$$

Therefore,

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y). \tag{6.7.52}$$

It is left as a homework assignment to show that

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y). \tag{6.7.53}$$

In general, $\mathrm{Cov}(X, Y)$ can be positive, negative, or zero. For it to be positive, X and Y decrease together or increase together. For a negative value, X would increase while Y decreases, or vice versa. If $\mathrm{Cov}(X, Y) > 0$, X and Y are *positively correlated.* If $\mathrm{Cov}(X, Y) < 0$, X and Y are *negatively correlated.* Finally, if $\mathrm{Cov}(X, Y) = 0$, X and Y are *uncorrelated.*

• Example 6.7.6

The following table gives a discrete joint density function:

$p_{XY}[x_i, y_j]$		x_i = 0	1	2	$p_Y[y_j]$
	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
y_j	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
$p_X[x_i]$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	

Because

$$E(XY) = \sum_{i=0}^{2}\sum_{j=0}^{2} x_i\, y_j\, p_{XY}[x_i, y_j] = \frac{3}{14}, \tag{6.7.54}$$

$$\mu_X = E(X) = \sum_{i=0}^{2} x_i\, p_X[x_i] = \frac{3}{4}, \qquad \text{and} \qquad \mu_Y = E(Y) = \sum_{j=0}^{2} y_j\, p_Y[y_j] = \frac{1}{2}, \tag{6.7.55}$$

then

$$\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{3}{14} - \frac{3}{4}\cdot\frac{1}{2} = -\frac{9}{56}. \tag{6.7.56}$$

Therefore, X and Y are *negatively* correlated. □

• Example 6.7.7

The random variables X and Y have the joint probability density function

$$p_{XY}(x, y) = \begin{cases} x + y, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{6.7.57}$$

Let us compute the covariance.

First, we must compute $p_X(x)$ and $p_Y(y)$. We find that

$$p_X(x) = \int_0^1 p_{XY}(x, y)\, dy = \int_0^1 (x + y)\, dy = x + \tfrac{1}{2} \tag{6.7.58}$$

for $0 < x < 1$, and

$$p_Y(y) = \int_0^1 p_{XY}(x, y)\, dx = \int_0^1 (x + y)\, dx = y + \tfrac{1}{2} \tag{6.7.59}$$

for $0 < y < 1$.

Because

$$E(XY) = \int_0^1 \int_0^1 xy(x + y)\, dx\, dy = \int_0^1 \left(y\frac{x^3}{3}\bigg|_0^1 + y^2\frac{x^2}{2}\bigg|_0^1 \right) dy \tag{6.7.60}$$

$$= \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2} \right) dy = \frac{y^2}{6} + \frac{y^2}{6}\bigg|_0^1 = \frac{1}{3}, \tag{6.7.61}$$

$$\mu_X = E(X) = \int_0^1 x\, p_X(x)\, dx = \int_0^1 \left(x^2 + \frac{x}{2}\right) dx = \frac{7}{12}, \tag{6.7.62}$$

$$\mu_Y = E(Y) = \int_0^1 y\, p_Y(y)\, dy = \int_0^1 \left(y^2 + \frac{y}{2}\right) dy = \frac{7}{12}, \tag{6.7.63}$$

then

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{3} - \frac{49}{144} = -\frac{1}{144}. \tag{6.7.64}$$

Therefore, X and Y are *negatively* correlated. □

Correlation

Although the covariance tells us how X and Y vary jointly, it depends upon the same units in which X and Y are measured. It is often better if we free ourselves of this nuisance, and we now introduce the concept of correlation.

Let X and Y be two random variables with $0 < \sigma_X^2 < \infty$ and $0 < \sigma_Y^2 < \infty$. The *correlation coefficient* $\rho(X,Y)$ between X and Y is given by

$$\rho(X,Y) = \text{Cov}\left[\frac{X - E(X)}{\sigma_X}, \frac{Y - E(Y)}{\sigma_Y}\right] = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}. \tag{6.7.65}$$

It is noteworthy that $|\rho(X,Y)| \le 1$.

Random Vectors

It is often useful to express our two random variables X and Y as a two-dimensional *random vector* $\mathbf{V} = (X\ \ Y)^T$. Then, the covariance can be written as a 2×2 *covariance matrix*, given by

$$\begin{pmatrix} \text{cov}(X,X) & \text{cov}(X,Y) \\ \text{cov}(Y,X) & \text{cov}(Y,Y) \end{pmatrix}.$$

These considerations can be generalized into the n-dimensional *random vector* consisting of n random variables that are all associated with the same events.

• Example 6.7.7

Using MATLAB, let us create two random variables by invoking `X = randn(N,1)` and `Y = randn(N,2)`, where `N` is the sample size. If `N = 10`, we would find that using the MATLAB command `cov(X,Y)` would yield

```
>> ans =

        3.1325   0.9748
        0.9748   1.4862
```

.

(If you do this experiment, you will also obtain a symmetric matrix but with different elements.) On the other hand, if `N = 1000`, we find that `cov(X,Y)` equals

```
>> ans =
```

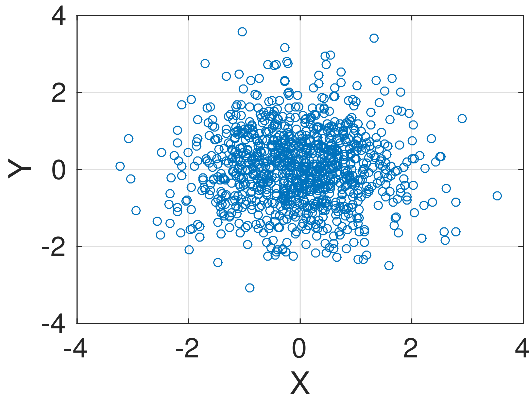


Figure 6.7.2: Scatter plot of points (X_i, Y_i) given by the random vector $\mathbf{V}$ in Example 6.7.7 when `N = 1000`.

```
  0.9793   -0.0100
 -0.0100    0.9927
```
.

The interpretation of the covariance matrix is as follows: The variance (or spread) of data given by **X** and **Y** is (essentially) unity. The correlation between **X** and **Y** is (essentially) zero. These results are confirmed in Figure 6.7.2 where we have plotted X and Y as the data points (X_i, Y_i) when `N = 1000`. We can see the symmetric distribution of data points.

Problems

1. A search committee of 5 is selected from a science department that has 7 mathematics professors, 8 physics professors, and 5 chemistry professors. If X and Y denote the number of mathematics and physics professors, respectively, that are selected, calculate the joint probability function.

2. In an experiment of rolling a fair die twice, let Z denote a random variable that equals the sum of the results. What is $p_Z[z_i]$? Hint: Let X denote the result from the first toss and Y denote the result from the second toss. What you must find is $Z = X + Y$.

3. Show that $\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X, Y)$.

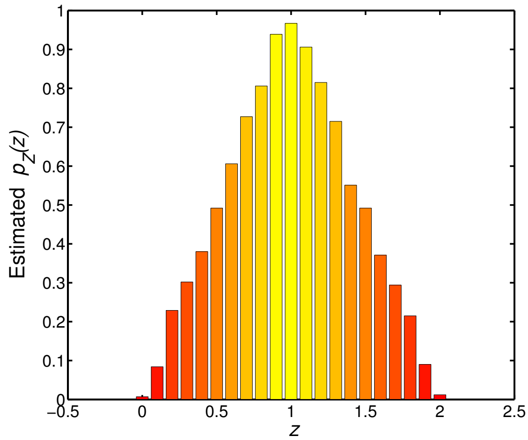
Project: Convolution

Consider two independent, uniformly distributed random variables (X, Y) that are summed to give $Z = X + Y$ with

$$p_X(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$p_Y(y) = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$



Convolution Project

Show that

$$p_Z(z) = \begin{cases} z, & 0 < z \le 1, \\ 2 - z, & 1 < z \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

Then confirm your results using MATLAB's intrinsic function `rand` to generate $\{x_i\}$ and $\{y_j\}$ and computing $p_Z(z)$. You may want to review Example 6.5.1 in my *Advanced Engineering Mathematics with MATLAB* to see how to compute a convolution analytically.

Further Readings

Beckmann, P., 1967: *Probability in Communication Engineering.* Harcourt, Brace & World, 511 pp. A presentation of probability as it applies to problems in communication engineering.

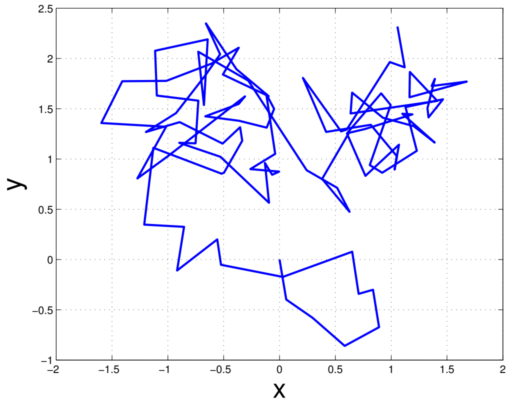
Ghahramani, S., 2000: *Fundamentals of Probability.* Prentice Hall, 511 pp. Nice introductory text on probability with a wealth of examples.

Hsu, H., 1997: *Probability, Random Variables, & Random Processes.* McGraw-Hill, 306 pp. Summary of results plus many worked problems.

Kay, S. M., 2006: *Intuitive Probability and Random Processes Using MATLAB*. Springer, 833 pp. A well-paced book designed for the electrical engineering crowd.

Ross, S. M., 2007: *Introduction to Probability Models.* Academic Press, 782 pp. An introductory undergraduate book in applied probability and stochastic processes.

Tuckwell, H. C., 1995: *Elementary Applications of Probability Theory.* Chapman & Hall, 292 pp. This book presents applications using probability theory, primarily from biology.



Chapter 7

Random Processes

In the previous chapter we introduced the concept of a random variable X. There X assumed various values of x according to a probability mass function $p_X[k]$ or probability density function $p_X(x)$. In this chapter we generalize the random variable so that it is also a function of time t. As before, the values of x assumed by the random variable $X(t)$ at a certain time is still unknown beforehand and unpredictable.

Our random, time-varying variable $X(t;\xi)$ is often used to describe a *stochastic* or *random process*. In that case, $X(t)$ is the *state* of the process at time t. The process can be either discrete or continuous in t.

A random process is not one function but a collection or family of functions, called *sample functions*, with some probability assigned to each. When we perform an experiment, we observe only one of these functions that is called a *realization* or *sample path* of the process. To observe more than a single function, we must repeat the experiment.

The *state space* of a random process is the set of *all* possible values that the random variable $X(t)$ can assume.

We can view random processes from many perspectives. First, it is a random function of time. This perspective is useful when we wish to relate an evolutionary physical phenomenon to its probabilistic model. Second, we can focus on its aspect as a random variable. This is useful in developing mathematical methods and tools to analyze random processes.

Another method for characterizing a random process examines its behavior as t and ξ vary or are kept constant. For example, if we allow t and ξ to vary, we obtain a family or *ensemble* of $X(t)$. If we allow t to vary while ξ is fixed, then $X(t)$ is simply a function of time and gives a sample function or *realization* for this particular random process. On the other hand, if we fix t and allow ξ to vary, $X(t)$ is a random variable equal to the state of the random process at time t. Finally, if we fix both t and ξ, then $X(t)$ is a number.

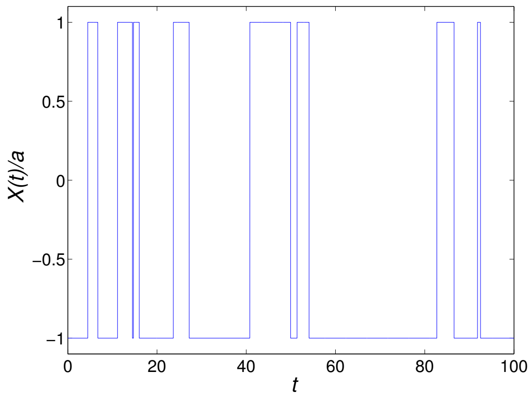


Figure 7.0.1: A realization of the random telegraph signal.

• **Example 7.0.1**

Consider a random process $X(t) = A$, where A is uniformly distributed in the interval $[0, 1]$. A plot of sample functions of $X(t)$ (a plot of $X(t)$ as a function of t) consists of horizontal straight lines that would cross the ordinate somewhere between 0 and 1. □

• **Example 7.0.2**

Consider the coin tossing experiment where the outcomes are either heads H or tails T. We can introduce the random process defined by

$$X(t; H) = \sin(t), \quad \text{and} \quad X(t; T) = \cos(t). \tag{7.0.1}$$

Note that the sample functions here are continuous functions of time. □

• **Example 7.0.3: Random telegraph signal**

Consider a signal that switches between $-a$ and $+a$ at random times. Suppose the process starts (at time $t = 0$) in the $-a$ state. It then remains in that state for a time interval T_1 at which point it switches to the state $X(t) = a$. The process remains in that state until $t = T_2$, then switches back to $X(t) = -a$. The switching time is given by a Poisson process, a random process that we discuss in Section 7.6. Figure 7.0.1 illustrates the random telegraph signal. □

Of all the possible random processes, a few are so useful in engineering and the physical sciences that they warrant special names. Some of them are:

• *Bernoulli process*

Imagine an electronics firm that produces electronic devices that either work (a success denoted by "S") or do not work (a failure or denoted "F"). We can model the production line as a series of independent, repeated events where p denotes the probability of producing a working device and $q = 1 - p$ is the probability of producing a faulty device. Thus, the production line can be modeled as a random process, called a *Bernoulli process*, which has discrete states and parameter space.

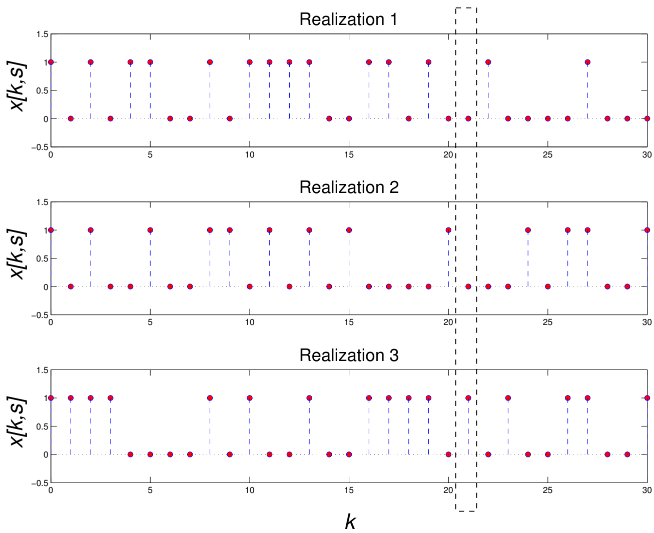


Figure 7.0.2: Three realization or sample functions of a Bernoulli random process with $p = 0.4$. The realization starts at $k = 0$ and continues forever. The dashed box highlights the values of the random variable $X[21, s]$.

If we denote each discrete trial by the integer k, a Bernoulli process generates successive outcomes at times $k = 0, 1, 2, \ldots$. Mathematically we can express this discrete random process by $X[k, s]$ where k denotes the time and s denotes the number of the realization or sample function. Furthermore, this random process maps the original experimental sample space $\{(F, F, S, \ldots), (S, F, F, \ldots), (F, F, F, \ldots), \ldots\}$ to the numerical sample space $\{(0, 0, 1, \ldots), (1, 0, 0, \ldots), (0, 0, 0, \ldots), \ldots\}$. Unlike the Bernoulli *trial* that we examined in the previous chapter, each simple event now becomes an infinite *sequence* of S's and F's.

Figure 7.0.2 illustrates three realizations or sample functions for a Bernoulli random variable when $p = 0.4$. In each realizations $s = 1, 2, \ldots$, the abscissa denotes time where each successive trial occurs at times $k = 0, 1, 2, \ldots$. When we fix the value of k, the quantity $X[k, s]$ is a random variable with a probability mass function of a Bernoulli random variable.

- *Markov process*

Communication systems transmit either the digits 0 or 1. Each transmitted digit often must pass through several stages. At each stage there is a chance that the digit that enters one stage will be changed by the time when it leaves.

A Markov process is a stochastic process that describes the probability that the digit will or will not be changed. It does this by computing the conditional distribution of any future state X_{n+1} by considering only the past states $X_0, X_1, \ldots, X_{n-1}$ and the present state X_n. In Section 7.4 we examine the simplest possible discrete Markov process, a Markov chain, when only the present and previous stages are involved. An example is the probabilistic description of birth and death, which is given in Section 7.5.

- *Poisson process*

The prediction of the total number of "events" that occur by time t is important to such diverse fields as telecommunications and banking. The most popular of these *counting processes* is the Poisson process. It occurs when:

1. the events occur "rarely,"
2. the events occur in nonoverlapping intervals of time that are independent of each other,
3. the events occur at a constant rate λ.

In Section 7.6 we explore this random process.

• *Wiener process*

A Wiener process W_t is a random process that is continuous in time and possesses the following three properties:

1. $W_0 = 0$,
2. W_t is almost surely continuous, and
3. W_t has independent increments with a distribution $W_t - W_s \sim N(0, t-s)$ for $0 \leq s \leq t$.

As a result of these properties, we have that

1. the expectation is zero, $E(W_t) = 0$,
2. the variance is $E(W_t^2) - E^2(W_t) = t$, and
3. the covariance is $\text{cov}(W_s, W_t) = \min(s, t)$.

Norbert Wiener (1894–1964) developed this process to rigorously describe the physical phenomena of Brownian motion—the apparent random motion of particles suspended in a fluid. In a Wiener process the distances traveled in Brownian motion are distributed according to a Gaussian distribution and the path is continuous but consists entirely of sharp corners.

Project: Gambler's Ruin Problem

Pete and John decide to play a coin-tossing game. Pete agrees to pay John 10 cents whenever the coin yields a "head" and John agrees to pay Pete 10 cents whenever it is a "tail." Let S_n denote the amount that John earns in n tosses of a coin. This game is a stochastic process with discrete time (number of tosses). The state space is $\{0, \pm 10, \pm 20, \cdots\}$ cents. A realization occurs each time that they play a new game.

Step 1: Create a MATLAB code to compute a realization of S_n. Plot several realizations (sample functions) of this random process. See Figure 7.0.3.

Step 2: Suppose Pete has 10 dimes. Therefore, there is a chance he will run out of dimes at some $n = N$. Modify your MATLAB code to construct a probability density function that gives the probability Pete will run out of money at time $n = N$. See Figure 7.0.3.

This problem is often formulated in terms of a gambler versus casino and called the *gambler's ruin problem*: A gambler enters a casino with \$n in cash and starts playing a game where he wins with probability p and loses with probability $1 - q$. The gambler plays the game repeatedly, betting \$1 in each round. He leaves the casino if his total fortune reaches \$$N$ or he runs out of money.

The gambler's ruin problem is also particularly popular because it a simple example of a important stochastic process called a *martingale*. In discrete time the martingale requires that the sequence $X_1, X_2, X_3, \ldots$ satisfies two conditions: $E(|X_n|) < \infty$ and $E(X_{n+1}|X_1, X_2, \ldots, X_n) = X_n$ for any time n, where $E(\cdot)$ denotes the expectation operator. If X_n is an observation, then we have a martingale if the conditional expected value of the next observation, given all the past observations, equals the most recent observation. To see that the gambler's run problem is a martingale, we compute

$$E(X_{n+1}|x_n) = \frac{1}{2}(X_n + 1) + \frac{1}{2}(X_n - 1) = X_n,$$

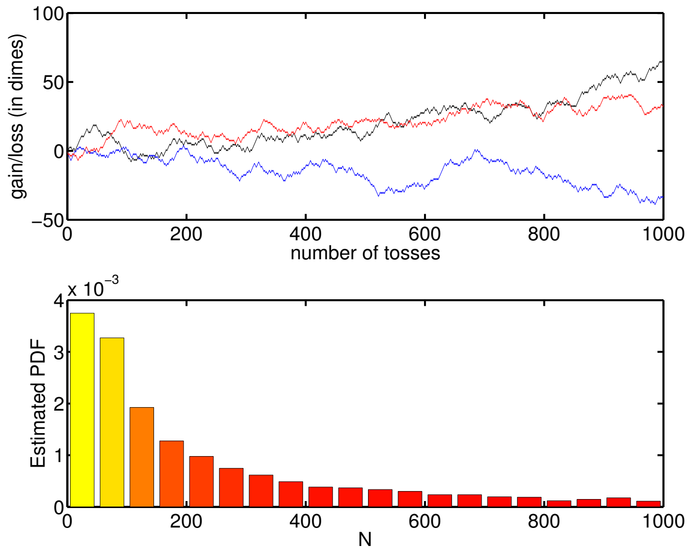


Figure 7.0.3: (a) Top frame: John's gains or losses as the result of the three different coin tossing games. (b) The probability density function for John's winning 10 dimes as a function of the number of tosses that are necessary to win 10 dimes.

where we denote the gambler's bankroll by X_n.

7.1 FUNDAMENTAL CONCEPTS

In Section 6.5 we introduced the concepts of mean (or expectation) and variance as they apply to discrete and continuous random variables. These parameters provide useful characterizations of a probability mass function or probability density function. Similar considerations hold in the case of random processes and we introduce them here.

Mean and variance

We define the mean of the random process $X(t)$ as the expected value of the process—that is, the expected value of the random variable defined by $X(t)$ for a fixed instant of time. Note that when we take the expectation, we hold the time as a nonrandom parameter and average only over the random quantities. We denote this mean of the random process by $\mu_X(t)$, since, in general, it may depend on time. The definition of the mean is just the expectation of $X(t)$:

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{\infty} x\, p_{X(t)}(t;x)\, dx. \tag{7.1.1}$$

In a similar vein, we can generalize the concept of variance so that it applies to random processes. Here variance also becomes a time-dependent function defined by

$$\sigma_X^2(t) = \text{Var}[X(t)] = E\left\{[X(t) - \mu_X(t)]^2\right\}. \tag{7.1.2}$$

• Example 7.1.1: Random linear trajectories

Consider the random process defined by

$$X(t) = A + Bt, \tag{7.1.3}$$

where A and B are uncorrelated random variables with means μ_A and μ_B. Let us find the mean of this random process.

From the linearity property of expectation, we have that

$$\mu_X(t) = E[X(t)] = E(A + Bt) = E(A) + E(B)t = \mu_A + \mu_B t. \tag{7.1.4}$$

□

• Example 7.1.2: Random sinusoidal signal

A random sinusoidal signal is one governed by $X(t) = A\cos(\omega_0 t + \Theta)$, where A and Θ are *independent* random variables, A has a mean μ_A and variance σ_A^2, and Θ has the probability density function $p_\Theta(x)$ that is nonzero only over the interval $(0, 2\pi)$.

The mean of $X(t)$ is given by

$$\mu_X(t) = E[X(t)] = E[A\cos(\omega_0 t + \Theta)] = E[A]E[\cos(\omega_0 t + \Theta)]. \tag{7.1.5}$$

We have used the property that the expectation of two independent random variables equals the product of the expectation of each of the random variables. Simplifying Equation 7.1.5,

$$\mu_X(t) = \mu_A \int_0^{2\pi} \cos(\omega_0 t + x) p_\Theta(x)\, dx. \tag{7.1.6}$$

A common assumption is that $p_\Theta(x)$ is uniformly distributed in the interval $(0, 2\pi)$, namely

$$p_\Theta(x) = \frac{1}{2\pi}, \qquad 0 < x < 2\pi. \tag{7.1.7}$$

Substituting Equation 7.1.7 into Equation 7.1.6, we find that

$$\mu_X(t) = \frac{\mu_A}{2\pi} \int_0^{2\pi} \cos(\omega_0 t + x)\, dx = 0. \tag{7.1.8}$$

□

• Example 7.1.3: Wiener random process or Brownian motion

A Wiener (random) process is defined by

$$X(t) = \int_0^t U(\xi)\, d\xi, \qquad t \ge 0, \tag{7.1.9}$$

where $U(t)$ denotes white Gaussian noise. It is often used to model Brownian motion. To find its mean, we have that

$$E[X(t)] = E\left[\int_0^t U(\xi)\, d\xi\right] = \int_0^t E[U(\xi)]\, d\xi = 0, \tag{7.1.10}$$

because the mean of white Gaussian noise equals zero. □

Autocorrelation function

When a random process is examined at two time instants $t = t_1$ and $t = t_2$, we obtain two random variables $X(t_1)$ and $X(t_2)$. A useful relationship between these two random variables is found by computing their correlation as a function of time instants t_1 and t_2. Because it is a correlation between the values of the same process sampled at two different instants of time, we shall call it the *autocorrelation function* of the process $X(t)$ and denote it by $R_X(t_1, t_2)$. It is defined in the usual way for expectations by

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)]. \tag{7.1.11}$$

Just as in the two random variables case, we can define the covariance and correlation coefficient, but here the name is slightly different. We define the *autocovariance function* as

$$\begin{aligned} C_X(t_1, t_2) &= E\{[X(t_1) - \mu_X(t_1)][X(t_2) - \mu_X(t_2)]\} && \textbf{(7.1.12)} \\ &= R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2). && \textbf{(7.1.13)} \end{aligned}$$

Note that the variance of the process and its average power (the names used for the average of $[X(t) - \mu_X(t)]^2$ and $[X(t)]^2$, respectively) can be directly obtained for the autocorrelation and the autocovariance functions, by simply using the same time instants for both t_1 and t_2:

$$E\{[X(t)]^2\} = R_X(t, t), \tag{7.1.14}$$

and

$$\sigma_X^2(t) = E\{[X(t) - \mu_X(t)]^2\} = C_X(t, t) = R_X(t, t) - \mu_X^2(t). \tag{7.1.15}$$

Therefore, the average power, Equation 7.1.14, and the variance, Equation 7.1.15, of the process follows directly from the definition of the autocorrelation and autocovariance functions.

• Example 7.1.4: Random linear trajectories

Let us continue Example 7.1.1 and find the autocorrelation of a random linear trajectory given by $X(t) = A + Bt$. From the definition of the autocorrelation,

$$\begin{aligned} R_X(t_1, t_2) &= E[X(t_1)X(t_2)] = E\{[A + Bt_1][A + Bt_2]\} && \textbf{(7.1.16)} \\ &= E(A^2) + E(AB)(t_1 + t_2) + E(B^2)t_1 t_2 && \textbf{(7.1.17)} \\ &= (\sigma_A^2 + \mu_A^2) + \mu_A\mu_B(t_1 + t_2) + (\sigma_B^2 + \mu_B^2)t_1 t_2, && \textbf{(7.1.18)} \end{aligned}$$

where σ_A^2 and σ_B^2 are the variances of the random variables A and B. We can easily find the autocovariance by

$$C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) = \sigma_A^2 + \sigma_B^2 t_1 t_2. \tag{7.1.19}$$

□

• Example 7.1.5: Random sinusoidal signal

We continue to examine the random sinusoidal signal given by $X(t) = A\cos(\omega_0 t + \Theta)$. The autocorrelation function is

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = E[A\cos(\omega_0 t_1 + \Theta)A\cos(\omega_0 t_2 + \Theta)] \tag{7.1.20}$$

$$= \tfrac{1}{2}E(A^2)E[\cos(\omega_0 t_2 - \omega_0 t_1) + \cos(\omega_0 t_2 + \omega_0 t_1 + 2\Theta)] \tag{7.1.21}$$

$$= \tfrac{1}{2}(\sigma_A^2 + \mu_A^2)\left\{\cos[\omega_0(t_2 - t_1)] + \int_0^{2\pi} \cos[\omega_0(t_2 + t_1) + 2x]p_\Theta(x)\,dx\right\}. \tag{7.1.22}$$

In our derivation we used (1) the property that the expectation of A^2 equals the sum of the variance and the square of the mean, and (2) the first term involving the cosine is *not* random because it is a function of only the time instants and the frequency. From Equation 7.1.22 we see that autocorrelation function may depend on both time instants if the probability density function of the phase angle is arbitrary. On the other hand, if $p_\Theta(x)$ is uniformly distributed, then the last term in Equation 7.1.22 vanishes because integrating the cosine function over the interval of one period is zero. In this case we can write the autocorrelation function as a function of only the time difference. The process also becomes wide-sense stationary with

$$R_X(\tau) = E[X(t)X(t+\tau)] = \tfrac{1}{2}(\sigma_A^2 + \mu_A^2)\cos(\omega_0 \tau). \tag{7.1.23}$$

□

Wide-sense stationary processes

The mathematical analysis of a random or stochastic process would appear to be hopeless because of the uncertainty of its time-dependent behavior at any instant of time. To circumvent this difficulty we will examine only those processes that have *certain* statistical properties at any instant. A wide-sense stationary process is one of the most popular.

A process is strictly stationary if its distribution and density functions do not depend on the absolute values of the time instants t_1 and t_2, but only on the difference of the time instants, $|t_1 - t_2|$. However, this is a very rigorous condition. If we are concerned only with the mean and autocorrelation function, then we can soften our definition of a stationary process to a limited form, and we call such processes wide-sense stationary processes. A *wide-sense stationary process* has a constant mean, and its autocorrelation function depends only on the time difference:

$$\mu_X(t) = E[X(t)] = \mu_X, \tag{7.1.24}$$

and

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = R_X(t_2 - t_1). \tag{7.1.25}$$

Because time does not appear in the mean, we simply write it as a constant mean value μ_X. Similarly, because the autocorrelation function is a function only of the time difference, we can write it as a function of a single variable, the time difference τ:

$$R_X(\tau) = E[X(t)X(t+\tau)]. \tag{7.1.26}$$

We can obtain similar expressions for the autocovariance function, which in this case depends only on the time difference as well:

$$C_X(\tau) = E\{[X(t) - \mu_X][X(t+\tau) - \mu_X]\} = R_X(\tau) - \mu_X^2. \qquad \textbf{(7.1.27)}$$

Finally, the average power and variance for a wide-sense stationary process are

$$E\{[X(t)]^2\} = R_X(0), \qquad \text{and} \qquad \sigma_X^2 = C_X(0) = R_X(0) - \mu_X^2, \qquad \textbf{(7.1.28)}$$

respectively. Therefore, a wide-sense stationary process has a constant average power and constant variance.

Problems

1. Find $\mu_X(t)$ and $\sigma_X^2(t)$ for the random process given by $X(t) = A\cos(\omega t)$, where ω is a constant and A is a random variable with the Gaussian (or normal) probability density function

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

2. Consider a sine-wave random process $X(t) = A\cos(\omega t + \Theta)$, $-\pi < t < \pi$, where A and ω are constants with $A > 0$. The phase function Θ is a random, uniform variable on the interval $[-\pi, \pi]$. Find the mean, variance and autocorrelation for this random function. Is this process wide-sense stationary?

3. Consider a countably infinite sequence $\{X_n, n = 0, 1, 2, 3, \ldots\}$ of a random variable defined by

$$X_n = \begin{cases} 1, & \text{for success in the } n\text{th trial,} \\ 0, & \text{for failure in the } n\text{th trial,} \end{cases}$$

with the probabilities $P(X_n = 0) = 1 - p$ and $P(X_n = 1) = p$. Thus, X_n is a Bernoulli process. For this process, $E(X_n) = p$ and $\operatorname{Var}(X_n) = p(1-p)$. Show that the autocorrelation is

$$R_X(t_1, t_2) = \begin{cases} p, & t_1 = t_2, \\ p^2, & t_1 \neq t_2; \end{cases}$$

and the autocovariance is

$$C_X(t_1, t_2) = \begin{cases} p(1-p), & t_1 = t_2, \\ 0, & t_1 \neq t_2. \end{cases}$$

Project: Computing the Autocorrelation Function

In most instances you must compute the autocorrelation function numerically. The purpose of this project is to explore this computation using the random telegraph signal. The exact solution is given by Equation 7.2.24. You will compute the autocorrelation two ways:

Step 1: Using Example 7.6.1, create MATLAB code that generates 500 realizations of the random telegraph signal.

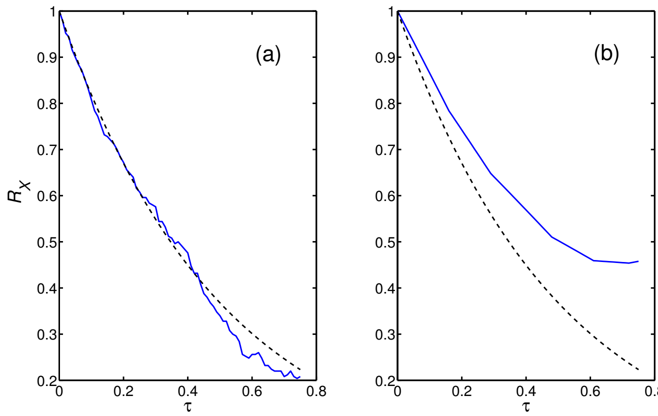


Figure 7.1.1: The autocorrelation function $R_X(\tau)$ for the random telegraph signal as a function of τ when $\lambda = 2$. The dashed line gives the exact solution. In frame (a) $X_k(t_S)X_k(t_S + \tau)$ has been averaged over 500 realizations when $t_S = 2$. In frame (b) $X_{200}(m\Delta t)X_{200}(m\Delta t + \tau)$ has been averaged with $M = 1200$ and $\Delta t = 0.01$.

Step 2: Choosing an arbitrary time t_S, compute $X_k(t_S)X_k(t_S + \tau)$ for $0 \leq 0 \leq \tau_{max}$ and $k = 1, 2, 3, \ldots, 500$. Then find the average value of $X_k(t_S)X_k(t_S + \tau)$. Plot $R_X(\tau)$ as a function of τ and include the exact answer for comparison. Does it matter how many sample functions you use?

Step 3: Now introduce a number of times $t_m = m\Delta t$, where $m = 0, 1, 2, \ldots, M$. Using only a *single realization* $k = K$ of your choice, compute $X_K(m\Delta t) \times X_K(m\Delta t + \tau)$. Then find the average value of $X_K(m\Delta t)X_K(m\Delta t + \tau)$ and plot this result as a function of τ. On the same plot, include the exact solution. Does the value of Δt matter? See Figure 7.1.1

7.2 POWER SPECTRUM

In earlier chapters we provided two alternative descriptions of signals, either in the time domain, which provides information on the shape of the waveform, or in the frequency domain, which provides information on the frequency content. Because random signals do not behave in any predictable fashion nor are they represented by a single function, it is unlikely that we can define the spectrum of a random signal by taking its Fourier transform. On the other hand, the autocorrelation of random signals describes in some sense whether the signal changes rapidly or slowly. In this section we explain and illustrate the concept of power spectrum of random signals.

For a wide-sense stationary random signal $X(t)$ with autocorrelation function $R_X(\tau)$, the *power spectrum* $S_X(\omega)$ of the random signal is the Fourier transform of the autocorrelation function:

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau)e^{-i\omega\tau}\, d\tau. \qquad \mathbf{(7.2.1)}$$

Consequently, the autocorrelation can be obtained from inverse Fourier transform of the power spectrum, or

$$R_X(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega)e^{i\omega\tau}\, d\omega. \qquad \mathbf{(7.2.2)}$$

As with any Fourier transform, it enjoys certain properties. They are:

1. The power spectrum is real and even: $S_X(-\omega) = S_X(\omega)$ and $S_X^*(\omega) = S_X(\omega)$, where $S_X^*(\omega)$ denotes the complex conjugate value of $S_X(\omega)$.
2. The power spectrum is nonnegative: $S_X(\omega) \geq 0$.
3. The average power of the random signal is equal to the integral of the power spectrum:

$$E\{[X(t)]^2\} = R_X(0) = \frac{1}{\pi}\int_0^\infty S_X(\omega)\,d\omega. \qquad \mathbf{(7.2.3)}$$

4. If the random signal has nonzero mean μ_X, then its power spectrum contains an impulse at zero frequency of magnitude $2\pi\mu_X^2$.
5. The Fourier transform of the autocovariance function of the random process is itself also a power spectrum and usually does not contain an impulse component in zero frequency.

Consider the following examples of the power spectrum:

• Example 7.2.1: Random sinusoidal signal

The sinusoidal signal is defined by

$$X(t) = A\cos(\omega_0 t + \Theta), \qquad \mathbf{(7.2.4)}$$

where the phase is uniformly distributed in the interval $[0, 2\pi]$. If the amplitude A has a mean of zero and a variance of σ^2, then the autocorrelation function is

$$R_X(\tau) = \tfrac{1}{2}\sigma^2\cos(\omega_0\tau) = R_X(0)\cos(\omega_0\tau). \qquad \mathbf{(7.2.5)}$$

The power spectrum of this signal is then

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(0)\cos(\omega_0\tau)e^{-i\omega\tau}\,d\tau = R_X(0)\pi\left[\delta(\omega-\omega_0) + \delta(\omega+\omega_0)\right]. \qquad \mathbf{(7.2.6)}$$

Because this signal contains only one frequency ω_0, its power spectrum is just two impulses, one at ω_0 and one at $-\omega_0$. Since the negative frequency appears due only to the even property of the power spectrum, it is clear that all power is concentrated at the frequency of the sinusoidal signal. While this is a very simple example, it does illustrate that the power spectrum indeed represents the way the power in the random signal is distributed among the various frequencies. We shall see later that if we also use linear systems in order to amplify or attenuate certain frequencies, the results mirror what we expect in the deterministic case. □

• Example 7.2.2: Modulated signal

Let us now examine a sinusoidal signal modulated by another random signal that contains low frequencies. This random process is described by

$$Y(t) = X(t)\cos(\omega_0 t + \Theta), \qquad \mathbf{(7.2.7)}$$

where the phase angle in Equation 7.2.7 is a random variable that is uniformly distributed in the interval $[0, 2\pi]$ and is independent of $X(t)$. Then the autocorrelation function of $Y(t)$ is given by

$$R_Y(\tau) = E[Y(t)Y(t+\tau)] = E\{X(t)\cos(\omega_0 t + \Theta)X(t+\tau)\cos[\omega_0(t+\tau)+\Theta]\} \tag{7.2.8}$$

$$= E[X(t)X(t+\tau)]E\{\cos(\omega_0 t + \Theta)\cos[\omega_0(t+\tau)+\Theta]\} = \tfrac{1}{2}R_X(\tau)\cos(\omega_0 t). \tag{7.2.9}$$

Let us take $R_X(\tau) = R_X(0)e^{-2\lambda|\tau|}$, the autocorrelation function for a random telegraph signal (see Equation 7.2.22). In this case,

$$R_Y(\tau) = \tfrac{1}{2}R_X(0)e^{-2\lambda|\tau|}\cos(\omega_0 t). \tag{7.2.10}$$

Turning to the power spectrum, the definition gives

$$S_Y(\omega) = \int_{-\infty}^{\infty} \tfrac{1}{2}R_X(\tau)\cos(\omega_0 t)e^{-i\omega\tau}\,d\tau \tag{7.2.11}$$

$$= \tfrac{1}{4}\int_{-\infty}^{\infty} R_X(\tau)\left(e^{i\omega_0\tau} + e^{-i\omega_0\tau}\right)e^{-i\omega\tau}\,d\tau \tag{7.2.12}$$

$$= \tfrac{1}{4}\left[S_X(\omega-\omega_0) + S_X(\omega+\omega_0)\right]. \tag{7.2.13}$$

Thus, the resulting power spectrum is shifted to the modulating frequency ω_0 and its negative value, with peak values located at both $\omega = \omega_0$ and $\omega = -\omega_0$. □

• Example 7.2.3: White noise

There are instances when we want to approximate random signals where the autocorrelation function is very narrow and very large about $\tau = 0$. In those cases we construct an idealization of the autocorrelation function by using the impulse or delta function $\delta(\tau)$.

In the present case when $R_X(\tau) = C\,\delta(\tau)$, the power spectrum is

$$S_X(\omega) = \int_{-\infty}^{\infty} C\,\delta(\tau)e^{-i\omega\tau}\,d\tau = C. \tag{7.2.14}$$

Thus, the power spectrum here is a flat spectrum whose value is equal to C. Because the power spectrum is flat for all frequencies, it is often called "white noise" since it contains all frequencies with equal weight.

An alternative derivation involves the random telegraph that we introduced in Example 7.0.3. As the switching rate becomes large and the rate λ approaches infinity, its amplitude increases as $\sqrt{\lambda}$. Because $R_X(0)$ increases linearly with λ, the autocorrelation function becomes

$$R_X(\tau) = C\lambda\exp(-2\lambda|\tau|). \tag{7.2.15}$$

The resulting power spectrum equals

$$S_X(\omega) = \lim_{\lambda\to\infty}\frac{4C\lambda^2}{\omega^2+4\lambda^2} = \lim_{\lambda\to\infty}\frac{C}{1+[\omega/(2\lambda)]^2} = C. \tag{7.2.16}$$

The power spectrum is again flat for all frequencies.

The autocorrelation for white noise is an idealization because it has infinite average power. Obviously no real signal has infinite power since in practice the power spectrum decays eventually. Nevertheless, white noise is still quite useful because the decay usually occurs at such high frequencies that we can tolerate the errors of introducing a flat spectrum. □

• Example 7.2.4: Random telegraph signal

In Example 7.0.3 we introduced the random telegraph signal: $X(t)$ equals either $+h$ or $-h$, changing its value from one to the other in Poisson-distributed moments of time. The probability of n changes in a time interval τ is

$$P_\tau(n) = \frac{(\lambda\tau)^n}{n!} e^{-\lambda\tau}, \tag{7.2.17}$$

where λ denotes the average frequency of changes.

To compute the power spectrum, we must first compute the correlation function via the product $X(t)X(t+\tau)$. This product equals h^2 or $-h^2$, depending on whether $X(t) = X(t+\tau)$ or $X(t) = -X(t+\tau)$, respectively. These latter relationships depend on the number of changes during the time interval. Now,

$$P[X(t) = X(t+\tau)] = P_\tau(n \text{ even}) = e^{-\lambda\tau} \sum_{n=1}^{\infty} \frac{(\lambda\tau)^{2n}}{(2n)!} = e^{-\lambda\tau}\cosh(\lambda\tau), \tag{7.2.18}$$

and

$$P[X(t) = -X(t+\tau)] = P_\tau(n \text{ odd}) = e^{-\lambda\tau} \sum_{n=1}^{\infty} \frac{(\lambda\tau)^{2n+1}}{(2n+1)!} = e^{-\lambda\tau}\sinh(\lambda\tau). \tag{7.2.19}$$

Therefore,

$$\begin{aligned} E[X(t)X(t+\tau)] &= h^2 P_\tau(n \text{ even}) - h^2 P_\tau(n \text{ odd}) && (7.2.20)\\ &= h^2 e^{-\lambda\tau}[\cosh(\lambda\tau) - \sinh(\lambda\tau)] && (7.2.21)\\ &= h^2 e^{-2\lambda|\tau|}. && (7.2.22) \end{aligned}$$

We have introduced the absolute value sign in Equation 7.2.24 because our derivation was based on $t_2 > t_1$ and the absolute value sign takes care of the case $t_2 < t_1$.

Using Problem 1, we have that

$$S_X(\omega) = 2h^2 \int_0^\infty e^{-2\lambda\tau}\cos(\lambda\tau)\,d\tau = \frac{4h^2\lambda}{\omega^2 + 4\lambda^2}. \tag{7.2.23}$$

Problems

1. Show that

$$S_X(\omega) = 2\int_0^\infty R_X(\tau)\cos(\omega\tau)\,d\tau.$$

7.3 TWO-STATE MARKOV CHAINS

A Markov chain is a probabilistic model in which the outcomes of successive trials depend only on its immediate predecessors. The mathematical description of a Markov chain involves the concepts of *states* and *state transition*. If $X_n = i$, then we have a process with state i and time n. Given a process in state i, there is a fixed probability P_{ij} that state i will transition into state j. In this section we focus on the situation of just two states.

Imagine that you want to predict the chance of rainfall tomorrow.[1] From close observation you note that the chance of rain tomorrow depends only on whether it is raining today and *not* on past weather conditions. From your observations you find that if it rains today, then it will rain tomorrow with probability α, and if it does not rain today, then the chance it will rain tomorrow is β. Assuming that these probabilities of changes are stationary (unchanging), you would like to answer the following questions:

1. Given that it is raining (or not raining), what are the chances of it raining in eight days?
2. Suppose the day is rainy (or dry). How long will the current weather remain before it changes for the first time?
3. Suppose it begins to rain during the week. How long does it take before it stops?

If the weather observation takes place at noon, we have a discrete parameter process; the two possible states of the process are rain and no rain. Let these be denoted by 0 for no rain and 1 for rain. The four possible transitions are $(0 \to 0)$, $(0 \to 1)$, $(1 \to 0)$, and $(1 \to 1)$. Let X_n be the state of the process at the nth time point. We have $X_n = 0, 1$. Clearly, $\{X_n, n = 0, 1, 2, \ldots\}$ is a two-state Markov chain. Therefore, questions about precipitation can be answered if all the properties of the two-state Markov chains are known. Let

$$P_{i,j}^{(m,n)} = P(X_n = j | X_m = i), \qquad i, j = 0, 1; \quad m \le n. \tag{7.3.1}$$

$P_{i,j}^{(m,n)}$ denotes the probability that the state of the process at the nth time point is j given that it was at state i at the mth time point. Furthermore, if this probability is larger for $i = j$ than when $i \neq j$, the system prefers to stay or *persist* in whatever state it is. When $n = m + 1$, we have that

$$P_{i,j}^{(m,m+1)} = P(X_{m+1} = j | X_m = i). \tag{7.3.2}$$

This is known as the one-step transition probability, given that the process is at i at time m.

There are two possibilities: either $P_{ij}^{(m,m+1)}$ depends on m or $P_{ij}^{(m,m+1)}$ is independent of m, where m is the initial value of the time parameter. Our precipitation model is an example of a second type of process in which the one-step transition probabilities do not change with time. Such processes are known as *time homogeneous*. Presently we shall restrict ourselves only to these processes. Consequently, without loss of generality we can use the following notation for the probabilities:

$$P_{ij} = P(X_{m+1} = j | X_m = i) \qquad \text{for all } m, \tag{7.3.3}$$

and

$$P_{ij}^{(n)} = P(X_{m+n} = j | X_m = i) \qquad \text{for all } m. \tag{7.3.4}$$

[1] See, for example, Gabriel, K. R., and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. R. Met. Soc.*, **88**, 90–95.

Chapman-Kolmogorov equation

The Chapman[2]-Kolmogorov[3] equations provide a mechanism for computing the transition probabilities after n steps. The n-step transition probabilities $P_{ij}^{(n)}$ denote the probability that a process in state i will be in state j after n transitions, or

$$P_{ij}^{(n)} = P[X_{n+k} = j | X_k = i], \qquad n \geq 0, \quad i, j \geq 0. \tag{7.3.5}$$

Therefore, $P_{ij}^{(1)} = P_{ij}$. The Chapman-Kolmogorov equations give a method for computing these n-step transition probabilities via

$$P_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P_{ik}^{(n)} P_{kj}^{(m)}, \qquad n, m \geq 0, \tag{7.3.6}$$

for all i and j. Here $P_{ik}^{(n)} P_{kj}^{(m)}$ represents the probability that the ith starting process will go to state j in $n+m$ transitions via a path that takes it into state k at the nth transition. Equation 7.3.6 follows from

$$P_{ij}^{(n+m)} = P[X_{n+m} = j | X_0 = i] = \sum_{k=0}^{\infty} P[X_{n+m} = j, X_n = k | X_0 = i] \tag{7.3.7}$$

$$= \sum_{k=0}^{\infty} P[X_{n+m} = j | X_n = k, X_0 = i] P[X_n = k | X_0 = i] = \sum_{k=0}^{\infty} P_{ik}^{(n)} P_{kj}^{(m)}. \tag{7.3.8}$$

Transmission probability matrix

Returning to the task at hand, we have that

$$P^{(2)} = P^{(1+1)} = P \cdot P = P^2, \tag{7.3.9}$$

and by induction

$$P^{(n)} = P^{(n-1+1)} = P^{(n-1)} \cdot P = P^n, \tag{7.3.10}$$

where $P^{(n)}$ denotes the transition matrix after n steps.

From our derivation, we see the following: (1) The one-step transition probability matrix completely defines the time-homogeneous two-state Markov chain. (2) All transition probability matrices show the important property that the elements in any of their rows add up to one. This follows from the fact that the elements of a row represent the probabilities of mutually exclusive and exhaustive events on a sample space.

---

[2] Chapman, S., 1928: On the Brownian displacements and thermal diffusion of grains suspended via non-uniform fluid. *Proc. R. Soc. London, Ser. A*, **119**, 34–54.

[3] Kolmogorov, A. N., 1931: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, **104**, 415–458.

Table 7.3.1: The Probability of Rain on the nth Day.

n	P_{00}	P_{10}	P_{01}	P_{11}
1	0.7000	0.2000	0.3000	0.8000
2	0.5500	0.3000	0.4500	0.7000
3	0.4750	0.3500	0.5250	0.6500
4	0.4375	0.3750	0.5625	0.6250
5	0.4187	0.3875	0.5813	0.6125
6	0.4094	0.3938	0.5906	0.6063
7	0.4047	0.3969	0.5953	0.6031
8	0.4023	0.3984	0.5977	0.6016
9	0.4012	0.3992	0.5988	0.6008
10	0.4006	0.3996	0.5994	0.6004
∞	0.4000	0.4000	0.6000	0.6000

For two-state Markov processes, this means that

$$P_{00}^{(n)} + P_{01}^{(n)} = 1, \quad \text{and} \quad P_{10}^{(n)} + P_{11}^{(n)} = 1. \tag{7.3.11}$$

Furthermore, with the one-step transmission probability matrix:

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad 0 \le a, b \le 1, \quad |1-a-b| < 1, \tag{7.3.12}$$

then the n-step transmission probability matrix is

$$\begin{pmatrix} P_{00}^{(n)} & P_{01}^{(n)} \\ P_{10}^{(n)} & P_{11}^{(n)} \end{pmatrix} = \begin{pmatrix} \frac{b}{a+b} + a\frac{(1-a-b)^n}{a+b} & \frac{a}{a+b} - a\frac{(1-a-b)^n}{a+b} \\ \frac{b}{a+b} - b\frac{(1-a-b)^n}{a+b} & \frac{a}{a+b} + b\frac{(1-a-b)^n}{a+b} \end{pmatrix}. \tag{7.3.13}$$

This follows from the Chapman-Kolmogorov equation that

$$P_{00}^{(1)} = 1 - a, \tag{7.3.14}$$

and

$$P_{00}^{(n)} = (1-a)P_{00}^{(n-1)} + bP_{01}^{(n-1)}, \qquad n > 1, \tag{7.3.15}$$

$$= b + (1-a-b)P_{00}^{(n-1)}, \tag{7.3.16}$$

since $P_{01}^{(n)} = 1 - P_{00}^{(n)}$. Solving these equations recursively for $n = 1, 2, 3, \ldots$ and simplifying, we obtain Equation 7.3.13 as long as both a and b do not equal zero.

• Example 7.3.1

Consider a precipitation model where the chance for rain depends only on whether it rained yesterday. If we denote the occurrence of rain by state 0 and state 1 denotes no rain, then observations might give you a transition probability that looks like:

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix}. \tag{7.3.17}$$

Given that the atmosphere starts today with any one of these states, the probability of finding that it is raining on the nth day is given by P^n. Table 7.3.1 illustrates the results as a function of n. Thus, regardless of whether it rains today or not, in ten days the chance for rain is 0.4 while the chance for no rain is 0.6. □

Limiting behavior

As Table 7.3.1 suggests, as our Markov chain evolves, it reaches some steady state. Let us explore this limit of $n \to \infty$ because it often provides a simple and insightful representation of a Markov process.

For large values of n it is possible to show that the limiting probability distribution of states is independent of the initial value. In particular, for $|1-a-b|<1$, we have that

$$\lim_{n\to\infty} P^{(n)} = \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{pmatrix}. \qquad \textbf{(7.3.18)}$$

This follows from $\lim_{n\to\infty}(1-a-b)^n \to 0$ since $|1-a-b|<1$. From Equation 7.3.13 the second term in each of the elements of the matrix tends to zero as $n \to \infty$.

Let us denote these limiting probabilities by $\pi_j = \lim_{n\to\infty} P_{ij}^{(n)}$. Then, from Equation 7.3.18,

$$\pi_{00} = \pi_{10} = \frac{b}{a+b} = \pi_0, \quad \text{and} \quad \pi_{01} = \pi_{11} = \frac{a}{a+b} = \pi_1, \qquad \textbf{(7.3.19)}$$

and these limiting distributions are independent of the initial state.

Number of visits to a certain state

When a random process visits several states, we would like to know the number of visits to a certain state. Let $N_{ij}^{(n)}$ denote the number of visits the two-state Markov chain $\{X_n\}$ makes to state j, starting initially at state i, in n time periods. If $\mu_{ij}^{(n)}$ denotes the expected number of visits that the process makes to state j in n steps after it originally started at state i, and the transition probability matrix P of the two-state Markov chain is

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \qquad \textbf{(7.3.20)}$$

with $|1-a-b|<1$, then

$$||\mu_{ij}^{(n)}|| = \begin{pmatrix} \frac{nb}{a+b} + \frac{a(1-a-b)[1-(1-a-b)^n]}{(a+b)^2} & \frac{na}{a+b} - \frac{a(1-a-b)[1-(1-a-b)^n]}{(a+b)^2} \\ \frac{nb}{a+b} - \frac{b(1-a-b)[1-(1-a-b)^n]}{(a+b)^2} & \frac{na}{a+b} + \frac{b(1-a-b)[1-(1-a-b)^n]}{(a+b)^2} \end{pmatrix} \qquad \textbf{(7.3.21)}$$

To prove Equation 7.3.21, we introduce a random variable $Y_{ij}^{(k)}$, where

$$Y_{ij}^{(n)} = \begin{cases} 1, & \text{if } X_k = j \text{ and } X_0 = i, \\ 0, & \text{otherwise,} \end{cases} \qquad \textbf{(7.3.22)}$$

for $i, j = 0, 1$. This random variable $Y_{ij}^{(n)}$ gives the time at which the process visits state j. The probability distribution of $Y_{ij}^{(n)}$ for fixed k is

$Y_{ij}^{(n)}$	0	1

Probability	$1 - P_{ij}^{(n)}$	$P_{ij}^{(n)}$

Thus, we have that

$$E[Y_{ij}^{(k)}] = P_{ij}^{(k)}, \quad i,j = 0,1; \quad k = 1,2,\ldots,n. \tag{7.3.23}$$

Because $Y_{ij}^{(k)}$ equals 1 whenever the process is in state j and 0 when it is not in j, the number of visits to j, starting originally from i, in n steps is

$$N_{ij}^{(n)} = Y_{ij}^{(1)} + Y_{ij}^{(2)} + \cdots + Y_{ij}^{(n)}. \tag{7.3.24}$$

Taking the expected values and using the property that the expectation of a sum is the sum of expectations,

$$\mu_{ij}^{(n)} = E\left[N_{ij}^{(n)}\right] = P_{ij}^{(1)} + P_{ij}^{(2)} + \cdots + P_{ij}^{(n)} = \sum_{k=1}^{n} P_{ij}^{(k)}. \tag{7.3.25}$$

From Equation 7.3.13, we substitute for each $P_{ij}^{(k)}$ and find

$$\mu_{00}^{(n)} = \sum_{k=1}^{n} P_{00}^{(k)} = \sum_{k=1}^{n} \left[\frac{b}{a+b} + \frac{a(1-a-b)^k}{a+b}\right], \tag{7.3.26}$$

$$\mu_{01}^{(n)} = \sum_{k=1}^{n} P_{01}^{(k)} = \sum_{k=1}^{n} \left[\frac{a}{a+b} - \frac{a(1-a-b)^k}{a+b}\right], \tag{7.3.27}$$

$$\mu_{10}^{(n)} = \sum_{k=1}^{n} P_{10}^{(k)} = \sum_{k=1}^{n} \left[\frac{b}{a+b} - \frac{b(1-a-b)^k}{a+b}\right], \tag{7.3.28}$$

and

$$\mu_{11}^{(n)} = \sum_{k=1}^{n} P_{11}^{(k)} = \sum_{k=1}^{n} \left[\frac{a}{a+b} + \frac{b(1-a-b)^k}{a+b}\right], \tag{7.3.29}$$

finally, noting that

$$\sum_{k=1}^{n} \frac{b}{a+b} = \frac{nb}{a+b}, \tag{7.3.30}$$

and

$$\sum_{k=1}^{n} \frac{a(1-a-b)^k}{a+b} = \frac{a}{a+b} \sum_{k=1}^{n} (1-a-b)^k \tag{7.3.31}$$

$$= \frac{a}{a+b} \left[(1-a-b) + (1-a-b)^2 + \cdots + (1-a-b)^n\right] \tag{7.3.32}$$

$$= \frac{a(1-a-b)}{a+b} \left[1 + (1-a-b) + \cdots + (1-a-b)^{n-1}\right] \tag{7.3.33}$$

$$= \frac{a(1-a-b)\left[1-(1-a-b)^n\right]}{(a+b)\left[1-(1-a-b)\right]}. \tag{7.3.34}$$

Here we used the property of a geometric series that

$$\sum_{k=0}^{n-1} x^k = \frac{1-x^n}{1-x}, \qquad |x| < 1. \tag{7.3.35}$$

• **Example 7.3.2**

Let us continue with our precipitation model that we introduced in Example 7.3.1. If we wish to know the expected number of days within a week that the atmosphere will be in a given state, we have from Equation 7.3.21 that

$$\mu_{00}^{(7)} = \frac{7b}{a+b} + \frac{a(1-a-b)[1-(1-a-b)^7]}{(a+b)^2} = 3.3953, \tag{7.3.36}$$

$$\mu_{10}^{(7)} = \frac{7b}{a+b} - \frac{b(1-a-b)[1-(1-a-b)^7]}{(a+b)^2} = 2.4031, \tag{7.3.37}$$

$$\mu_{01}^{(7)} = \frac{7a}{a+b} - \frac{a(1-a-b)[1-(1-a-b)^7]}{(a+b)^2} = 3.6047, \tag{7.3.38}$$

and

$$\mu_{11}^{(7)} = \frac{7a}{a+b} + \frac{b(1-a-b)[1-(1-a-b)^7]}{(a+b)^2} = 4.5969, \tag{7.3.39}$$

since $a = 0.3$ and $b = 0.2$. □

Duration of stay

In addition to computing the number of visits to a certain state, it would also be useful to know the fraction of the discrete time that a process stays in state j out of n when the process started in state i. These fractions are:

$$\lim_{n\to\infty} \frac{\mu_{00}^{(n)}}{n} = \lim_{n\to\infty} \frac{\mu_{10}^{(n)}}{n} = \pi_0, \tag{7.3.40}$$

and

$$\lim_{n\to\infty} \frac{\mu_{01}^{(n)}}{n} = \lim_{n\to\infty} \frac{\mu_{11}^{(n)}}{n} = \pi_1. \tag{7.3.41}$$

Thus, the limiting probabilities also give the fraction of time that the process spends in the two states in the long run.

If the process is in state i ($i = 0, 1$) at some time, let us compute the number of additional time periods it stays in state i until it moves out of that state. We now want to show that this probability distribution α_i, $i = 0, 1$, is

$$P(\alpha_0 = n) = a(1-a)^n, \tag{7.3.42}$$

and

$$P(\alpha_1 = n) = b(1-b)^n, \tag{7.3.43}$$

where $n = 1, 2, 3, \ldots$. Furthermore,

$$E(\alpha_0) = (1-a)/a, \qquad E(\alpha_1) = (1-b)/b, \tag{7.3.44}$$

and

$$\text{Var}(\alpha_0) = (1-a)/a^2, \qquad \text{Var}(\alpha_1) = (1-b)/b^2, \tag{7.3.45}$$

where the transition probability matrix P of the Markov chain $\{X_n\}$ equals

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \qquad (\mathbf{7.3.46})$$

with $|1-a-b| < 1$. Clearly a or b cannot equal zero.

To prove this we note that at every step, the process has two choices: either to stay in the same state or to move to the other state. Suppose the process is in state 0 at some time. The probability of a sequence of outcomes of the type $\{0\,\underbrace{0\cdots0}_{n}\,1\}$ is required. Because of the property of Markov-dependence, we therefore have the realization of a Bernoulli process with n consecutive outcomes of one type followed by an outcome of the other type. Therefore, the probability distribution of α_0 is geometric with $(1-a)$ as the probability of "failure," and the distribution of α_1 is geometric with $(1-b)$ as the probability of failure. Thus, from Equation 6.6.5, we have that

$$P(\alpha_0 = n) = a(1-a)^n, \qquad (\mathbf{7.3.47})$$

and

$$P(\alpha_1 = n) = b(1-b)^n, \qquad (\mathbf{7.3.48})$$

where $n = 0, 1, 2, \ldots$. The expressions for the mathematical expectation and variance of α_0 and α_1 easily follow from the corresponding expressions for the geometric distribution.

• Example 7.3.3

Let us illustrate our expectation and variance expressions for our precipitation model. From Equation 7.3.44 and Equation 7.3.45, we have that

$$E(\alpha_0) = (1-a)/a = 2.3333, \qquad E(\alpha_1) = (1-b)/b = 4, \qquad (\mathbf{7.3.49})$$

and

$$\operatorname{Var}(\alpha_0) = (1-a)/a^2 = 7.7777, \qquad \operatorname{Var}(\alpha_1) = (1-b)/b^2 = 20, \qquad (\mathbf{7.3.50})$$

since $a = 0.3$ and $b = 0.2$. ☐

• Example 7.3.4: Gambler's ruin problem

At the beginning of this chapter we introduced the gambler's ruin problem as an example of a random process. Here we wish to redo that problem as a Markov chain

Our particular version of the game is as follows: A gambler plays a game involving the flipping of a coin. The probability of the coin coming up heads is p while the probability of the coin coming up tails is $q = 1-p$. He enters the game with some initial amount of money and plays until (1) he has lost all of his money or (2) he has gained N units of money. We would like to describe this game at the j flip of the coin.

Let x_i denote the probability that the gambler has i units of money. At the j flip, these probabilities will be affected by the states x_{i+1} and x_{i-1} according to

$$x_i^{j+1} = p x_{i+1}^j + q x_{i-1}^j, \qquad i = 1, 2, \ldots, N-1. \qquad (\mathbf{7.3.51})$$

Equation 7.3.51 does not describe the states $i = 0$ and $i = N$, the absorbing states. State $i = 0$ corresponds to the gambler losing all of his money and quitting the game while state $i = N$ corresponds to the gambler winning N units of money and calling it a night. Once

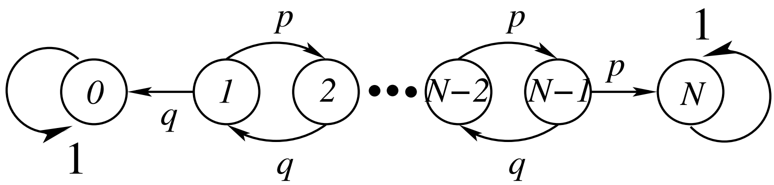


Figure 7.3.1: Markov chain diagram for the gambler's ruin problem.

these absorbing states are attained, there is no way of going to another state: $x_0^{j+1} = x_0^j$ and $x_N^{j+1} = x_N^j$. Since these absorbing states can be eventually reached from any other state, the game will eventually reach a steady state. We have illustrated this Markov chain in Figure 7.3.1.

The most convenient way of computing $\mathbf{x}$ is via matrix algebra. Using matrix notation, we can compute the probabilities from:

$$\mathbf{x}^{j+1} = \mathbf{x}^j P, \tag{7.3.52}$$

where

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & q & 0 & p & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{pmatrix} \tag{7.3.53}$$

and $\mathbf{x}$ is the row vector $[x_0 \; x_1 \; \cdots \; x_{N-1} x_N]$.

To illustrate the evolution of the gambler's ruin problem, let us set $N = 3$, $p = q = 0.5$, and $\mathbf{x}_0 = [0\ 1\ 0\ 0]^T$. Then.

$$\mathbf{x}_1 = \mathbf{x}_0 P = [0.5\ 0\ 0.5\ 0] \tag{7.3.54}$$

$$\mathbf{x}_2 = \mathbf{x}_0 P^2 = [0.5\ 0.25\ 0\ 0.25] \tag{7.3.55}$$

$$\mathbf{x}_3 = \mathbf{x}_0 P^3 = [0.625\ 0\ 0.125\ 0.25] \tag{7.3.56}$$

$$\mathbf{x}_{10} = \mathbf{x}_0 P^{10} = [0.66601562\ 0.00097656\ 0\ 0.33300781] \tag{7.3.57}$$

$$\mathbf{x}_{100} = \mathbf{x}_0 P^{100} = [0.6660666\ 0.00000000\ 0.00000000\ 0.3333333]. \tag{7.3.58}$$

The interpretation of these results is straightforward. After 100 games, the probability that the gambler, with an initial bankroll of one unit of money, will lose all his money is 2/3 while the chance that he will go home with 3 units of money is 1/3. There are no other outcomes to the game.

Problems

1. Given

$$P = \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{pmatrix},$$

(a) compute P^n and (b) find $\lim_{n\to\infty} P^n$.

2. Suppose you want to model how your dog learns a new trick. Let Fido be in state 0 if he learns the new trick and in state 1 if he fails to learn the trick. Suppose that if he learns

the trick, he will retain the trick. If he has yet to learn the trick, there is a probability α of him learning it with each training session. (a) Write down the transition matrix. (b) Compute $P^{(n)}$ where n is the number of training sessions. (c) What is the steady-state solution? Interpret your result. (d) Compute the expected amount of time that Fido will spend in each state during n training sessions.

7.4 BIRTH AND DEATH PROCESSES

In the previous section we considered two-state Markov chains that undergo n steps. As the time interval between steps tends to zero, the Markov process becomes continuous in time. In this section and the next, we consider two independent examples of continuous Markov processes.

We began Chapter 6 by showing that the deterministic description of birth and death is inadequate to explain the extinction of species. Here we will fill out the details of our analysis and extend them to population dynamics and chemical kinetics. Deterministic models lead to first-order ordinary differential equations, and this description fails when the system initially contains a small number of particles.

Consider a population of organisms that multiply by the following rules:

1. The sub-populations generated by two co-existing individuals develop completely independently of one another;
2. an individual existing at time t has a chance $\lambda\, dt + o(dt)$ of multiplying by binary fission during the following time interval of length dt;
3. the "birth rate" λ is the same for all individuals in the population at any time t;
4. an individual existing at time t has a chance $\mu\, dt + o(dt)$ of dying in the following time interval of length dt; and
5. the "death rate" μ is the same for all individuals at any time t.

Rule 3 is usually interpreted in the sense that in each birth, just one new member is added to the population, but of course mathematically (and because the age structure of the population is being ignored) it is not possible to distinguish between this and an alternative interpretation in which one of the parents dies when the birth occurs and is replaced by two children.

Let n_0 be the number of individuals at the initial time $t = 0$ and let $p_n(t)$ denote the probability that the population size $N(t)$ has the value n at the time t. Then

$$\frac{dp_n}{dt} = (n-1)\lambda p_{n-1} - n(\lambda+\mu)p_n + \mu(n+1)p_{n+1}, \qquad n \ge 1, \tag{7.4.1}$$

and

$$\frac{dp_0(t)}{dt} = \mu p_1(t), \tag{7.4.2}$$

subject to the initial condition that

$$p_n(0) = \begin{cases} 1, & n = n_0, \\ 0, & n \neq n_0. \end{cases} \tag{7.4.3}$$

Equation 7.4.1 through Equation 7.4.3 constitute a system of linear ordinary equations. The question now turns on how to solve them most efficiently. To this end we introduce a probability-generating function:

$$\phi(z,t) = \sum_{n=0}^{\infty} z^n p_n(t). \tag{7.4.4}$$

Summing Equation 7.4.1 from $n = 1$ to ∞ after we multiplied it by z^n and using Equation 7.4.2, we obtain

$$\sum_{n=0}^{\infty} z^n \frac{dp_n}{dt} = \lambda \sum_{n=1}^{\infty} (n-1) z^n p_{n-1}(t) - (\lambda+\mu) \sum_{n=1}^{\infty} n z^n p_n(t) + \mu \sum_{n=0}^{\infty} (n+1) z^n p_{n+1}(t). \quad \textbf{(7.4.5)}$$

Because

$$\sum_{n=0}^{\infty} z^n \frac{dp_n}{dt} = \frac{\partial \phi}{\partial t}, \quad \textbf{(7.4.6)}$$

$$\sum_{n=1}^{\infty} n z^n p_n(t) = z \sum_{n=0}^{\infty} n z^{n-1} p_n(t) = z \frac{\partial \phi}{\partial z}, \quad \textbf{(7.4.7)}$$

$$\sum_{n=1}^{\infty} (n-1) z^n p_{n-1}(t) = \sum_{k=0}^{\infty} k z^{k+1} p_k(t) = z^2 \sum_{k=0}^{\infty} k z^{k-1} p_k(t) = z^2 \frac{\partial \phi}{\partial z}, \quad \textbf{(7.4.8)}$$

and

$$\sum_{n=0}^{\infty} (n+1) z^n p_{n+1}(t) = \sum_{k=1}^{\infty} k z^{k-1} p_k(t) = \sum_{k=0}^{\infty} k z^{k-1} p_k(t) = \frac{\partial \phi}{\partial z}, \quad \textbf{(7.4.9)}$$

Equation 7.4.5 becomes the first-order partial differential equation

$$\frac{\partial \phi}{\partial t} = (\lambda z - \mu)(z-1) \frac{\partial \phi}{\partial z}, \quad \textbf{(7.4.10)}$$

subject to the initial condition

$$\phi(z, 0) = z^{n_0}. \quad \textbf{(7.4.11)}$$

Equation 7.4.10 is an example of a first-order partial differential equation of the general form

$$P(x,y) \frac{\partial u}{\partial x} + Q(x,y) \frac{\partial u}{\partial y} = 0. \quad \textbf{(7.4.12)}$$

This equation has solutions[4] of the form $u(x,y) = f(\xi)$ where $f(\cdot)$ is an arbitrary function that is differentiable and $\xi(x,y) =$ constant are solutions to

$$\frac{dx}{P(x,y)} = \frac{dy}{Q(x,y)}. \quad \textbf{(7.4.13)}$$

In the present case,

$$\frac{dt}{1} = -\frac{dz}{(\lambda z - \mu)(z-1)} = -\frac{dz}{(\lambda - \mu)(z-1)} + \frac{dz}{(\lambda - \mu)(z - \mu/\lambda)}. \quad \textbf{(7.4.14)}$$

Integrating Equation 7.4.14,

$$-(\lambda - \mu)t + \ln[\psi(z)] = \ln(\xi), \quad \textbf{(7.4.15)}$$

or

$$\xi(z,t) = \psi(z) e^{-(\lambda-\mu)t}, \quad \textbf{(7.4.16)}$$

---

[4] See Webster, A. G., 1966: *Partial Differential Equations of Mathematical Physics*. Dover, 446 pp. See Section 22.

where

$$\psi(z) = \frac{\lambda z - \mu}{z - 1}. \tag{7.4.17}$$

Therefore, the general solution is

$$\phi(z,t) = f\left[\psi(z)e^{-(\lambda-\mu)t}\right]. \tag{7.4.18}$$

Our remaining task is to find $f(\cdot)$. From the initial condition, Equation 7.4.11, we have that

$$\phi(z,0) = f[\psi(z)] = z^{n_0}. \tag{7.4.19}$$

Because $z = [\mu - \psi(z)]/[\lambda - \psi(z)]$, then

$$f(\psi) = \left(\frac{\mu - \psi}{\lambda - \psi}\right)^{n_0}. \tag{7.4.20}$$

Therefore,

$$\phi(z,t) = \left[\frac{\mu - \psi(z)e^{-(\lambda-\mu)t}}{\lambda - \psi(z)e^{-(\lambda-\mu)t}}\right]^{n_0}. \tag{7.4.21}$$

Once we find $\phi(z,t)$, we can compute the probabilities of each of the species from the probability generating function. For example,

$$P\{N(t) = 0|N(0) = n_0\} = p_0(t) = \phi(0,t). \tag{7.4.22}$$

From Equation 7.4.17 we have $\psi(0) = \mu$ and

$$\phi(0,t) = \left\{\frac{\mu\left[1 - e^{-(\lambda-\mu)t}\right]}{\lambda - \mu e^{-(\lambda-\mu)t}}\right\}^{n_0}, \qquad \lambda \neq \mu, \tag{7.4.23}$$

and

$$\phi(0,t) = \left(\frac{\lambda t}{1 + \lambda t}\right)^{n_0}, \qquad \lambda = \mu. \tag{7.4.24}$$

An important observation from Equation 7.4.23 and Equation 7.4.24 is that

$$\lim_{t\to\infty} p_0(t) = 1, \qquad \lambda \le \mu, \tag{7.4.25}$$

and

$$\lim_{t\to\infty} p_0(t) = \left(\frac{\mu}{\lambda}\right)^{n_0}, \qquad \lambda > \mu. \tag{7.4.26}$$

This limit can be interpreted as the probability of extinction of the population in a finite time. Consequently, there will be "almost certain" extinction whenever $\lambda \le \mu$. These results, which are true whatever the initial number of individuals may be, show very clearly the inadequacy of the deterministic description of population dynamics.

Finally, let us compute the mean and variance for the birth and death process. The expected number of individuals at time t is

$$m(t) = E[N(t)] = \sum_{n=0}^{\infty} n\, p_n(t) = \sum_{n=1}^{\infty} n\, p_n(t). \tag{7.4.27}$$

Now

$$\frac{dm}{dt} = \sum_{n=1}^{\infty} n \frac{dp_n}{dt} = \sum_{n=1}^{\infty} n \left[(n-1)\lambda p_{n-1} - n(\lambda+\mu)p_n + \mu(n+1)p_{n+1}\right] \qquad \textbf{(7.4.28)}$$

$$= \lambda \sum_{n=1}^{\infty}(n-1)^2 p_{n-1} + \lambda \sum_{n=1}^{\infty}(n-1)p_{n-1} - (\lambda+\mu)\sum_{n=1}^{\infty} n^2 p_n + \mu \sum_{n=1}^{\infty}(n+1)^2 p_{n+1}$$

$$- \mu \sum_{n=1}^{\infty}(n+1)p_{n+1} \qquad \textbf{(7.4.29)}$$

$$= -(\lambda+\mu)\sum_{n=1}^{\infty} n^2 p_n + \lambda \sum_{i=0}^{\infty} i^2 p_m + \mu \sum_{k=2}^{\infty} k^2 p_k + \lambda \sum_{i=0}^{\infty} i\, p_i - \mu \sum_{k=2}^{\infty} k\, p_k. \qquad \textbf{(7.4.30)}$$

In the first three sums in Equation 7.4.30, terms from $i, k, n = 2$, and onward cancel and leave $-(\lambda+\mu)p_1 + \lambda p_1 = -\mu p_1$. Therefore,

$$\frac{dm}{dt} = -\mu p_1 + \lambda \sum_{i=0}^{\infty} i\, p_i - \mu \sum_{k=2}^{\infty} k\, p_k = (\lambda-\mu)\sum_{n=0}^{\infty} n\, p_n, = (\lambda-\mu)m. \qquad \textbf{(7.4.31)}$$

If we choose the initial condition $m(0) = n_0$, the solution is

$$m(t) = n_0 e^{(\lambda-\mu)t}. \qquad \textbf{(7.4.32)}$$

This is the same as the deterministic result with the birth rate $\overline{b}$ replaced by λ and the death rate $\overline{d}$ replaced by μ. Furthermore, if $\lambda = \mu$, the mean size of the population is constant.

The second moment of $N(t)$ is

$$M(t) = \sum_{n=0}^{\infty} n^2 p_n(t). \qquad \textbf{(7.4.33)}$$

Proceeding as before, we have that

$$\frac{dM}{dt} = \sum_{n=1}^{\infty} n^2 \frac{dp_n}{dt} = \sum_{n=1}^{\infty} n^2 \left[\lambda(n-1)p_{n-1} - (\lambda+\mu)n\, p_n + \mu(n+1)p_{n+1}\right] \qquad \textbf{(7.4.34)}$$

$$= \lambda \sum_{n=1}^{\infty}(n-1)^3 p_{n-1} + 2\lambda \sum_{n=1}^{\infty}(n-1)^2 p_{n-1} + \lambda \sum_{n=1}^{\infty}(n-1)p_{n-1} - (\lambda+\mu)\sum_{n=1}^{\infty} n^3 p_n$$

$$+ \mu \sum_{n=1}^{\infty}(n+1)^3 p_{n+1} - 2\mu \sum_{n=1}^{\infty}(n+1)^2 p_{n+1} + \mu \sum_{n=1}^{\infty}(n+1)p_{n+1} \qquad \textbf{(7.4.35)}$$

$$= \lambda \sum_{k=1}^{\infty} k^3 p_k + 2\lambda \sum_{k=1}^{\infty} k^2 p_k + \lambda \sum_{k=1}^{\infty} k\, p_k - (\lambda+\mu)\sum_{n=1}^{\infty} n^3 p_n$$

$$+ \mu \sum_{i=2}^{\infty} i^3 p_i - 2\mu \sum_{i=2}^{\infty} i^2 p_i + \mu \sum_{i=2}^{\infty} i\, p_i. \qquad \textbf{(7.4.36)}$$

The three sums, which contain either i^3 or k^3 or n^3 in them, cancel when $i, k, n = 2$ and onward; these three sums reduce to $-\mu p_1$. The sums that involve i^2 or k^2 can be written in terms of $M(t)$. Finally, the sums involving i and k can be expressed in terms of $m(t)$. Therefore, Equation 7.4.36 becomes the first-order ordinary differential equation

$$\frac{dM}{dt} - 2(\lambda - \mu)M = (\lambda + \mu)m(t) = (\lambda + \mu)n_0 e^{(\lambda-\mu)t}, \qquad \textbf{(7.4.37)}$$

with $M(0) = n_0^2$.

Equation 7.4.37 can be solved exactly using the technique of integrating factors. Its solution is

$$M(t) = n_0^2 e^{2(\lambda-\mu)t} + \frac{\lambda + \mu}{\lambda - \mu} n_0 e^{(\lambda-\mu)t}\left[e^{(\lambda-\mu)t} - 1\right]. \qquad \textbf{(7.4.38)}$$

From the definition of variance, Equation 6.6.5, the variance of the population in the birth and death process equals

$$\operatorname{Var}[N(t)] = n_0 \frac{(\lambda + \mu)}{(\lambda - \mu)} e^{(\lambda-\mu)t}\left[e^{(\lambda-\mu)t} - 1\right], \qquad \lambda \neq \mu, \qquad \textbf{(7.4.39)}$$

or

$$\operatorname{Var}[N(t)] = 2\lambda n_0 t, \qquad \lambda = \mu. \qquad \textbf{(7.4.40)}$$

• Example 7.4.1: Chemical kinetics

The use of Markov processes to describe birth and death has become quite popular. Indeed, it can be applied to any phenomena where something is being created or destroyed. Here we illustrate its application in chemical kinetics.

Let the random variable $X(t)$ be the number of A molecules in a unimolecular reaction $A \to B$ (such as radioactive decay) at time t. A stochastic model that describes the decrease of A can be constructed from the following assumptions:

1. The probability of transition from n to $n - 1$ in the time interval $(t, t + \Delta t)$ is $n\lambda\Delta t + o(\Delta t)$ where λ is a constant and $o(\Delta t)$ denotes that $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$.
2. The probability of a transition from n to $n - j$, $j > 1$, in the time interval $(t, t + \Delta t)$ is at least $o(\Delta t)$ because the time interval is so small that only one molecule undergoes a transition.
3. The reverse reaction occurs with probability zero.

The equation that governs the probability that $X(t) = n$ is

$$p_n(t + \Delta t) = (n + 1)\lambda\Delta t p_{n+1}(t) + (1 - \lambda n\Delta t)p_n(t) + o(\Delta t). \qquad \textbf{(7.4.41)}$$

Transposing $p_n(t)$ from the right side, dividing by Δt, and taking the limit $\Delta t \to 0$, we obtain the differential-difference equation[5]

$$\frac{dp_n}{dt} = (n + 1)\lambda p_{n+1}(t) - n\lambda p_n(t). \qquad \textbf{(7.4.42)}$$

Equation 7.4.42 is frequently called the *stochastic master equation.* The first term on the right side of this equation vanishes when $n = n_0$.

[5] McQuarrie, D. A., 1963: Kinetics of small systems. I. *J. Chem. Phys.*, **38**, 433–436.

The solution of Equation 7.4.42 once again involves introducing a generating function for $p_n(t)$, namely

$$F(z,t) = \sum_{n=0}^{n_0} p_n(t) z^n, \qquad |z| < 1. \tag{7.4.43}$$

Summing Equation 7.4.42 from $n = 0$ to n_0 after multiplying it by z^n, we find

$$\sum_{n=0}^{n_0} z^n \frac{dp_n}{dt} = \lambda \sum_{n=0}^{n_0-1} (n+1) z^n p_{n+1}(t) - \lambda \sum_{n=1}^{n_0} n z^n p_n(t). \tag{7.4.44}$$

Because

$$\sum_{n=0}^{n_0} z^n \frac{dp_n}{dt} = \frac{\partial F}{\partial t}, \tag{7.4.45}$$

$$\sum_{n=0}^{n_0} n z^n p_n(t) = z \sum_{n=0}^{n_0} n z^{n-1} p_n(t) = z \frac{\partial F}{\partial z}, \tag{7.4.46}$$

and

$$\sum_{n=0}^{n_0-1} (n+1) z^n p_{n+1}(t) = \sum_{k=1}^{n_0} k z^{k-1} p_k(t) = \sum_{k=0}^{n_0} k z^{k-1} p_k(t) = \frac{\partial F}{\partial z}, \tag{7.4.47}$$

Equation 7.4.44 becomes the first-order partial differential equation

$$\frac{\partial F}{\partial t} = \lambda (1 - z) \frac{\partial F}{\partial z}. \tag{7.4.48}$$

The solution of Equation 7.4.48 follows the method used to solve Equation 7.4.10. Here we find $\xi(z,t)$ via

$$\frac{dt}{1} = \frac{dz}{\lambda(z-1)}, \tag{7.4.49}$$

or

$$\xi(z,t) = (z-1)e^{-\lambda t}. \tag{7.4.50}$$

Therefore,

$$F(z,t) = f\left[(z-1)e^{-\lambda t}\right]. \tag{7.4.51}$$

To find $f(\cdot)$, we use the initial condition that $F(z,0) = z^{n_0}$. This yields $f(y) = (1+y)^{n_0}$ and

$$F(z,t) = \left[1 + (z-1)e^{-\lambda t}\right]^{n_0}. \tag{7.4.52}$$

Once again, we can compute the mean and variance of this process. Because

$$\left.\frac{\partial F}{\partial z}\right|_{z=1} = \sum_{n=0}^{n_0} n p_n(t), \tag{7.4.53}$$

the mean is given by

$$E[X(t)] = \frac{\partial F(1,t)}{\partial z}. \tag{7.4.54}$$

To compute the variance, we first compute the second moment. Since

$$z \frac{\partial F}{\partial z} = \sum_{n=0}^{n_0} n z^n p_n(t), \tag{7.4.55}$$

and

$$\frac{\partial}{\partial z}\left(z\frac{\partial F}{\partial z}\right) = \sum_{n=0}^{n_0} n^2 z^{n-1} p_n(t), \tag{7.4.56}$$

we have that

$$\sum_{n=0}^{n_0} n^2 p_n(t) = \frac{\partial^2 F(1,t)}{\partial z^2} + \frac{\partial F(1,t)}{\partial z}. \tag{7.4.57}$$

From Equation 6.6.5, the final result is

$$\text{Var}[X(t)] = \frac{\partial^2 F(1,t)}{\partial z^2} + \frac{\partial F(1,t)}{\partial z} - \left[\frac{\partial F(1,t)}{\partial z}\right]^2. \tag{7.4.58}$$

Upon substituting Equation 7.4.52 into Equations 7.4.54 and 7.4.58, the mean and variance for this process are

$$E[X(t)] = n_0 e^{-\lambda t}, \qquad \text{and} \qquad \text{Var}[X(t)] = n_0 e^{-\lambda t}\left(1 - e^{-\lambda t}\right). \tag{7.4.59}$$

Because the expected value of the stochastic representation also equals the deterministic result, the two representations are "consistent in the mean." Further study shows that this is true only for unimolecular reactions. Upon expanding Equation 7.4.52, we find that

$$p_n(t) = \binom{n_0}{n} e^{-n\lambda t}\left(1 - e^{-\lambda t}\right)^{n_0 - n}. \tag{7.4.60}$$

An alternative method to the generating function involves Laplace transforms.[6] To illustrate this method, we again examine the reaction $A \to B$. The stochastic master equation is

$$\frac{dp_n}{dt} = (n-1)\lambda p_{n-1}(t) - n\lambda p_n(t), \qquad n_0 \le n < \infty, \tag{7.4.61}$$

$p_n(t) = 0$ for $0 < n < n_0$, where $p_n(t)$ denotes the probability that we have n particles of B at time t. The initial condition is that $p_{n_0}(0) = 1$ and $p_m(0) = 0$ for $m \neq n_0$ where n_0 denotes the initial number of molecules of B.

Taking the Laplace transform of Equation 7.4.61, we find that

$$sP_n(s) = (n-1)\lambda P_{n-1}(s) - n\lambda P_n(s), \qquad n_0 < n < \infty, \tag{7.4.62}$$

and

$$sP_{n_0}(s) - 1 = -n\lambda P_{n_0}(s). \tag{7.4.63}$$

Therefore, solving for $P_n(s)$,

$$P_n(s) = \frac{(n-1)\lambda}{s + n\lambda} P_{n-1}(s) = \frac{\lambda^{n-n_0}(n-1)!}{(n_0-1)!} \prod_{k=n_0}^{n} (s + k\lambda)^{-1}. \tag{7.4.64}$$

From partial fractions,

$$P_n(s) = \frac{(n-1)!}{(n_0-1)!} \sum_{k=n_0}^{n} \frac{(-1)^{k-n_0}}{(k-n_0)!(n-k)!(s+k\lambda)}. \tag{7.4.65}$$

---

[6] Ishida, K., 1969: Stochastic model for autocatalytic reaction. *Bull. Chem. Soc. Japan*, **42**, 564–565.

Taking the inverse Laplace transform,

$$p_n(t) = \frac{(n-1)!}{(n_0-1)!(n-n_0)!} \sum_{k=n_0}^{n} \frac{(-1)^{k-n_0}(n-n_0)!}{(k-n_0)!(n-k)!} e^{-\lambda k t} \tag{7.4.66}$$

$$= \frac{(n-1)! e^{-\lambda n_0 t}}{(n_0-1)!(n-n_0)!} \sum_{j=0}^{n-n_0} \frac{(-1)^j (n-n_0)!}{j!(n-n_0-j)!} e^{-\lambda j t} \tag{7.4.67}$$

$$= \frac{(n-1)! e^{-\lambda n_0 t}}{(n_0-1)!(n-n_0)!} \left(1-e^{-\lambda t}\right)^{n-n_0}, \tag{7.4.68}$$

where we introduced $j = k - n_0$ and eliminated the summation via the binomial theorem. Equation 7.4.68 is identical to results[7] given by Delbrück using another technique. □

• **Example 7.4.2**

In the chemical reaction $rA \underset{\mu}{\overset{\lambda}{\rightleftarrows}} B$, r molecules of A combine to form one molecule of B. If $X(t) = n$ is the number of B molecules, then the probability $p_n(t) = P\{X(t) = n\}$ of having n molecules of B is given by

$$\frac{dp_n}{dt} = -\left[n\mu + (N-n)\lambda\right] p_n + (N-n+1)\lambda p_{n-1} + (n+1)\mu p_{n+1}, \tag{7.4.69}$$

where $0 \le n \le N$, rN is the total number of molecules of A, λ is the rate at which r molecules of A combine to produce B, and μ is the rate at which B decomposes into A.

Multiplying Equation 7.4.69 by z^n and summing from $n = -1$ to $N+1$,

$$\begin{aligned}\sum_{n=-1}^{N+1} z^n \frac{dp_n}{dt} &= -N\lambda \sum_{n=-1}^{N+1} z^n p_n + (\lambda - \mu) \sum_{n=-1}^{N+1} n z^n p_n + N\lambda \sum_{n=-1}^{N+1} z^n p_{n-1} \\ &\quad - \lambda \sum_{n=-1}^{N+1} (n-1) z^n p_{n-1} + \mu \sum_{n=-1}^{N+1} (n+1) z^n p_{n+1}.\end{aligned} \tag{7.4.70}$$

Defining

$$F(z,t) = \sum_{n=-1}^{N+1} p_n(t) z^n, \qquad |z| < 1, \tag{7.4.71}$$

with $p_{-1} = p_{N+1} = 0$, we have that

$$\frac{\partial F}{\partial t} = \sum_{n=-1}^{N+1} z^n \frac{dp_n}{dt}, \tag{7.4.72}$$

$$\frac{\partial F}{\partial z} = \sum_{n=-1}^{N+1} n z^{n-1} p_n = \sum_{i=-2}^{N} (i+1) z^i p_{i+1} = \sum_{i=-1}^{N+1} (i+1) z^i p_{i+1}, \tag{7.4.73}$$

---

[7] Delbrück, M., 1940: Statistical fluctuations in autocatalytic reactions. *J. Chem. Phys.*, **8**, 120–124. See his Equation 7.

$$z\frac{\partial F}{\partial z} = \sum_{n=-1}^{N+1} nz^n p_n, \tag{7.4.74}$$

$$z^2\frac{\partial F}{\partial z} = \sum_{n=-1}^{N+1} nz^{n+1} p_n = \sum_{i=0}^{N+2}(i-1)z^i p_{i-1} = \sum_{i=-1}^{N+1}(i-1)z^i p_{i-1}, \tag{7.4.75}$$

and

$$F = \sum_{n=-1}^{N+1} z^{n+1}p_n = \sum_{i=0}^{N+2} z^i p_{i-1} = \sum_{i=-1}^{N+1} z^i p_{i-1}. \tag{7.4.76}$$

Therefore, the differential-difference equation, Equation 7.4.70, can be replaced by

$$\frac{\partial F}{\partial t} = N\lambda(z-1)F + \left[\mu - (\mu-\lambda)z - \lambda z^2\right]\frac{\partial F}{\partial z}. \tag{7.4.77}$$

Using the same technique as above, this partial differential equation can be written as

$$\frac{dt}{-1} = \frac{dz}{(1-z)(\mu+\lambda z)} = \frac{dF}{-N\lambda(z-1)}. \tag{7.4.78}$$

Equation 7.4.78 yields the independent solutions

$$\frac{1-z}{\mu+\lambda z}e^{-(\mu+\lambda)t} = \xi(z,t) = \text{constant}, \tag{7.4.79}$$

and

$$(\mu+\lambda)^{-N}F(z,t) = \eta(z,t) = \text{another constant}, \tag{7.4.80}$$

where $f(\cdot)$ is an arbitrary, differentiable function. If there are m units of B at $t=0$, $0 \le m \le N$, the initial condition is $F(z,0) = z^m$. Then,

$$f\left(\frac{1-z}{\mu+\lambda z}\right) = \frac{z^m}{(\mu+\lambda z)^N}, \tag{7.4.81}$$

or

$$f(x) = \frac{(1-\mu x)^m}{(\mu+\lambda)^N}(1+\lambda x)^{N-m}. \tag{7.4.82}$$

After some algebra, we finally find that

$$\begin{aligned} F(z,t) = \frac{1}{(\mu+\lambda)^N}&\left\{\mu\left[1-e^{-(\mu+\lambda)t}\right] + z\left[\lambda + \mu e^{-(\mu+\lambda)t}\right]\right\}^m \\ &\times\left\{\mu + \lambda e^{-(\mu+\lambda)t} + \lambda z\left[1-e^{-(\mu+\lambda)t}\right]\right\}^{N-m}. \end{aligned} \tag{7.4.83}$$

Computing the mean and variance, we obtain

$$E(X) = \frac{m}{\mu+\lambda}\left[\lambda + \mu e^{-(\mu+\lambda)t}\right] + \frac{(N-m)\lambda}{\mu+\lambda}\left[1-e^{-(\mu+\lambda)t}\right], \tag{7.4.84}$$

and

$$\begin{aligned} \operatorname{Var}(X) = {}& \frac{m\mu}{(\mu+\lambda)^2}\left[\lambda + \mu e^{-(\mu+\lambda)t}\right]\left[1-e^{-(\mu+\lambda)t}\right] \\ &+ \frac{(N-m)\lambda}{(\mu+\lambda)^2}\left[\mu + \lambda e^{-(\mu+\lambda)t}\right]\left[1-e^{-(\mu+\lambda)t}\right]. \end{aligned} \tag{7.4.85}$$

Problems

1. During their study of growing cancerous cells (with growth rate α), Bartoszyński et al.[8] developed a probabilistic model of a tumor that has not yet metastasized. In their mathematical derivation a predictive model gives the probability $p_n(t)$ that certain nth type of cells (out of N) will develop. This probability can change in two ways: (1) Each of the existing cells has the probability $\lambda n\Delta t + o(\Delta t)$ of mutating to another type between t and $t+\Delta t$. (2) The probability that cells in state n at time t will shed a metastasis between t and $t + \Delta t$ is $\mu nce^{t/\alpha}\Delta t + o(\Delta t)$, where μ is a constant and c is the size of a single cell. Setting $\rho = \lambda c/N$ and $\nu = \mu c$, the governing equations for $p_n(t)$ are

$$\frac{dp_n}{dt} = -(\rho+\nu)ne^{t/\alpha}p_n + \rho(n+1)e^{t/\alpha}p_{n+1}, \qquad n = 0, 1, 2, \ldots, N-1,$$

and

$$\frac{dp_N}{dt} = -(\rho+\nu)Ne^{t/\alpha}p_N,$$

with the initial conditions $p_N(0) = 1$ and $p_n(0) = 0$ if $n \neq N$.

Step 1: Introducing the generating function

$$\phi(z,t) = \sum_{n=0}^{N} z^n p_n(t), \qquad 0 \leq z \leq 1,$$

show that our system of linear differential-difference equations can be written as the first-order partial differential equation

$$\frac{\partial \phi}{\partial t} = [\rho - (\rho+\nu)z]e^{t/\alpha}\frac{\partial \phi}{\partial z}$$

with $\phi(z,0) = z^N$.

Step 2: Solve the partial differential equation in Step 1 and show that

$$\phi(z,t) = \left(\frac{\rho}{\rho+\nu}\right)^N \left\{1 - \left(1 - \frac{\rho+\nu}{\rho}z\right)\exp\left[-\alpha(\rho+\nu)\left(e^{t/a}-1\right)\right]\right\}^N.$$

Project: Stochastic Simulation of Chemical Reactions

Most stochastic descriptions of chemical reactions cannot be attacked analytically and numerical simulation is necessary. The purpose of this project is to familiarize you with some methods used in the stochastic simulation of chemical reactions. In particular, we will use the Lokta reactions given by the reaction equations:

$$A + X \xrightarrow{k_1} 2X, \tag{1}$$

$$X + Y \xrightarrow{k_2} 2Y, \tag{2}$$

$$Y \xrightarrow{k_3} Z. \tag{3}$$

---

[8] Bartoszyński, R., B. F. Jones, and J. P. Klein, 1985: Some stochastic models of cancer metastases. *Commun. Statist.-Stochastic Models*, **1**, 317–339.

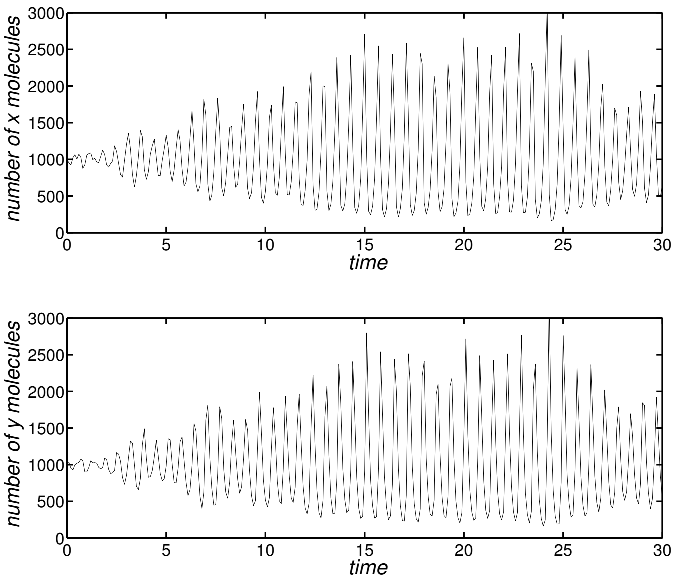


Figure 7.4.1: The temporal variation of the molecules in a Lokta reaction when $\Delta t = 10^{-5}$, $k_1 a = 10$, $k_2 = 0.01$, $k_3 = 10$, and $x(0) = y(0) = 1000$.

Surprisingly, simple numerical integration of the master equation is not fruitful. This occurs because of the number and nature of the independent variables; there is only one master equation but N reactants and time for independent variables.

An alternative to integrating the master equation is a direct stochastic simulation. In this approach, the (transition) probability for each reaction is computed: $p_1 = k_1 ax\Delta t$, $p_2 = k_2 xy\Delta t$, and $p_3 = k_3 y\Delta t$, where Δt is the time between each consecutive state and a is the constant number of molecules of A. The obvious question is: Which of these probabilities should we use?

Our first attempt follows Nakanishi:[9] Assume that Δt is sufficiently small so that $p_1 + p_2 + p_3 < 1$. Using a normalized uniform distribution, such as MATLAB's `rand`, compute a random variable `r` for each time step. Then march forward in time. At each time step, there are four possibilities. If $0 < \mathtt{r} \le p_1$, then the first reaction occurs and $x(t + \Delta t) = x(t) + 1$, $y(t + \Delta t) = y(t)$. If $p_1 < \mathtt{r} \le p_1 + p_2$, then the second reaction occurs and $x(t + \Delta t) = x(t) - 1$, $y(t + \Delta t) = y(t) + 1$. If $p_1 + p_2 < \mathtt{r} \le p_1 + p_2 + p_3$, then the third reaction occurs and $x(t + \Delta t) = x(t)$, $y(t + \Delta t) = y(t) - 1$. Finally, if $p_1 + p_2 + p_3 < \mathtt{r} \le 1$, then no reaction occurs and $x(t + \Delta t) = x(t)$, $y(t + \Delta t) = y(t)$.

For the first portion of this project, create MATLAB code to simulate our chemical reaction using this simulation technique. Explore how your results behave as you vary $x(0)$, $y(0)$ and especially Δt. See Figure 7.4.1.

One of the difficulties in using Nakanishi's method is the introduction of Δt. What value should we choose to ensure that $p_1 + p_2 + p_3 < 1$? Several years later, Gillespie[10]

---

[9] This is the technique used by Nakanishi, T., 1972: Stochastic analysis of an oscillating chemical reaction. *J. Phys. Soc. Japan*, **32**, 1313–1322.

[10] Gillespie, D. T., 1976: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, **22**, 403–434; Gillespie, D. T., 1977: Exact stochastic

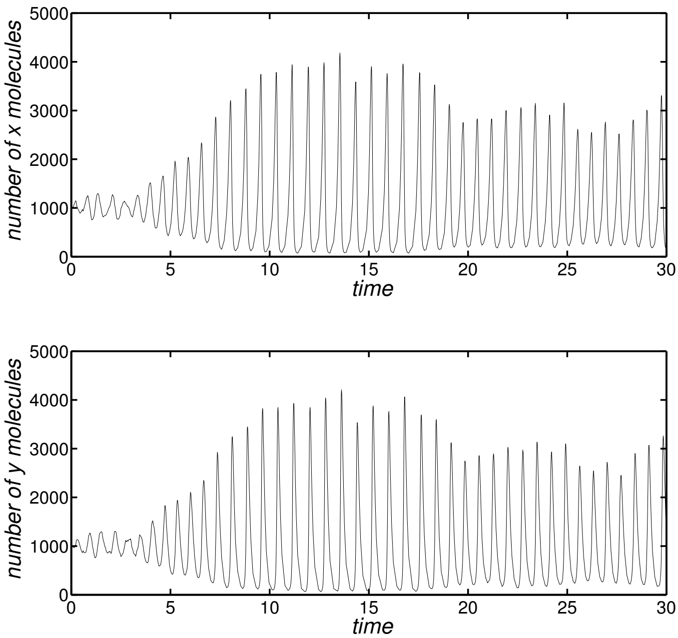


Figure 7.4.2: Same as Figure 7.4.1 except that Gillespie's method has been used.

developed a similar algorithm. He introduced three parameters, $a_1 = k_1ax$, $a_2 = k_2xy$, and $a_3 = k_3y$, along with $a_0 = a_1 + a_2 + a_3$. These parameters a_1, a_2, and a_3 are similar to the probabilities p_1, p_2, and p_3. Similarly, he introduced a random number r_2 that is chosen from a normalized uniform distribution. Then, if $0 < r_2a_0 \leq a_1$, the first reaction occurs and $x(t+\Delta t) = x(t)+1$, $y(t+\Delta t) = y(t)$. If $a_1 < r_2a_0 \leq a_1 + a_2$, then the second reaction occurs and $x(t+\Delta t) = x(t)-1$, $y(t+\Delta t) = y(t)+1$. If $a_1 + a_2 < r_2a_0 \leq a_0$, then the third reaction occurs and $x(t+\Delta t) = x(t)$, $y(t+\Delta t) = y(t) - 1$. Because of his selection criteria for the reaction that occurs during a time step, one of the three reactions must take place. See Figure 7.4.2.

The most radical difference between the Nakanishi and Gillespie schemes involves the time step. It is no longer constant but varies with time and equals $\Delta t = \ln(1/r_1)/a_0$, where r_1 is a random variable selected from a normalized uniform distribution. The theoretical justification for this choice is given in Section III of his paper.

For the second portion of this project, create MATLAB code to simulate our chemical reaction using Gillespie's technique. You might like to plot $x(t)$ vs $y(t)$ and observe the patterns that you obtain.

Finally, for a specific time, compute the probability density function that gives the probability that x and y molecules exist. See Figure 7.4.3.

7.5 POISSON PROCESSES

The Poisson random process is a counting process that counts the number of occurrences of some particular event as time increases. In other words, for each value of t, there

simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361

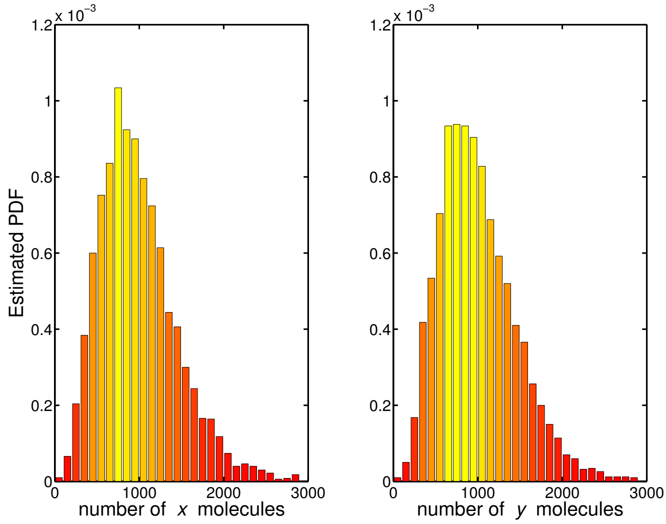


Figure 7.4.3: The estimated probability density function for the chemical reactions given by Equations (1) through (3) (for X on the left, Y on the right) at time $t = 10$. Five thousand realizations were used in these computations.

is a number $N(t)$, which gives the number of events that occurred during the interval $[0, t]$. For this reason $N(t)$ is a discrete random variable with the set of possible values $\{0, 1, 2, \ldots\}$. Figure 7.5.1 illustrates a sample function. We can express this process mathematically by

$$N(t) = \sum_{n=0}^{\infty} H(t - T[n]), \tag{7.5.1}$$

where $T[n]$ is the time to the nth arrival, a random sequence of times. The question now becomes how to determine the values of $T[n]$. The answer involves three rather physical assumptions. They are:

1. $N(0) = 0$.
2. $N(t)$ has independent and stationary increments. By stationary we mean that for any two equal time intervals Δt_1 and Δt_2, the probability of n events in Δt_1 equals the probability of n events in Δt_2. By independent we mean that for any time interval $(t, t + \Delta t)$ the probability of n events in $(t, t + \Delta t)$ is independent of how many events have occurred earlier or how they have occurred.
3. $$P[N(t+\Delta t) - N(t) = k\,] = \begin{cases} 1 - \lambda\Delta t, & k = 0, \\ \lambda\Delta t, & k = 1, \\ 0, & k > 1, \end{cases} \tag{7.5.2}$$

   for all t. Here λ equals the expected number of events in an interval of unit length of time. Because $E[N(t)] = \lambda$, it is the average number of events that occur in one unit of time and in practice it can be measured experimentally.

We begin our analysis of Poisson processes by finding $P[N(t) = 0\,]$ for any $t > 0$. If there are no arrivals in $[0, t]$, then there must be no arrivals in $[0, t - \Delta t]$ and also no arrivals

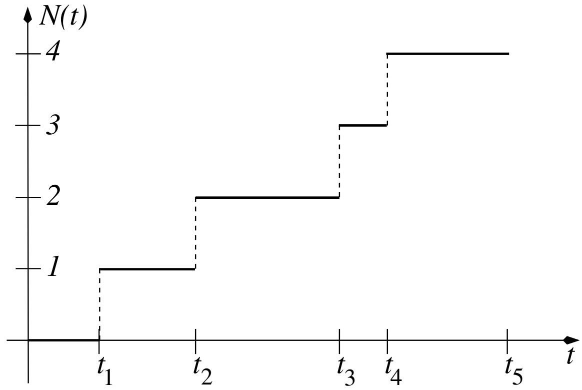


Figure 7.5.1: Schematic of a Poisson process.

in $(t - \Delta t, t]$. Therefore,

$$P[N(t) = 0\,] = P[N(t - \Delta t) = 0, N(t) - N(t - \Delta t) = 0\,]. \qquad \textbf{(7.5.3)}$$

Because $N(t)$ is independent,

$$P[N(t) = 0\,] = P[N(t - \Delta t) = 0\,]P[N(t) - N(t - \Delta t) = 0\,]. \qquad \textbf{(7.5.4)}$$

Furthermore, since $N(t)$ is stationary,

$$P[N(t) = 0\,] = P[N(t - \Delta t) = 0\,]P[N(t + \Delta t) - N(t) = 0\,]. \qquad \textbf{(7.5.5)}$$

Finally, from Equation 7.5.2,

$$P[N(t) = 0\,] = P[N(t - \Delta t) = 0\,](1 - \lambda \Delta t). \qquad \textbf{(7.5.6)}$$

Let us denote $P[N(t) = 0\,]$ by $P_0(t)$. Then,

$$P_0(t) = P_0(t - \Delta t)(1 - \lambda \Delta t), \qquad \textbf{(7.5.7)}$$

or

$$\frac{P_0(t) - P_0(t - \Delta t)}{\Delta t} = -\lambda P_0(t - \Delta t). \qquad \textbf{(7.5.8)}$$

Taking the limit as $\Delta t \to 0$, we obtain the (linear) differential equation

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t). \qquad \textbf{(7.5.9)}$$

The solution of Equation 7.5.9 is

$$P_0(t) = Ce^{-\lambda t}, \qquad \textbf{(7.5.10)}$$

where C is an arbitrary constant. To evaluate C, we have the initial condition $P_0(0) = P[N(0) = 0\,] = 1$ from Axion 1. Therefore,

$$P[N(t) = 0\,] = P_0(t) = e^{-\lambda t}. \qquad \textbf{(7.5.11)}$$

Next, let us find $P_1(t) = P[N(t) = 1\,]$. We either have no arrivals in $[0, t - \Delta t]$ and one arrival in $(t - \Delta t, t]$ or one arrival in $[0, t - \Delta t]$ and no arrivals in $(t - \Delta t, t]$. These are the only two possibilities because there can be at most one arrival in a time interval Δt. The two events are mutually exclusive. Therefore,

$$
\begin{aligned}
P[N(t) = 1\,] &= P[N(t - \Delta t) = 0, N(t) - N(t - \Delta t) = 1\,] \\
&\quad + P[N(t - \Delta t) = 0, N(t) - N(t - \Delta t) = 0\,] && \textbf{(7.5.12)} \\
&= P[N(t - \Delta t) = 0\,]P[N(t) - N(t - \Delta t) = 1\,] \\
&\quad + P[N(t - \Delta t) = 1\,]P[N(t) - N(t - \Delta t) = 0\,] && \textbf{(7.5.13)} \\
&= P[N(t - \Delta t) = 0\,]P[N(t + \Delta t) - N(t) = 1\,] \\
&\quad + P[N(t - \Delta t) = 1\,]P[N(t + \Delta t) - N(t) = 0\,]. && \textbf{(7.5.14)}
\end{aligned}
$$

Equation 7.5.13 follows from independence while Equation 7.5.14 follows from stationarity. Introducing $P_1(t)$ in Equation 7.5.14 and using Axion 3,

$$P_1(t) = P_0(t - \Delta t)\lambda \Delta t + P_1(t - \Delta t)(1 - \lambda \Delta t), \qquad \textbf{(7.5.15)}$$

or

$$\frac{P_1(t) - P_1(t - \Delta t)}{\Delta t} = -\lambda P_1(t - \Delta t) + \lambda P_0(t - \Delta t). \qquad \textbf{(7.5.16)}$$

Taking the limit as $\Delta t \to 0$, we obtain

$$\frac{dP_1(t)}{dt} + \lambda P_1(t) = \lambda P_0(t). \qquad \textbf{(7.5.17)}$$

In a similar manner, we can prove that

$$\frac{dP_k(t)}{dt} + \lambda P_k(t) = \lambda P_{k-1}(t), \qquad \textbf{(7.5.18)}$$

where $k = 1, 2, 3, \ldots$ and $P_k(t) = P[N(t) = k]$.

This set of simultaneous linear equations can be solved recursively. Its solution is

$$P_k(t) = \exp(-\lambda t)\frac{(\lambda t)^k}{k!}, \qquad k = 0, 1, 2, \ldots, \qquad \textbf{(7.5.19)}$$

which is the Poisson probability mass function. Here λ is the average number of arrivals per second.

In the realization of a Poisson process, one of the important quantities is the *arrival time*, t_n, shown in Figure 7.5.1. Of course, the arrival time is also a random process and will change with each new realization. A related quantity $Z_i = t_i - t_{i-1}$, the time intervals between two successive occurrences (interoccurrence times) of Poisson events. We will now show that the random variables Z_1, Z_2, etc., are independent and identically distributed with

$$P(Z_n \le x) = 1 - e^{-\lambda x}, \qquad x \ge 0, \quad n = 1, 2, 3, \ldots \qquad \textbf{(7.5.20)}$$

We begin by noting that

$$P(Z_1 > t) = P[N(t) = 0] = e^{-\lambda t} \qquad \textbf{(7.5.21)}$$

from Equation 7.5.19. Therefore, Z_1 has an exponential distribution.

Let us denote its probability density by $p_{Z_1}(z_1)$. From the joint conditional density function,

$$P(Z_2 > t) = \int_0^{\xi_1} P(Z_2 > t|Z_1 = z_1)p_{Z_1}(z_1)\,dz_1, \qquad \textbf{(7.5.22)}$$

where $0 < \xi_1 < t$. If $Z_1 = z_1$, then $Z_2 > t$ if and only if $N(t + z_1) - N(z_1) = 0$. Therefore, using the independence and stationary properties,

$$P\{Z_2 > t|Z_1 = P[N(t + z_1) - N(z_1) = 0]\} = P[N(t) = 0] = e^{-\lambda t}. \qquad \textbf{(7.5.23)}$$

Consequently,

$$P(Z_2 > t) = e^{-\lambda t}, \qquad \textbf{(7.5.24)}$$

showing that Z_2 is also exponential. Also, Z_2 is independent of Z_1. Now, let us introduce $p_{Z_2}(z_2)$ as the probability density of $Z_1 + Z_2$. By similar arguments we can show that Z_3 is also exponential. The final result follows by induction.

• Example 7.5.1: Random telegraph signal

We can use the fact that interoccurrence times are independent and identically distributed to realize the Poisson process. An important application of this is in the generation of the random telegraph signal: $X(t) = (-1)^{N(t)}$. However, no one uses this definition to compute the signal; they use the arrival times to change the signal from $+1$ to -1 or vice versa.

We begin by noting that $T_i = T_{i-1} + Z_i$, with $i = 1, 2, \ldots$, $T_0 = 0$, and T_i is the ith arrival time. Each Z_i has the same exponent probability density function. From Equation 6.4.17,

$$Z_i = \frac{1}{\lambda} \ln\left(\frac{1}{1 - U_i}\right), \qquad \textbf{(7.5.25)}$$

where the U_i's are from a uniform distribution. The realization of a random telegraphic signal is given by the MATLAB code:

```
clear
N = 100; % number of switches in realization
lambda = 0.15; % switching rate
X = [ ];
% generate N uniformly distributed random variables
S = rand(1,N);
% transform S into an exponential random variable
T = - log(S)/lambda;
V = cumsum(T); % compute switching times
t = [0.01:0.01:100]; % create time array
icount = 1; amplitude = -1; % initialize X(t)
for k = 1:10000
  if ( t(k) >= V(icount) ) % at each switching point
    icount = icount + 1;
    amplitude = - amplitude; % switch sign
  end
  X(k) = amplitude; % generate X(t)
end
```

```
plot(t,X) % plot results
xlabel('\it t','FontSize',25);
ylabel('\it X(t)/a','FontSize',25);
axis([0 max(t) -1.1 1.1])
```

This was the MATLAB code that was used to generate Figure 7.5.2. □

• Example 7.0.2

It takes a workman an average of one hour to put a widget together. Assuming that the task can be modeled as a Poisson process, what is the probability that a workman can build 12 widgets during an eight-hour shift?

The probability that n widgets can be constructed by time t is

$$P[N(t) = n] = e^{-\lambda t}\frac{(\lambda t)^n}{n!}. \tag{7.5.26}$$

Therefore, the probability that 12 or more widgets can be constructed in eight hours is

$$P[N(t) \geq 12] = e^{-8}\sum_{n=12}^{\infty}\frac{8^n}{n!} = 0.1119, \tag{7.5.27}$$

since $\lambda = 1$.

We could have also obtained our results by creating 12 exponentially distributed time periods and summed them together using MATLAB:

```
t_uniform = rand(1,12);
T = - log(1-t_uniform);
total_time = sum(T);
```

Then, by executing this code a large number `N` of times and counting the number `icount` of times that `total_time <= 8`, the probability equals `icount / N`.

Problems

1. Use the generating function

$$F(z,t) = \sum_{n=0}^{\infty} p_n(t) z^n, \qquad |z| < 1,$$

with $F(z,0) = 1$ to solve Equation 7.5.18 by showing that $F(z,t) = e^{\lambda t(z-1)}$. Then, by expanding $F(z,t)$, recover Equation 7.5.19.

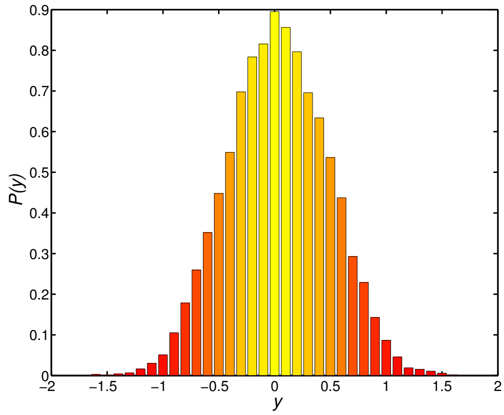


Figure 7.5.1: The probability density $P(y)$ of the output from an ideal integrator with finite memory when the input is a random telegraphic signal when $\Delta t = 0.01$, $\lambda = 2$, and $\tau_1 = 10$.

Project: Output from a Filter When the Input Is a Random Telegraphic Signal[11]

In the study of many systems, such as linear filters, the output $y(\cdot)$ can be written as

$$y(t) = \int_{-\infty}^{t} W(t-\tau)x(\tau)\,d\tau,$$

where $W(\cdot)$ is the weight function and $x(\cdot)$ is the input. The purpose of this project is to explore the probability density $P(y)$ of the output when $x(t)$ is the random telegraphic signal, a Poisson random process. You will filter this input two ways: (1) ideal integrator with finite memory: $W(t) = H(t) - H(t-\tau_1)$, $\tau_1 > 0$, and (2) simple $RC = 1$ low-pass filter $W(t) = e^{-t}H(t)$.

Step 1: Use MATLAB to code $x(t)$ where the expected time between the zeros is λ.

Step 2: Develop MATLAB code to compute $y(t)$ for each of the weight functions $W(t)$.

Step 3: Compute $P(y)$ for both filters. How do your results vary as λ varies?

Further Readings

Beckmann, P., 1967: *Probability in Communication Engineering.* Harcourt, Brace & World, 511 pp. A presentation of probability as it applies to problems in communication engineering.

Gillespie, D. T., 1991: *Markov Processes: An Introduction for Physical Scientists.* Academic Press, 592 pp. For the scientist who needs an introduction to the details of the subject.

Hsu, H., 1997: *Probability, Random Variables, & Random Processes.* McGraw-Hill, 306 pp. Summary of results plus many worked problems.

[11] Suggested by a paper by McFadden, J. A., 1959: The probability density of the output of a filter when the input is a random telegraphic signal: Differential-equation approach. *IRE Trans. Circuit Theory*, **6**, 228–233.

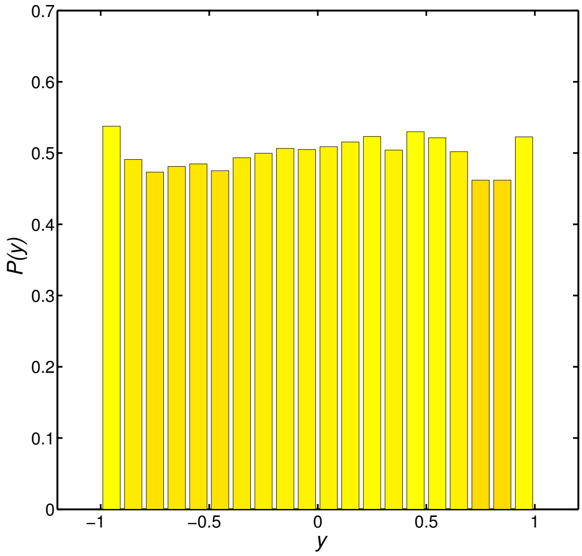
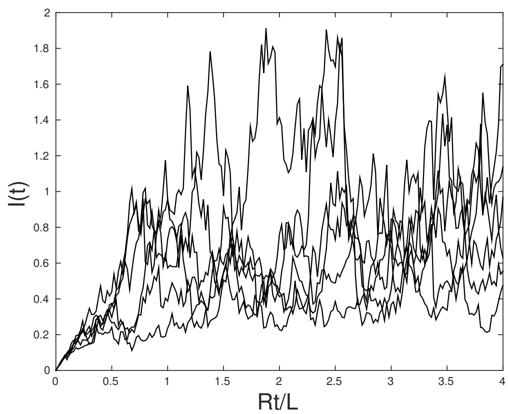


Figure 7.5.2: The probability density $P(y)$ of the output from a simple $RC = 1$ filter, when the input is a random telegraphic signal, when $\lambda = 1$, and $\Delta t = 0.05$.

Kay, S. M., 2006: *Intuitive Probability and Random Processes Using MATLAB*. Springer, 833 pp. A well-paced book designed for the electrical engineering crowd.

Ross, S. M., 2007: *Introduction to Probability Models.* Academic Press, 782 pp. An introductory undergraduate book in applied probability and stochastic processes.

Tuckwell, H. C., 1995: *Elementary Applications of Probability Theory.* Chapman & Hall, 292 pp. This book presents applications using probability theory, primarily from biology.



Chapter 8
Itô's Stochastic Calculus

In elementary differential equation classes, students study the solution to first-order ordinary differential equations

$$\frac{dx}{dt} = a(t,x), \qquad x(0) = x_0. \tag{8.0.1}$$

There we showed that Equation 8.0.1 has the solution

$$x(t) = x(0) + \int_0^t a[\eta, x(\eta)]\, d\eta. \tag{8.0.2}$$

Consider now the analogous *stochastic differential equation*:

$$dX(t) = a[t, X(t)]\, dt, \qquad X(0) = X_0. \tag{8.0.3}$$

Although Equations 8.0.1 and 8.0.3 formally appear the same, an immediate question is what is meant by $dX(t)$. In elementary calculus, the concept of the infinitesimal involves limits, continuity, and so forth. As we shall see in Section 8.2, Brownian motion, a very common stochastic process, is nowhere differentiable. Here we can merely say that $dX(t) = X(t+dt) - X(t)$.

Consider now a modification of Equation 8.0.3 where we introduce a random forcing:

$$dX(t) = a[t, X(t)]\, dt + b[t, X(t)]\, dB(t), \qquad X(0) = X_0. \tag{8.0.4}$$

Here $dB(t) = B(t+dt) - B(t)$, $B(t)$ denotes Brownian motion and $a[t, X(t)]$ and $b[t, X(t)]$ are deterministic functions. Consequently, changes to $X(t)$ result from (1) the effects of the initial conditions and (2) noise generated by Brownian motion (the driving force). Stochastic processes governed by Equation 8.0.4 are referred to as *Itô processes*.

Following the methods used to derive 8.0.2, we can formally write the solution to Equation 8.0.4 as

$$X(t) = X_0 + \int_0^t a[\eta, X(\eta)]\, d\eta + \int_0^t b[\eta, X(\eta)]\, dB(\eta). \tag{8.0.5}$$

The first integral in Equation 8.0.5 is the conventional Riemann integral from elementary calculus and is well understood. The second integral, however, is new and must be treated with care. It is called *Itô's stochastic integral* and treated in Section 8.3.

In summary, a simple analog to first-order ordinary differential equations for a single random variable $X(t)$ raises several important questions. What is meant by the infinitesimal and the integral in stochastic calculus? In this chapter we will focus on Itô processes and the associated calculus. Although Itô's calculus is an important discipline, it is not the only form of stochastic calculus. The interested student is referred elsewhere for further study.

Problems

1. The Poisson random process $N(t)$ is defined by

$$N(t) = \sum_{n=1}^{\infty} H(t - t_n),$$

where t_n is a sequence of independent and identically distributed inter-arrival times t_n. A graphical representation of $N(t)$ would consist of ever-increasing steps with the edges located at $t = t_n$. Use the definition of $dN(t) = N(t + dt) - N(t)$ to show that

$$dN(t) = \begin{cases} 1, & \text{for } t = t_n, \\ 0, & \text{otherwise.} \end{cases}$$

2. The telegraph signal is defined by $X(t) = (-1)^{N(t)}$, where $N(t)$ is given by the Poisson random distribution in Problem 1. Show[1] that

$$dX(t) = X(t + dt) - X(t) = (-1)^{N(t)} \left[(-1)^{dN(t)} - 1\right] = -2X(t)\, dN(t).$$

Hint: Consider $dN(t)$ at various times.

3. If $X(t)$ and $Y(t)$ denote two stochastic processes, use the definition of the derivative to show that (a) $d[cX(t)] = c\, dX(t)$, where c is a constant, (b) $d[X(t) \pm Y(t)] = dX(t) \pm dY(t)$, and (c) $d[X(t)Y(t)] = X(t)\, dY(t) + Y(t)\, dX(t) + dX(t)\, dY(t)$.

8.1 RANDOM DIFFERENTIAL EQUATIONS

A large portion of this book has been devoted to solving differential equations. Here we examine the response of differential equations to random forcing where the differential

---

[1] Taken from Janaswamy, R., 2013: On random time and on the relation between wave and telegraph equation. *IEEE Trans. Antennas Propag.*, **61**, 2735–2744.

equation describes a nonrandom process. This is an important question in the sciences and engineering because noise, a random phenomenon, is ubiquitous in nature.

Because the solution to random differential equations can be found by conventional techniques, we can use them to study the effect of randomness on the robustness of a solution to a differential equation subject to small changes of the initial condition. Although this may be of considerable engineering interest, it is really too simple to develop a deep understanding of stochastic differential equations.

• Example 8.1.1: LR circuit

One of the simplest differential equations involves the mathematical model for an LR electrical circuit:

$$L\frac{dI}{dt} + RI = E(t), \tag{8.1.1}$$

where $I(t)$ denotes the current within an electrical circuit with inductance L and resistance R, and $E(t)$ is the mean electromotive force. If we solve this first-order ordinary differential equation using an integrating factor, its solution is

$$I(t) = I(0)\exp\left(-\frac{Rt}{L}\right) + \frac{1}{L}\exp\left(-\frac{Rt}{L}\right)\int_0^t F(\tau)\exp\left(\frac{R\tau}{L}\right)\,d\tau. \tag{8.1.2}$$

Clearly, if the electromotive forcing is random, so is the current.

In the previous chapter we showed that the mean and variance were useful parameters in characterizing a random variable. This will also be true here. If we find the mean of the solution,

$$E[I(t)] = I(0)\exp\left(-\frac{Rt}{L}\right) \tag{8.1.3}$$

provided $E[F(t)] = 0$. Thus, the mean of the current is the same as that for an ideal LR circuit.

Turning to the variance,

$$\begin{aligned}
\sigma_X^2(t) &= E[I^2(t)] - \{E[I(t)]\}^2 &\quad (8.1.4)\\
&= E\left[I^2(0)\exp\left(-\frac{2Rt}{L}\right)\right]\\
&\quad + \frac{2I(0)}{L}\exp\left(-\frac{2Rt}{L}\right)\int_0^t E[F(\tau)]\exp\left(\frac{R\tau}{L}\right)\,d\tau\\
&\quad + \frac{1}{L^2}\exp\left(-\frac{2Rt}{L}\right)\int_0^t\int_0^t E[F(\tau)F(\tau')]\exp\left[\frac{R(\tau+\tau')}{L}\right]\,d\tau'\,d\tau &\quad (8.1.5)\\
&\quad - I^2(0)\exp\left(-\frac{2Rt}{L}\right)\\
&= \frac{1}{L^2}\exp\left(-\frac{2Rt}{L}\right)\int_0^t\int_0^t E[F(\tau)F(\tau')]\exp\left[\frac{R(\tau+\xi)}{L}\right]\,d\tau'\,d\tau. &\quad (8.1.6)
\end{aligned}$$

To proceed further we need the autocorrelation $E[F(\tau)F(\tau')]$. In papers by Ornstein et al.[2] and Jones and McCombie,[3] they adopted a random process with the autocorrelation

[2] Ornstein, L. S., H. C. Burger, J. Taylor, and W. Clarkson, 1927: The Brownian movement of a galvanometer and the influence of the temperature of the outer circuit. *Proc. Roy. Soc. London, Ser. A*, **115**, 391–406.

[3] Jones, R. V., and C. W. McCombie, 1952: Brownian fluctuations in galvanometer and galvanometer amplifiers. *Phil. Trans. Roy. Soc. London, Ser. A*, **244**, 205–230.

function

$$E[F(\tau)F(\tau')] = 2D\delta(\tau - \tau'). \tag{8.1.7}$$

The advantage of this process is that it is mathematically the simplest because it possesses a white power spectrum. Unfortunately this random process can never be physically realized because it would possess infinite mean square power. All physically realizable processes involve a power spectrum that tends to zero at sufficiently high frequencies. If $\Phi(\omega)$ denotes the power spectrum, this condition can be expressed as

$$\int_0^\infty \Phi(\omega)\, d\omega < \infty. \tag{8.1.8}$$

In view of these considerations, let us adopt the autocorrelation

$$R_X(\tau - \tau') = \int_0^\infty \Phi(\omega) \cos[\omega(\tau - \tau')]\, d\omega, \tag{8.1.9}$$

where $\Phi(\omega)$ is the power spectrum of $F(\tau)$. Therefore, the variance becomes

$$\sigma_X^2(t) = \frac{1}{L^2} \int_0^t \int_0^t \int_0^\infty \Phi(\omega) \exp\left[-\frac{R(t-\tau)}{L}\right] \exp\left[-\frac{R(t-\tau')}{L}\right] \cos[\omega(\tau - \tau')]\, d\omega\, d\tau\, d\tau'. \tag{8.1.10}$$

Reversing the ordering of integration,

$$\sigma_X^2(t) = \frac{1}{L^2} \int_0^\infty \Phi(\omega) \int_0^t \int_0^t \exp\left[-\frac{R(2t - \tau - \tau')}{L}\right] \cos[\omega(\tau - \tau')]\, d\tau\, d\tau'\, d\omega. \tag{8.1.11}$$

We can evaluate the integrals involving τ and τ' exactly. Equation 8.1.11 then becomes

$$\sigma_X^2(t) = \int_0^\infty \frac{\Phi(\omega)}{\omega^2 + R^2/L^2} \left[1 + e^{-2Rt/L} - 2e^{-Rt/L} \cos(\omega t)\right] d\omega. \tag{8.1.12}$$

Let us now consider some special cases. As $t \to 0$, $\sigma_X^2(t) \to 0$ and the variance is initially small. On the other hand, as $t \to \infty$,

$$\sigma_X^2(t) = \int_0^\infty \frac{\Phi(\omega)}{\omega^2 + R^2/L^2}\, d\omega. \tag{8.1.13}$$

Thus, the variance grows to a constant value, which we would have found by using Fourier transforms to solve the differential equation.

Consider now the special case $\Phi(\omega) = 2D/\pi$, a forcing by white noise. Ignoring the defects in this model, we can evaluate the integrals in Equation 8.1.13 exactly and find that

$$\sigma_X^2(t) = \frac{DL}{R} \left(1 - e^{-2Rt/L}\right). \tag{8.1.14}$$

These results are identical to those found by Uhlenbeck and Ornstein[4] in their study of a free particle in Brownian motion. □

---

[4] Uhlenbeck, G. E., and L. S. Ornstein, 1930: On the theory of the Brownian motion. *Phys. Review*, **36**, 823–841. See the top of their page 828.

• **Example 8.1.2: Damped harmonic motion**

Another classic differential equation that we can excite with a random process is the damped harmonic oscillator:

$$y'' + 2\xi\omega_0 y' + \omega_0^2 y = F(t), \tag{8.1.15}$$

where $0 \leq \xi < 1$, y denotes the displacement, t is time, $\omega_0^2 = k/m$, $2\xi\omega_0 = \beta/m$, m is the mass of the oscillator, k is the linear spring constant, and β denotes the constant of a viscous damper. The solution to this second-order ordinary differential equation is

$$y(t) = y(0)e^{-\xi\omega_0 t}\left[\cos(\omega_1 t) + \frac{\xi\omega_0}{\omega_1}\sin(\omega_1 t)\right] + \frac{y'(0)}{\omega_1}e^{-\xi\omega t}\sin(\omega_1 t) + \int_0^t h(t-\tau)F(\tau)\,d\tau, \tag{8.1.16}$$

where $\omega_1 = \omega_0\sqrt{1-\xi^2}$, and

$$h(t) = \frac{e^{-\xi\omega_0 t}}{\omega_1}\sin(\omega_1 t)H(t). \tag{8.1.17}$$

Again we begin by finding the mean of Equation 8.1.16. It is

$$E[y(t)] = y(0)e^{-\xi\omega_0 t}\left[\cos(\omega_1 t) + \frac{\xi\omega_0}{\omega_1}\sin(\omega_1 t)\right] + \frac{y'(0)}{\omega_1}e^{-\xi\omega t}\sin(\omega_1 t) + \int_0^t h(t-\tau)E[F(\tau)]\,d\tau. \tag{8.1.18}$$

If we again choose a random process where $E[F(t)] = 0$, the integral vanishes and the stochastic mean of the motion only depends on the initial conditions.

Turning to the variance,

$$\sigma_X^2(t) = E[y^2(t)] - \{E[y(t)]\}^2 = \int_0^t \int_0^t h(t-\tau)h(t-\tau')E[F(\tau)F(\tau')]\,d\tau\,d\tau'. \tag{8.1.19}$$

If we again adopt the autocorrelation function

$$R_X(\tau - \tau') = \int_0^\infty \Phi(\omega)\cos[\omega(\tau-\tau')]\,d\omega, \tag{8.1.20}$$

where $\Phi(\omega)$ is the power spectrum of $F(\tau)$, then

$$\sigma_X^2(t) = \int_0^\infty \frac{\Phi(\omega)}{\omega_1^2}\int_0^t \int_0^t e^{-\xi\omega_0(2t-\tau-\tau')}\sin[\omega_1(t-\tau)]\sin[\omega_1(t-\tau')]\cos[\omega(\tau-\tau')]\,d\tau\,d\tau'\,d\omega. \tag{8.1.21}$$

Carrying out the integrations in τ and τ', we finally obtain

$$\begin{aligned}\sigma_X^2(t) = \int_0^\infty \frac{\Phi(\omega)}{|\Omega(\omega)|^2}\Bigg(1 + e^{-2\xi\omega_0 t}\Bigg\{1 + \frac{2\xi\omega_0}{\omega_1}\sin(\omega_1 t)\cos(\omega_1 t)\\ - e^{\omega_0\xi t}\left[2\cos(\omega_1 t) + \frac{2\xi\omega_0}{\omega_1}\sin(\omega_1 t)\right]\cos(\omega t) - e^{\xi\omega_0 t}\frac{2\omega}{\omega_1}\sin(\omega_1 t)\sin(\omega t)\\ + \frac{\xi^2\omega_0^2 - \omega_1^2 + \omega^2}{\omega_1^2}\sin^2(\omega_1 t)\Bigg\}\Bigg)\,d\omega,\end{aligned} \tag{8.1.22}$$

where $|\Omega(\omega)|^2 = (\omega_0^2 - \omega^2)^2 + 4\omega^2\omega_0^2\xi^2$.

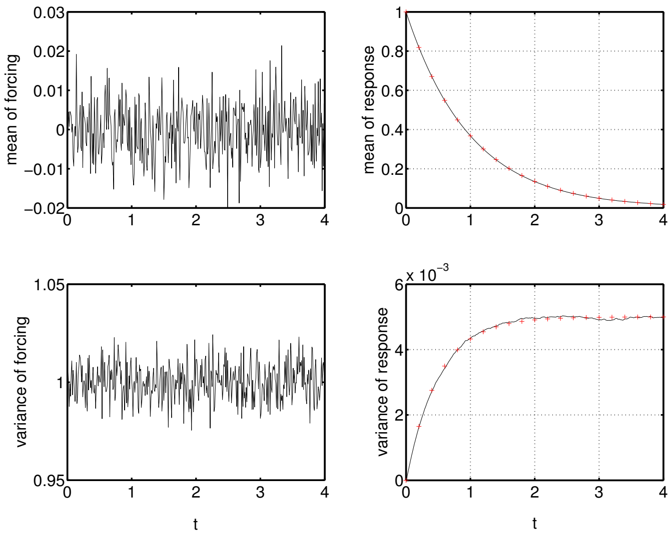


Figure 8.1.1: The mean and variance of the response for the differential equation $y' + y = f(t)$ when forced by Gaussian random noise. The parameters used are $y(0) = 1$ and $\Delta\tau = 0.01$.

As in the previous example, $\sigma_X^2(t) \to 0$ as $t \to 0$ and the variance is initially small. The steady-state variance now becomes

$$\sigma_X^2(t) = \int_0^\infty \frac{\Phi(\omega)}{|\Omega(\omega)|^2}\, d\omega. \tag{8.1.23}$$

Finally, for the special case $\Phi(\omega) = 2D/\pi$, the variance is

$$\sigma_X^2(t) = \frac{D}{2\xi\omega_0^2}\left\{1 - \frac{e^{-2\xi\omega_0 t}}{\omega_1^2}\left[\omega_1^2 + \omega_0\omega_1\xi\sin(2\omega_1 t) + 2\xi^2\omega_0^2\sin^2(\omega_1 t)\right]\right\}. \tag{8.1.24}$$

These results are identical to those found by Uhlenbeck and Ornstein[5] in their study of a harmonically bound particle in Brownian motion.

Project: Low-Pass Filter with Random Input

Consider the initial-value problem

$$y' + y = f(t), \qquad y(0) = y_0.$$

It has the solution

$$y(t) = y_0 e^{-t} + e^{-t}\int_0^t e^\tau f(\tau)\, d\tau.$$

This differential equation is identical to that governing an RC electrical circuit. This circuit has the property that it filters out high-frequency disturbances. Here we explore the case when $f(t)$ is a random process.

---

[5] Ibid. See their pages 834 and 835.

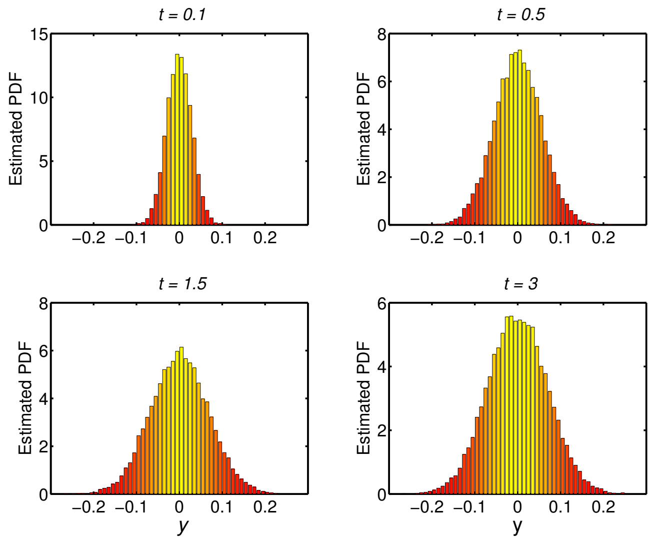


Figure 8.1.2: The probability density function for the response to the differential equation $y' + y = f(t)$ when $f(t)$ is a Gaussian distribution. Twenty thousand realizations were used to compute the density function. Here the parameters used are $y(0) = 0$ and $\Delta\tau = 0.01$.

Step 1: Using the MATLAB intrinsic function `randn`, generate a stationary white noise excitation of length `N`. Let `deltat` denote the time interval Δt between each new forcing so that `n = 1` corresponds to `t = 0` and `n = N` corresponds to the end of the record `t = T`.

Step 2: Using the Gaussian random forcing that you created in Step 1, develop a MATLAB code to compute $y(t)$ given $y(0)$ and $f(t)$.

Step 3: Once you have confidence in your code, modify it so that you can generate many realizations of $y(t)$. Save your solution as a function of t and realization. Use MATLAB's intrinsic functions `mean` and `var` to compute the mean and variance as a function of time. Figure 8.1.1 shows the results when 2000 realizations were used. For comparison the mean and variance of the forcing have also been included. Ideally this mean and variance should be zero and one, respectively. We have also included the exact mean and variance, given by Equation 8.1.3 and Equation 8.1.14, when we set $L = R = 1$ and $D = \Delta t/2$.

Step 4: Now generalize your MATLAB code so that you can compute the probability density function of finding $y(t)$ lying between y and $y + dy$ at various times. Figure 8.1.2 illustrates four times when $y(0) = 0$ and $\Delta\tau = 0.01$.

Step 5: Modify your MATLAB code so that you can compute the autocovariance. See Figure 8.1.3.

Project: First-Passage Problem with Random Vibrations[6]

In the design of devices, it is often important to know the chance that the device will exceed its design criteria. In this project you will examine how often the amplitude of a

[6] Based on a paper by Crandall, S. H., K. L. Chandiramani, and R. G. Cook, 1966: Some first-passage problems in random vibration. *J. Appl. Mech.*, **33**, 532–538.

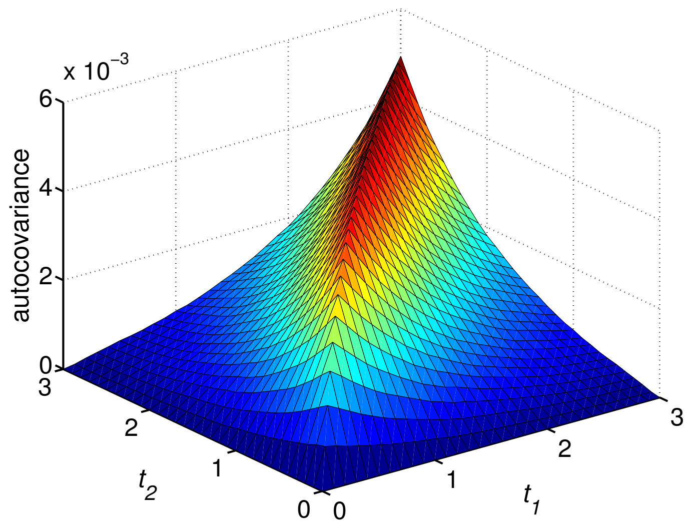


Figure 8.1.3: The autocovariance function for the differential equation $y' + y = f(t)$ when $f(t)$ is a Gaussian distribution. Twenty thousand realizations were used. The parameters used here are $y(0) = 0$ and $\Delta\tau = 0.01$.

simple, slightly damped harmonic oscillator

$$y'' + 2\zeta\omega_0 y' + \omega_0^2 y = f(t), \qquad 0 < \zeta \ll 1, \tag{8.1.25}$$

will exceed a certain magnitude when forced by white noise. In the physical world this transcending of a barrier or passage level leads to "bottoming" or "short circuiting."

Step 1: Using the MATLAB command `randn`, generate a stationary white noise excitation of length `N`. Let `deltat` denote the time interval Δt between each new forcing so that `n = 1` corresponds to `t = 0` and `n = N` corresponds to the end of the record `t = T`.

Step 2: The exact solution to Equation 8.1.25 is

$$\begin{aligned} y(t) = {} & y(0)e^{-\zeta\omega_0 t}\left[\cos\left(\sqrt{1-\zeta^2}\,\omega_0 t\right) + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t\right)\right] \\ & + \frac{y'(0)}{\omega_0\sqrt{1-\zeta^2}}e^{-\zeta\omega_0 t}\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t\right) \\ & + \int_0^t e^{-\zeta\omega_0(t-\tau)}\frac{\sin\left[\sqrt{1-\zeta^2}\,\omega_0(t-\tau)\right]}{\sqrt{1-\zeta^2}}\frac{f(\tau)}{\omega_0^2}\,d(\omega_0\tau) \end{aligned} \tag{8.1.26}$$

$$\begin{aligned} = {} & y(0)e^{-\zeta\omega_0 t}\left[\cos\left(\sqrt{1-\zeta^2}\,\omega_0 t\right) + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t\right)\right] \\ & + \frac{y'(0)}{\omega_0\sqrt{1-\zeta^2}}e^{-\zeta\omega_0 t}\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t\right) \\ & + e^{-\zeta\omega_0 t}\frac{\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t\right)}{\sqrt{1-\zeta^2}}\int_0^t e^{\zeta\omega_0\tau}\cos\left(\sqrt{1-\zeta^2}\,\omega_0\tau\right)\frac{f(\tau)}{\omega_0^2}\,d(\omega_0\tau) \end{aligned} \tag{8.1.27}$$

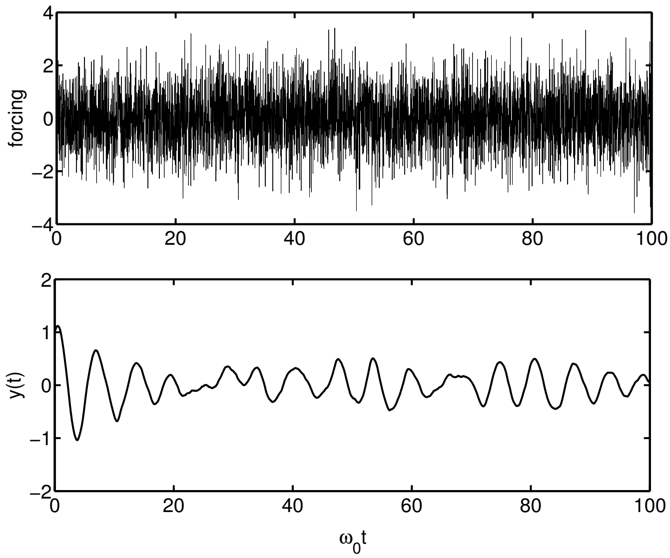


Figure 8.1.4: A realization of the random function $y(t)$ governed by Equation (1) when forced by the Gaussian random forcing shown in the top frame. The parameters used here are $y(0) = 1$, $y'(0) = 0.5$, $\zeta = 0.1$, and $\omega_0\Delta\tau = 0.02$.

$$- e^{-\zeta\omega_0 t}\frac{\cos\left(\sqrt{1-\zeta^2}\,\omega_0 t\right)}{\sqrt{1-\zeta^2}}\int_0^t e^{\zeta\omega_0\tau}\sin\left(\sqrt{1-\zeta^2}\,\omega_0\tau\right)\frac{f(\tau)}{\omega_0^2}\,d(\omega_0\tau).$$

Because you will be computing numerous realizations of $y(t)$ for different $f(t)$'s, an efficient method for evaluating the integrals must be employed. Equation 8.1.27 is more efficient than Equation 8.1.26.

Using the Gaussian random forcing that you created in Step 1, develop a MATLAB code to compute $y(t)$ given $y(0)$, $y'(0)$, ζ and $f(t)$. Figure 8.1.4 illustrates a realization where the trapezoidal rule was used to evaluate the integrals in Equation 8.1.27.

Step 3: Now that you can compute $y(t)$ or `y(n)` for a given Gaussian random forcing, generalize your code so that you can compute `irun` realizations and store them in `y(n,m)` where `m = 1:irun`. For a specific `n` or $\omega_0 t$, you can use MATLAB's commands `mean` and `var` to compute the mean $\mu_X(t)$ and the variance $\sigma_X^2(t)$. Figure 8.1.5 shows the results when 1000 realizations were used. For comparison the mean and variance of the forcing have also been included. Ideally this mean and variance should be zero and one, respectively. The crosses give the exact results that

$$\mu_X(t) = y(0)e^{-\zeta\omega_0 t}\left[\cos\left(\sqrt{1-\zeta^2}\,\omega_0 t\right) + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t\right)\right]$$
$$+ \frac{y'(0)}{\omega_0\sqrt{1-\zeta^2}}e^{-\zeta\omega_0 t}\sin\left(\sqrt{1-\zeta^2}\,\omega_0 t\right)$$

and Equation 8.1.24 when $D = \omega_0\Delta t/2$.

Step 4: Finally, generalize your MATLAB code so that you store the time `T(m)` that the solution `y(n)` exceeds a certain amplitude $b > 0$ for the *first* time during the realization `m`.

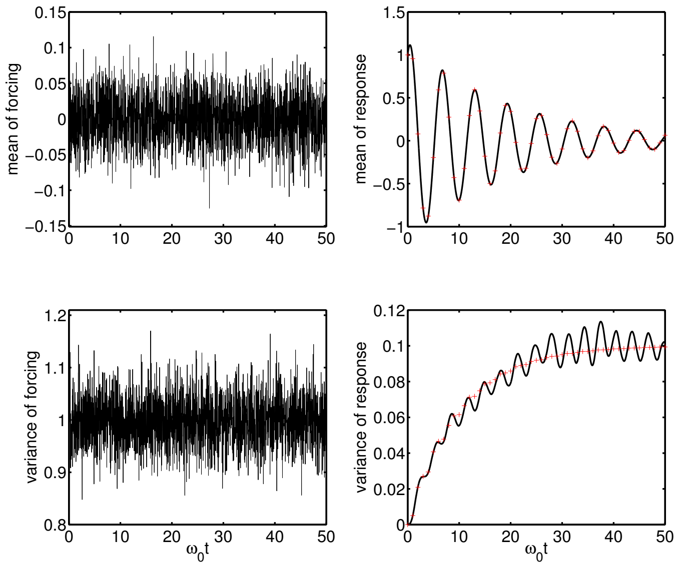


Figure 8.1.5: The mean $\mu_X(t)$ and variance $\sigma_X^2(t)$ of a slightly damped simple harmonic oscillator when forced by the Gaussian random noise. The parameters used here are $y(0) = 1$, $y'(0) = 0.5$, $\zeta = 0.1$, and $\omega_0\Delta\tau = 0.02$.

Of course, you can do this for several different b's during a particular realization. Once you have this data you can estimate the probability density function using `histc`. Figure 8.1.6 illustrates four probability density functions for $b = 0.4$, $b = 0.8$, $b = 1.2$, and $b = 1.6$.

Project: Wave Motion Generated by Random Forcing[7]

In the previous projects we examined ordinary differential equations that we forced with a random process. Here we wish to extend our investigation to the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = \cos(\omega t)\delta[x - X(t)],$$

subject to the boundary conditions

$$\lim_{|x|\to\infty} u(x,t) \to 0, \qquad 0 < t,$$

and initial conditions

$$u(x,0) = u_t(x,0) = 0, \qquad -\infty < x < \infty.$$

Here ω is a constant and $X(t)$ is a stochastic process.

In Example 5.4.4 we show that the solution to this problem is

$$u(x,t) = \tfrac{1}{2}\int_0^t H[t - \tau - |X(\tau) - x|]\cos(\omega\tau)\,d\tau.$$

[7] Based on a paper by Knowles, J. K., 1968: Propagation of one-dimensional waves from a source in random motion. *J. Acoust. Soc. Am.*, **43**, 948–957.

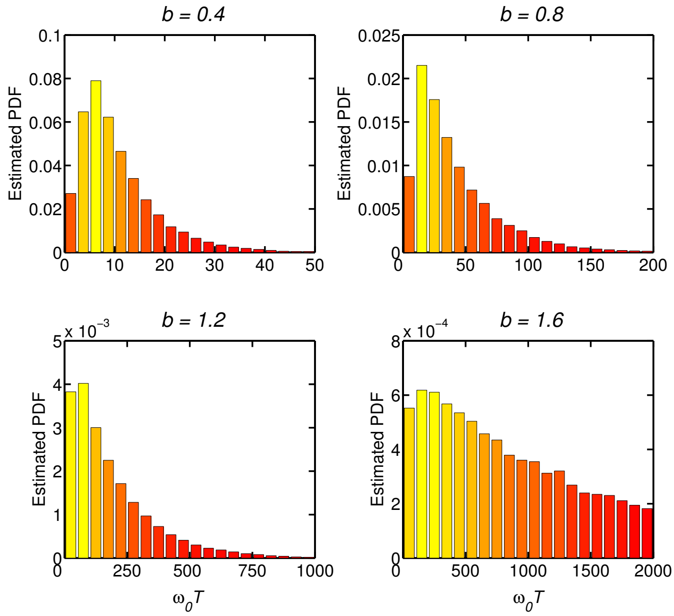


Figure 8.1.6: The probability density function that a slightly damped oscillator exceeds b at the time $\omega_0 T$. Fifty thousand realizations were used to compute the density function. The parameters used here are $y(0) = 0$, $y'(0) = 0$, $\zeta = 0.05$, and $\omega_0 \Delta\tau = 0.05$. The mean value of $\omega_0 T$ is 10.7 when $b = 0.4$, 41.93 when $b = 0.8$, 188.19 when $b = 1.2$, and 1406.8 when $b = 1.6$.

When the stochastic forcing is absent $X(t) = 0$, we can evaluate the integral and find that

$$u(x,t) = \frac{1}{2\omega} H(t - |x|) \sin[\omega(t - |x|)].$$

Step 1: Invoking the MATLAB command `randn`, use this Gaussian distribution to numerically generate an excitation $X(t)$.

Step 2: Using the Gaussian distribution from Step 1, develop a MATLAB code to compute $u(x,t)$. Figure 8.1.7 illustrates one realization where the trapezoidal rule was used to evaluate the integral.

Step 3: Now that you can compute $u(x,t)$ for a particular random forcing, generalize your code so that you can compute `irun` realizations. Then, for particular values of x and t, you can compute the corresponding mean and variance from the `irun` realizations. Figure 8.1.8 shows the results when 10,000 realizations were used.

Step 4: Redo your calculations but use a sine wave with random phase: $X(t) = A\sin(\Omega t + \xi)$, where A and Ω are constants and ξ is a random variable with a uniform distribution on $[0, 2\pi]$.

8.2 RANDOM WALK AND BROWNIAN MOTION

In 1827 the Scottish botanist Robert Brown (1773–1858) investigated the fertilization process in a newly discovered species of flower. Brown observed under the microscope that when the pollen grains from the flower were suspended in water, they performed a "rapid

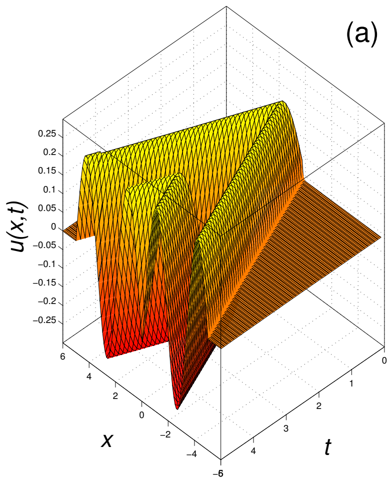


Figure 8.1.7: The solution (realization) of the wave equation when forced by a Gaussian distribution and $\omega = 2$. In frame (a), there is no stochastic forcing $X(t) = 0$. Frame (b) shows one realization.

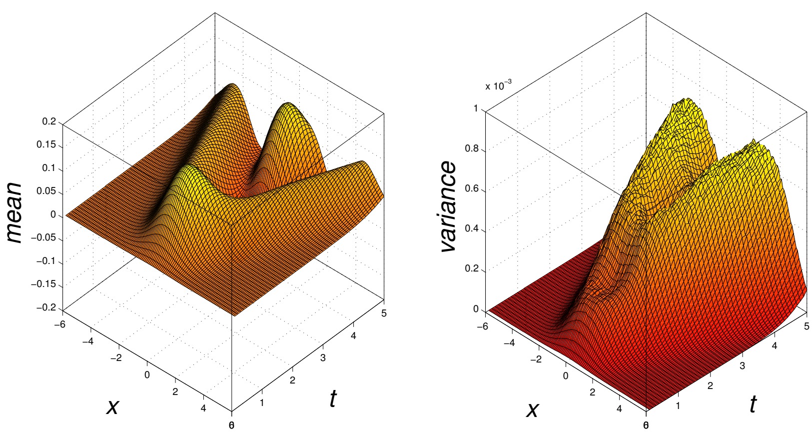


Figure 8.1.8: The mean and variance when the wave equation is forced by the stochastic forcing $\cos(\omega t)\delta[x - X(t)]$, where $\omega = 2$ and $X(t)$ is a Gaussian distribution.

oscillation motion." This motion, now known as *Brownian motion*, results from the random kinetic strikes on the pollen grain by water molecules. Brownian motion is an example of a random process known as *random walk*. This process has now been discovered in diverse disciplines, from biology[8] to finance. In this section we examine its nature.

Consider a particle that moves along a straight line in a series of steps of equal length. Each step is taken, either forwards or backwards, with equal probability $\frac{1}{2}$. After taking N steps, the particle *could* be at any one (let us denote it m) of the following points:

---

[8] Codling, E. A., M. J. Plank, and S. Benhamou, 2008: Random walk models in biology. *J. R. Soc. Interface*, **5**, 813–834.

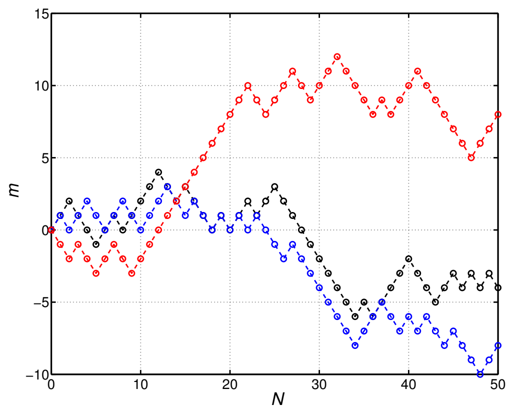


Figure 8.2.1: Three realizations of a one-dimensional random walk where $N = 50$.

$-N, -N+1, \ldots, -1, 0, 1, \ldots, N-1$ and N. Here m is a random variable.

We can generate realizations of one-dimensional Brownian motion using the MATLAB code:

```
clear

NN = 50; % select the number of steps for the particle to take
t = (0:1:NN); % create ``time'' as the particle moves

% create an array to give the position at each time step
m = zeros(size(t));
m(1) = 0; % initialize the position of particle

for N = 1:NN % now move the particle
  x = rand(1); % generate a random variable lying between [0,1]
  if (x <= 0.5) step = 1; % if less then 0.5, make it a ``head''
  else step = -1; end % otherwise it is a ``tail''
% move the particle one step to the right or left
  m(N+1) = m(N) + step;
end

%     plot the results

hold on
plot(t,m,'--ko','LineWidth',2,'MarkerSize',8)
xlabel('N','FontSize',25); ylabel('m','FontSize',25)
grid on % add a grid to axes
```

Figure 8.2.1 illustrates three such realizations.

A natural question would now be: What are the quantitative properties of random walk? In particular, what is the probability $P(m, N)$ that the particle is at point m after N displacements? We begin by noting the probability of any given sequence of N steps is $\left(\frac{1}{2}\right)^N$. The desired probability $P(m, N)$ equals $\left(\frac{1}{2}\right)^N$ times the number of distinct sequences of steps that will lead to the point m after N steps. To reach m, we must take $(N+m)/2$ steps in the positive direction and $(N-m)/2$ in the negative direction since $(N+m)/2-(N-m)/2 = m$. (Note both m *and* N must be even or odd.) The number of these distinct sequences is

$$\frac{N!}{\left[\frac{1}{2}(N+m)\right]!\left[\frac{1}{2}(N-m)\right]!}. \tag{8.2.1}$$

Therefore,

$$P(m,N) = \frac{N!}{\left[\frac{1}{2}(N+m)\right]!\left[\frac{1}{2}(N-m)\right]!}\left(\frac{1}{2}\right)^N. \tag{8.2.2}$$

Comparing these results with Equation 6.6.14, we see that $P(m,N)$ is simply a binomial distribution. For this reason, we immediately know $E(m) = 0$ and $\text{Var}(m) = N$. That is, the *average* position is the origin and the *spread* of the Brownian motion occurs as the square root of steps taken increases.

The case of greatest interest arises when N is large and $m \ll N$. Then we can approximate $P(m,N)$ by the Poisson distribution,

$$P(m,N) \approx \sqrt{\frac{2}{\pi N}}\exp\left(-\frac{m^2}{2N}\right). \tag{8.2.3}$$

Let us reexpress Equation 8.2.3 in terms of x and t where $x = m\Delta x$ and $t = N\Delta t$. Using these definitions, our equation becomes

$$P(x,t) = \frac{1}{2\sqrt{\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right), \tag{8.2.4}$$

if we define $D = (\Delta x)^2/(2\Delta t)$. The attentive student will note that $P(x,t)$ is the Green's function for the heat function, Example 5.5.1.

An alternative approach to this problem would be to compute many random walks and then calculate the probability density function from these computations. We can construct a MATLAB code to do this. First we would realize many random walks (here 2000) and count the number of times that they end at position m:

```
clear

NN = 100; % set the end point of the random walks
% introduce intermediate positions along the random walk
t = (0:1:NN);
% initialize array ‘‘m’’ which gives the position at any time
m = zeros(size(t));

for icount = 1:2000 % now perform many random walks

  m(1) = 0; % initial position of particle in each walk

  for N = 1:NN
    x = rand(1); % create a random variable between [0,1]
% if ’’x’’ less than 0.5, we have a ’’heads’’
    if (x <= 0.5) step = 1;
    else step = -1; end % otherwise we have a ’’tail’’
```

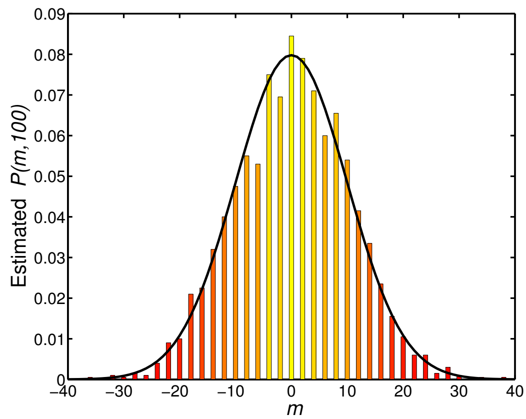


Figure 8.2.2: Numerical computation of $P(m, 100)$ using 2000 random walks. The black line gives Equation 8.2.3.

```
    m(N+1) = m(N) + step; % now take a step forward or backward
  end

% set up array that tracks of the final position of the particle
  location(icount) = m(N+1);

end

xx = -40:1:40;
% now count the particles that ended somewhere
%     between -40 and 40
[n,xout] = hist(location,xx)
% for comparison, compute Equation 8.2.3
w_exact = sqrt(2/(pi*NN))*exp(-xout.*xout/(2*NN));
n = n / 2000; % now compute the mass probability function

%     plot the results
bar_h = bar(xout,n)
bar_child = get(bar_h,'Children')
set(bar_child,'CData',n)
colormap(Autumn)
hold on
plot(xout,w_exact,'-k','LineWidth',3)
xlabel('\it m','FontSize',25)
ylabel('Estimated \it P(m,100)','FontSize',25)
```

Figure 8.2.2 illustrates the results of simulating random walk.

• Example 8.2.1: On the probability of striking a barrier

An important question in engineering is what is the probability that a given random system will exceed its design constraints. Here we ask a similar question about one-dimensional

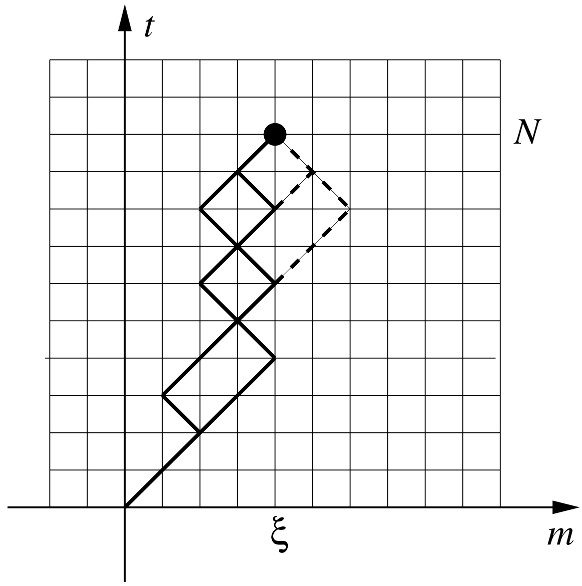


Figure 8.2.3: Several random walks from the origin to point (ξ, N). All of these walks would be excluded from our calculations because they either cross or touch the line $m = \xi$ before the final step.

Brownian motion: What is the probability that after taking N steps the particle arrives at ξ without ever having touched or crossed the line $m = \xi$ at *any* earlier step? We will do it exactly and then confirm our results using MATLAB.

The arrival of the particle at ξ after N steps implies that its position after $N-1$ steps must have been either $\xi - 1$ or $\xi + 1$. However, a trajectory from $(\xi+1, N-1)$ to (ξ, N) is not allowed because it must have crossed the line $m = \xi$ earlier. On the other hand, not all trajectories arriving at (ξ, N) from $(\xi - 1, N - 1)$ are acceptable because a certain number will have touched or crossed the line $m = \xi$ earlier than its last step. See Figure 8.2.3. Thus the number of permitted ways of arriving at ξ for the first time after N steps equals all possible ways of arriving at ξ *minus* any arrivals from $(\xi - 1, N - 1)$ *and* any arrivals that crossed or touched the line $m = \xi$ earlier than the $N - 1$.

From our previous work, the number of possible ways from the origin to (ξ, N) is

$$\frac{N!}{\left[\frac{1}{2}(N+\xi)\right]!\left[\frac{1}{2}(N-\xi)\right]!}. \tag{8.2.5}$$

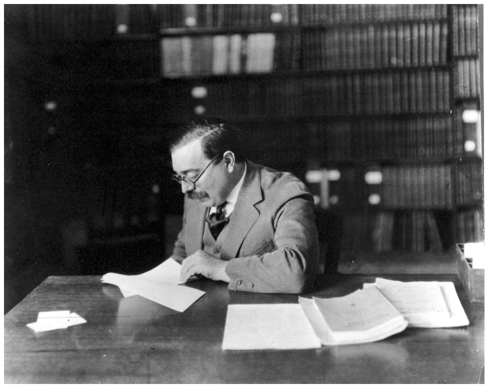
The number of possible ways from the origin to $(\xi + 1, N - 1)$ is

$$\frac{(N-1)!}{\left[\frac{1}{2}(N+\xi)\right]!\left[\frac{1}{2}(N-\xi-2)\right]!}. \tag{8.2.6}$$

Finally, the number of trajectories arriving at $(\xi - 1, N - 1)$ but having an earlier contact with, or a crossing of, the line $m = \xi$ is also

$$\frac{(N-1)!}{\left[\frac{1}{2}(N+\xi)\right]!\left[\frac{1}{2}(N-\xi-2)\right]!}, \tag{8.2.7}$$

since it equals the number of trajectories that arrive at $(\xi + 1, N - 1)$. From Figure 8.2.3 we see that, due to symmetry, the trajectory that leads to $(\xi + 1, N - 1)$ also leads to



One of the great mathematicians of the twentieth century, Norbert Wiener (1894–1964) graduated from high school at the age of 11 and Tufts at 14. Obtaining a doctorate in mathematical logic at 18, he repeatedly traveled to Europe for further education. His work extends over an extremely wide range from stochastic processes to harmonic analysis to cybernetics. (Photo courtesy of the MIT Museum with permission.)

$(\xi - 1, N - 1)$. Consequently the number of trajectories from the origin to (ξ, N) that have never touched or crossed $m = \xi$ is

$$\frac{N!}{\left[\frac{1}{2}(N+\xi)\right]!\left[\frac{1}{2}(N-\xi)\right]!} - 2\frac{(N-1)!}{\left[\frac{1}{2}(N+\xi)\right]!\left[\frac{1}{2}(N-\xi-2)\right]!}, \tag{8.2.8}$$

or

$$\frac{\xi}{N}\frac{N!}{\left[\frac{1}{2}(N+\xi)\right]!\left[\frac{1}{2}(N-\xi)\right]!}. \tag{8.2.9}$$

The probability $P(\xi, N)$ that we are seeking is

$$P(\xi, N) = \frac{\xi}{N}\frac{N!}{\left[\frac{1}{2}(N+\xi)\right]!\left[\frac{1}{2}(N-\xi)\right]!}\left(\frac{1}{2}\right)^N. \tag{8.2.10}$$

For large N, $P(\xi, N)$ is approximately given by

$$P(\xi, N) = \frac{\xi}{N}\sqrt{\frac{2}{\pi N}}\exp\left(-\frac{\xi^2}{2N}\right). \tag{8.2.11}$$

We can also compute this probability using the MATLAB code given above. In this code we replace the counting process `location(icount) = m(N+1);` by

```
b = sort(m);

if ( (m(NN+1) == b(NN+1)) & (b(NN+1)>b(NN)) )
jcount = jcount + 1;
location(jcount) = m(NN+1);
end
```

where we initialize `jcount = 0` at the beginning. The idea behind this code is as follows: For each of the `icount` trajectories, we use the MATLAB routine `sort` to arrange them from

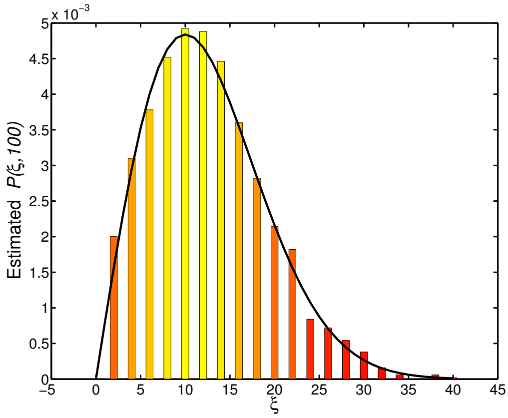


Figure 8.2.4: The probability $P(\xi, N)$ that a particle will reach the point $m = \xi$ without the particle ever crossing or touching the line $m = \xi$ earlier than $N = 100$. The solid line is the theoretical probability given by Equation 8.2.11. Here $50,000$ random walks were taken.

the left-most to the right-most position. To be included in the count of particles reaching $m = \xi$ at step N, the last position of the particle must be (ξ, N) *and* it may never have reached or crossed $m = \xi$. The `if` condition ensures that both conditions are met. If they are, that particular walk is accepted. Once again, the various right-most positions are binned and the probability is computed. Figure 8.2.4 illustrates this process using 5000 random walks and this result is compared with the probability given by Equation 8.2.11. □

• Example 8.2.2: Wiener process

Consider the time interval $(0, t]$ and let us subdivide it into subintervals of length Δt so that there are $t/\Delta t$ subintervals. Suppose now that a particle, initially at $x = 0$, takes a step (in one space dimension) at the times $\Delta t, 2\Delta t, \ldots$ and that the size of the step is either Δx or $-\Delta x$, with a probability of $\frac{1}{2}$ that the step is to the left or right. The position of the particle $X(t)$ at time t is a random walk, which has executed $t/\Delta t$ steps. Because the position depends on the choice of Δt and Δx, $X(t)$ depends upon t, Δt and Δx.

Mathematically we can describe this random process by

$$X(t) = \sum_{n=1}^{t/\Delta t} Z_i, \tag{8.2.12}$$

where the Z_i's are independent and identically distributed with

$$P(Z_i = \Delta x) = P(Z_i = -\Delta x) = \tfrac{1}{2}, \tag{8.2.13}$$

and $n = 1, 2, \ldots$. For each Z_i,

$$E(Z_i) = 0, \qquad \text{and} \qquad \mathrm{Var}(Z_i) = E(Z_i^2) = (\Delta x)^2. \tag{8.2.14}$$

From Equation 8.2.12, we see that

$$E[X(t)] = 0, \qquad \text{and} \qquad \mathrm{Var}[X(t)] = E(Z_i^2) = \frac{t}{\Delta t}\mathrm{Var}(Z_i) = \frac{t(\Delta x)^2}{\Delta t}. \tag{8.2.15}$$

Presently we have said nothing about the relationship between Δt and Δx except that both are small. However, we cannot have just any relationship between them because the variance would be either zero or infinite. The only reasonable choice is $\Delta x = \sqrt{\Delta t}$, which makes $\text{Var}[X(t)] = t$ for all values of Δt. In the limit $\Delta t \to 0$ the random variable $X(t)$ converges into a random variable, hereafter denoted by $B(t)$, with the properties that $E[B(t)] = 0$ and $\text{Var}[B(t)] = t$. The collection of random variables $\{B(t), t > 0\}$ is a continuous process in time and called a *Wiener process*. □

Our previous example shows that Brownian motion and the Wiener process are very closely linked. Because Brownian motion occurs in so many physical and biological processes, we shall focus on that motion (and the corresponding Wiener process) exclusively from now on. We define the *standard Brownian motion* (or *Wiener process*) $B(t)$ as a stochastic process that has the following properties:

1. It starts at zero: $B(0) = 0$.
2. Noting that $B(t) - B(s) \sim N(0, t-s)$, $E\{[B(t)-B(s)]^2\} = t-s$ and $\text{Var}\{[B(t)-B(s)]^2\} = 2(t-s)^2$. Replacing t with $t + dt$ and s with t, we find that $E\{[dB(t)]^2\} = dt$.
3. It has stationary and independent increments. Stationary increments means that $B(t+h) - B(\eta + h) = B(t) - B(\eta)$ for all h. An independent increment means $B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ are independent random variables.
4. Because increments of Brownian motion on adjacent intervals are independent regardless of the *length of the interval*, the derivative will oscillate wildly as $\Delta x \to 0$ and never converge. Consequently, Brownian motion is *nowhere differentiable*.
5. It has continuous sample paths, i.e., "no jumps."
6. The expectation values for the moments are given by

$$E[B^{2n}(t)] = \frac{(2n)!t^n}{n!2^n}, \qquad \text{and} \qquad E[B^{2n-1}(t)] = 0, \tag{8.2.16}$$

where $n > 0$. See Problem 1 at the end of Section 8.4.

Problems

1. Show that $E\{\sin[aB(t)]\} = 0$, where a is a real.

2. Show that

$$E\{\cos[aB(t)]\} = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!}(a^2 t)^n,$$

where a is a real.

3. Show that $E\{\exp[aB(t)]\} = \exp(a^2t/2)$, where a is a real.

Project: Probabilistic Solutions of Laplace's Equation

Laplace's equation can be solved using finite difference or finite element methods, respectively. During the 1940s, the apparently unrelated fields of random processes and potential theory were shown to be in some sense mathematically equivalent.[9] As a result, it is possible to use Brownian motion to solve Laplace's equation, as you will discover in this project. The present numerical method is useful for the following reasons: (1) the entire region need not be solved in order to determine potentials at relatively few points, (2) computation time is not lengthened by complex geometries, and (3) a probabilistic potential theory computation is more topologically efficient than matrix manipulations for problems in two and three spatial dimensions.

To understand this technique,[10] consider the following potential problem:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \qquad 0 < x < 1, \quad 0 < y < 1, \tag{8.2.17}$$

subject to the boundary conditions

$$u(x,0) = 0, \qquad u(x,1) = x, \qquad 0 < x < 1, \tag{8.2.18}$$

and

$$u(0,y) = u(1,y) = 0, \qquad 0 < y < 1. \tag{8.2.19}$$

If we introduce a uniform grid with $\Delta x = \Delta y = \Delta s$, then the finite difference method yields the difference equation:

$$4u(i,j) = u(i+1,j) + u(i-i,j) + u(i,j+1) + u(i,j-1), \tag{8.2.20}$$

with $i, j = 1, N-1$ and $\Delta s = 1/N$.

Consider now a random process of the Markov type in which a large number N_1 of noninteracting particles are released at some point (x_1, y_1) and subsequently perform Brownian motion in steps of length Δs each unit of time. At some later time, when a few arrive at point (x,y), we define a probability $P(i,j)$ of any of them reaching the boundary $y = 1$ with potential u_k at any subsequent time in the future. Whenever one of these particles does (later) arrive on $y = 1$, it is counted and removed from the system. Because $P(i,j)$ is defined over an infinite time interval of the diffusion process, the probability of any particles leaving (x,y) and arriving along some other boundary (where the potential equals 0) at some future time is $1 - P(i,j)$. Whenever a particle arrives along these boundaries it is also removed from the square.

Having defined $P(i,j)$ for an arbitrary (x,y), we now compute it in terms of the probabilities of the neighboring points. Because the process is Markovian, where a particle jumps from a point to a neighbor with no memory of the past,

$$\begin{aligned} P(i,j) = {} & p(i+1,j|i,j)P(i+1,j) + p(i-1,j|i,j)P(i-1,j) \\ & + p(i,j+1|i,j)P(i,j+1) + p(i,j-1|i,j)P(i,j-1), \end{aligned} \tag{8.2.21}$$

---

[9] See Hersh, R., and R. J. Griego, 1969: Brownian motion and potential theory. *Sci. Amer.*, **220**, 67–74.

[10] For the general case, see Bevensee, R. M., 1973: Probabilistic potential theory applied to electrical engineering problems. *Proc. IEEE*, **61**, 423–437.

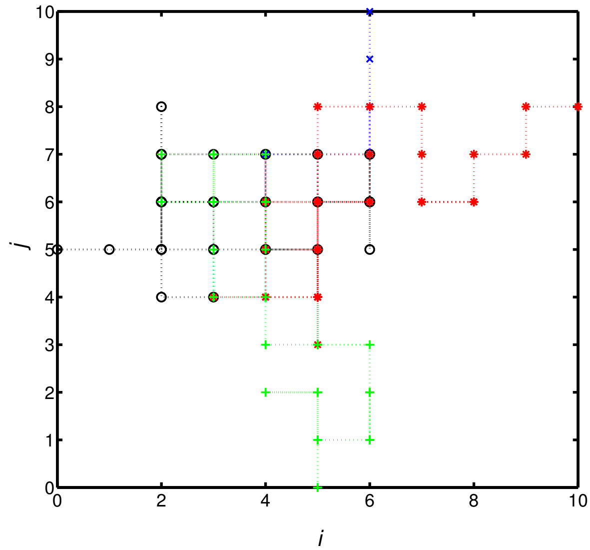


Figure 8.2.5: Four Brownian motions within a square domain with $\Delta x = \Delta y$. All of the random walks begin at grid point $i = 4$ and $j = 6$.

where $p(i+1, j|i, j)$ is the conditional probability of jumping to $(x + \Delta s, y)$, given that the particle is at (x, y). Equation 8.2.21 evaluates $P(i, j)$ as the sum of the probabilities of reaching $y = 1$ at some future time by various routes through the four neighboring points around (x, y). The sum of all the p's is exactly one because a particle at (x, y) must jump to a neighboring point during the next time interval.

Let us now compare Equation 8.2.20 and Equation 8.2.21. The potential $u(i, j)$ in Equation 8.2.20 and $P(i, j)$ becomes an identity if we take the conditional probabilities as

$$p(i+1, j|i, j) = p(i-1, j|i, j) = p(i, j+1|i, j) = p(i, j-1|i, j) = \tfrac{1}{4},$$

and if we also force $u(i, N) = P(i, N) = i$, $u(i, 0) = P(i, 0) = 0$, $u(0, j) = P(0, j) = 0$, and $u(N, j) = P(N, j) = 0$. Both the potential u and the probability P become continuous functions in the space as $\Delta s \to 0$, and both are well behaved as (x, y) approaches a boundary point. A particle starting along $y = 1$, where the potential is u_k, has a probability u_k of arriving there; a particle starting on the remaining boundaries, where the potential is zero, is immediately removed with no chance of arriving along $y = 1$. From these considerations, we have

$$u(i, j) \equiv P(i, j) = \lim_{N\to\infty} \frac{1}{N} \sum_k N_k u_k,$$

where N is the number of particles starting at (x, y) and N_k equals the number of particles that eventually—after infinite time—arrive along the entire boundary at potential u_k. This sum includes the boundary $y = 1$ and (trivially) the remaining boundaries.

Step 1: Develop a MATLAB code to perform two-dimensional Brownian motion. Let U be a uniformly distributed random variable lying between 0 and 1. You can use `rand`. If $0 < U \le \frac{1}{4}$, take one step to the right; $\frac{1}{4} < U \le \frac{1}{2}$, take one step to the left; if $\frac{1}{2} < U \le \frac{3}{4}$, take one step downward; and if $\frac{3}{4} < U \le 1$, take one step upward. For the arbitrary point `i,j` located on a grid of $N \times N$ points with $2 \le i, j \le N-1$, repeatedly take a random step

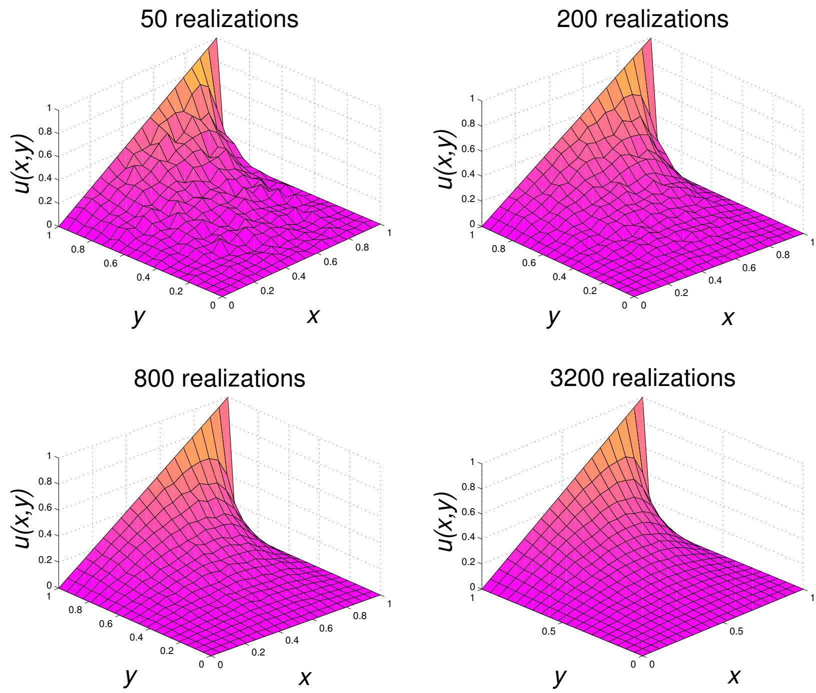


Figure 8.2.6: Solution to Equation 8.2.17 through Equation 8.2.19 using the probabilistic solution method.

until you reach one of the boundaries. Record the value of the potential at the boundary point. Let us call this result `u_k(1)`. Figure 8.2.5 illustrates four of these two-dimensional Brownian motions.

Step 2: Once you have confidence in your two-dimensional Brownian motion code, generalize it to solve Equation 8.2.17 through Equation 8.2.19 using `runs` realizations at some interior grid point. Then the solution `u(i,j)` is given by

$$\mathtt{u(i,j)} = \frac{1}{\mathtt{runs}} \sum_{n=1}^{\mathbf{runs}} \mathtt{u\_k}(n).$$

Step 3: Finally, plot your results. Figure 8.2.6 illustrates the potential field for different values of `runs`. What are the possible sources of error in using this method?

8.3 ITÔ'S STOCHASTIC INTEGRAL

In the previous section we noted that Brownian motion (the Wiener process) is nowhere differentiable. An obvious question is what is meant by the integral of a stochastic variable.

Consider the interval $[a, b]$, which we subdivide so that $a = t_0 < t_1 < t_2 < \cdots < t_n = b$. The earliest and simplest definition of the integral is

$$\int_a^b f(t)\, dt = \lim_{\Delta t \to 0} \sum_{i=1}^{n} f(\tau_i)\, \Delta t_i, \tag{8.3.1}$$

where $t_{i-1} \leq \tau_i \leq t_i$ and $\Delta t_i = t_i - t_{i-1}$. In the case of the classic integral, the integration is with regards to the increment dt.

Itô's integral is an integral where the infinitesimal increment involves Brownian motion $dB(t)$, which is a random variable. Before we can define this integral, we must introduce two important concepts. The first one is nonanticipating processes: A process $F(t)$ is a *nonanticipating process* if $F(t)$ is independent of any future increment $B(s) - B(t)$ for any s and t where $s > t$. Nonanticipating processes are important because Itô's integral applies only to them.

The second important concept is convergence in the mean square sense. It is defined by

$$\lim_{n\to\infty} E\left\{\left[S_n - \int_a^b F(t)\,dB(t)\right]^2\right\} = 0, \tag{8.3.2}$$

where S_n is the partial sum

$$S_n = \sum_{i=1}^{n} F(t_{i-1})\left[B(t_i) - B(t_{i-1})\right]. \tag{8.3.3}$$

We are now ready to define the Itô integral: It is the limit of the partial sum S_n:

$$\operatorname*{ms-lim}_{n\to\infty} S_n = \int_a^b F(t)\,dB(t), \tag{8.3.4}$$

where we denoted the limit in the mean square sense by ms−lim. Combining Equation 8.3.3 and Equation 8.3.4 together, we find that

$$\int_a^b f[t, B(t)]\,dB(t) = \lim_{\Delta t\to 0} \sum_{i=1}^{n} f[t_{i-1}, B(t_{i-1})]\left[B(t_i) - B(t_{i-1})\right], \tag{8.3.5}$$

where $t_i = i\Delta t$ and $\Delta t = (b-a)/N$. As one might suspect,

$$\int_a^b dB(t) = B(b) - B(a). \tag{8.3.6}$$

Because $F(t)$ and $dB(t)$ are random variables, so is Itô's integral.

The results from Equation 8.3.6 would be misunderstood if we think about them as we do in conventional calculus. We cannot evaluate the right side of Equation 8.3.6 by looking up $B(t)$ in some book entitled "Tables of Brownian Motion." This equation only holds true for a particular realization (sample path).

• **Example 8.3.1**

Let us use the definition of the Itô integral to evaluate Itô integral $\int_0^t B(x)\,dB(x)$. In the present case,

$$S_n = \sum_{i=1}^{n} B(x_{i-1})\left[B(x_i) - B(x_{i-1})\right], \tag{8.3.7}$$

where $x_i = it/n$. Because $2a(b-a) = b^2 - a^2 - (b-a)^2$,

$$S_n = \frac{1}{2}\sum_{i=1}^{n} B^2(x_i) - \frac{1}{2}\sum_{i=1}^{n} B^2(x_{i-1}) - \frac{1}{2}\sum_{i=1}^{n} [B(x_i) - B(x_{i-1})]^2 \tag{8.3.8}$$

$$= \tfrac{1}{2}B^2(t) - \frac{1}{2}\sum_{i=1}^{n} [B(x_i) - B(x_{i-1})]^2 . \tag{8.3.9}$$

Therefore,

$$\operatorname*{ms-lim}_{n\to\infty} S_n = \tfrac{1}{2}B^2(t) - \tfrac{1}{2} \operatorname*{ms-lim}_{n\to\infty} \sum_{i=1}^{n} [B(x_i) - B(x_{i-1})]^2 \tag{8.3.10}$$

$$= \tfrac{1}{2}B^2(t) - \frac{t}{2}. \tag{8.3.11}$$

As a consequence,

$$\int_0^t B(\eta)\, dB(\eta) = \tfrac{1}{2}B^2(t) - \frac{t}{2}, \tag{8.3.12}$$

or

$$\int_a^b B(t)\, dB(t) = \tfrac{1}{2}[B^2(b) - B^2(a)] - \frac{b-a}{2}. \tag{8.3.13}$$

Consider now the derivative of $B^2(t)$,

$$d[B^2(t)] = [B(t+dt) - B(t)]^2 = 2B(t)\, dB(t) + dB(t)\, dB(t). \tag{8.3.14}$$

In order for Equation 8.3.12 and Equation 8.3.14 to be consistent, we arrive at the very important result that

$$[dB(t)]^2 = dt \tag{8.3.15}$$

in the mean square sense. We will repeatedly use this result in the remaining portions of the chapter. □

Because the Itô integral is a random variable, two important quantities are its mean and variance. Let us turn first to the computation of the expectation of $\int_a^b f[t, B(t)]\, dB(t)$. From Equation 8.3.5 we find that

$$E\left\{\int_a^b f[t, B(t)]\, dB(t)\right\} = \lim_{\Delta t\to 0} E\left\{\sum_{i=1}^{n} f[t_{i-1}, B(t_{i-1})]\Delta B_i\right\} \tag{8.3.16}$$

$$= \lim_{\Delta t\to 0} \sum_{i=1}^{n} E\{f[t_{i-1}, B(t_{i-1})]\} E[\Delta B_i] = 0. \tag{8.3.17}$$

Therefore

$$E\left\{\int_a^b f[t, B(t)]\, dB(t)\right\} = 0. \tag{8.3.18}$$

To compute the variance, we begin by noting that

$$\left\{\int_a^b f[t, B(t)]\, dB(t)\right\}^2 = \lim_{\Delta t\to 0}\left\{\sum_{i=1}^n f[t_{i-1}, B(t_{i-1})]\Delta B_i\right\}^2 \tag{8.3.19}$$

$$= \lim_{\Delta t\to 0}\sum_{i=1}^n f^2[t_{i-1}, B(t_{i-1})](\Delta B_i)^2 \tag{8.3.20}$$

$$+ 2\sum_{\substack{i=1\\ i\neq j}}^n \sum_{j=1}^n f[t_{i-1}, B(t_{i-1})]\Delta B_i\, f[t_{j-1}, B(t_{j-1})]\Delta B_j.$$

Taking the expectation of both sides of Equation 8.3.20, we have that

$$E\left[\left\{\int_a^b f[t, B(t)]\, dB(t)\right\}^2\right] = \lim_{\Delta t\to 0}\sum_{i=1}^n E\{f^2[t_{i-1}, B(t_{i-1})]\}E[(\Delta B_i)^2] \tag{8.3.21}$$

$$+ 2\lim_{\Delta t\to 0}\sum_{\substack{i=1\\ i\neq j}}^n \sum_{j=1}^n E\{f(t_{i-1}, B(t_{i-1}))\}E[\Delta B_i]\, E\{f(t_{j-1}, B(t_{j-1}))\}E[\Delta B_j]$$

$$= \lim_{\Delta t\to 0}\sum_{i=1}^n E\{f^2[t_{i-1}, B(t_{i-1})]\}(t_{i+1} - t_i). \tag{8.3.22}$$

The double summation vanishes because of the independence of Brownian motion. Therefore, the final result is

$$E\left\{\int_a^b f[t, B(t)]\, dB(t)\right\}^2 = \int_a^b E\{f^2[t, B(t)]\}\, dt. \tag{8.3.23}$$

□

• **Example 8.3.2**

Consider the random number $X = \int_a^b \sqrt{t}\sin[B(t)]\, dB(t)$. Let us find $E(X)$ and $E(X^2)$. From Equation 8.3.18, we have that $E(X) = 0$. For that reason, $\text{var}(X) = E(X^2)$ and

$$\text{Var}(X) = E(X^2) = \int_a^b E\left\{\left|\sqrt{t}\,\sin[B(t)]\right|^2\right\}\, dt = \int_a^b t\, E\left\{\sin^2[B(t)]\right\}\, dt \tag{8.3.24}$$

$$= \int_a^b (t/2)E\{1 - \cos[2B(t)]\}\, dt = \int_a^b \frac{t}{2}\left[1 - \sum_{n=0}^\infty \frac{(-1)^n 2^{2n} t^n}{2^n n!}\right]\, dt \tag{8.3.25}$$

$$= -\frac{1}{2}\int_a^b \sum_{n=1}^\infty \frac{(-1)^n 2^n}{n!} t^{n+1}\, dt = \frac{1}{2}\sum_{n=1}^\infty \frac{(-1)^{n+1} 2^n}{(n+2)n!}\left(b^{n+2} - a^{n+2}\right). \tag{8.3.26}$$

The value of $E\{\cos[2B(t)]\}$ follows from Problem 2 at the end of the last section. □

Table 8.3.1 gives a list of Itô stochastic integrals. Most of these results were not derived from the definition of the Itô stochastic integral but from Itô lemma, to which we now turn.

Problems

Consider the random variable $X = \int_a^b f[t, B(t)]\, dB$. Find $E(X)$ and $\text{Var}(X)$ for the following $f[t, B(t)]$:

1. $f[t, B(t)] = t$
2. $f[t, B(t)] = tB(t)$
3. $f[t, B(t)] = |B(t)|$
4. $f[f, B(t)] = \sqrt{t}\, \exp[B(t)]$

5. If $X = \int_a^b f(t)\{\sin[B(t)] + \cos[B(t)]\}\, dB(t)$, show that $\text{var}(X) = \int_a^b f^2(t)\, dt$, if $f(t)$ is a real function.

Project: Numerical Integration of Itô's Integral

Equation 8.3.5 is useful for numerically integrating the Itô integral

$$\int_0^t f[x, B(x)]\, dB(x).$$

Write a MATLAB script to check Example 8.3.1 for various values of n when $t = 1$. How does the error vary with n?

Project: Numerical Check of Equations 8.3.18 and 8.3.23

Using the script from the previous project, develop MATLAB code to compute Equation 8.3.18 and Equation 8.3.23. Using a million realizations (sample paths), compare your numerical results with the exact answer when $a = 1$, $b = 1$, and $f[t, B(t)] = \sqrt{t}\, \sin[B(t)]$.

8.4 ITÔ'S LEMMA

Before we can solve stochastic differential equations, we must derive a key result in stochastic calculus: *Itô's formula or lemma*. This is stochastic calculus's version of the chain rule.

Consider a function $f(t)$ that is twice differentiable. Using Taylor's expansion,

$$df(B) = f(B + dB) - f(B) = f'(B)\, dB + \tfrac{1}{2} f''(B)\, (dB)^2 + \cdots, \tag{8.4.1}$$

where $B(t)$ denotes Brownian motion. Integrating Equation 8.4.1 from s to t, we find that

$$\int_s^t df(B) = f[B(t)] - f[B(s)] = \int_s^t f'(B)\, dB + \tfrac{1}{2} \int_s^t f''(B)\, dx + \cdots, \tag{8.4.2}$$

because $[dB(x)]^2 = dx$. The first integral on the right side of Equation 8.4.2 is an Itô's stochastic integral while the second one can be interpreted as the Riemann integral of $f''(B)$. Therefore, Itô's lemma or formula is

$$f[B(t)] - f[B(s)] = \int_s^t f'(B)\, dB + \tfrac{1}{2} \int_s^t f''(B)\, dx \tag{8.4.3}$$

Table 8.3.1: A Table of Itô Stochastic Integrals with $t > 0$ and $b > a > 0$

1.
$$\int_a^b dB(t) = B(b) - B(a)$$

2.
$$\int_0^t B(\eta)\, dB(\eta) = \tfrac{1}{2}[B^2(t) - t]$$

3.
$$\int_0^t [B^2(\eta) - \eta]\, dB(\eta) = \tfrac{1}{3}B^2(t) - tB(t)$$

4.
$$\int_0^t \eta\, dB(\eta) = tB(t) - \int_0^t B(\eta)\, d\eta$$

5.
$$\int_0^t B^2(\eta)\, dB(\eta) = \tfrac{1}{3}B^3(t) - \int_0^t B(\eta)\, d\eta$$

6.
$$\int_0^t e^{\lambda^2\eta/2} \cos[\lambda B(\eta)]\, dB(\eta) = \frac{1}{\lambda} e^{\lambda^2 t/2} \sin[\lambda B(t)]$$

7.
$$\int_0^t e^{\lambda^2\eta/2} \sin[\lambda B(\eta)]\, dB(\eta) = \frac{1}{\lambda} \left\{1 - e^{\lambda^2 t/2} \cos[\lambda B(t)]\right\}$$

8.
$$\int_0^t \exp\left[-\tfrac{1}{2}\lambda^2\eta \pm \lambda B(\eta)\right]\ dB(\eta) = \pm\frac{1}{\lambda} \left\{\exp\left[-\tfrac{1}{2}\lambda^2 t \pm \lambda B(t)\right] - 1\right\}$$

9.
$$\int_a^b B(\eta) \exp\left[\frac{B^2(\eta)}{2\eta}\right]\ \frac{dB(\eta)}{\eta^{3/2}} = b^{-1/2} \exp\left[\frac{B^2(b)}{2b}\right] - a^{-1/2} \exp\left[\frac{B^2(a)}{2a}\right]$$

10.
$$\int_a^b f(\eta)\, dB(\eta) = f(t)B(t)\Big|_a^b - \int_a^b f'(\eta)B(\eta)\, d\eta$$

11.
$$\int_a^b g'[B(\eta)]\, dB(\eta) = g[B(t)]\Big|_a^b - \frac{1}{2}\int_a^b g''[B(\eta)]\, d\eta$$

for $t > s$.

• Example 8.4.1

Consider the case when $f(t) = t^2$ and $s = 0$. Then, Itô's formula yields

$$B^2(t) - B^2(0) = 2\int_0^t B(x)\,dB(x) - \int_0^t dx. \tag{8.4.4}$$

Evaluating the second integral and noting that $B(0) = 0$, we again obtain Equation 8.3.12, that

$$\int_0^t B(x)\,dB(x) = \tfrac{1}{2}[B^2(t) - t]. \tag{8.4.5}$$

□

• Example 8.4.2

Consider the case when $f(t) = e^{at}$ and $s = 0$. Then, Itô's formula yields

$$e^{aB(t)} - 1 = a\int_0^t e^{aB(x)}\,dB(x) + \frac{a^2}{2}\int_0^t e^{aB(x)}\,dx. \tag{8.4.6}$$

Computing the expectation of both sides,

$$E\left[e^{aB(t)}\right] - 1 = \frac{a^2}{2}\int_0^t E\left[e^{aB(x)}\right]\,dx. \tag{8.4.7}$$

Solving this integral equation, we find that $E[e^{aB(t)}] = e^{a^2t/2}$, a result that we found earlier in Problem 2, Section 8.2. □

• Example 8.4.3

If $f(t) = \sin(\lambda t)$, $\lambda > 0$, then Itô's formula gives

$$\sin[\lambda B(t)] = \lambda\int_0^t \cos[\lambda B(\eta)]\,dB(\eta) - \tfrac{1}{2}\lambda^2\int_0^t \sin[\lambda B(\eta)]\,d\eta. \tag{8.4.8}$$

Taking the expectation of both sides of Equation 8.4.8, we find that

$$E\{\sin[\lambda B(t)]\} = -\tfrac{1}{2}\lambda^2\int_0^t E\{\sin[\lambda B(\eta)]\}\,d\eta. \tag{8.4.9}$$

Setting $g(t) = E\{\sin[\lambda B(t)]\}$, then

$$g(t) = -\tfrac{1}{2}\lambda^2\int_0^t g(\eta)\,d\eta. \tag{8.4.10}$$

The solution to this integral equation is $g(t) = 0$. Therefore, $E\{\sin[\lambda B(t)]\} = 0$. □

Educated at the Imperial University of Tokyo, Kiyoshi Itô (1915–2008) applied the techniques of differential and integral to stochastic processes. Much of Itô's original work from 1938 to 1945 was done while he worked for the Japanese National Statistical Bureau. After receiving his doctorate, Itô became a professor at the University of Kyoto from 1952 to 1979. (Author: Konrad Jacobs, Source: Archives of the Mathematisches Forschungsinstitut Oberwolfach.)

The second version of Itô's lemma begins with the second-order Taylor expansion of the function $f(t,x)$:

$$\begin{aligned} f[t+dt, B(t+dt)] - f[t, B(t)] = {} & f_t[t, B(t)]\, dt + f_x[t, B(t)]\, dB(t) \\ & + \tfrac{1}{2}\big\{ f_{tt}[t, B(t)]\, (dt)^2 + f_{xt}[t, B(t)]\, dt\, dB(t) \\ & + f_{xx}[t, B(t)]\, [dB(t)]^2 \big\} + \cdots . \end{aligned} \tag{8.4.11}$$

Here we assume that $f[t, B(t)]$ has continuous partial derivatives of at least second order. Neglecting higher-order terms in Equation 8.4.11, which include the terms with factors such as $(dt)^2$ and $dt\, dB(t)$ but *not* $[dB(t)]^2$ because $[dB(t)]^2 = dt$, our second version of Itô's lemma is

$$f[t, B(t)] - f[s, B(s)] = \int_s^t \big\{ f_t[\eta, B(\eta)] + \tfrac{1}{2} f_{xx}[\eta, B(\eta)] \big\}\, d\eta + \int_s^t f_x[\eta, B(\eta)]\, dB(\eta) \tag{8.4.12}$$

if $t > s$.

• Example 8.4.4

Consider the function $f(t,x) = e^{x-t/2}$. Then,

$$f_t(t,x) = -\tfrac{1}{2}e^{x-t/2}, \quad f_x(t,x) = e^{x-t/2}, \quad \text{and} \quad f_{xx}(t,x) = e^{x-t/2}. \tag{8.4.13}$$

Therefore, from Itô's lemma, we have that

$$e^{B(t)-t/2} - e^{B(s)-s/2} = \int_s^t e^{-\eta/2} e^{B(\eta)}\, dB(\eta). \tag{8.4.14}$$

□

• Example 8.4.5: Integration by parts

Consider the case when $F(t,x) = f(t)g(x)$. The Itô formula gives

$$d[f(t)g(x)] = \left\{f'(t)g[B(t)] + \tfrac{1}{2}f(t)g''[B(t)]\right\}\, dt + f(t)g'[B(t)]\, dB(t). \tag{8.4.15}$$

Integrating both sides of Equation 8.4.15, we find that

$$\int_a^b f(t)g'[B(t)]\, dB(t) = f(t)g[B(t)]\Big|_a^b - \int_a^b f'(t)g[B(t)]\, dt - \tfrac{1}{2}\int_a^b f(t)g''[B(t)]\, dt, \tag{8.4.16}$$

which is the stochastic version of integration by parts.

For example, let us choose $f(t) = e^{\alpha t}$ and $g(x) = \sin(x)$. Equation 8.4.16 yields

$$\int_0^t e^{\alpha\eta}\cos[B(\eta)]\, dB(\eta) = e^{\alpha\eta}\sin[B(\eta)]\Big|_0^t - \alpha\int_0^t e^{\alpha\eta}\sin[B(\eta)]\, d\eta + \tfrac{1}{2}\int_0^t e^{\alpha\eta}\sin[B(\eta)]\, d\eta \tag{8.4.17}$$

$$= e^{\alpha t}\sin[B(t)] - \left(\alpha - \tfrac{1}{2}\right)\int_0^t e^{\alpha\eta}\sin[B(\eta)]\, d\eta. \tag{8.4.18}$$

In the special case of $\alpha = \frac{1}{2}$, Equation 8.4.18 simplifies to

$$\int_0^t e^{\alpha\eta}\cos[B(\eta)]\, dB(\eta) = e^{t/2}\sin[B(t)]. \tag{8.4.19}$$

□

An important extension of Itô's lemma involves the function $f[t, X(t)]$ where $X(t)$ is no longer simply Brownian motion but is given by the first-order stochastic differential equation

$$dX(t) = cX(t)\, dt + \sigma X(t)\, dB(t), \tag{8.4.20}$$

where c and σ are real. The second-order Taylor expansion of the function $f[t, X(t)]$ becomes

$$\begin{aligned} f[t+dt, X(t+dt)] - f[t, X(t)] &= f_t[t, X(t)]\, dt + f_x[t, X(t)]\, dX(t) \\ &+ \tfrac{1}{2}\left\{f_{tt}[t, X(t)]\,(dt)^2 + f_{xt}[t, X(t)]\, dt\, dX(t) + f_{xx}[t, X(t)]\,[dX(t)]^2\right\} + \cdots. \end{aligned} \tag{8.4.21}$$

Next, we substitute for $dX(t)$ using Equation 8.4.20, neglect terms involving $(dt)^2$ and $dt\, dB(t)$, and substitute $[dB(t)]^2 = dt$. Consequently,

$$df = f[t + dt, X(t + dt)] - f[t, X(t)] \tag{8.4.22}$$

$$= \sigma X(t) f_x[t, X(t)]\, dB(t) + \left\{ f_t[t, X(t)] + cX(t) f_x[t, X(t)] + \tfrac{1}{2}\sigma^2 X^2(t) f_{xx}[t, X(t)] \right\} dt. \tag{8.4.23}$$

The present extension of Itô's lemma reads

$$\begin{aligned} f[t, X(t)] - f[s, X(s)] &= \int_s^t \left\{ f_t[\eta, X(\eta)] + cX(\eta) f_x[\eta, X(\eta)] + \tfrac{1}{2}\sigma^2 X^2(\eta) f_{xx}[\eta, X(\eta)] \right\} d\eta \\ &\quad + \int_s^t \sigma X(\eta) f_x[\eta, X(\eta)]\, dB(\eta) \qquad (8.4.24) \\ &= \int_s^t \left\{ f_t[\eta, X(\eta)] + \tfrac{1}{2}\sigma^2 X^2(\eta) f_{xx}[\eta, X(\eta)] \right\} d\eta \\ &\quad + \int_s^t f_x[\eta, X(\eta)]\, dX(\eta), \qquad (8.4.25) \end{aligned}$$

where

$$dX(\eta) = cX(\eta)\, d\eta + \sigma X(\eta)\, dB(\eta) \tag{8.4.26}$$

and $t > s$.

We can finally generalize Itô's formula to the case of several Itô processes with respect to the *same* Brownian motion. For example, let $X(t)$ and $Y(t)$ denote two Itô processes governed by

$$dX(t) = A^{(1,1)}(t)\, dt + A^{(2,1)}(t)\, dB(t), \tag{8.4.27}$$

and

$$dY(t) = A^{(1,2)}(t)\, dt + A^{(2,2)}(t)\, dB(t). \tag{8.4.28}$$

For stochastic process $f[t, X(t), Y(t)]$, the Taylor expansion is

$$\begin{aligned} df[t, X(t), Y(t)] &= f_t[t, X(t), Y(t)]\, dt + f_x[t, X(t), Y(t)]\, dX(t) \\ &\quad + f_y[t, X(t), Y(t)]\, dY(t) \qquad (8.4.29) \\ &\quad + \tfrac{1}{2} f_{xx}[t, X(t), Y(t)] A^{(2,1)}(t) A^{(2,1)}(t)\, dt + \tfrac{1}{2} f_{xy}[t, X(t), Y(t)] A^{(2,1)}(t) A^{(2,2)}(t)\, dt \\ &\quad + \tfrac{1}{2} f_{yx}]t, X(t), Y(t)] A^{(2,2)}(t) A^{(2,1)}(t)\, dt + \tfrac{1}{2} f_{yy}[t, X(t), Y(t)] A^{(2,2)}(t) A^{(2,2)}(t)\, dt. \end{aligned}$$

• **Example 8.4.6: Product rule**

Consider the special case $f(t, x, y) = xy$. Then $f_t = 0$, $f_x = y$, $f_y = x$, $f_{xx} = f_{yy} = 0$, and $f_{xy} = f_{yx} = 1$. In this case, Equation 8.4.29 simplifies to

$$d[X(t)Y(t)] = Y(t)\, dX(t) + X(t)\, dY(t) + A^{(2,1)}[t, X(t), Y(t)] A^{(2,2)}[t, X(t), Y(t)]\, dt. \tag{8.4.30}$$

A very important case occurs when $A^{(2,1)}[t, X(t), Y(t)] = 0$ and $X(t) = g(t)$ is purely deterministic. In this case,

$$d[g(t)Y(t)] = Y(t)\, dg(t) + g(t)\, dY(t). \tag{8.4.31}$$

This is exactly the product rule from calculus.

Problems

1. (a) Use Equation 8.4.3 and $f(t) = t^n$ to show that

$$B^n(t) = n \int_0^t B^{n-1}(x)\, dB(x) + \frac{n(n-1)}{2} \int_0^t B^{n-2}(x)\, dx.$$

(b) Show that

$$E[B^n(t)] = \frac{n(n-1)}{2} \int_0^t E[B^{n-2}(x)]\, dx.$$

(c) Because $E[B(t)] = 0$ and $E[B^2(t)] = t$, show that

$$E\big[B^{2k+1}(t)\big] = 0, \qquad \text{and} \qquad E\big[B^{2k}(t)\big] = \frac{(2k)!}{2^k k!} t^k.$$

2. Let $f(t,x) = x^2 t$ and use Itô's formula to show that

$$\int_0^t B^2(\eta)\, dt + 2 \int_0^t \eta B(\eta)\, dB(\eta) = tB^2(t) - t^2/2.$$

3. Let $f(t,x) = x^{3/2}$ and use Itô's formula to show that

$$\int_0^t B^{1/2}(\eta)\, dB(\eta) = \frac{2}{3} B^{3/2}(t) - \frac{1}{4} \int_0^t B^{-1/2}(\eta)\, dt.$$

4. Let $f(t,x) = x^3/3 - tx$ and use Itô's formula to show that

$$\int_0^t [B^2(\eta) - \eta]\, dB(\eta) = \tfrac{1}{3} B^3(t) - t\, B(t).$$

5. If $f(x)$ is any continuously differentiable function, use Equation 8.4.29 to show that

$$\int_0^t f(\eta)\, dB(\eta) = f(t)B(t) - \int_0^t f'(\eta) B(\eta)\, d\eta.$$

6. If $f(t) = e^t$, use the previous problem to show that

$$\int_0^t e^\eta\, dB(\eta) = e^t B(t) - \int_0^t e^\eta B(\eta)\, d\eta.$$

7. Let $G(x)$ denote the antiderivative of $g(x)$. Use Equation 8.4.3 to show that

$$\int_a^b g[B(t)]\, dB(t) = G[B(t)]\Big|_a^b - \tfrac{1}{2} \int_a^b g'[B(t)]\, dt.$$

8. (a) If $g(x) = xe^x$, use Problem 7 to show that

$$\int_0^t B(\eta)e^{B(\eta)}\,dB(\eta) = [B(t)-1]e^{B(t)} + 1 - \tfrac{1}{2}\int_0^t [B(\eta)+1]e^{B(\eta)}\,d\eta.$$

(b) Use Equation 8.3.18 to show that

$$\begin{aligned} E\Big[B(t)e^{B(t)}\Big] &= E\Big[e^{B(t)}\Big] - 1 + \tfrac{1}{2}\int_0^t \Big\{E\Big[B(\eta)e^{B(\eta)}\Big] + E\Big[e^{B(\eta)}\Big]\Big\}\,d\eta \\ &= e^{t/2} - 1 + \tfrac{1}{2}\int_0^t \Big\{e^{\eta/2} + E\Big[B(\eta)e^{B(\eta)}\Big]\Big\}\,d\eta. \end{aligned}$$

(c) Setting $g(t) = E\big[B(t)e^{B(t)}\big]$, use Laplace transforms to show that

$$E\Big[B(t)e^{B(t)}\Big] = te^{t/2}.$$

9. (a) If $g(x) = 1/(1+x^2)$, use Problem 7 to show that

$$\int_0^t \frac{dB(\eta)}{1+B^2(\eta)} = \arctan[B(t)] + \int_0^t \frac{B(\eta)}{[1+B^2(\eta)]^2}\,d\eta.$$

(b) Use Equation 8.3.18 to show that

$$\int_0^t E\left\{\frac{B(\eta)}{[1+B^2(\eta)]^2}\right\}\,d\eta = -E\{\arctan[B(t)]\}\,.$$

(c) Because

$$-\frac{3\sqrt{3}}{16} \le \frac{x}{(1+x^2)^2} \le \frac{3\sqrt{3}}{16}, \quad \text{or} \quad -\frac{3\sqrt{3}}{16}t \le \int_0^t \frac{B(\eta)}{[1+B^2(\eta)]^2}\,d\eta \le \frac{3\sqrt{3}}{16}t,$$

show that

$$-\frac{3\sqrt{3}}{16}t \le E\{\arctan[B(t)]\} \le \frac{3\sqrt{3}}{16}t.$$

10. If $g(x) = x/(1+x^2)$, use Problem 7 to show that

$$\int_0^t \frac{B(\eta)}{1+B^2(\eta)}\,dB(\eta) = \tfrac{1}{2}\log[1+B^2(t)] - \tfrac{1}{2}\int_0^t \frac{1-B^2(\eta)}{[1+B^2(\eta)]^2}\,d\eta.$$

11. Use integration by parts with $f(t) = e^{\beta t}$ and $g(x) = -\cos(x)$ to show that

$$\int_0^t e^{\beta\eta}\sin[B(\eta)]\,dB(\eta) = 1 - e^{\beta t}\cos[B(t)] + \left(\beta - \tfrac{1}{2}\right)\int_0^t e^{\beta\eta}\cos[B(\eta)]\,d\eta.$$

Then, take $\beta = \frac{1}{2}$ and show that

$$\int_0^t e^{\eta/2} \sin[B(\eta)]\, dB(\eta) = 1 - e^{t/2} \cos[B(t)].$$

12. Redo Example 8.4.3 and show that $E\{\cos[\lambda B(t)]\} = e^{-\lambda^2 t/2}$, $\lambda > 0$.

13. Use trigonometric double angle formulas to show that

(a) $$E\{\sin[t + \lambda B(t)]\} = e^{-\lambda^2 t/2} \sin(t),$$

and

(b) $$E\{\cos[t + \lambda B(t)]\} = e^{-\lambda^2 t/2} \cos(t),$$

when $\lambda > 0$.

14. Following Example 8.4.4 with $f(t, x) = \pm\lambda \exp\bigl(\pm\lambda x - \lambda^2 t/2\bigr)$, $\lambda > 0$, show that

$$\int_0^t \exp\left[\pm\lambda B(\eta) - \frac{\lambda^2 \eta}{2}\right] dB(\eta) = \pm\frac{1}{\lambda}\left\{\exp\left[\pm\lambda B(t) - \frac{\lambda^2 t}{2}\right] - 1\right\}.$$

15. Following Example 8.4.4 with $f(t, x) = \exp(\lambda^2 t/2) \sin(\lambda x)$, $\lambda > 0$, show that

$$\int_0^t \exp\left(\frac{\lambda^2 \eta}{2}\right) \cos[\lambda B(\eta)]\, dB(\eta) = \frac{1}{\lambda} \exp\left(\frac{\lambda^2 t}{2}\right) \sin[\lambda B(t)].$$

16. Following Example 8.4.4 with $f(t, x) = -\exp(\lambda^2 t/2) \cos(\lambda x)$, $\lambda > 0$, show that

$$\int_0^t \exp\left(\frac{\lambda^2 \eta}{2}\right) \sin[\lambda B(\eta)]\, dB(\eta) = \frac{1}{\lambda}\left\{1 - \exp\left(\frac{\lambda^2 t}{2}\right) \cos[\lambda B(t)]\right\}.$$

17. Following Example 8.4.4 with $f(t, x) = t^{-1/2} \exp[x^2/(2t)]$, show that

$$\int_a^b B(t) \exp\left[\frac{B^2(t)}{2t}\right] \frac{dB(t)}{t^{3/2}} = b^{-1/2} \exp\left[\frac{B^2(b)}{2b}\right] - a^{-1/2} \exp\left[\frac{B^2(a)}{2a}\right].$$

18. The average of geometric Brownian motion on $[0, t]$ is defined by

$$G(t) = \frac{1}{t} \int_0^t e^{B(\eta)}\, d\eta.$$

Use the product rule to find $dG(t)$. Hint: Take the time derivative of $tG(t) = \int_0^t e^{B(\eta)}\, d\eta$.

8.5 STOCHASTIC DIFFERENTIAL EQUATIONS

We have reached the point where we can examine stochastic differential equations. Of all the possible stochastic differential equations, we will focus on Langevin's equation[11]—a model of the velocity of Brownian particles. We will employ this model in a manner similar to that played by simple harmonic motion in the study of ordinary differential equations. It illustrates many of the aspects of stochastic differential equations without being overly complicated.

• Example 8.5.1

Before we consider the general stochastic differential equation, consider the following cases where we can make clever use of the product rule. For example, let us solve

$$dX(t) = [t + B^2(t)]\,dt + 2tB(t)\,dB(t), \qquad X(0) = X_0. \tag{8.5.1}$$

In the present case, we can find the solution by noting that

$$dX(t) = B^2(t)\,dt + t[2B(t)\,dB(t) + dt] = B^2(t)\,dt + t\,d[B^2(t)] = d[tB^2(t)]. \tag{8.5.2}$$

Integrating both sides of Equation 8.5.2, we find that the solution to Equation 8.5.1 is

$$X(t) = tB^2(t) + X_0. \tag{8.5.3}$$

Similarly, let us solve the stochastic differential equation

$$dX(t) = \frac{b - X(t)}{1-t}\,dt + dB(t), \qquad 0 \le t < 1, \tag{8.5.4}$$

with $X(0) = X_0$.

We begin by writing Equation 8.5.4 as

$$\frac{d[b - X(t)]}{1-t} + \frac{b - X(t)}{(1-t)^2}\,dt = -\frac{dB(t)}{1-t}. \tag{8.5.5}$$

Running the product rule backwards,

$$d\left[\frac{b - X(t)}{1-t}\right] = -\frac{dB(t)}{1-t}. \tag{8.5.6}$$

Integrating both sides of Equation 8.5.6 from 0 to t, we find that

$$\frac{b - X(t)}{1-t} = b - X(0) - \int_0^t \frac{dB(\eta)}{1-\eta}. \tag{8.5.7}$$

Solving for $X(t)$, we obtain the final result that

$$X(t) = b - [b - X(0)](1-t) + (1-t)\int_0^t \frac{dB(\eta)}{1-\eta}. \tag{8.5.8}$$

---

[11] Langevin, P., 1908: Sur la théorie du mouvement brownien. *C. R. Acad. Sci. Paris*, **146**, 530–530. English translation: Langevin, P., 1997: On the theory of Brownian motion. *Am. J. Phys.*, **65**, 1079–1081.

In the present case we cannot simplify the integral in Equation 8.5.8 and must apply numerical quadrature if we wish to have numerical values. □

In the introduction we showed that the solution to Langevin's equation:

$$dX(t) = cX(t)\,dt + \sigma\,dB(t), \qquad X(0) = X_0, \tag{8.5.9}$$

is

$$X(t) = X_0 + c\int_0^t X(\eta)\,d\eta + \sigma\int_0^t dB(\eta). \tag{8.5.10}$$

An obvious difficulty in understanding this solution is the presence of $X(s)$ in the first integral on the right side of Equation 8.5.21.

Let us approach its solution by considering the function $f(t,x) = e^{-ct}x$. Then, by Itô's lemma, Equation 8.4.16,

$$\begin{aligned} f[t,X(t)] - X(0) = &\int_0^t \left\{ f_t[\eta,X(\eta)] + cX(\eta)f_x[\eta,X(\eta)] + \tfrac{1}{2}\sigma^2 f_{xx}[\eta,X(\eta)] \right\} d\eta \\ &+ \int_0^t \sigma f_x[\eta,X(\eta)]\,dB(\eta), \end{aligned} \tag{8.5.11}$$

because $f[0,X(0)] = X(0)$. Direct substitution of $f(t,x)$ into Equation 8.5.11 yields

$$e^{-ct}X(t) - X_0 = \sigma\int_0^t e^{-c\eta}\,dB(\eta). \tag{8.5.12}$$

Finally, solving for $X(t)$, we obtain

$$X(t) = X(0)e^{ct} + \sigma e^{ct}\int_0^t e^{-c\eta}\,dB(\eta), \tag{8.5.13}$$

an explicit expression for $X(t)$. For the special case when X_0 is constant, $X(t)$ is known as an *Ornstein-Uhlenbeck process.*[12]

An alternative derivation begins by multiplying Equation 8.5.9 by the integrating factor e^{-ct} so that the equation now reads

$$e^{-ct}\,dX(t) - ce^{-ct}X(t)\,dt = \sigma e^{-ct}\,dB(t). \tag{8.5.14}$$

Running the product rule, Equation 8.4.23, backwards, we have that

$$d\big[e^{-ct}X(t)\big] = \sigma e^{-ct}\,dB(t). \tag{8.5.15}$$

Integrating both sides of Equation 8.5.15, we obtain Equation 8.5.12.

• Example 8.5.2: Exact stochastic differential equation

Consider the stochastic differential equation

$$X(t) = X(0) + c\int_0^t X(s)\,ds + \sigma\int_0^t X(s)\,dB(s), \tag{8.5.16}$$

[12] Uhlenbeck and Ornstein, op. cit.

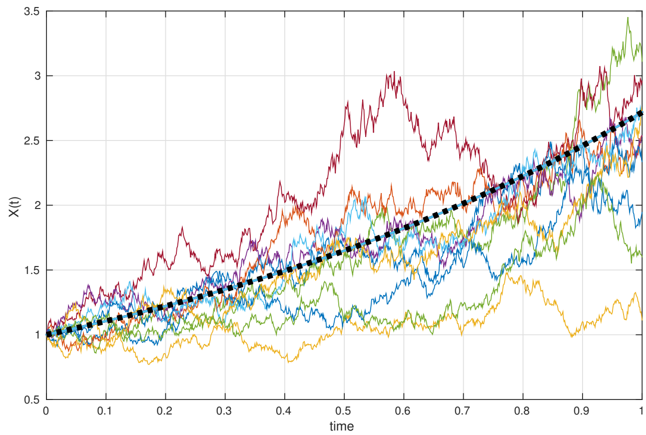


Figure 8.5.1: Ten realizations (sample paths) of geometric Brownian motion when $c = 0.1$, $\sigma = 0.5$, and $X(0) = 1$. The heavy line is the mean of $X(t)$.

with $c, \sigma > 0$.

If $X(t) = f[t, B(t)]$, then by Itô's lemma, Equation 8.4.9,

$$X(t) = X(0) + \int_0^t \left\{ f_t[s, B(s)] + \tfrac{1}{2} f_{xx}[s, B(s)] \right\} ds + \int_0^t f_x[s, B(s)]\, dB(s). \qquad \textbf{(8.5.17)}$$

Comparing Equation 8.5.16 and Equation 8.5.17, we find that

$$cf(t, x) = f_t(t, x) + \tfrac{1}{2} f_{xx}(t, x), \qquad \textbf{(8.5.18)}$$

and

$$\sigma f(t, x) = f_x(t, x). \qquad \textbf{(8.5.19)}$$

From Equation 8.5.19,

$$f_{xx}(t, x) = \sigma f_x(t, x) = \sigma^2 f(t, x). \qquad \textbf{(8.5.20)}$$

Therefore, Equation 8.5.18 can be replaced by

$$\left(c - \tfrac{1}{2}\sigma^2\right) f(t, x) = f_t(t, x). \qquad \textbf{(8.5.21)}$$

Equation 8.5.19 and Equation 8.5.21 can be solved using separation of variables, which yields

$$f(t, x) = f(0, 0) \exp\left[\left(c - \tfrac{1}{2}\sigma^2\right) t + \sigma x\right], \qquad \textbf{(8.5.22)}$$

or

$$X(t) = f[t, B(t)] = X(0) \exp\left[\left(c - \tfrac{1}{2}\sigma^2\right) t + \sigma B(t)\right]. \qquad \textbf{(8.5.23)}$$

Thus, a stochastic differential equation can sometimes be solved as the solution of a deterministic partial differential equation. In the present case, this solution is called *geometric Brownian motion*. For its solution numerically, see Example 8.6.1. See Figure 8.5.1. □

• Example 8.5.3: Homogeneous linear equation

Consider the homogeneous linear stochastic differential equation

$$dX(t) = c_1(t)X(t)\,dt + \sigma_1(t)X(t)\,dB(t). \tag{8.5.24}$$

Let us introduce $f(t,x) = \ln(x)$. Then by Itô's lemma, Equation 8.4.21,

$$df = d[\ln(X)] = \left[c_1(t) - \tfrac{1}{2}\sigma_1^2(t)\right]\,dt + \sigma_1(t)\,dB(t), \tag{8.5.25}$$

because $f_t = 0$, $f_x = 1/x$ and $f_{xx} = -1/x^2$. Integrating both sides of Equation 8.5.25 and exponentiating the resulting expression, we obtain

$$X(t) = X(0)\exp\left[\int_0^t \left[c_1(\eta) - \tfrac{1}{2}\sigma_1^2(\eta)\right]\,d\eta + \int_0^t \sigma_1(\eta)\,dB(\eta)\right\}. \tag{8.5.26}$$

□

• Example 8.5.4: General case

Consider the homogeneous linear stochastic differential equation

$$dX(t) = [c_1(t)X(t) + c_2(t)]\,dt + [\sigma_1(t)X(t) + \sigma_2(t)]\,dB(t). \tag{8.5.27}$$

Our analysis begins by considering the homogeneous linear stochastic differential equation

$$dY(t) = c_1(t)Y(t)\,dt + \sigma_1(t)Y(t)\,dB(t), \qquad Y(0) = 1. \tag{8.5.28}$$

From the previous example,

$$Y(t) = \exp\left\{\int_0^t \left[c_1(\eta) - \tfrac{1}{2}\sigma_1^2(\eta)\right]\,d\eta + \int_0^t \sigma_1(\eta)\,dB(\eta)\right\}. \tag{8.5.29}$$

Next, let us introduce two random variables, $X_1 = 1/Y$ and $X_2 = X$. Using Itô lemma $f(t,x) = 1/x$, then

$$dX_1 = df(t,Y) = d\left(\frac{1}{Y}\right) = \left[\sigma_1^2(t) - c_1(t)\right]\,\frac{dt}{Y} - \sigma_1(t)\frac{dB(t)}{Y} \tag{8.5.30}$$

$$= \left[\sigma_1^2(t) - c_1(t)\right]\,X_1(t)\,dt - \sigma_1(t)X_1(t)\,dB(t), \tag{8.5.31}$$

since $f_t = 0$, $f_x = -1/x^2$ and $f_{xx} = 2/x^3$.

Using Equation 8.4.30, where X_1 is governed by Equation 8.5.31 and X_2 is governed by 8.5.27 because $X_2 = X$,

$$d(X_1X_2) = [c_2(t) - \sigma_1(t)\sigma_2(t)]\,X_1(t)\,dt + \sigma_2(t)X_1(t)\,dB(t). \tag{8.5.32}$$

Upon integrating both sides of Equation 8.5.32, we have

$$X_1X_2 - X_1(0) = \int_0^t [c_2(\eta) - \sigma_1(\eta)\sigma_2(\eta)]\,\frac{d\eta}{Y(\eta)} + \int_0^t \sigma_2(\eta)\,\frac{dB(\eta)}{Y(\eta)}. \tag{8.5.33}$$

Consequently, our final result is

$$X(t) = Y(t) \left\{ X(0) + \int_0^t [c_2(\eta) - \sigma_1(\eta)\sigma_2(\eta)] \, \frac{d\eta}{Y(\eta)} + \int_0^t \sigma_2(\eta) \, \frac{dB(\eta)}{Y(\eta)} \right\}, \qquad \mathbf{(8.5.34)}$$

where $Y(t)$ is given by Equation 8.5.29. □

• **Example 8.5.5: Stochastic Verhulst equation**

The stochastic Verhulst equation is

$$dX(t) = aX(t)[M - X(t)]\,dt + bX(t)\,dB(t), \qquad X(0) = X_0. \qquad \mathbf{(8.5.35)}$$

We begin its solution by introducing $\Phi(t) = 1/X(t)$. Then by Itô's lemma, Equation 8.4.21 with $f(x) = 1/x$,

$$d\Phi(t) = -\Phi(t)[(aM - b^2)\,dt + b\,dB(t)] + a\,dt, \qquad \Phi(0) = 1/X_0. \qquad \mathbf{(8.5.36)}$$

To solve Equation 8.5.36, we use the results from Example 8.5.4 with $c_1(t) = b^2 - aM$, $c_2(t) = a$, $\sigma_1(t) = -b$, and $\sigma_2(t) = 0$. Denoting $\epsilon(t) = (aM - b^2/2)t + bB(t)$, we can write Equation 8.5.34 as

$$\Phi(t)e^{\xi(t)} - \Phi(0) = a \int_0^t e^{\xi(\eta)}\,d\eta, \qquad \mathbf{(8.5.37)}$$

or

$$\frac{e^{\xi(t)}}{X(t)} - \frac{1}{X_0} = a \int_0^t e^{\xi(\eta)}\,d\eta. \qquad \mathbf{(8.5.38)}$$

Solving for $X(t)$, we obtain the final result that

$$X(t) = \frac{X_0 \exp[\xi(t)]}{1 + aX_0 \int_0^t \exp[\xi(\eta)]\,d\eta}. \qquad \mathbf{(8.5.39)}$$

Problems

1. Solve the stochastic differential equation

$$dX(t) = \tfrac{1}{2}e^{t/2}B(t)\,dt + e^{t/2}\,dB(t), \qquad X(0) = X_0,$$

by running the product rule backwards.

2. Solve the stochastic differential equation

$$dX(t) = e^{2t}[1 + 2B^2(t)]\,dt + 2e^{2t}B(t)\,dB(t), \qquad X(0) = X_0,$$

by running the product rule backwards. Hint: Rewrite the differential equation $dX(t) = e^{2t}[2B(t)\,dB(t) + dt] + (2e^{2t}\,dt)B^2(t)$.

3. Solve the stochastic differential equation

$$dX(t) = [1 + B(t)]\,dt + [t + 2B(t)]\,dB(t), \qquad X(0) = X_0,$$

by running the product rule backwards. Hint: Rewrite the differential equation $dX(t) = 2B(t)\,dB(t) + dt + B(t)\,dt + t\,dB(t)$.

4. Solve the stochastic differential equation

$$dX(t) = [3t^2 + B(t)]\,dt + t\,dB(t), \qquad X(0) = X_0,$$

by running the product rule backwards. Hint: Rewrite the differential equation $dX(t) = 3t^2\,dt + [B(t)\,dt + t\,dB(t)]$.

5. Solve the stochastic differential equation

$$dX(t) = B^2(t)\,dt + 2tB(t)\,dB(t), \qquad X(0) = X_0,$$

by running the product rule backwards. Hint: Rewrite the differential equation $dX(t) = t[2B(t)\,dB(t) + dt] + B^2(t)\,dt - t\,dt$.

6. Find the integrating factor and solution to the stochastic differential equation

$$dX(t) = [\beta - \alpha X(t)]\,dt + \sigma\,dB(t), \qquad X(0) = X_0,$$

where $B(t)$ is Brownian motion and α, β and σ are constants.

7. Find the integrating factor and solution to the stochastic differential equation

$$dX(t) = [1 + 2X(t)]\,dt + e^{2t}\,dB(t), \qquad X(0) = X_0,$$

where $B(t)$ is Brownian motion.

8. Find the integrating factor and solution to the stochastic differential equation

$$dQ(t) + \frac{Q(t)}{RC}\,dt = \frac{V(t)}{R}\,dt + \frac{\alpha(t)}{R}\,dB(t), \qquad Q(0) = Q_0,$$

where R and C are real, positive constants, and $B(t)$ is Brownian motion.

9. Find the integrating factor and solution to the stochastic differential equation[13]

$$dX(t) = 2tX(t)\,dt + e^{-t}\,dt + dB(t), \qquad t \in [0, 1],$$

with $X(0) = X_0$, and $B(t)$ is Brownian motion.

10. Find the integration factor and solution to the stochastic differential equation

$$dX(t) = [4X(t) - 1]\,dt + 2\,dB(t), \qquad X(0) = X_0,$$

where $B(t)$ is Brownian motion.

---

[13] Khodabin, M., and M. Rostami, 2015: Mean square numerical solution of stochastic differential equations by fourth order Runge-Kutta method and its applications in the electric circuits with noise. *Adv. Diff. Eq.*, **2015**, 62.

11. Find the integration factor and solution to the stochastic differential equation

$$dX(t) = [2 - X(t)]\,dt + e^{-t}B(t)\,dB(t), \qquad X(0) = X_0,$$

where $B(t)$ is Brownian motion.

12. Find the integration factor and solution to the stochastic differential equation

$$dX(t) = [1 + X(t)]\,dt + e^{t}B(t)\,dB(t), \qquad X(0) = X_0,$$

where $B(t)$ is Brownian motion.

13. Find the integration factor and solution to the stochastic differential equation

$$dX(t) = \left[\tfrac{1}{2}X(t) + 1\right]\,dt + e^{t}\cos[B(t)]\,dB(t), \qquad X(0) = X_0,$$

where $B(t)$ is Brownian motion.

14. Find the integration factor and solution to the stochastic differential equation

$$dX(t) = \left[t + \tfrac{1}{2}X(t)\right]\,dt + e^{t}\sin[B(t)]\,dB(t), \qquad X(0) = X_0,$$

where $B(t)$ is Brownian motion.

15. Following Example 8.5.2, solve the exact stochastic differential equation:

$$dX(t) = e^{t}\left[1 + B^2(t)\right]\,dt + \left[1 + 2e^{t}B(t)\right]\,dB(t), \qquad X(0) = X_0.$$

Step 1: Show that $f_t + \frac{1}{2}f_{xx} = e^t(1 + x^2)$, and $f_x = 1 + 2e^t x$.

Step 2: Show that $f(t, x) = x + e^t x^2 + g(t)$.

Step 3: Show that $g(t) = X_0$ and $X(t) = B(t) + e^t B^2(t) + X_0$.

16. Following Example 8.5.2, solve the exact stochastic differential equation:

$$dX(t) = \left\{2tB^2(t) + 3t^2\left[1 + B(t)\right]\right\}\,dt + \left[1 + 3t^2B^2(t)\right]\,dB, \qquad X(0) = X_0.$$

Step 1: Show that $f_t + \frac{1}{2}f_{xx} = 2tx^3 + 3t^2(1 + x)$, and $f_x = 3t^2x^2 + 1$.

Step 2: Show that $f(t, x) = t^2x^3 + x + g(t)$.

Step 3: Show that $g'(t) = 3t^2$.

Step 4: Show that $X(t) = t^2[B^3(t) + t] + B(t) + X_0$.

Using Equation 8.5.26, solve the following stochastic differential equations:

17. $dX(t) = t^2X(t)\,dt + tX(t)\,dB(t), \qquad X(0) = X_0$

18. $dX(t) = \cos(t)X(t)\,dt + \sin(t)X(t)\,dB(t), \qquad X(0) = X_0$

19. $dX(t) = \ln(t+1)X(t)\,dt + \sqrt{\ln(t+1)}\,X(t)\,dB(t), \qquad X(0) = X_0$

20. $dX(t) = \ln(t+1)X(t)\,dt + tX(t)\,dB(t), \qquad X(0) = X_0$

21. Following Example 8.5.5, solve the stochastic differential equation

$$dX(t) = [aX^n(t) + bX(t)]\,dt + cX(t)\,dB(t), \qquad X(0) = X_0,$$

where $n > 1$.

Step 1: Setting $\Phi(t) = X^{1-n}(t)$, use Itô's lemma Equation 8.4.21 with $f(x) = 1/x^{n-1}$ to show that

$$d\Phi(t) = (1-n)\Phi(t)\left[\left(b - \tfrac{1}{2}nc^2\right)\,dt + c\,dB(t)\right] + (1-n)a\,dt.$$

Step 2: Setting $c_1(t) = (1-n)b - n(1-n)c^2/2$, $c_2(t) = (1-n)a$, $\sigma_1(t) = (1-n)c$, and $\sigma_2(t) = 0$, show that

$$\frac{\exp[(n-1)\xi(t)]}{X^{n-1}(t)} - \frac{1}{X_0^{n-1}} = (1-n)a\int_0^t \exp[(n-1)\xi(\eta)]\,d\eta,$$

or

$$\frac{\exp[(n-1)\xi(t)]}{X^{n-1}(t)} = \frac{1}{X_0^{n-1}} + (1-n)a\int_0^t \exp[(n-1)\xi(\eta)]\,d\eta,$$

where $\xi(t) = (b - c^2/2)t + cB(t)$.

22. Following Example 8.5.5, solve the stochastic Ginzburg-Landau equation:

$$dX(t) = \left[ae^{cX(t)} + b\right]\,dt + \sigma\,dB(t), \qquad X(0) = X_0.$$

Step 1: Setting $\Phi(t) = \exp[-cX(t)]$, use Itô's lemma Equation 8.4.21 with $f(x) = e^{-cx}$ to show that

$$d\Phi(t) = -\left(bc - \tfrac{1}{2}\sigma^2c^2\right)\Phi(t)\,dt - \sigma c\Phi(t)\,dB(t) - ac\,dt.$$

Step 2: Setting $c_1(t) = \sigma^2c^2/2 - bc$, $c_2(t) = -ac$, $\sigma_1(t) = -\sigma c$, and $\sigma_2(t) = 0$, show that

$$X(t) = X_0 + bt + \sigma B(t) - \frac{1}{c}\ln\left\{1 - ac\int_0^t \exp[cX_0 + bc\xi + \sigma cB(\xi)]\,d\xi\right\}.$$

23. Following Example 8.5.5, solve the stochastic differential equation:

$$dX(t) = \left\{[1 + X(t)][1 + X^2(t)]\right]\,dt + [1 + X^2(t)]\,dB(t), \qquad X(0) = X_0.$$

Step 1: Setting $\Phi(t) = \tan^{-1}[X(t)]$, use Itô's lemma Equation 8.4.21 with $f(x) = \tan^{-1}(x)$ to show that $d\Phi(t) = dt + dB(t)$.

Step 2: Solving the stochastic differential equation in Step 1, show that

$$X(t) = \tan[\tan^{-1}(X_0) + t + B(t)].$$

8.6 NUMERICAL SOLUTION OF STOCHASTIC DIFFERENTIAL EQUATIONS

In this section we construct numerical schemes for integrating the stochastic differential equation

$$dX(t) = a[X(t),t]\,dt + b[X(t),t]\,dB(t) \tag{8.6.1}$$

on $t_0 \le t \le T$ with the initial-value $X(t_0) = X_0$.

Our derivation begins by introducing the grid $t_0 < t_1 < t_2 < \cdots < t_n < \cdots < t_N = T$. For simplicity we assume that all of the time increments are the same and equal to $0 < \Delta t < 1$ although our results can be easily generalized when this is not true. Now

$$X_{n+1} = X_n + \int_{t_n}^{t_{n+1}} a[X(\eta),\eta]\,d\eta + \int_{t_n}^{t_{n+1}} b[X(\eta),\eta]\,dB(\eta). \tag{8.6.2}$$

The crudest approximation to the integrals in Equation 8.6.2 is

$$\int_{t_n}^{t_{n+1}} a[X(\eta),\eta]\,d\eta \approx a[X(t_n),t_n]\Delta t_n, \tag{8.6.3}$$

and

$$\int_{t_n}^{t_{n+1}} b[X(\eta),\eta]\,dB(\eta) \approx b[X(t_n),t_n]\Delta B_n. \tag{8.6.4}$$

Substituting these approximations into Equation 8.6.2 yields the *Euler-Marugama approximation*.[14] For the Itô process $X(t) = \{X(t), t_0 \le t \le T\}$:

$$X_{n+1} = X_n + a(t_n, X_n)\,(t_{n+1} - t_n) + b(t_n, X_n)\left(B_{t_{n+1}} - B_{t_n}\right) \tag{8.6.5}$$

for $n = 0, 1, 2, \ldots, N-1$ with the initial value X_0.

When $b = 0$, the stochastic iterative scheme reduces to the conventional Euler scheme for ordinary differential equations. When $b \ne 0$, we have an extra term generated by the random increment $\Delta B_n = B(t_{n+1}) - B(t_n)$ where $n = 0, 1, 2, \ldots, N-1$ for Brownian motion (the Wiener process) $B(t) = B(t), t \ge 0$. Because these increments are independent Gaussian random variables, the mean equals $E(\Delta B_n) = 0$ while the variance is $E[(\Delta B_n)^2] = \Delta t$. We can generate ΔB_n using the MATLAB function `randn`.

An important consideration in the use of any numerical scheme is the rate of convergence. During the numerical simulation of a realization, at time t there will be a difference between the exact solution $X(t)$ and the numerical approximation $Y(t)$. This difference $e(t) = X(t) - Y(t)$ will also be a random variable. A stochastic differential equation scheme *converges strongly with order m*, if for any time t, $E(|e(t)|) = O[(\Delta t)^m]$ for sufficiently small time step Δt. The strong order for the Euler-Marugama method can be proven to be $\frac{1}{2}$.

To construct a strong order 1 approximation to Equation 8.6.1, we return to Equation 8.6.2. Using Equation 8.4.12, we have

$$\begin{aligned} X_{n+1} - X_n &= \int_{t_n}^{t_{n+1}} \left[a[X_n(\eta),\eta] + \int_{t_n}^{\eta} \left(aa_x + \tfrac{1}{2}b^2 a_{xx}\right)\,d\xi + \int_{t_n}^{\eta} ba_x\,dB(\xi)\right]\,d\eta \\ &\quad + \int_{t_n}^{t_{n+1}} \left[b[X_n(\eta),\eta] + \int_{t_n}^{\eta} \left(ab_x + \tfrac{1}{2}b^2 b_{xx}\right)\,d\xi + \int_{t_n}^{\eta} bb_x\,dB(\xi)\right]\,d\eta \qquad (8.6.6) \\ &= a[X(t_n),t_n]\Delta t + b[X(t_n),t_n)\Delta B_n + R_n, \qquad (8.6.7) \end{aligned}$$

---

[14] Maruyama, G., 1955: Continuous Markov processes and stochastic equations. *Rend. Circ. Math. Palermo, Ser. 2,*, **4**, 48–90.

where

$$R_n = \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{\eta} bb_x \, dB(\xi) \right] dB(\eta) + \text{higher-order terms.} \tag{8.6.8}$$

Dropping the higher-order terms,

$$R_n \approx b[X(t_n), t_n] b_x[X(t_n), t_n] \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{\eta} dB(\xi) \right] dB(\eta). \tag{8.6.9}$$

Consider now the double integrals

$$(\Delta B_n)^2 = \left(\int_{t_n}^{t_{n+1}} dB(\eta) \right) \left(\int_{t_n}^{t_{n+1}} dB(\eta) \right) = \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t_{n+1}} dB(\xi) \right] dB(\eta). \tag{8.6.10}$$

Now,

$$\int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{t_{n+1}} dB(\xi) \right] dB(\eta) = \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{\eta} dB(\xi) \right] dB(\eta) + \int_{t_n}^{t_{n+1}} \left[\int_{\eta}^{t_{n+1}} dB(\xi) \right] dB(\eta)$$
$$+ \int_{t_n}^{t_{n+1}} [dB(\eta)]^2 \tag{8.6.11}$$
$$= 2 \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{\eta} dB(\xi) \right] dB(\eta) + \int_{t_n}^{t_{n+1}} [dB(\eta)]^2 \tag{8.6.12}$$
$$= 2 \int_{t_n}^{t_{n+1}} \left[\int_{t_n}^{\eta} dB(\xi) \right] dB(\eta) + \Delta t, \tag{8.6.13}$$

because

$$\int_{t_n}^{t_{n+1}} [dB(\eta)]^2 = \int_{t_n}^{t_{n+1}} d\eta = \Delta t. \tag{8.6.14}$$

Combining Equation 8.6.9, Equation 8.6.10, and Equation 8.6.13 yields

$$R_n \approx b[X(t_n), t_n] b_x[X(t_n), t_n] \left[(\Delta B_n)^2 - \Delta t \right]. \tag{8.6.15}$$

Finally, substituting Equation 8.6.15 into Equation 8.6.7 gives the final result, the Milstein method:[15]

$$X_{n+1} = X_n + a(X_n, t_n)\, \Delta t_n + b(X_n, t_n)\, \Delta B_n + \tfrac{1}{2} b(X_n, t_n) \frac{\partial b(X_n, t_n)}{\partial x} \left[(\Delta B_n)^2 - \Delta t \right]. \tag{8.6.16}$$

• Example 8.6.1

Consider the Itô process $X(t)$ defined by the linear stochastic differential equation

$$dX(t) = aX(t)\, dt + bX(t)\, dB(t), \tag{8.6.17}$$

---

[15] Milstein, G., 1974: Approximate integration of stochastic differential equations. *Theory Prob. Applic.*, **19**, 557–562.

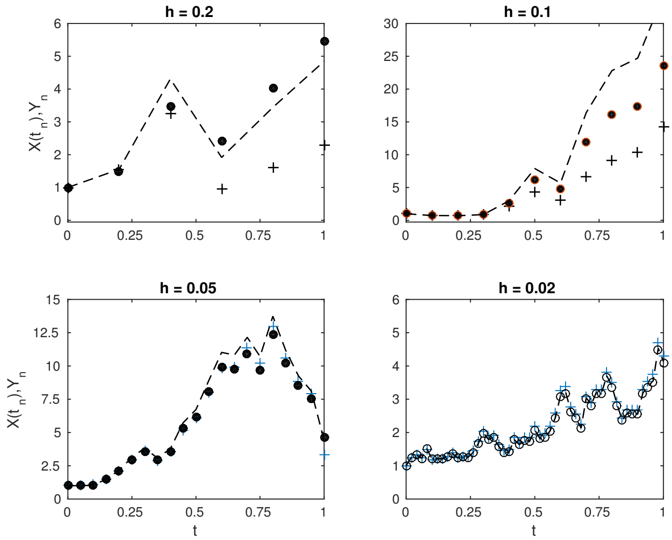


Figure 8.6.1: The numerical solution of the stochastic differential equation, Equation 8.6.17, using the Euler-Marugama (crosses) and the Milstein (circles) methods for various time steps h. The dashed line gives the exact solution.

for $t \in [0,T]$. If this Itô process has the drift $a(x,t) = ax$ and the diffusion coefficient $b(x,t) = bx$, the exact solution (see Equation 8.5.16) is

$$X(t) = X_0 \exp\left[\left(a - \frac{b^2}{2}\right) t + bB(t)\right] \tag{8.6.18}$$

for $t \in [0,T]$. Figure 8.6.1 compares the numerical solution of this stochastic differential equation using the Euler-Marugama and Milstein method against the exact solution. Note that each frame has a different solution because the Brownian forcing changes with each realization. □

Although a plot of various realizations can give an idea of how the stochastic processes affect the solution, two more useful parameters are the sample mean and standard deviation at time t_n:

$$\overline{X}(t_n) = \frac{1}{J}\sum_{j=1}^{J} X_j(t_n), \tag{8.6.19}$$

and

$$\sigma^2(t_n) = \frac{1}{J-1}\sum_{j=1}^{J}\left[X_j(t_n) - \overline{X}(t_n)\right]^2, \tag{8.6.20}$$

where J are the number of realizations and $X_j(t_n)$ is the value of the random variable at time t_n of the jth realization.

In many physical problems, "noise" is the origin of the stochastic process and we suspect that we have a normal distribution $N(\mu, \sigma^2)$ where μ and σ are the population mean and standard deviation, respectively. Then, using the sample statistics, Equations 8.6.20 and

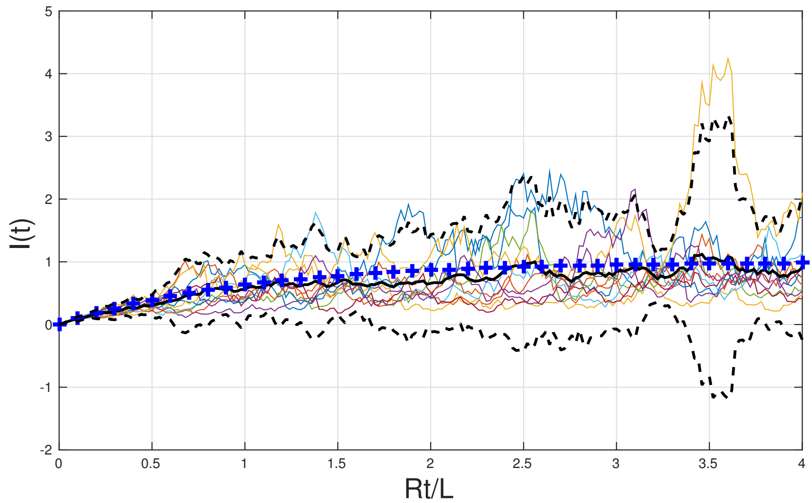


Figure 8.6.2: Eleven realizations as a function of the nondimensional time Rt/L of the numerical solution of Equation 8.6.24 using the Euler-Marugama method when $h = 0.02$, $\alpha/L = 1$, $\beta/L = 0$, $I_0 = 0$, and $v(t) = R$. The mean and 95% confidence interval (here $t_{student} = 2.228$) are given by the heavy solid and dashed lines, respectively. Finally, the crosses (+) give the deterministic solution.

8.6.21, a two-sided confidence interval can be determined as

$$\left[\overline{X}(t_n) - \tau_{student}\frac{\sigma(t_n)}{\sqrt{J}}, \overline{X}(t_n) + \tau_{student}\frac{\sigma(t_n)}{\sqrt{J}}\right]$$

based on the student-τ distribution with $J-1$ degrees of freedom.

Project: RL Electrical Circuit with Noise

An important component of contemporary modeling is the mixture of deterministic and stochastic aspects of a physical system. In this project you will see how this is done using a simple electrical system.[16]

Consider a simple electrical circuit consisting of a resistor with resistance R and an inductor with inductance L. If the circuit is driven by a voltage source $v(t)$, the current $I(t)$ at a given time t is given by the first-order ordinary differential equation

$$L\frac{dI}{dt} + RI = v(t), \qquad I(0) = I_0. \tag{8.6.21}$$

Step 1: Using classical methods, show that the deterministic solution to Equation 8.6.21 is

$$I(t) = I_0 e^{-Rt/L} + \frac{1}{L}\int_0^t \exp\left[\frac{R}{L}(\tau - t)\right] v(\tau)\, d\tau. \tag{8.6.22}$$

[16] See Kolářová, E., 2005: Modeling RL electrical circuits by stochastic differential equations. *Proc. Int. Conf. Computers as Tool*, Belgrade (Serbia and Montenegro), **IEEE R8**, 1236–1238.

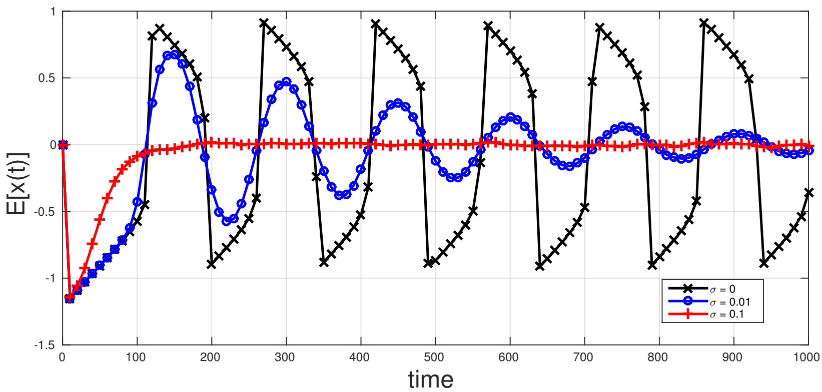


Figure 8.6.3: Plot of $E[x(t)]$ versus time for the FitzHugh-Nagamo model for three values of σ. The value of the parameters are $a = 0.8$, $m = 1.2$, and $\tau = 100$. The Euler method was used with a time step of 0.1.

There are two possible ways that randomness can enter this problem. First, the power supply could introduce some randomness so that the right side of Equation 8.6.21 could read $v(t) + \alpha\, dB_2(t)/dt$. Second, some physical process within the resistor could cause randomness so that the resistance would now equal $R + \beta\, dB_1(t)/dt$. Here $B_1(t)$ and $B_2(t)$ denote two independent white noise processes and α, β are nonnegative constants. In this case the governing differential equation would now read

$$\frac{dI}{dt} + \frac{1}{L}\left[R + \alpha\frac{dB_1}{dt}\right] = \frac{1}{L}\left[v(t) + \beta\frac{dB_2}{dt}\right], \qquad I(0) = I_0. \tag{8.6.23}$$

Converting Equation 8.6.23 into the standard form of a stochastic ordinary differential equation, we have that

$$dI = \frac{1}{L}\left[v(t) - RI(t)\right] dt - \frac{\alpha}{L} I(t)\, dB_1(t) + \frac{\beta}{L}\, dB_2(t), \qquad I(0) = I_0. \tag{8.6.24}$$

Step 2: Using MATLAB, create a script to numerically integrate Equation 8.6.24 for a given set of α, β, $I_0 = 0$, R, L, and $v(t)$. Plot $I(t)$ as a function of the nondimensional time Rt/L for many realizations (say 20). See Figure 8.6.2.

Step 3: Although some idea of the effect of randomness is achieved by plotting several realizations, a better way would be to compute the mean and standard deviation at a given time. On the plot from the previous step, plot the mean and standard deviation of your solution as a function of nondimensional time. How does it compare to the deterministic solution?

Project: Relaxation Oscillator with Brownian Motion Forcing

The FitzHugh-Nagamo[17] model describes excitable systems such as a neuron. We will modify it so that the forcing is due to Brown motion. The governing equations are

$$dx = -x(x^2 - a^2)\, dt - y\, dt + \sigma\, dB_1(t),$$

[17] FitzHugh, R., 1961: Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.*, **1**, 445-466; Nagumo, J., S. Arimoto, and S. Yoshizawa, 1962: An active pulse transmission line simulating nerve axon. *Proc. IRE*, **50**, 2061–2070.

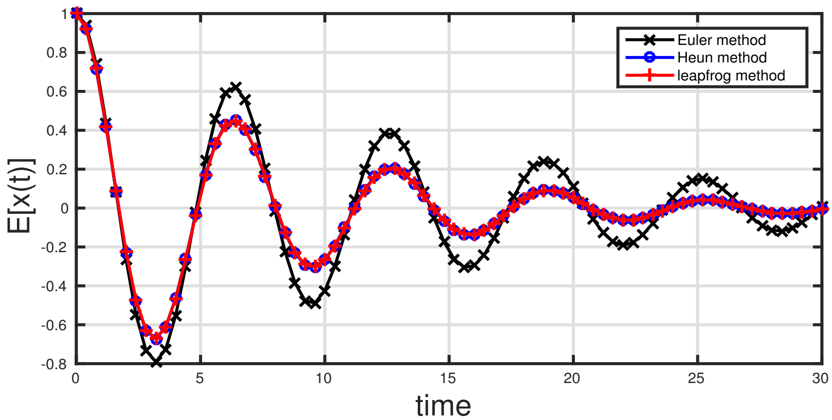


Figure 8.6.4: Plot of $E[x(t)]$ versus time for the damped harmonic oscillator forced by Brownian motion. The value of the parameters are $k = 1$, $\gamma = 0.25$, and $alpha = \Delta t = 0.1$. Five thousand realization were performed.

and

$$dy = (x - my)\, dt/\tau + \sigma\, dB_2(t),$$

where a, m, σ, and τ are parameters.

Write a MATLAB script to numerically integrate this modified FitzHugh-Nagamo model for various values of σ. Using many simulations, compute $E[x(t)]$ as a function of time t. See Figure 8.6.3. What is the effect of the Brownian motion forcing?

Project: Stochastically Damped Harmonic Oscillator

The damped stochastic harmonic oscillator is governed by the stochastic differential equations:

$$dv(t) = -\gamma v(t)\, dt - k^2 x(t)\, dt - \alpha x(t)\, dB(t), \qquad \text{and} \qquad dx(t) = v(t)\, dt,$$

where k, α and γ are real constants. This system of equations is of interest for two reasons: (1) The system is forced by Brownian motion. (2) The noise is multiplicative rather than additive because the forcing term is $x(t)\, dB(t)$ rather than just $dB(t)$.

We could solve both equations numerically using Euler's method.[18] The purpose of this project is to introduce you to the Heun method. In the Heun method we first compute an estimate of the solution x^* and v^* by taking a Euler-like time step:

$$x^* = x_i + v_i \Delta t, \quad \text{and} \quad v^* = v_i - \gamma v_i \Delta t - k^2 x_i \Delta t - \alpha x_i \Delta B_i,$$

where x_i and v_i denote the displacement and velocity at time $t_i = i\Delta t$, Δt is the time step, and $i = 0, 1, 2, \ldots$. With these estimates we compute the value for x_{i+1} and v_{i+1} using

$$x_{i+1} = x_i + \tfrac{1}{2}(v_i + v^*)\Delta t, \quad \text{and} \quad v_{i+1} = v_i - \tfrac{1}{2}\gamma(v_i + v^*)\Delta t - \tfrac{1}{2}k^2(x_i + x^*)\Delta t - \alpha x_i \Delta B_i.$$

---

[18] For further details, see Greiner, A., W. Strittmatter, and J. Honerkamp, 1988: Numerical integration of stochastic differential equations. *J. Stat. Phys.*, **51**, 95–108.

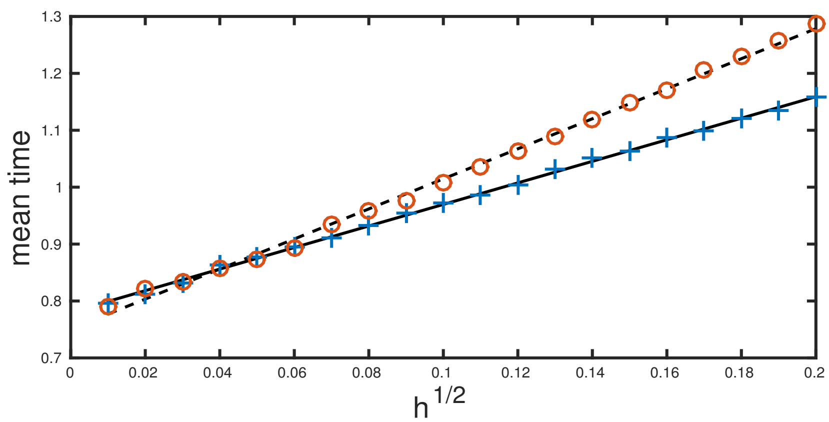


Figure 8.6.5: The mean time that it takes a particle to travel from $X(0) = -1$ to $X = 0$ in the double-well potential stated in the text. Sixty thousand realizations were used with a time step h. Two differential numerical schemes were used: the Euler-Marugama (crosses) and the Milstein (circles) methods. The curves are linear least-squares fits through the results.

Qiang and Habib[19] developed a leapfrog algorithm to solve this problem. Because the algorithm is rather complicated, the interested student is referred to their paper.

Write a MATLAB script to use the Euler and Huen methods to numerically integrate the stochastic harmonic oscillator when $10\alpha = 4\gamma = k = 1$ and $x(0) = v(0) = 0$. Using many simulations, compute $E[x(t)]$ as a function of time t. See Figure 8.6.4. What happens to the accuracy of the solution for larger values of Δt?

Project: Mean First Passage Time

The stochastic differential equation

$$dX(t) = [X(t) - X^3(t)]\,dt + \frac{4}{X^2(t)+1}\,dB(t)$$

describes the motion of a particle in a double-well potential $V(x) = x^4/4 - x^2/2$, subject to a spatially dependent random forcing when the acceleration $X''(t)$ can be neglected. An important question is what is the average (mean) time that it takes a particle initially located at a minimum $X(0) = -1$ to reach the local maximum $X(t) = 0$.

Write MATLAB code that computes $X(t)$ as a function of time t. Using this code and creating N realizations, compute the length of time that it takes the particle to reach $X(t) = 0$ in each realization. Then compute the mean from those times and plot the results as a function $\sqrt{h}$, the square root of the time step. See Figure 8.6.5. We used $\sqrt{h}$ rather than h following the suggestions of Seeßelberg and Petruccione.[20]

---

[19] Qiang, J., and S. Habib, 2000: Second-order stochastic leapfrog algorithm for multiplicative noise Brownian motion. *Phys. Review*, **62**, 7430–7437.

[20] Seeßelberg, M., and F. Petruccione, 1993: An improved algorithm for the estimation of the mean first passage of ordinary stochastic differential equations. *Comput. Phys. Commun.*, **74**, 247–255.

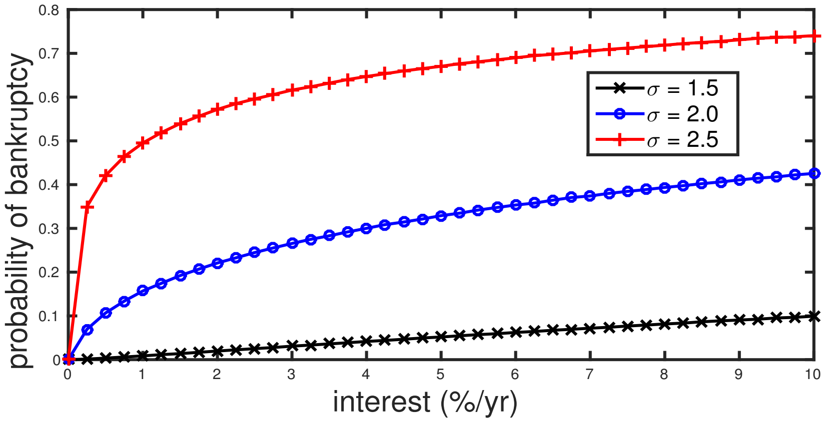


Figure 8.6.6: The probability of bankruptcy over a three-year period as a function of interest rate of a firm with initial wealth $X_0 = 500$ and debt $D = 100$. Other parameters are $h = 0.01$ yr and $\mu = 1.001$/year. The units on σ is year$^{-1/2}$. Five hundred thousand realizations were used to compute the probability.

Project: Bankruptcy of a Company

The stochastic differential equation[21]

$$dX(t) = [\mu X(t) - iD]\,dt + \sigma X(t)\,dB(t), \qquad 0 < t < T,$$

with $X(0) = X_0$, describes the evolution with time t of the wealth $X(t)$ of a firm. Here μ and σ denote the deterministic and stochastic evolution of the firm's wealth, respectively, X_0 is the initial wealth of the firm, and iD gives the amount of payment to a financier (bank) who initially loaned the firm the amount D at the interested rate i. Write a MATLAB code to simulate the wealth of a firm during its lifetime T given a known D, i and X_0 with $\mu = 1.001$/year, and various values of σ.

During the simulation there is a chance that the firm goes bankrupt at time $t = \tau < T$. This occurs when the stochastic process hits the barrier $X(\tau) = 0$. If n denotes the number of times that bankruptcy occurs in N simulations, the probability of bankruptcy is $P[X(\tau) = 0] = n/N$. Using your code for simulating a firm's wealth, compute the probability of bankruptcy as a function of interest rate for a small ($D = 20$, $X_0 = 100$), medium ($D = 100$, $X_0 = 500$), and large ($D = 200$, $X_0 = 1000$) firm. See Figure 8.6.6. How does the average value of τ vary with interest rate?

Further Readings

Kloeden, P. E., and E. Platen, 1992: *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, 632 pp. A solid book covering numerical schemes for solving stochastic differential equations.

Mikosch, T., 1998: *Elementary Stochastic Calculus with Finance in View*. World Scientific, 212 pp. Very well-crafted book on stochastic calculus.

[21] See Cerqueti, R., and A. G. Quaranta, 2012: The perspective of a bank in granting credit: An optimization model. *Optim. Lett.*, **6**, 867–882.

Answers to the Odd-Numbered Problems

Section 1.1

1. $1+2i$ 3. $-2/5$ 5. $2+2i\sqrt{3}$ 7. $z=e^{3\pi i/2}$ 9. $z=4e^{\pi i/3}$ 11. $z=2\sqrt{2}e^{7\pi i/4}$

Section 1.2

1. $\pm\sqrt{2}, \quad \pm\sqrt{2}\left(\frac{1}{2}+\frac{\sqrt{3}i}{2}\right), \quad \pm\sqrt{2}\left(-\frac{1}{2}+\frac{\sqrt{3}i}{2}\right)$ 3. $i, \quad \pm\frac{\sqrt{3}}{2}-\frac{i}{2}$

5. $\pm\frac{1}{\sqrt{2}}\left(-\sqrt{\sqrt{a^2+b^2}+a}+i\sqrt{\sqrt{a^2+b^2}+a}\right)$ 7. $\pm(1+i), \quad \pm 2(1-i)$

Section 1.3

1. $u=2-y,\ v=x$ 3. $u=x^3-3xy^2,\ v=3x^2y-y^3$

5. $f'(z)=3z(1+z^2)^{1/2}$ 7. $f'(z)=2(1+4i)z-3$

9. $f'(z)=-3i(iz-1)^{-4}$ 11. $-1/4$

13. $(-1)^n/\pi$ 15. $v(x,y)=2xy+\text{constant}$

17. $v(x,y)=x\sin(x)e^{-y}+ye^{-y}\cos(x)+\text{constant}$

Section 1.4

1. 0 3. $2i$ 5. $14/15-i/3$

Section 1.5

1. $(e^{-2} - e^{-4})/2$ 3. $\pi/2$ 5. $17/6 + 52i/3$ 7. $-\sinh(1)i/3$

Section 1.6

1. $\pi i/32$ 3. $\pi i/2$ 5. $-2\pi i$ 7. $2\pi i$ 9. -6π 11. $2\pi i/3$

Section 1.7

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

3. $f(z) = z^{10} - z^9 + \dfrac{z^8}{2} - \dfrac{z^7}{6} + \cdots - \dfrac{1}{11!z} + \cdots$
We have an essential singularity and the residue equals $-1/11!$

5. $f(z) = \dfrac{1}{2!} + \dfrac{z^2}{4!} + \dfrac{z^4}{6!} + \cdots$
We have a removable singularity where the value of the residue equals zero.

7. $f(z) = -\dfrac{2}{z} - 2 - \dfrac{7z}{6} - \dfrac{z^2}{2} - \cdots$
We have a simple pole and the residue equals -2.

9. $f(z) = \dfrac{1}{2}\dfrac{1}{z-2} - \dfrac{1}{4} + \dfrac{z-2}{8} - \cdots$
We have a simple pole and the residue equals $1/2$.

Section 1.8

1. $-3\pi i/4$ 3. $-2\pi i$.   5. $2\pi i$ 7. $2\pi i$ 9. $-2i$

Section 1.11

3. $z = C\sqrt{\tau} - \pi + \pi\, i$ 5. $z = \tau^{7/4}$ or $\tau = z^{4/7}$

7. $z = a\, \cosh^{-1}(\tau)/\pi, \quad 0 \le \Im[\cosh^{-1}(\tau)] \le \pi$

Section 2.1

1. $f(t) = e^{-a|t|}/(2a)$ 3. $f(t) = ite^{-a|t|}/(4a)$ 5. $f(t) = -2e^{-3t/2}\sin\left(\sqrt{3}\,t/2\right)H(t)/\sqrt{3}$

7.
$$f(t) = \begin{cases} \dfrac{e^{-a|t|}\cosh\left(\sqrt{a^2-1}|t|\right)}{4a} - \dfrac{e^{-a|t|}\sinh\left(\sqrt{a^2-1}|t|\right)}{4\sqrt{a^2-1}}, & a > 1, \\ \dfrac{e^{-a|t|}\cos\left(\sqrt{1-a^2}|t|\right)}{4a} - \dfrac{e^{-a|t|}\sin\left(\sqrt{1-a^2}|t|\right)}{4\sqrt{1-a^2}}, & 0 < a < 1. \end{cases}$$

Section 2.2

1. $f(t) = (2-t)e^{-2t} - 2e^{-3t}$ 3. $f(t) = \left(t^2/4 - t/4 + 1/8\right)e^{2t} - 1/8$

5. $f(t) = \left[(t-1)/2 - 1/4 + e^{-2(t-1)}/4\right]H(t-1)$

Section 2.3

1. $f(t) = 1 + 2t$
3. $f(t) = t + t^2/2$
5. $f(t) = t^3 + t^5/20$
7. $f(t) = t^2 - t^4/3$
9. $f(t) = 5e^{2t} - 4e^t - 2te^t$
11. $f(t) = (1-t)^2 e^{-t}$
13. $f(t) = e^{2t} - e^{-t}\left[\cos\left(\sqrt{3}\,t\right)\right) + \sqrt{3}\,\sin\left(\sqrt{3}\,t\right)\right]$
15. $f(t) = 4 + 5t^2/2 + t^4/24$
17. $x(t) = 2A\sqrt{t}/(\pi C) - Bt/(2C)$

Section 3.1

1. $F(z) = 2z/(2z-1)$ if $|z| > 1/2$
3. $F(z) = (z^6 - 1)/(z^6 - z^5)$ if $|z| > 0$
5. $F(z) = (a^2 + a - z)/[z(z-a)]$ if $|z| > a$.

Section 3.2

1. $F(z) = zTe^{aT}/(ze^{aT} - 1)^2$
3. $F(z) = z(z+a)/(z-a)^3$
5. $F(z) = [z - \cos(1)]/\{z[z^2 - 2z\cos(1) + 1]\}$
7. $F(z) = z[z\sin(\theta) + \sin(\omega_0 T - \theta)]/[z^2 - 2z\cos(\omega_0 T) + 1]$
9. $F(z) = z/(z+1)$
11. $f_n * g_n = n + 1$
13. $f_n * g_n = 2^n/n!$

Section 3.3

1. $f_0 = 0.007143,\ f_1 = 0.08503,\ f_2 = 0.1626,\ f_3 = 0.2328$
3. $f_0 = 0.09836,\ f_1 = 0.3345,\ f_2 = 0.6099,\ f_3 = 0.7935$
5. $f_n = 8 - 8\left(\frac{1}{2}\right)^n - 6n\left(\frac{1}{2}\right)^n$
7. $f_n = (1 - \alpha^{n+1})/(1-\alpha)$
9. $f_n = \left(\frac{1}{2}\right)^{n-10} H_{n-10} + \left(\frac{1}{2}\right)^{n-11} H_{n-11}$
11. $f_n = \frac{1}{9}(6n-4)(-1)^n + \frac{4}{9}\left(\frac{1}{2}\right)^n$
13. $f_n = a^n/n!$

Section 3.4

1. $y_n = 1 + \frac{1}{6}n(n-1)(2n-1)$
3. $y_n = \frac{1}{2}n(n-1)$
5. $y_n = \frac{1}{6}\left[5^n - (-1)^n\right]$
7. $y_n = (2n-1)\left(\frac{1}{2}\right)^n + \left(-\frac{1}{2}\right)^n$
9. $y_n = 2^n - n - 1$
11. $x_n = 2 + (-1)^n; y_n = 1 + (-1)^n$
13. $x_n = 1 - 2(-6)^n; y_n = -7(-6)^n$

Section 3.5

1. marginally stable
3. unstable

Section 4.1

7. $\widehat{x}(t) = \dfrac{1}{\pi}\ln\left|\dfrac{t+a}{t-a}\right|$

Section 4.2

5. $w(t) = u(t) * v(t) = \pi e^{-1}\sin(t)$

Section 4.3

1. $z(t) = e^{i\omega t}$

Section 4.4

3. $x(t) = \dfrac{1-t^2}{(1+t^2)^2}$; $\quad \widehat{x}(t) = \dfrac{2t}{(1+t^2)^2}$

Section 5.2

1. $G(s) = 1/(s+k)$ $\quad g(t|0) = e^{-kt}$
   $g(t|\tau) = e^{-k(t-\tau)}H(t-\tau)$ $\quad a(t) = \left(1 - e^{-kt}\right)/k$

3. $G(s) = 1/(s^2+4s+3)$ $\quad g(t|0) = \frac{1}{2}(e^{-t} - e^{-3t})$
   $g(t|\tau) = \frac{1}{2}\left[e^{-(t-\tau)} - e^{-3(t-\tau)}\right]H(t-\tau)$
   $a(t) = \frac{1}{6}e^{-3t} - \frac{1}{2}e^{-t} + \frac{1}{3}$

5. $G(s) = 1/[(s-2)(s-1)]$ $\quad g(t|0) = e^{2t} - e^{t}$
   $g(t|\tau) = \left[e^{2(t-\tau)} - e^{t-\tau}\right]H(t-\tau)$ $\quad a(t) = \frac{1}{2} + \frac{1}{2}e^{2t} - e^{t}$

7. $G(s) = 1/(s-9)^2$ $\quad g(t|0) = \frac{1}{3}\sinh(3t)$
   $g(t|\tau) = \frac{1}{3}\sinh[3(t-\tau)]H(t-\tau)$ $\quad a(t) = \frac{1}{9}\left[\cosh(3t) - 1\right]$

9. $G(s) = 1/[s(s-1)]$ $\quad g(t|0) = e^{t} - 1$
   $g(t|\tau) = \left[e^{t-\tau} - 1\right]H(t-\tau)$ $\quad a(t) = e^{t} - t - 1$

11.
$$g(x|\xi) = \frac{(1+x_<)(L-1-x_>)}{L},$$

and
$$g(x|\xi) = -\frac{2e^{x+\xi}}{e^{2L}-1} + \frac{2L^3}{\pi^2}\sum_{n=1}^{\infty}\frac{\varphi_n(\xi)\varphi_n(x)}{n^2(n^2\pi^2+L^2)},$$

where $\varphi_n(x) = \sin(n\pi x/L) + n\pi\cos(n\pi x/L)/L$.

13.
$$g(x|\xi) = \frac{\sinh(kx_<)\sinh[k(L-x_>)]}{k\,\sinh(kL)},$$

and
$$g(x|\xi) = 2L\sum_{n=1}^{\infty}\frac{\sin(n\pi\xi/L)\sin(n\pi x/L)}{n^2\pi^2+k^2L^2}.$$

15.
$$g(x|\xi) = \frac{\sinh(kx_<)\{k\cosh[k(x_>-L)] - \sinh[k(x_>-L)]\}}{k\sinh(kL) + k^2\cosh(kL)},$$

and
$$g(x|\xi) = 2\sum_{n=1}^{\infty}\frac{(1+k_n^2)\sin(k_n\xi)\sin(k_n x)}{[1+(1+k_n^2)L](k_n^2+k^2)},$$

where k_n is the nth root of $\tan(kL) = -k$.

17.
$$g(x|\xi) = \frac{[a\sinh(kx_<) - k\cosh(kx_<)]\cosh[k(L-x_>)]}{k[a\cosh(kL) - k\sinh(kL)]},$$

and
$$g(x|\xi) = 2\sum_{n=1}^{\infty} \frac{(a^2+k_n^2)\cos[k_n(\xi - L)]\cos[k_n(x-L)]}{[(a^2+k_n^2)L - a](k_n^2+k^2)},$$

where k_n is the nth root of $k\,\tan(kL) = -a$.

Section 5.4

3.

$$g(x,t|\xi,\tau) = \frac{t-\tau}{L}H(t-\tau) + \frac{2}{\pi}H(t-\tau)\sum_{n=1}^{\infty}\frac{1}{n}\cos\left(\frac{n\pi\xi}{L}\right)\cos\left(\frac{n\pi x}{L}\right)\sin\left[\frac{n\pi(t-\tau)}{L}\right]$$

5.

$$\begin{aligned} u(x,t) = {} & 2\sum_{n=1}^{\infty}\sin\left(\frac{n\pi x}{L}\right)\left\{\frac{n\pi}{L^2+n^2\pi^2}\left[e^{-t} - \cos\left(\frac{n\pi t}{L}\right)\right] + \frac{L}{L^2+n^2\pi^2}\sin\left(\frac{n\pi t}{L}\right)\right\} \\ & + 2\sin\left(\frac{\pi x}{L}\right)\cos\left(\frac{\pi t}{L}\right) \\ & + \frac{4L}{\pi^2}\sum_{m=1}^{\infty}\frac{1}{(2m-1)^2}\sin\left[\frac{(2m-1)\pi x}{L}\right]\sin\left[\frac{(2m-1)\pi t}{L}\right] \end{aligned}$$

7.
$$u(x,t) = 1 - \frac{t^2}{2L} - \frac{2L}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\cos\left(\frac{n\pi x}{L}\right)\left[1 - \cos\left(\frac{n\pi t}{L}\right)\right]$$

Section 5.5

3.

$$\begin{aligned} g(x,t|\xi,\tau) = \frac{2}{L}\left\{\sum_{n=1}^{\infty}\sin\left[\frac{(2n-1)\pi\xi}{2L}\right]\sin\left[\frac{(2n-1)\pi x}{2L}\right]\exp\left[-\frac{(2n-1)^2\pi^2(t-\tau)}{4L^2}\right]\right\} \\ \times H(t-\tau) \end{aligned}$$

5.
$$\begin{aligned} u(x,t) = {} & 2\pi\sum_{n=1}^{\infty}\frac{n}{n^2\pi^2 - L^2}\sin\left(\frac{n\pi x}{L}\right)\left[e^{-t} - \exp\left(-\frac{n^2\pi^2 t}{L^2}\right)\right] \\ & + \frac{4}{\pi}\sum_{m=1}^{\infty}\frac{1}{2m-1}\sin\left[\frac{(2m-1)\pi x}{L}\right]\exp\left[-\frac{(2m-1)^2\pi^2 t}{L^2}\right] \end{aligned}$$

7.
$$u(x,t) = 1 - \frac{t}{L} - \frac{2L}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\cos\left(\frac{n\pi x}{L}\right)\left[1-\exp\left(-\frac{n^2\pi^2 t}{L^2}\right)\right]$$

Section 5.6

1.
$$g(x,y|\xi,\eta) = \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\exp\left(-\frac{n\pi}{a}|y-\eta|\right)\sin\left(\frac{n\pi\xi}{a}\right)\sin\left(\frac{n\pi x}{a}\right)$$

5.
$$g(r,\theta|\rho,\theta') = \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}r_<^{n\pi/\beta}r_>^{-n\pi/\beta}\sin\left(\frac{n\pi\theta'}{\beta}\right)\sin\left(\frac{n\pi\theta}{\beta}\right)$$

7.
$$g(r,z|\rho,\zeta) = \frac{2}{\pi a^2 L}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{J_0(k_m\rho/a)J_0(k_m r/a)}{\pi a^2 L J_1^2(k_m)(k_m^2/a^2+n^2\pi^2/L^2)}\sin\left(\frac{n\pi\zeta}{L}\right)\sin\left(\frac{n\pi z}{L}\right)$$

Section 6.2

1. (a) $S = \{HH, HT, TH, TT\}$ (b) $S = \{ab, ac, ba, bc, ca, cb\}$

(c) $S = \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$ (d) $S = \{bbb, bbg, bgb, bgg, ggb, ggg, gbb, gbg\}$

(e) $S = \{bbb, bbg, bgb, bgg, ggb, ggg, gbb, gbg\}$

3. 1/3 5. 1/3 7. 2/13 9. 1/720, 1/120

11. 1/2 13. 1/2 15. 9/16

Section 6.3

1. $F_X(x) = \begin{cases} 0, & x < 0, \\ 1-p, & 0 \le x < 1, \\ 1, & 1 \le x. \end{cases}$ 3. 27

Section 6.4

1. $F_X(x) = \begin{cases} 0, & x \le 0, \\ 1-e^{-\lambda x}, & 0 < x. \end{cases}$ 3. $F_X(x) = \begin{cases} 0, & x < -1, \\ (1+x)^2/2, & -1 \le x < 0, \\ 1-(x-1)^2/2, & 0 \le x < 1, \\ 1, & 1 \le x. \end{cases}$

Section 6.5

1. $E(X) = \frac{1}{2}$, and $\mathrm{Var}(X) = \frac{1}{4}$ 3. $k = 3/4$, $E(X) = 1$, and $\mathrm{Var}(X) = \frac{1}{5}$

5. $\phi_X(\omega) = \left(pe^{i\omega} + q\right)^n$, $\mu_X = np$, $\mathrm{Var}(X) = npq$

7. $\phi_X(\omega) = p/(1 - qe^{\omega i})$, $\mu_X = q/p$, $\mathrm{Var}(X) = q/p^2$

Section 6.6

1. (a) 1/16, (b) 1/4, (c) 15/16, (d) 1/16 5. $P(X > 0) = 0.01$, and $P(X > 1) = 9 \times 10^{-5}$

7. $P(T < 150) = \frac{1}{3}$, and $P(X = 3) = 0.1646$

Section 6.7

1.
$$p_{XY}[x_i, y_j] = \frac{\binom{7}{x_i}\binom{8}{y_j}\binom{5}{5 - x_i - y_j}}{\binom{20}{5}},$$

where $x_i = 0, 1, 2, 3, 4, 5$, $y_j = 0, 1, 2, 3, 4, 5$ and $0 \le x_i + y_j \le 5$.

Section 7.1

1. $\mu_X(t) = 0$, and $\sigma_X^2(t) = \cos(\omega t)$

3. For $t_1 = t_2$, $R_X(t_1, t_2) = p$; for $t_1 \neq t_2$, $R_X(t_1, t_2) = p^2$. For $t_1 = t_2$, $C_X(t_1, t_2) = p(1-p)$; for $t_1 \neq t_2$, $C_X(t_1, t_2) = 0$.

Section 7.4

1.
$$P^n = \begin{pmatrix} 2/3 + (1/3)(1/4)^n & 1/3 - (1/3)(1/4)^n \\ 2/3 - (2/3)(1/4)^n & 1/3 + (2/3)(1/4)^n \end{pmatrix}. \qquad P^\infty = \begin{pmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{pmatrix}.$$

Section 8.3

1. $E(X) = 0, \qquad \mathrm{Var}(X) = E(X^2) = (b^3 - a^3)/3$

3. $E(X) = 0, \qquad \mathrm{Var}(X) = E(X^2) = (b^2 - a^2)/2$

Section 8.5

1. $X(t) = e^{t/2}B(t) + X_0$ 3. $X(t) = B^2(t) + tB(t) + X_0$ 5. $X(t) = tB^2(t) - t^2/2 + X_0$

7. $X(t) = X(0)e^{2t} + \frac{1}{2}\left(e^{2t} - 1\right) + e^{2t}B(t)$

9. $X(t) = e^{t^2}X(0) + e^{t^2}\int_0^t e^{-\eta^2-\eta}\,d\eta + e^{t^2}\int_0^t e^{-\eta^2}\,dB(\eta)$

11. $X(t) = X_0e^{-t} + 2(1 - e^{-t}) + \frac{1}{2}e^{-t}[B^2(t) - t]$

13. $X(t) = X_0e^{t/2} + 2\left(e^{t/2} - 1\right) + e^t\sin[B(t)]$

17. $X(t) = X_0\exp\left[\frac{t^3}{6} + \int_0^t \eta\,dB(\eta)\right]$

19. $X(t) = X_0\exp\left[\frac{1}{2}t\ln(t+1) + \frac{1}{2}\ln(t+1) - \frac{1}{2}t + \int_0^t \sqrt{\ln(\eta+1)}\,dB(\eta)\right]$

Index