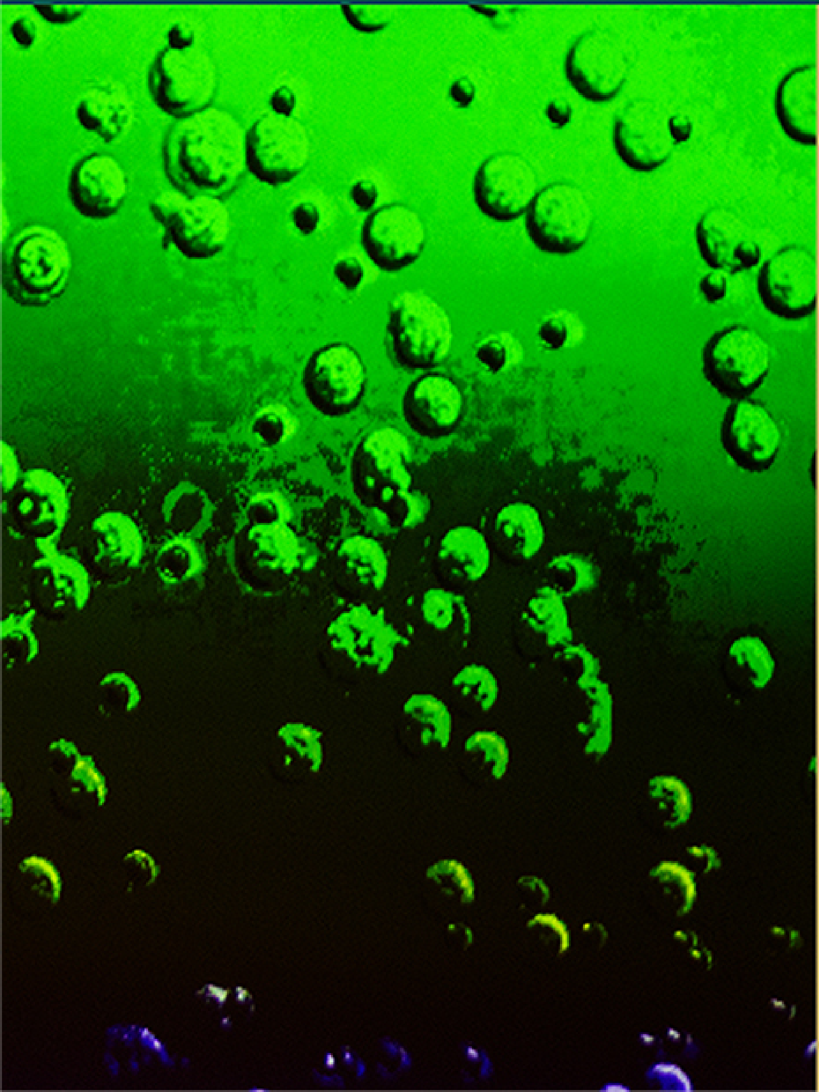
8-9-2022

Mathematics for Biomedical Physics

Jogindra M. Wadehra
Wayne State University, ad5541@wayne.edu

Mathematics for Biomedical Physics



Jogindra M. Wadehra

WAYNE STATE
University Library System

Mathematics for Biomedical Physics

Jogindra M. Wadehra

Department of Physics and Astronomy, Wayne State University

Digital Publishing
Wayne State University Library System
5150 Anthony Wayne Drive
Detroit, MI, 48202

Mathematics for Biomedical Physics
Jogindra M. Wadehra
Published 2022
ISBN-13 979-8-9857754-0-2

About the Author

Jogindra M. Wadehra is a Professor in the Department of Physics and Astronomy of Wayne State University in Detroit, Michigan, USA, and a Fellow of the American Physical Society. He received his Ph.D. in theoretical atomic and molecular physics from New York University. Before joining Wayne State University, he worked at University of Pittsburgh and Los Alamos National Laboratory as Research Associate, and at Texas A&M University as Assistant Professor. He has also done research at Air Force Wright Aeronautical Laboratory and Lawrence Berkeley National Laboratory.

Acknowledgments

I am particularly indebted to Dr. P. J. Drallos for his assistance in making majority of the figures in the book. I also thank my faculty colleagues, Professors P. Hoffmann, Z. Huang, and R. Naik, for encouraging me to turn my hand-written notes into a textbook. I am grateful to C. Hayes who critically read the whole manuscript and made several important suggestions for revising it. Finally, I thank J. Neds-Fox and C. E. Ball for their considerable editorial assistance in turning the texts into their publishable form.

Table of Contents

PREFACE ---------- X

CHAPTER 1: DIFFERENTIAL CALCULUS ---------- 1

1.1 DERIVATIVE OF A FUNCTION OF A SINGLE VARIABLE ---------- 1
1.2 FUNCTIONS OF MULTIPLE VARIABLES: PARTIAL DERIVATIVES ---------- 6
1.3 WAVE EQUATION ---------- 17
1.4 IMPLICIT DERIVATIVES ---------- 22
PROBLEMS FOR CHAPTER 1 ---------- 27

CHAPTER 2: INTEGRAL CALCULUS ---------- 30

2.1 INDEFINITE INTEGRALS ---------- 30
2.2 DEFINITE INTEGRALS ---------- 32
2.3 DIFFERENTIATION OF INTEGRALS: LEIBNIZ'S RULE ---------- 36
2.4 MULTIPLE INTEGRALS ---------- 38
2.5 BAG OF TRICKS ---------- 39
2.6 EXPONENTIAL AND GAUSSIAN INTEGRALS ---------- 46
PROBLEMS FOR CHAPTER 2 ---------- 49

CHAPTER 3: INFINITE SERIES ---------- 52

3.1 AN INFINITE SERIES ---------- 52
3.2 TESTS FOR CONVERGENCE ---------- 54
3.3 SERIES OF FUNCTIONS ---------- 60
PROBLEMS FOR CHAPTER 3 ---------- 68

INTERLUDE ---------- 71

PROBLEMS FOR INTERLUDE ---------- 75

CHAPTER 4: FOURIER SERIES ---------- 76

4.1 PERIODIC FUNCTIONS ---------- 76
4.2 FOURIER SERIES ---------- 79
4.3 FOURIER TRANSFORMS ---------- 89
PROBLEMS FOR CHAPTER 4 ---------- 96

CHAPTER 5: COMPLEX VARIABLES ---------- 98

5.1 COMPLEX NUMBERS AND COMPLEX ALGEBRA ---------- 98
5.2 PROPERTIES OF COMPLEX NUMBERS ---------- 100
5.3 POWERS OF COMPLEX NUMBERS ---------- 103
5.4 ROOTS OF A COMPLEX NUMBER ---------- 105
5.5 DEMOIVRE'S FORMULA ---------- 109
PROBLEMS FOR CHAPTER 5 ---------- 112

CHAPTER 6: DETERMINANTS ---------- 113

6.1 DEFINITION OF A DETERMINANT ---------- 113
6.2 PROPERTIES OF DETERMINANT $\mathbb{A}$ ---------- 116
PROBLEMS FOR CHAPTER 6 ---------- 124

CHAPTER 7: MATRICES ---------- 125

7.1 A GENERAL MATRIX ---------- 125
7.2 PROPERTIES OF MATRICES ---------- 127
7.3 SOLVING INHOMOGENEOUS ALGEBRAIC EQUATIONS ---------- 132
7.4 SPECIAL MATRICES ---------- 133
7.5 CHARACTERISTIC OR SECULAR EQUATION OF A MATRIX ---------- 135
PROBLEMS FOR CHAPTER 7 ---------- 139

CHAPTER 8: VECTOR ANALYSIS ---------- 141

8.1 VECTORS AND THEIR ADDITION/SUBTRACTION ---------- 141
8.2 MULTIPLICATION OF VECTORS: SCALAR PRODUCT ---------- 145
8.3 MULTIPLICATION OF VECTORS: VECTOR PRODUCT ---------- 147
8.4 TRIPLE SCALAR PRODUCT ---------- 150
8.5 TRIPLE VECTOR PRODUCT ---------- 151
PROBLEMS FOR CHAPTER 8 ---------- 153

CHAPTER 9: CURVILINEAR COORDINATES AND MULTIPLE INTEGRALS ---------- 156

9.1 CARTESIAN COORDINATES ---------- 156
9.2 SPHERICAL COORDINATES ---------- 157
9.3 CYLINDRICAL COORDINATES ---------- 159
9.4 PLANE POLAR COORDINATES ---------- 160
9.5 MULTIPLE INTEGRALS ---------- 161
9.6 SOLIDS OF REVOLUTION ---------- 163
9.7 CENTER OF MASS ---------- 164
9.8 MOMENT OF INERTIA ABOUT AN AXIS ---------- 166
PROBLEMS FOR CHAPTER 9 ---------- 170

CHAPTER 10: VECTOR CALCULUS ---------- 172

10.1 VECTOR FUNCTIONS ---------- 172
10.2 LINE INTEGRALS ---------- 177
10.3 SURFACE INTEGRALS ---------- 179
10.4 VOLUME INTEGRALS ---------- 180
10.5 FLUX OF A VECTOR FUNCTION: GAUSS'S (OR DIVERGENCE) THEOREM ---------- 181
10.6 CIRCULATION OF A VECTOR FUNCTION: STOKE'S THEOREM ---------- 185
PROBLEMS FOR CHAPTER 10 ---------- 190

CHAPTER 11: FIRST ORDER DIFFERENTIAL EQUATIONS ---------- 192

11.1 A FIRST ORDER DIFFERENTIAL EQUATION ---------- 192

11.2 INTEGRATING FACTOR ---- 194
11.3 APPLICATIONS OF NUCLEAR PHYSICS ---- 199
PROBLEMS FOR CHAPTER 11 ---- 203

CHAPTER 12: DIFFUSION EQUATION ---- 205

12.1 FICK'S LAW ---- 205
12.2 DIFFUSION EQUATION IN THREE DIMENSIONS ---- 207
12.3 DIFFUSION EQUATION IN ONE DIMENSION ---- 209
12.4 FINAL COMMENTS ---- 210
PROBLEMS FOR CHAPTER 12 ---- 211

CHAPTER 13: PROBABILITY DISTRIBUTION FUNCTIONS ---- 212

13.1 A DISCRETE PROBABILITY DISTRIBUTION FUNCTION ---- 212
13.2 BINOMIAL DISTRIBUTION FUNCTION ---- 215
13.3 POISSON DISTRIBUTION FUNCTION ---- 217
13.4 MAXWELL-BOLTZMANN (CONTINUOUS) DISTRIBUTION FUNCTION ---- 218
13.5 NORMAL OR GAUSSIAN (CONTINUOUS) DISTRIBUTION FUNCTION ---- 220
13.6 A FINAL REMARK ---- 225
PROBLEMS FOR CHAPTER 13 ---- 226

APPENDIX A: LIMITING VALUE OF $\mathrm{SIN}\theta/\theta$ AS $\theta \to 0.$ ---- 228
APPENDIX B: MNEMONIC FOR MAXWELL'S RELATIONS OF THERMODYNAMICS ---- 230
APPENDIX C: PROOF OF THE EPSILON-DELTA IDENTITY ---- 232
APPENDIX D: AREAS UNDER THE STANDARD NORMAL CURVE ---- 233

Preface

A few years ago, our department started a new undergraduate degree program in Biomedical Physics. The prerequisites to enter this program included two semesters of physics and two semesters of calculus, both differential and integral. A group of participating educators (including some members from the departments of physics, mathematics, and biology as well as from the School of Medicine) developed the courses and their contents for this new program. For one of these courses, Mathematics for Biomedical Physics, all the required topics were not covered in any single textbook available. As the instructor of this course, I had to prepare my own personal notes for some of the topics that I shared with students. I felt that it would be great if a single textbook can be produced which covered all the mathematical topics that were needed for the biomedical physics program. I wanted to make this textbook freely available to all students everywhere using internet. Access to education is a basic human right which should not be hindered by any lack of money or dearth of educational resources. With this thought in mind, I started turning the course materials and lectures that I had been using in my course into an open educational resource (OER). After transcribing all the notes in Word, I gave the notes away to students in my next class, seeking their feedback and comments. I have carefully included responses to these comments in the book.

Even though the prerequisites for this textbook are two semester-long courses in calculus, which cover functions of single variables, the first two chapters of this textbook are devoted to differential calculus and integral calculus. These chapters lead the reader from calculus of functions of a single variable to calculus of multivariable functions. In this scheme, the ideas related to partial derivatives as well as multiple integrals are revealed quite naturally. Throughout the textbook I have attempted to start with something that a typical student may be familiar with and end up with something that will be entirely new for the reader. I would appreciate receiving (at wadehra@wayne.edu) constructive feedback, from students and faculty as well as other readers, who are using this book as a whole or in parts.

The textbook is geared to introduce several mathematical topics at the rudimentary level so that students can appreciate the applications of mathematics to the interdisciplinary field of biomedical physics. Most of the topics are presented in their simplest but rigorous form so that students can easily understand the advanced form of these topics when the need arises. Several end-of-chapter problems and chapter examples relate the applications of mathematics to biomedical physics. After mastering the topics of this book, the students would be ready to embark on quantitative thinking in various topics of biology and medicine. The famous renaissance philosopher and astronomer, Galileo Galilei, is quoted to say, “Mathematics is the language in which Nature has written the book of Universe”. This textbook is an endeavor to teach the language of Nature in a careful, yet simple, manner to undergraduate students.

Chapter 1: Differential Calculus

We will start with a review of differential calculus, assuming that the reader has already seen derivatives for functions of a single variable. After a brief review of simple derivatives, we will introduce the derivatives of functions of multiple variables, also known as partial derivatives.

1.1 DERIVATIVE OF A FUNCTION OF A SINGLE VARIABLE

$F(x)$ is a function of a single *independent* variable x as shown in Figure 1.1. An independent variable means that its value can be assigned at will, without any constraining conditions. By definition, the derivative of $F(x)$ is $\frac{dF}{dx}$, given by

$$\frac{dF}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \qquad Eq. (1.1)$$

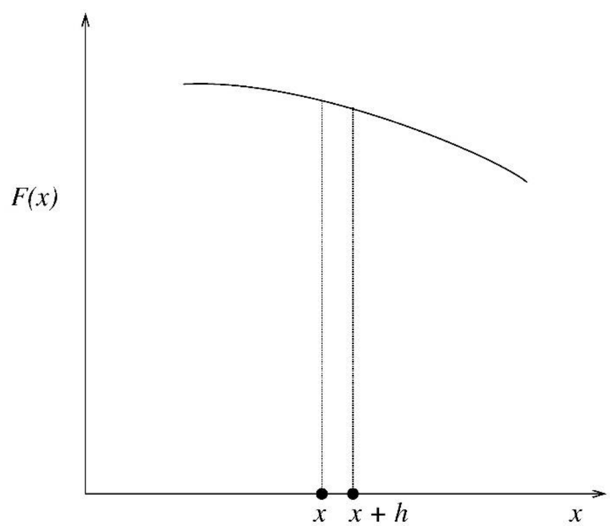


Figure 1.1. $\boldsymbol{F(x)}$ is a function of a single independent variable $\boldsymbol{x}$.

As a simple example, consider the derivative of $F(x) = \sin(x)$,

$$\frac{dF}{dx} = \lim_{h \to 0} \frac{\sin(x+h) - \sin(x)}{h} .$$

In the numerator, we use the trigonometric identity

$$\sin A - \sin B = 2 \sin\left(\frac{A-B}{2}\right) \cos\left(\frac{A+B}{2}\right)$$

to get

$$\frac{d\sin(x)}{dx} = \lim_{h\to 0}\frac{2\sin\left(\frac{h}{2}\right)\cos\left(x+\frac{h}{2}\right)}{h} = \lim_{h\to 0}\frac{\sin(h/2)}{h/2}\lim_{h\to 0}\cos\left(x+\frac{h}{2}\right) .$$

Or,

$$\frac{d\sin(x)}{dx} = \cos(x)\lim_{h\to 0}\frac{\sin(h/2)}{h/2} .$$

The limiting value of this expression can be easily obtained, using simple geometry and trigonometry (see Appendix A), to be 1. Using this limiting value, we get the result

$$\frac{d\sin(x)}{dx} = \cos(x) .$$

A similar procedure can be used to determine derivatives of other well-known functions. Here is a compilation of derivatives of several commonly used functions:

$$\frac{dC}{dx} = 0, \quad C \text{ is a constant}$$

$$\frac{d\,x^n}{dx} = n\,x^{n-1}$$

$$\frac{d\exp(x)}{dx} = \exp(x)$$

$$\frac{d\ln(x)}{dx} = \frac{1}{x}$$

$$\frac{d\cos(x)}{dx} = -\sin(x)$$

$$\frac{d\tan(x)}{dx} = \sec^2 x$$

$$\frac{d\arcsin x}{dx} = \frac{1}{\sqrt{1-x^2}}$$

$$\frac{d\arccos x}{dx} = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{d\arctan x}{dx} = \frac{1}{1+x^2}$$

Higher Order Derivatives

Assume that the derivative $\frac{dF}{dx}$ of the function $F(x)$ is represented by a new function $F_1(x)$; that is, $F_1(x) = \frac{dF}{dx}$. Then, $\frac{dF_1}{dx}$ is the second derivative of $F(x)$, namely, $\frac{d^2F}{dx^2}$. Similarly, if $F_2(x) = \frac{dF_1}{dx}$, then $\frac{dF_2}{dx}$ is the third derivative of $F(x)$, namely, $\frac{d^3F}{dx^3}$. Using this procedure, higher derivatives of any order can be determined for a function of a single variable.

Chain Rule

When a function F of variable x can be expressed as a function of another function as $F[g(x)]$, then the derivative of F can be most conveniently evaluated using the chain rule as

$$\frac{dF}{dx} = \frac{dF}{dg}\frac{dg}{dx} \ . \qquad Eq.(1.2)$$

Example: Using the chain rule, determine the derivative of the function $\boldsymbol{F(x) = \sqrt{x^2 + a^2}}$.

Solution: To determine the derivative $\frac{dF}{dx}$, first define a new function $g(x) = x^2 + a^2$. Then, $F(x)$ can be written as a function of $g(x)$ as $F(x) = \sqrt{g(x)} = g^{1/2}$. Using chain rule,

$$\frac{dF}{dx} = \frac{dF}{dg}\frac{dg}{dx} = \frac{1}{2}\,g^{-1/2}\,(2x) = \frac{x}{\sqrt{x^2 + a^2}} \ .$$

Product and Quotient Rules

When a function F of variable x can be expressed as a product of two simpler functions as $F(x) = f(x)\,g(x)$, then the derivative of $F(x)$ is given by

$$\frac{dF}{dx} = f(x)\frac{dg(x)}{dx} + \frac{df(x)}{dx}g(x) \ . \qquad Eq.(1.3)$$

In words, if $F(x) = (FIRST)(\,SECOND)$, then

$$\frac{dF}{dx} = (FIRST)\frac{d\,(SECOND)}{dx} + \frac{d\,(FIRST)}{dx}(SECOND).$$

Example: Using the product rule, determine the derivative of $F(x) = x^2 \ln x$.

Solution: Derivative of $F(x)$ using the product rule is

$$\frac{dF}{dx} = x^2 \frac{d \ln x}{dx} + \ln x \frac{dx^2}{dx} = x^2 \frac{1}{x} + \ln x \ (2x) = x + 2x \ln x \ .$$

A special case of the product rule is the quotient rule, which is used when the function $F(x)$ can be expressed as $F(x) = \frac{f(x)}{g(x)}$. In this case,

$$\frac{dF}{dx} = \frac{1}{g^2}\left[g(x) \frac{df}{dx} - f(x) \frac{dg}{dx}\right] . \qquad Eq. (1.4)$$

Example: Using the quotient rule, determine the derivative of $F(x) = \exp(ax)/x$.

Solution: The derivative of $F(x)$ using the quotient rule is

$$\frac{dF}{dx} = \frac{1}{x^2}\left[x \frac{d \exp{(ax)}}{dx} - \exp(ax) \frac{dx}{dx}\right] = \frac{a}{x}\exp(ax) - \frac{1}{x^2}\exp(ax) = \exp(ax)\frac{ax-1}{x^2} \ .$$

Interpretations of a Derivative

The derivative $\frac{dF}{dx}$ of the function $F(x)$ has two distinct interpretations.

First, the value of a derivative, $\frac{dF(x)}{dx}$, at $x = a$ is the slope of the line that is tangent to the function $F(x)$ at $x = a$. Now, since the slope of a function at a point where the function has its minimum or maximum value is zero, it follows that the derivative of the function at its extremum points (that is, points with minimum or maximum values) is also zero. So, the extremum (minimum or maximum) points of a function can be determined by setting its derivative equal to zero. The equation $\frac{dF(x)}{dx} = 0$ can be solved to obtain the values of $x = x_{min}$ where the function is minimum or $x = x_{max}$ where the function is maximum. In the vicinity of x_{min}, the values of $\frac{dF(x)}{dx}$ are negative for $x < x_{min}$, zero for $x = x_{min}$, and positive for $x > x_{min}$. In other words, $\frac{d^2F}{dx^2}$ is positive at $x = x_{min}$. Similarly, in the vicinity of x_{max}, the values of $\frac{dF(x)}{dx}$ are positive for $x < x_{max}$, zero for $x = x_{max}$, and negative for $x > x_{max}$. In other words, $\frac{d^2F}{dx^2}$ is negative at $x = x_{max}$.

Second, the derivative $\frac{dF(x)}{dx}$ represents the rate of variation of function $F(x)$ with x. Suppose that when independent variable x changes by a small amount Δx, then the corresponding change in the value of the function is ΔF. Using this interpretation of the derivative,

$$(Change\ in\ F) = (Rate\ of\ change\ of\ F\ with\ x) \cdot (Change\ in\ x)$$

or,

$$(\Delta F) = \frac{dF}{dx}\,(\Delta x)\ . \qquad\qquad Eq. (1.5)$$

The idea about the rate of change of a quantity appears in many diverse areas of knowledge. For example, in physiology we talk about the rate at which blood flows in veins; in geography we talk about the rate at which population grows in a certain area; in meteorology we talk about the rate at which pressure varies with altitude; in medicine we talk about the rate at which a cancerous tumor grows or a contagious virus spreads; in sociology we talk about the rate at which a news or some rumor spreads; in psychology we talk about the rate at which different people learn certain skills (that is, the learning curve); in physics we talk about the rate at which the velocity of an accelerating automobile changes, etc. No matter in which context we talk about the rate of change of a quantity, its mathematical description would always be presented in the form of a derivative.

Example: As an example of the first interpretation of a derivative, let us look at the case of a young mom, sitting on a bench on a concrete patio, watching her small child playing on the grass nearby, as shown in the Figure 1.2. The lengths d_p and d_g are the shortest distances of mom and child, respectively, from the patio-grass boundary. Mom can run on the patio with speed v_p and on the grass with speed v_g. When an emergency arises, the mom would like to reach the child in the shortest possible time. Which path should she take during an emergency?

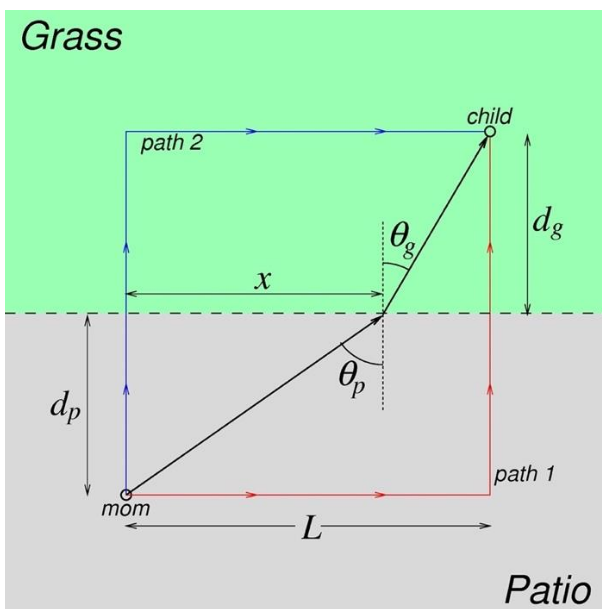


Figure 1.2. The quickest path a mom can take to reach her child in distress.

Solution: The mom can either take path 1 or path 2 (shown in the figure) or any path in between. Consider an arbitrary path, as shown in the figure, taken by mom to rush to her child. Total time taken by mom using this path is

$$t = \frac{\sqrt{d_p^2 + x^2}}{v_p} + \frac{\sqrt{d_g^2 + (L - x)^2}}{v_g} .$$

Here L is the separation between the mom and the child along the patio-grass boundary. The distance x, shown in the figure, will vary depending on the specific path taken by the rushing mom. The path with the shortest time, according to the above interpretation of the derivative, will be the one for which $\frac{dt}{dx}$ is equal to zero. Thus,

$$\frac{dt}{dx} = \frac{2x}{2v_p\sqrt{d_p^2 + x^2}} + \frac{2(L - x)(-1)}{2v_g\sqrt{d_g^2 + (L - x)^2}} = 0 ,$$

or

$$\frac{1}{v_p}\frac{x}{\sqrt{d_p^2 + x^2}} = \frac{1}{v_g}\frac{(L - x)}{\sqrt{d_g^2 + (L - x)^2}} ,$$

or, in terms of the angles θ_p and θ_g shown in the Figure 1.2,

$$\frac{\sin\theta_p}{v_p} = \frac{\sin\theta_g}{v_g} .$$

This relationship describes the path that mom should take to reach the child in the shortest possible time. It is, essentially, Snell's law of refraction in optics that concerns the bending of light as it travels from one medium (number 1) into another medium (number 2). Mathematically, Snell's law is stated as $n_1 \sin\theta_1 = n_2 \sin\theta_2$ where n_1 and n_2 are the indices of refraction of the two media. Light travels with different speeds in media with different indices of refraction with $v_1 = c/n_1$ and $v_2 = c/n_2$ (c being the speed of light in vacuum). Thus, Snell's law can be expressed as $\sin\theta_1/v_1 = \sin\theta_2/v_2$. Stated differently, Snell's law implies that when light travels from a point in the first medium to another point in the second medium, it takes a path that minimizes its time of travel.

1.2 FUNCTIONS OF MULTIPLE VARIABLES: PARTIAL DERIVATIVES

Recall that for an independent variable, any value can be assigned to it without any constraining conditions. If F is a function of two or more *independent* variables, then the derivative of F with respect to one of the variables, while holding the other variable fixed, is called the partial derivative. Formally, if $F(x, y)$ is a function of two independent variables x and y, then

$$\frac{\partial F(x,y)}{\partial x} = \lim_{h \to 0} \frac{F(x+h,y) - F(x,y)}{h} \qquad Eq. (1.6a)$$

$$\frac{\partial F(x,y)}{\partial y} = \lim_{h \to 0} \frac{F(x,y+h) - F(x,y)}{h} \qquad Eq. (1.6b)$$

are the partial derivatives of F. Note the use of $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ for partial derivatives versus $\frac{d}{dx}$ for derivatives of a function of only one independent variable. A common notation for writing partial derivatives also includes $\left(\frac{\partial F}{\partial x}\right)_y$ and $\left(\frac{\partial F}{\partial y}\right)_x$ in which the fixed variable is written as a subscript.

Example: Determine the partial derivatives of the function $\boldsymbol{F(x,y) = x^3 - x\,y^2 + y}$.

Solution: Since x and y are two independent variables, the partial derivatives of $F(x,y)$ with respect to x and y are:

$$\frac{\partial F(x,y)}{\partial x} = 3\,x^2 - y^2 \ ,$$

$$\frac{\partial F(x,y)}{\partial y} = -2xy + 1 \ .$$

As before, $\frac{\partial F}{\partial x}$ (or $\frac{\partial F}{\partial y}$) represents the rate of variation of F with x (or y). If both x and y vary independently, then

$$(Change\ in\ F) = (Rate\ of\ change\ of\ F\ with\ x) \cdot (Change\ in\ x) + (Rate\ of\ change\ of\ F\ with\ y) \cdot (Change\ in\ y) \ ,$$

or

$$(\Delta F) = \frac{\partial F}{\partial x}\,(\Delta x) + \frac{\partial F}{\partial y}\,(\Delta y) \ . \qquad Eq. (1.7)$$

Higher Order Partial Derivatives: Clairaut's Theorem

If $F_x = \frac{\partial F}{\partial x}$ and $F_y = \frac{\partial F}{\partial y}$, then $\frac{\partial F_x}{\partial x} = \frac{\partial^2 F}{\partial x^2}, \frac{\partial F_x}{\partial y} = \frac{\partial^2 F}{\partial y \partial x}, \frac{\partial F_y}{\partial x} = \frac{\partial^2 F}{\partial x \partial y}$ and $\frac{\partial F_y}{\partial y} = \frac{\partial^2 F}{\partial y^2}$ are the second derivatives of $F(x,y)$. Higher order partial derivatives can be defined in an analogous manner. The $Clairaut's\ theorem$ states that

$$\frac{\partial F_x}{\partial y} = \frac{\partial F_y}{\partial x} \quad \text{or} \quad \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right) \ .$$

In words, it means that for a function of several independent variables, the multiple partial derivatives can be carried out in any order.

Example: Verify Clairaut's theorem for the function $\boldsymbol{F(x,y) = x^3 - x\,y^2 + \,y}$.

Solution: In the previous example, we calculated $\frac{\partial F}{\partial x} = 3\,x^2 - \,y^2$ and $\frac{\partial F}{\partial y} = -2xy + 1$. So now

$$\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right) = \,-2y\ \ ,$$

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right) = \,-2y\ \ .$$

so the Clairaut's theorem is indeed satisfied.

Example: Check whether the Clairaut's theorem for the function $\boldsymbol{F(x,y) = \exp(ax)\sin(by)}$ is satisfied.

Solution: In this case, there are two independent variables x and y. So, first partial derivatives are

$$\frac{\partial F(x,y)}{\partial x} = a\,\exp(ax)\,\sin(by)\ \ ,$$

$$\frac{\partial F(x,y)}{\partial y} = \exp(ax)\,b\,\cos(by)\ \ .$$

The second derivative gives

$$\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right) = \ a\,\exp(ax)\,b\,\cos(by)\ \ ,$$

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right) = \ a\,\exp(ax)\,b\,\cos(by)\ \ .$$

that verifies Clairaut's theorem.

If F is a function of n independent variables, that is, $F = F(x_1, x_2, x_3, \dots x_n)$, then $\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots \frac{\partial F}{\partial x_n}$ represent the rates of variation of F with $x_1, x_2, \dots x_n$, respectively. Since all n variables can change independently, then

$$(Change\ in\ F) = (Rate\ of\ change\ of\ F\ with\ x_1)\cdot(Change\ in\ x_1) + \ (Rate\ of\ change\ of\ F\ with\ x_2) \cdot(Change\ in\ x_2) + \cdots + (Rate\ of\ change\ of\ F\ with\ x_n)\cdot(Change\ in\ x_n)\ \ ,$$

or,

$$(\Delta F) = \frac{\partial F}{\partial x_1}\,(\Delta x_1) + \frac{\partial F}{\partial x_2}\,(\Delta x_2) + \cdots + \frac{\partial F}{\partial x_n}\,(\Delta x_n)\ . \qquad Eq.(1.8)$$

(ΔF) is the total change in function F when all n variables are changed independently. It is called *total differential*. At a point where F is extremum (namely, a minimum or a maximum), $\Delta F = 0$. Since $x_1, x_2, \dots x_n$ are independent variables, they may be chosen so that all but one of the Δx, in turn, are zero. It follows that

$$\frac{\partial F}{\partial x_1} = 0,\quad \frac{\partial F}{\partial x_2} = 0,\quad \dots \frac{\partial F}{\partial x_{n-1}} = 0 \ \text{ and } \frac{\partial F}{\partial x_n} = 0\ . \qquad Eq.(1.9)$$

These n algebraic equations can be solved to find the values of variables $x_1, x_2, \dots x_n$ that make the function F extremum.

The idea of a total differential, introduced above, is very useful in estimating the largest possible error in the measurement of a physical quantity that is a function of several variables. An example will help in understanding the concept.

Example: According to the Poiseuille's equation, the total volume of blood flowing through a blood vessel of radius $\boldsymbol{R}$ and length $\boldsymbol{L}$ per unit time is given by $\boldsymbol{F = k\frac{R^4}{L}}$, where constant $\boldsymbol{k}$ is independent of the geometry of the blood vessel. If the relative error, $\boldsymbol{\Delta R}$, in the measurement of the radius is 2% and the relative error, $\boldsymbol{\Delta L}$, in the measurement of length is 4%, then what is the largest possible relative error, $\boldsymbol{\Delta F}$, in the measurement of flux of blood, $\boldsymbol{F}$, through this blood vessel?

Solution: The relative error in the measurement of a quantity means error in measuring that quantity divided by the actual value of that quantity. In other words, the ratios $\frac{\Delta R}{R}, \frac{\Delta L}{L}$ and $\frac{\Delta F}{F}$ are, respectively, the relative errors in the measurements of the radius, length and flux of the blood through a blood vessel. These ratios are determined by first taking the natural log of the Poiseuille's equation and then taking the derivatives of both sides as

$$\ln F = \ln k + 4 \ln R - \ln L$$

and

$$\frac{\Delta F}{F} = 4\,\frac{\Delta R}{R} - \frac{\Delta L}{L}\ .$$

Since error in the measurement of radius is 2%, it means that the value of $\frac{\Delta R}{R}$ ranges between -0.02 and $+0.02$. Similarly, the value of $\frac{\Delta L}{L}$ ranges between -0.04 and $+0.04$. Thus, the largest possible value of $\frac{\Delta F}{F}$ is $4(+0.02) - (-0.04) = 0.12$. Or, the largest possible relative error in the measurement of flux of blood is 12%.

Now, going back to the function $F = F(x_1, x_2, x_3, \dots x_n)$ of n variables, if $x_1, x_2, \dots x_n$ are constrained by one or more relationships of the form,

$$\Phi(x_1, x_2, \dots x_n) = \text{ constant } ,$$

then not all n variables are independent. In fact, if F is a function of n variables and if there are m $(m < n)$ constraining relationships among variables, then the number of independent variables is only $n - m$. In this case one can use the *method of Lagrange multipliers* to find the extremum value of the function F.

Method of Lagrange Multipliers

Instead of considering a general function of n variables, let us focus our attention on a function $F(x, y, z)$ of three variables, x, y and z. These variables are constrained by the relation

$$\Phi(x, y, z) = \text{ constant } .$$

Because of this constraint only two out of three variables are independent. Let us choose x and y to be the independent variables. Since we wish to determine the extremum value of $F(x, y, z)$, we take its derivative and set it equal to zero:

$$\frac{\partial F}{\partial x}\,\Delta x + \frac{\partial F}{\partial y}\,\Delta y + \frac{\partial F}{\partial z}\,\Delta z = 0 \ . \qquad Eq. (1.10a)$$

From the constraint equation, $\Phi = $ constant, we get

$$\frac{\partial \Phi}{\partial x}\,\Delta x + \frac{\partial \Phi}{\partial y}\,\Delta y + \frac{\partial \Phi}{\partial z}\,\Delta z = 0 \ . \qquad Eq. (1.10b)$$

We multiply Eq. (1.10b) by a multiplier, $-\lambda$, and add it to Eq. (1.10a) to get

$$\left(\frac{\partial F}{\partial x} - \lambda\,\frac{\partial \Phi}{\partial x}\right)\Delta x + \left(\frac{\partial F}{\partial y} - \lambda\,\frac{\partial \Phi}{\partial y}\right)\Delta y + \left(\frac{\partial F}{\partial z} - \lambda\,\frac{\partial \Phi}{\partial z}\right)\Delta z = 0 \ .$$

Since the multiplier λ is yet undetermined and can be chosen at will, we choose it so that

$$\frac{\partial F}{\partial z} - \lambda\,\frac{\partial \Phi}{\partial z} = 0 \ .$$

This choice of λ will remove the term containing Δz, leaving only the two independent variables x and y. As x and y are independent variables, we can choose them, in turn, so that $\Delta x = 0$ and $\Delta y = 0$ separately. It then follows that

$$\frac{\partial F}{\partial x} - \lambda \frac{\partial \Phi}{\partial x} = 0 \ ,$$

and

$$\frac{\partial F}{\partial y} - \lambda \frac{\partial \Phi}{\partial y} = 0 \ .$$

The last three equations, along with the equation of constraint, provide the values of the multiplier λ and the optimal values of the variables x, y and z that make the function F extremum. The constant λ is called Lagrange's undetermined multiplier.

Example: A box (sides a, b, c) of fixed volume V is to be designed so that its surface area is minimum. Find the optimal values of a, b, c.

Solution:

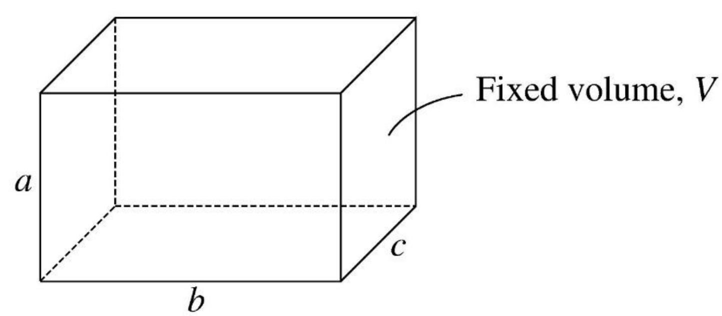


Figure 1.3. A box of fixed volume V whose surface area is minimized.

In this case, the equation of constraint is

$$V = a\,b\,c = \text{constant} \ .$$

Because of this constraint, only two sides of the box can be changed independently—the third side will be determined by the equation of constraint. We choose a and b as the independent variables.

We need to find the minimum value of the surface area of this box, which is

$$S = 2(ab + bc + ca) \ .$$

Setting the derivative of S equal to zero, we get

$$\Delta S = \frac{\partial S}{\partial a}\,\Delta a + \frac{\partial S}{\partial b}\,\Delta b + \frac{\partial S}{\partial c}\,\Delta c = 0 \ ,$$

or, after dividing by the common factor of 2,

$$(b + c)\,\Delta a + \ (c + a)\,\Delta b + (a + b)\,\Delta c = 0 \ . \qquad\qquad Eq. (1.11a)$$

Also, the volume V is constant. So,

$$\Delta V = \frac{\partial V}{\partial a}\,\Delta a + \frac{\partial V}{\partial b}\,\Delta b + \frac{\partial V}{\partial c}\,\Delta c = 0 \ ,$$

or

$$(bc)\,\Delta a + (ca)\,\Delta b + (ab)\,\Delta c = 0 \ . \qquad\qquad Eq. (1.11b)$$

Multiply Eq. (1.11b) by undetermined multiplier $-\lambda$ and add it to Eq. (1.11a) to get

$$[(b + c) - \lambda(bc)]\,\Delta a + [(c + a) - \lambda(ca)]\,\Delta b + [(a + b) - \lambda(ab)]\,\Delta c = 0 \ . \qquad Eq. (1.11c)$$

The multiplier λ is still undetermined and can be chosen at will. We choose $\lambda = (a + b)/(ab)$, or

$$(a + b) - \lambda(ab) = 0$$

so that term containing Δc is removed in Eq. (1.11c). Also, since a and b are chosen as independent variables, it follows from Eq. (1.11c) that

$$(b + c) - \lambda(bc) = 0 \ ,$$

and

$$(c + a) - \lambda(ca) = 0 \ .$$

From these three relationships, we get

$$\lambda(abc) = a(b + c) = b(c + a) = c(a + b) \ .$$

Or

$$a = b = c \ .$$

Thus, a box of a fixed volume V and a minimum surface area is a cube.

Example: Determine the ratio of radius, r, and height, h, of a right circular cylinder of fixed volume V, that will make the surface area, S, of the cylinder a minimum.

Solution:

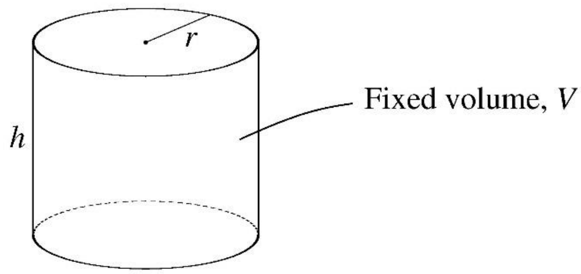


Figure 1.4. A cylinder of fixed volume V whose surface area is minimized.

In this case, $V = \pi r^2 h =$ constant is the equation of constraint. Both r and h can vary, but only one of them is independent. Let us choose r to be the independent variable. The surface area, S, of the cylinder consists of two end circles and the curved surface of the cylinder. Thus,

$$S = 2\pi rh + 2\pi r^2 \ .$$

From the constraint condition

$$\Delta V = \frac{\partial V}{\partial r}\,\Delta r + \frac{\partial V}{\partial h}\,\Delta h = 0 \ ,$$

or

$$2\pi rh\,\Delta r + \pi r^2\,\Delta h = 0 \ ,$$

or

$$2h\,\Delta r + r\,\Delta h = 0 \ . \qquad Eq.(1.12a)$$

Also, when the surface area, S, is a minimum, then

$$\Delta S = \frac{\partial S}{\partial r}\,\Delta r + \frac{\partial S}{\partial h}\,\Delta h = 0 \ ,$$

or

$$(2\pi h + 4\pi r)\,\Delta r + \ 2\pi r\,\Delta h = 0 \ \ ,$$

or

$$(h + 2r)\,\Delta r + \ r\,\Delta h = 0 \ \ . \qquad\qquad Eq. (1.12b)$$

On multiplying Eq. (1.12b) by undetermined multiplier $-\lambda$ and adding it to Eq. (1.12a) we get

$$[2h - \lambda(h + 2r)]\,\Delta r + \ [r - \lambda r]\,\Delta h = 0 \ \ . \qquad\qquad Eq. (1.12c)$$

The multiplier λ, which can be chosen at will, is taken as $\lambda = 1$. This choice of λ removes the term containing Δh in Eq. (1.12c). Also, since r is an independent variable, we can choose $\Delta r \neq 0$ which makes

$$2h - (h + 2r) = \ 0 \ \ ,$$

or

$$h = 2r \ \ .$$

Thus, the right circular cylinder of a fixed volume V will have the least surface area when the height of the cylinder is equal to its diameter.

Let us try to construct some common solids of different shapes, using playdough of a fixed volume V, such that the solid has a minimum surface area, S.

If we construct a playdough cube of side length L, then $V = \ L^3$ and $S = 6\,L^2 = 6\,V^{2/3}$.

On the other hand, a playdough cylinder with a minimum surface area will have a height equal to its diameter. Thus, if R is the radius of this cylinder, then $V = \ \pi R^2(2R) = \ 2\pi R^3$. The surface area is $S = 2(\pi R^2) + (2\pi R)(2R) = \ 6\pi R^2 = 6\pi \left(\frac{V}{2\pi}\right)^{2/3} = \ 5.54\,V^{2/3}$.

Finally, a playdough sphere of radius R will have its volume as $V = \left(\frac{4\pi}{3}\right) R^3$ and its surface area as $S = 4\pi R^2 = 4\pi \left(\frac{3V}{4\pi}\right)^{2/3} = 4.84\,V^{2/3}$.

Thus, among these common solids, all of the same volume V, the sphere will have the smallest surface area. This may partly explain why all the celestial bodies – stars, planets, and moons – are spherical in shape. It also explains why soap bubbles and raindrops tend to be spheres. In other words, Mother Nature prefers to make shapes with least surface area, for a fixed volume of its own playdough.

Chain Rule for Partial Derivatives

Recall that chain rule of calculus is applicable when we need to take derivative of a function of another function.

Case I: The function $F(x)$ is a function of a single variables x which itself is a function of two other variables, s and t, that is, $x(s,t)$. Then, in principle, F is a function of two variables, s and t. The derivatives of F with respect to s and t are

$$\frac{\partial F}{\partial s} = \frac{dF}{dx}\frac{\partial x}{\partial s} ,$$

and

$$\frac{\partial F}{\partial t} = \frac{dF}{dx}\frac{\partial x}{\partial t} .$$

Example: Given the function $\boldsymbol{F(x) = x^2}$ with $\boldsymbol{x(s,t) = s + t}$, determine the derivatives $\boldsymbol{\frac{\partial F}{\partial s}}$ and $\boldsymbol{\frac{\partial F}{\partial t}}$.

Solution:

$$\frac{\partial F}{\partial s} = (2x)(1) = 2(s+t) ,$$

and

$$\frac{\partial F}{\partial t} = (2x)(1) = 2(s+t) .$$

Alternatively,

$$F(s,t) = (s+t)^2$$

and, directly, the partial derivatives of $F(s,t)$ with respect to s or t provide the same results as those obtained by using the chain rule for partial derivatives.

Case II: $F(x,y)$ is a function of two variables x and y, and x and y themselves are functions of a single variable t. Then, in principle, F is a function of a single variable t and the derivative of F with respect to t is,

$$\frac{dF}{dt} = \frac{\partial F}{\partial x}\frac{dx}{dt} + \frac{\partial F}{\partial y}\frac{dy}{dt} \; .$$

Example: Given the function $\boldsymbol{F(x,y) = \ln(2x+3y)}$ with $\boldsymbol{x = t^2}$ and $\boldsymbol{y = \sin(3t)}$, determine the derivative $\boldsymbol{\frac{dF}{dt}}$.

Solution: Using the chain rule for partial derivatives,

$$\frac{dF}{dt} = \frac{\partial F}{\partial x}\frac{dx}{dt} + \frac{\partial F}{\partial y}\frac{dy}{dt} = \frac{2}{2x+3y}\,(2t) + \frac{3}{2x+3y}\,[3\cos(3t)] = \frac{4t + 9\cos(3t)}{2t^2 + 3\sin(3t)} \; .$$

Alternatively, we could write

$$F(x,y) = \ln[2t^2 + 3\sin(3t)]$$

and then use the rule, for taking derivative of a function of a single variable, to get the same result.

Case III: The function $F(x,y)$ is a function of two variables x and y, and the two variables x and y themselves are functions of two other variables, s and t; that is, $x(s,t)$ and $y(s,t)$. Then, in principle, F is a function of two variables, s and t. The derivatives of F with respect to s and t are

$$\frac{\partial F}{\partial s} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial s} \; ,$$

and

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial t} \; .$$

Example: Given the function $\boldsymbol{F(x,y) = x^2 + xy + y^2}$ with $\boldsymbol{x(s,t) = s + t}$ and $\boldsymbol{y(s,t) = st}$, determine the derivatives $\boldsymbol{\frac{\partial F}{\partial s}}$ and $\boldsymbol{\frac{\partial F}{\partial t}}$.

Solution:

Using the chain rule for partial derivatives,

$$\frac{\partial F}{\partial s} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial s} = (2x+y)(1) + (x+2y)(t) = (2s + 2t + st) + (s + t + 2st)t \; ,$$

and

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial t} = (2x + y)(1) + (x + 2y)(s) = (2s + 2t + st) + (s + t + 2st)s \ .$$

Alternatively, one could express $F(x, y)$ as $F(s, t)$ in the form

$$F(s, t) = (s + t)^2 + (s + t)st + s^2t^2 \ .$$

Taking the partial derivatives of $F(s, t)$ with respect to s or t will lead to the same results as those obtained by using the chain rule for partial derivatives.

1.3 WAVE EQUATION

As an application of partial derivatives, we will derive an equation, the wave equation, which describes a periodic function. A function is periodic when it repeats itself either in time or in space. Periodic phenomena are very common in a variety of fields including biology, physics, chemistry, astronomy, and so on. In biology, periodic phenomena include cell cycle, cardiac cycle, circadian rhythms, ovarian cycle, and metabolic cycle. In physics, examples of periodic phenomena include a swinging pendulum, mass on a spring, and water waves (or ripples). In chemistry, periodicity is found in the properties of chemical elements and in oscillating chemical reactions. In astronomy, there are abundant examples of periodic phenomena such as motion of satellites and the Moon around the Earth, motion of planets around the Sun, and the associated periodic occurrences of seasons, tides, and the day-and-night cycle.

One of the simplest mathematical functions that repeats itself is the sine (or cosine) function. For example, $\sin(kx)$ is a periodic function that repeats itself in spatial coordinate x. In fact, the value of this function at some point x_0 is same as its value at $x_0 + \frac{2\pi}{k}$. In other words, the wavelength [or, the length over which the function repeats itself] of this function is $\lambda = \frac{2\pi}{k}$. Similarly, $\sin(\omega t)$ is a periodic function that repeats itself in time t. The value of this function at some time t_0 is the same as its value at $t_0 + \frac{2\pi}{\omega}$. In other words, the period [or, the time over which the function repeats itself] of this function is $T = \frac{2\pi}{\omega}$. Note that $\frac{1}{T}$ measures the number of times the function repeats itself in a unit time. Therefore, $\frac{1}{T}$, that measures how frequently the function repeats itself, is called the frequency, f, of the wave. Thus, $f = \frac{1}{T} = \frac{\omega}{2\pi}$.

Now, a wave, according to its dictionary meaning, refers to “a disturbance on the surface of a liquid body, as the sea or a lake.” We can set up a water wave (commonly called a ripple) by throwing a large stone in a lake. This wave will look like a series of surges which are progressively moving outwardly away from the point where stone touched the water. If we throw a bottle cork in the disturbed water, we will observe that the cork will be simply bobbing up and down at its fixed location, without moving horizontally with the wave, as surges pass by it. It indicates that, in case of ripples in water, the wave is moving outward while the water itself is not moving with the wave. We can investigate the behavior of this wave either as a function of spatial coordinate x or as a function of time t. Using a camera, if we take the photograph of this wave, it will look like Figure 1.5 which shows, at a fixed time [the time at which photograph was taken], the wave as a function of x. On the other hand, we could focus our attention at some fixed point, say the bobbing cork with spatial coordinate x, and measure its displacement with respect to the horizontal level of calm water as a function of time t. Figure 1.6 shows this displacement, for a fixed value of x [location of cork], as a function of time t. From Figures 1.5 and 1.6 we note that this wave is periodic in both x and t. If we represent this wave mathematically by a sine function, we get a sinusoidal or a harmonic wave.

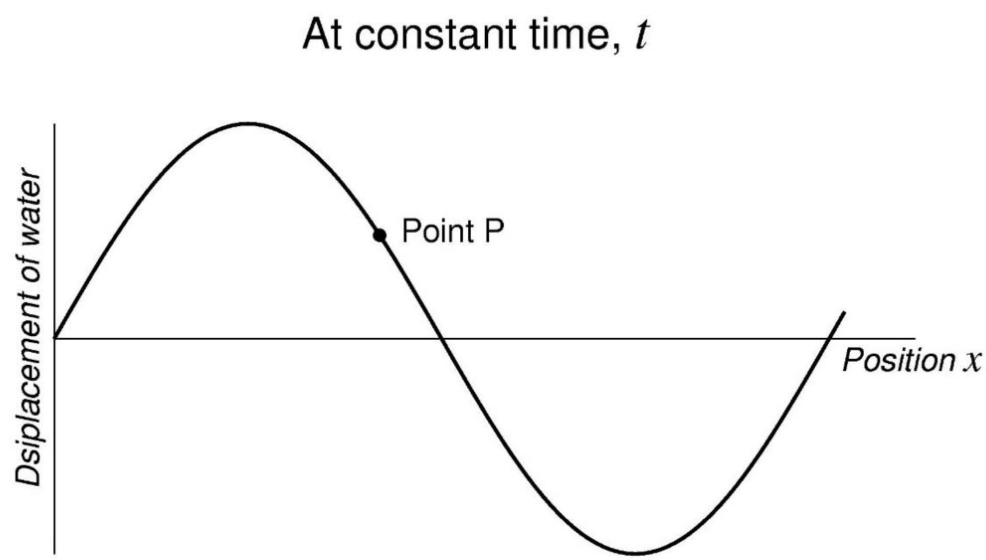


Figure 1.5. Periodic function $sin(kx - \omega t)$ as a function of x for a constant value of t.

We note, in passing, that because of the identity

$$\sin\left(\frac{\pi}{2} + \theta\right) = \cos(\theta) \ ,$$

for a given angle θ, the sine function and the cosine function look the same except the sine function is ahead of cosine function by a phase of $\pi/2$. In general, if a function looks like a sine or a cosine function, we will refer to it as a **sinusoidal function**.

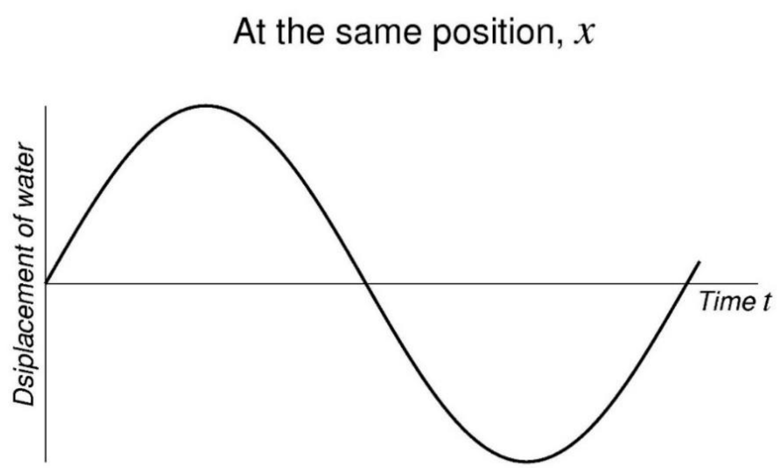


Figure 1.6. Periodic function $sin(kx - \omega t)$ as a function of t for a constant value of x.

The mathematical function

$$F(x,t) = A\sin(kx - \omega t) \qquad Eq. \ (1.13a)$$

represents the sinusoidal wave of figures 1.5 and 1.6. The function $F(x,t)$ represents the displacement, at time t, of a point in water located at position x. The largest value of the displacement is A, which is called the amplitude of the wave. The wavenumber k is related to the wavelength λ as $k = \frac{2\pi}{\lambda}$. The angular frequency ω is related to the period T as $\omega = \frac{2\pi}{T}$. The argument of the sine function, namely $(kx - \omega t)$, is called the phase of the wave. Each point on the wave, such as point P in Figure 1.5, has a fixed constant value of phase which does not change as the point P moves along with the wave. From Eq. (1.13a), as t increases, x also must increase to keep the phase constant. Thus, this wave travels along the $+x$ direction. Similarly, a wave of the form

$$F(x,t) = A\sin(kx + \omega t) \qquad Eq. (1.13b)$$

travels along the $-x$ direction.

The phase velocity refers to the speed of an arbitrary point, like P of some fixed phase, on the wave. For point P,

$$\text{phase} = \ kx - \omega t = \text{constant} \ .$$

Thus,

$$k\frac{dx}{dt} - \omega = 0$$

or

$$\frac{dx}{dt} = \frac{\omega}{k} = v \ . \qquad Eq. (1.14)$$

Here v is called the phase velocity of the wave. Thus,

$$F_+(x, t) = A \sin(kx - \omega t)$$

and

$$F_-(x, t) = A \sin(kx + \omega t)$$

are the sinusoidal (or harmonic) waves travelling along $+x$ and $-x$ direction, respectively. These waves are moving with a (phase) velocity of

$$v = \frac{\omega}{k} = \left(\frac{\lambda}{2\pi}\right)\left(\frac{2\pi}{T}\right) = \frac{\lambda}{T} \ .$$

Using partial derivatives of

$$F(x, t) = A \sin(kx - \omega t)$$

we get

$$\frac{\partial F}{\partial x} = kA \cos(kx - \omega t)$$

$$\frac{\partial^2 F}{\partial x^2} = -k^2 A \sin(kx - \omega t) = -k^2 F$$

$$\frac{\partial F}{\partial t} = -\,\omega A \cos(kx - \omega t)$$

$$\frac{\partial^2 F}{\partial t^2} = -\omega^2 A \sin(kx - \omega t) = -\omega^2 F$$

Combining these equations, we get

$$-k^2 \omega^2 F = \omega^2 \frac{\partial^2 F}{\partial x^2} = k^2 \frac{\partial^2 F}{\partial t^2} \ .$$

Finally, using $\omega = kv$, where v is the phase velocity, we have

$$\frac{\partial^2 F}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 F}{\partial t^2} = 0 \ . \qquad\qquad Eq. (1.15)$$

This is known as the $wave\ equation$. Equations of this kind, which relate various partial derivatives, are called *partial differential equations.*

In the above discussion, we derived the wave equation by starting from sinusoidal (or, sine- and cosine-like) functions. However, the general solutions of the wave equation are not necessarily sinusoidal. We will now show that general solutions of the wave equation are not merely functions of variables x and t, but are functions of the combinations $x + vt$ and $x - vt$. To show this, we make a change of variables from x and t to $r = x + vt$ and $s = x - vt$. Using chain rule for partial derivatives,

$$\frac{\partial F}{\partial x} = \frac{\partial F}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial F}{\partial s}\frac{\partial s}{\partial x} = \frac{\partial F}{\partial r} + \frac{\partial F}{\partial s} = \left(\frac{\partial}{\partial r} + \frac{\partial}{\partial s}\right) F \ ,$$

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial r}\frac{\partial r}{\partial t} + \frac{\partial F}{\partial s}\frac{\partial s}{\partial t} = \frac{\partial F}{\partial r} v + \frac{\partial F}{\partial s}(-v) = v\left(\frac{\partial}{\partial r} - \frac{\partial}{\partial s}\right) F \ .$$

Or, in operator notation,

$$\frac{\partial}{\partial x} \equiv \left(\frac{\partial}{\partial r} + \frac{\partial}{\partial s}\right) \ ,$$

$$\frac{\partial}{\partial t} \equiv v\left(\frac{\partial}{\partial r} - \frac{\partial}{\partial s}\right) \ .$$

Then,

$$\frac{\partial^2 F}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial x}\right) = \left(\frac{\partial}{\partial r} + \frac{\partial}{\partial s}\right)\left(\frac{\partial F}{\partial r} + \frac{\partial F}{\partial s}\right) = \frac{\partial^2 F}{\partial r^2} + 2\frac{\partial^2 F}{\partial r \partial s} + \frac{\partial^2 F}{\partial s^2} \ ,$$

$$\frac{\partial^2 F}{\partial t^2} = \frac{\partial}{\partial t}\left(\frac{\partial F}{\partial t}\right) = v\left(\frac{\partial}{\partial r} - \frac{\partial}{\partial s}\right) v\left(\frac{\partial F}{\partial r} - \frac{\partial F}{\partial s}\right) = v^2\left(\frac{\partial^2 F}{\partial r^2} - 2\frac{\partial^2 F}{\partial r \partial s} + \frac{\partial^2 F}{\partial s^2}\right) \ .$$

Thus, the wave equation becomes,

$$\frac{\partial^2 F}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 F}{\partial t^2} = 4\ \frac{\partial^2 F}{\partial r \partial s} = 0 \ .$$

It implies that

$$\frac{\partial}{\partial r}\left(\frac{\partial F}{\partial s}\right) = 0 \ ,$$

as well as

$$\frac{\partial}{\partial s}\left(\frac{\partial F}{\partial r}\right) = 0 \ .$$

In words, it implies that $\frac{\partial F}{\partial s}$ is independent of r and $\frac{\partial F}{\partial r}$ is independent of s. Thus, the function F can be expressed as a sum of two separate functions, g and h, such that g is a function of variable r only and h is a function of variable s only. So, a general solution of the wave equation is of the form,

$$F = g(r) + h(s) \ ,$$

or

$$F(x,t) = g(x+vt) + h(x-vt) \ .$$

We note in passing that the two sinusoidal functions that we encountered previously are indeed functions of variables $r = x + vt$ and $s = x - vt$. Explicitly, using $\omega = kv$,

$$F_+(x,t) = A\sin(kx - \omega t) = A\sin k(x - vt) \ , \qquad Eq.(1.16a)$$

and

$$F_-(x,t) = A\sin(kx + \omega t) = A\sin k(x + vt) \ . \qquad Eq.(1.16b)$$

1.4 IMPLICIT DERIVATIVES

Consider the relationship $y^2 + \sin x \sin y = x$ between two variables x and y. Both x and y can vary, though not independently since a change in x leads to a change in y and vice versa. Thus, either y can be treated as a function of a single variable x, or x can be treated as a function of the variable y. We can determine $\frac{dy}{dx}$, the rate of change of y with x, or $\frac{dx}{dy}$, the rate of change of x with y. We note that it is not easy to write either x as a function of y or y as a function of x. Starting with the relationship between x and y, we first differentiate it with respect to x to get

$$2y\frac{dy}{dx} + \cos x \sin y + \sin x \cos y \ \frac{dy}{dx} = 1 \ ,$$

or

$$(2y + \sin x \cos y)\ \frac{dy}{dx} = 1 - \cos x \sin y \ ,$$

or

$$\frac{dy}{dx} = \frac{1 - \cos x \sin y}{2y + \sin x \cos y} \ .$$

Again, starting with the relationship between x and y, we next differentiate it with respect to y to get

$$2y + \cos x \sin y \frac{dx}{dy} + \sin x \cos y \ = \frac{dx}{dy} \ ,$$

or

$$(1 - \cos x \sin y) \frac{dx}{dy} = 2y + \sin x \cos y \ ,$$

or

$$\frac{dx}{dy} = \frac{2y + \sin x \cos y}{1 - \cos x \sin y} \ .$$

These kinds of derivatives, which contain both variables x and y, are called implicit derivatives. Note in passing, in this example,

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} \ , \qquad Eq. (1.17)$$

which is a general property of derivatives of a function of a single variable.

As another **example**, consider the relationship $x + y = \exp(xy)$ between variables x and y. We first differentiate this relationship with respect to x to get

$$1 + \frac{dy}{dx} = \exp(xy) \left\{ x \frac{dy}{dx} + y \right\} \ ,$$

or

$$\{x \exp(xy) - 1\} \frac{dy}{dx} = 1 - y \exp(xy) \ ,$$

or

$$\frac{dy}{dx} = \frac{1 - y \exp(xy)}{x \exp(xy) - 1} \ .$$

Next, differentiate the original relationship with respect to y to get

$$\frac{dx}{dy} + 1 = \exp(xy)\left\{x + \frac{dx}{dy}y\right\} \ ,$$

or

$$\{1 - y\exp(xy)\}\,\frac{dx}{dy} = x\exp(xy) - 1 \ ,$$

or

$$\frac{dx}{dy} = \frac{x\exp(xy) - 1}{1 - y\exp(xy)} \ .$$

Again, the reciprocity relationship of Eq. (1.17), namely,

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} \ ,$$

is satisfied since we are dealing with a function of a single variable here. Now, let us explore whether this reciprocity relationship also works for partial derivatives. As an example, consider the coordinates of a point in a plane. The Cartesian (x, y) and the plane polar (ρ, ϕ) coordinates of the point are related to each other. The x and y coordinates can be expressed as functions of variables ρ and ϕ,

$$x = \rho\cos\phi, \ y = \rho\sin\phi \ .$$

Conversely, the ρ and ϕ coordinates can be considered as functions of variables x and y,

$$\rho = \sqrt{x^2 + y^2}, \phi = \tan^{-1}\left(\frac{y}{x}\right) \ .$$

From here,

$$\frac{\partial\phi}{\partial x} = \frac{-\frac{y}{x^2}}{1 + \left(\frac{y}{x}\right)^2} = -\frac{y}{x^2 + y^2} \ ,$$

and

$$\frac{\partial x}{\partial\phi} = -\rho\sin\phi = -y \ .$$

Clearly, $\frac{\partial\phi}{\partial x} \neq \frac{1}{\frac{\partial x}{\partial\phi}}$. The reason is that the variables being held fixed are different for each of the two cases. In the above example $\frac{\partial\phi}{\partial x}$ is actually $\left(\frac{\partial\phi}{\partial x}\right)_y$, that is, variable y is held fixed while evaluating this derivative. Similarly, $\frac{\partial x}{\partial\phi}$ is $\left(\frac{\partial x}{\partial\phi}\right)_\rho$ since variable ρ is fixed during evaluation of this derivative.

Legendre Transformation

Suppose $F(x,y)$ is a function of two independent variables x and y. It is possible to define two new variables u and v, which are combinations of x and y [for example, $u = x + y$ and $v = x - y$], and, inversely, x and y are functions of variables u and v. Now if x and y in F are replaced by u and v, then F will become a function of two new variables u and v. Thus,

$$F(x,y) \to F[x(u,v), y(u,v)] \equiv G(u,v) \ .$$

The procedure for replacing one set of independent variables in a function by another set of independent variables is accomplished, in general, by *Legendre transformation*. The change of independent variables is helpful in classical mechanics in discussion of canonically conjugate variables, and in thermodynamics in the discussion of Maxwell relations. The transformation procedure is described below.

For $F(x,y)$, we can write

$$dF = \frac{\partial F}{\partial x}\, dx + \frac{\partial F}{\partial y}\, dy = p\, dx + q\, dy \ ,$$

where $p(x,y) = \frac{\partial F}{\partial x}$ and $q(x,y) = \frac{\partial F}{\partial y}$. Using Clairaut theorem

$$\left(\frac{\partial p}{\partial y}\right)_x = \left(\frac{\partial q}{\partial x}\right)_y$$

Now, define three new functions,

$$f = F(x,y) - p(x,y)\, x \ ,$$

$$g = F(x,y) - q(x,y)\, y \ ,$$

$$\text{and} \qquad h = F(x,y) - p(x,y)x - q(x,y)y \ .$$

Then,

$$df = dF - p\, dx - x\, dp = -x\, dp + q\, dy \ ,$$

$$dg = dF - q\,dy - y\,dq = p\,dx - y\,dq \ ,$$

$$dh = dF - p\,dx - x\,dp - q\,dy - y\,dq = -x\,dp - y\,dq \ .$$

Thus, f is a function of variables p and y, g is a function of variables x and q, and h is a function of variables p and q. The partial derivatives of $f(p,y)$, $g(x,q)$ and $h(p,q)$ are

$$\frac{\partial f(p,y)}{\partial p} = -x \quad , \qquad \frac{\partial f(p,y)}{\partial y} = +q \ ,$$

$$\frac{\partial g(x,q)}{\partial x} = +p \quad , \qquad \frac{\partial g(x,q)}{\partial q} = -y \ ,$$

$$\frac{\partial h(p,q)}{\partial p} = -x \quad , \qquad \frac{\partial h(p,q)}{\partial q} = -y \ .$$

Finally, using the Clairaut theorem,

$$-\left(\frac{\partial x}{\partial y}\right)_p = +\left(\frac{\partial q}{\partial p}\right)_y \ , \qquad Eq.(1.18a)$$

$$+\left(\frac{\partial p}{\partial q}\right)_x = -\left(\frac{\partial y}{\partial x}\right)_q \ , \qquad Eq.(1.18b)$$

$$-\left(\frac{\partial x}{\partial q}\right)_p = -\left(\frac{\partial y}{\partial p}\right)_q \ . \qquad Eq.(1.18c)$$

These three relations along with the original relation

$$+\left(\frac{\partial p}{\partial y}\right)_x = +\left(\frac{\partial q}{\partial x}\right)_y \ , \qquad Eq.(1.18d)$$

are the basis of Maxwell's relations in thermodynamics. Appendix B describes an easy mnemonic device to remember Maxwell's relations with correct signs of terms.

PROBLEMS FOR CHAPTER 1

1. Given $F = \exp(x)\sin y$, $x = uv^2, y = u^2 v$, use the chain rule to find

$$u\frac{\partial F}{\partial u} + v\frac{\partial F}{\partial v} \ .$$

Write the result in terms of variables x and y only.

2. Consider a function of a single variable x, $F(x) = x^3 - 6x^2 + 9x + 4$.

(a) Determine the values of x at which the function $F(x)$ is an extremum.

(b) At the extremum points, does the function $F(x)$ have a minimum or a maximum value?

3. **Biomedical Physics Application.** A common inhabitant of human intestines is the bacterium Escherichia coli. A single cell of this bacterium in a nutrient-broth medium divides into two cells every twenty minutes. The initial population of a culture is 250 cells.

(a) Find an expression for the number of cells after t hours.

(b) Find the number of cells after 5 hours.

(c) When will the population reach 128,000 cells?

4. **Biomedical Physics Application.** At 8:00 AM in the morning, a biologist starts a six-hour experiment with bacterium Escherichia coli. As mentioned in problem 3, a cell of this bacterium in a nutrient-broth medium divides into two cells every twenty minutes. At 12:00 PM the biologist measures the cell population to be 2,048,000.

(a) What was the initial population of the bacterium cells in the beginning of the experiment at 8:00 AM?

(b) What will be the final population of the bacterium cells at the end of the experiment at 2:00 PM?

5. If $V(r,\theta) = (\alpha r^n + \beta r^{-n})\cos(n\theta)$ where α, β and n are constants, find

$$\frac{\partial V}{\partial r}, \frac{\partial^2 V}{\partial r^2} \text{ and } \frac{\partial^2 V}{\partial \theta^2} \ .$$

Hence show that

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \theta^2} = 0 \ .$$

6. The pressure P, volume V. and temperature T of a certain gas are related by the van der Waals' equation of state,

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \ ,$$

where a, b and R are all constants. Find the values of P, V and T for which $\left(\frac{\partial P}{\partial V}\right)_T = 0$ and $\left(\frac{\partial^2 P}{\partial V^2}\right)_T = 0$ are satisfied simultaneously. These particular values of pressure, volume and temperature are called critical values (P_c, V_c and T_c). One can define reduced parameters as

$$P_r = \frac{P}{P_c}, V_r = \frac{V}{V_c} \text{ and } T_r = \frac{T}{T_c} \ .$$

Show that the van der Waals' equation can be recast, in terms of the reduced parameters, in the following invariant form:

$$\left(P_r + \frac{3}{V_r^2}\right)\left(V_r - \frac{1}{3}\right) = \frac{8T_r}{3} \ .$$

7. **Biomedical Physics Application.** In all mammalians, including human beings, the rate of growth of skull is known to be different from the rate of growth of backbone. The allometric relationship between skull size S(t) and backbone length B(t), at age t, is

$$S(t) = a\,B(t)^b \ ,$$

where $a = 1.16$ and $b = 0.93$.

(a) Determine the relationship between the relative growth rates, $\frac{1}{S}\frac{dS}{dt}$ and $\frac{1}{B}\frac{dB}{dt}$, of skull and backbone, respectively.

(b) Which part of the body, skull or backbone, grows faster than the other?

8. If $f(x, y, z) = 1/(x^2 + y^2 + z^2)$, calculate $\nabla^2 f$, where $\boldsymbol{\nabla}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is called the Laplacian operator.

9. For $f(x, y) = x^3 - y^3 - 2xy + 6$, find the values of $\partial^2 f/\partial x^2$ and $\partial^2 f/\partial y^2$ at the points where $\frac{\partial f}{\partial x} = 0$ as well as $\frac{\partial f}{\partial y} = 0$.

10. Show explicitly that arbitrary functions $F(x - vt)$ and $F(x + vt)$ are solutions of the wave equation

$$\frac{\partial^2 F}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 F}{\partial t^2} = 0 \ .$$

11. **Biomedical Physics Application.** All human beings belong to one of the four main blood groups (types of blood) – A, B, AB and O. Blood group O is the most common type of blood in the world. Mixing blood groups can lead to a life-threatening situation. The blood type of a person is determined by three alleles (a variant form of a gene), A, B and O that are inherited from parents, one from father and other from mother. The inherited genes join to form the blood groups A (joining A with A or A with O), B (joining B with B or B with O), AB (joining A with B) and O (joining O with O). According to the Hardy-Weinberg Law of genetics, the fraction of population that carries two different alleles is

$$P = 2xy + 2yz + 2zx$$

where x, y and z are the fractions of alleles A, B and O in the population. Using the fact that $x + y + z = 1$, show using the Method of Lagrange Multipliers that P can be at most 2/3.

12. Assume that a and b are the side lengths of a right-angle triangle. The size of the hypotenuse, h, of this triangle is fixed at $5\sqrt{2}\ \mathrm{m}$. Using the Method of Lagrange Multipliers, determine the lengths of the sides a and b when the area of the triangle is extremum.

13. **Biomedical Physics Application.** The drug response function $R(t)$, describing the level of medication in the bloodstream after a drug is administered, can be represented as

$$R(t) = R_0\, t^{-3/2} \exp(-a/t)$$

where R_0 and a depend on the nature of the drug. For a particular drug $R_0 = 0.01$ and $a = 0.138$, and t is measured in minutes. Determine the time, t, at which the level of the medication in the bloodstream is maximum.

14. **Biomedical Physics Application.** In the angioplasty procedure, a "balloon" is inflated inside a partially clogged artery to restore the normal blood flow. The volume of blood flowing per unit time past a given point, F, is proportional to the fourth power of the radius, R, of the artery carrying the blood (Poiseuille's equation),

$$F = k\, R^4 \ .$$

What will be the relative change in F (that is, $\frac{dF}{F}$) when an artery is constricted by a 2% change in its radius due to clogging?

Chapter 2: Integral Calculus

In this chapter we will first review integral calculus, assuming that the reader has already seen integrals of functions of a single variable. After a short review, we will introduce the multiple integrals.

2.1 INDEFINITE INTEGRALS

An indefinite integral is defined as an antiderivative in the following sense. If a known function, $F(x)$, is represented as the derivative of an unknown function, $f(x)$, that is, $F(x) = \frac{df}{dx}$, then

$$\int F(x)\, dx = f(x) \ .$$

The function $f(x)$ is the antiderivative, or integral, of $F(x)$. In this case x is the $variable\ of\ integration$ and the known function $F(x)$ is the $integrand$. Since $\frac{dC}{dx} = 0$ if C is a constant, it is customary to write

$$\int F(x)\, dx = f(x) + C \ , \qquad Eq.(2.1)$$

where C is called the constant of integration. Some well-known indefinite integrals are

$$\int x^n\, dx = \frac{x^{n+1}}{n+1} + C \ ,$$

$$\int \exp(x)\ dx = \exp(x) + C \ ,$$

$$\int \frac{1}{x}\, dx = \ln x + C \ ,$$

$$\int \sin x\ dx = -\cos x + C \ ,$$

$$\int \cos x\ dx = \sin x + C \ ,$$

$$\int \sec^2 x\ dx = \tan x + C \ ,$$

$$\int \frac{1}{1+x^2}\, dx = \arctan x + C \ ,$$

$$\int \frac{1}{\sqrt{1 - x^2}}\, dx = \arcsin x + C \ .$$

The product rule of differential calculus [see Eq. (1.3)] corresponds to $integration\ by\ parts$ in integral calculus. In differential calculus, if $F(x) = f(x)\, g(x)$, then

$$\frac{dF}{dx} = f(x)\frac{dg(x)}{dx} + \frac{df(x)}{dx} g(x) = \frac{d[f(x)g(x)]}{dx} \ .$$

Using the definition of an indefinite integral as an antiderivative, it can be written as

$$\int f(x)\frac{dg(x)}{dx}\, dx + \int \frac{df(x)}{dx} g(x)\, dx = f(x)g(x) \ .$$

On rewriting this equation as

$$\int f(x)\frac{dg(x)}{dx}\, dx = f(x)g(x) - \int \frac{df(x)}{dx} g(x)\, dx \ , \qquad Eq. (2.2a)$$

we get the rule for *integration by parts,* which can be written in verbose form as

$$\int [FIRST][SECOND]\, dx$$

$$= [FIRST][Integral\ of\ SECOND] - \int [Derivative\ of\ FIRST][Integral\ of\ SECOND]\, dx \ .$$

Whenever the integrand can be expressed as a product of two functions, $FIRST$ and $SECOND$, it is convenient to identify as $FIRST$ the function whose derivatives are comparatively simpler than the function itself. Similarly, it is convenient to identify as $SECOND$ that function whose integrals are simpler than the function itself.

An alternate way of writing the $integration\ by\ parts$ procedure is

$$\int f(x)\, dg(x) = f(x)g(x) - \int g(x)\, df(x) \ . \qquad Eq. (2.2b)$$

Example: Evaluate $\boldsymbol{I = \int x^2 \exp x \ dx}$ using integration by parts.

Solution: In this case, the integrand is a product of two functions, x^2 and $\exp x$. Since x^2 becomes simpler on differentiation while $\exp x$ stays the same on differentiation or integration, we choose $FIRST$ as x^2 and $SECOND$ as $\exp x$. Then,

$$I = x^2 \exp x - \int (2x)\ \exp x\ dx \ .$$

Now, I is converted into a simpler integral, which can be evaluated using integration by parts one more time,

$$I = x^2 \exp x - 2\left[x \exp x - \int \exp x \, dx\right] = x^2 \exp x - 2x \exp x + 2 \exp x \ .$$

Finally, after including the constant of integration, the integral is evaluated as

$$I = (x^2 - 2x + 2) \exp x + C \ .$$

Example: Evaluate $I = \int x^2 \ln x \, dx$ using integration by parts.

Solution: In this case, $\ln x$ becomes simpler, compared to x^2, on taking derivatives. So, we choose $FIRST$ as $\ln x$ and $SECOND$ as x^2. Then,

$$I = \ln x \frac{x^3}{3} - \int \frac{1}{x}\frac{x^3}{3} \, dx = \ln x \frac{x^3}{3} - \frac{1}{3}\int x^2 \, dx = \ln x \frac{x^3}{3} - \frac{x^3}{9} \ .$$

After including the constant of integration, the integral is evaluated as

$$I = \frac{x^3}{9} [3 \ln x - 1] + C \ .$$

2.2 DEFINITE INTEGRALS

In a definite integral the range of the values of x, the variable of integration, is provided. If the range is $a \le x \le b$, then

$$\int_a^b F(x) \, dx = f(x)|_{x=a}^{x=b} = f(b) - f(a) \ , \qquad Eq. (2.3)$$

where $F(x) = \frac{df}{dx}$. Thus, we first determine the function $f(x)$ which is antiderivative is $F(x)$. Then, the value of the definite integral is the difference between the values of $f(x)$ at the upper limit and at the lower limit.

There is an alternate way of interpreting a definite integral. If $F(x)$ is a continuous function in the whole range of x, $a \le x \le b$, as shown in Figure 2.1, then we can divide the range into n equal intervals, each of width $\Delta x = (b - a)/n$. Also, if x_i, for $i = 1,2, \ldots n$ is the location of the midpoint of the ith interval, then the definite integral is defined as a limit of a sum as follows:

$$\int_a^b F(x) \, dx \to \lim_{n\to\infty} \sum_{i=1}^{n} F(x_i)\Delta x \ .$$

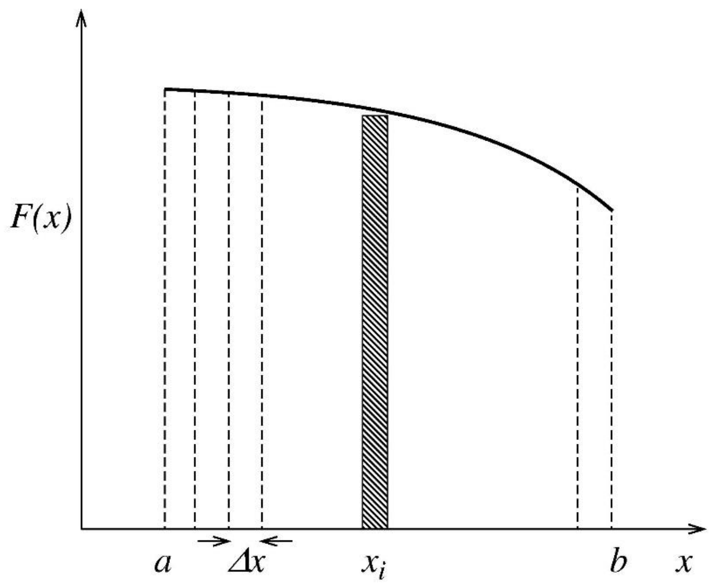


Figure 2.1. Area under the $\boldsymbol{F(x)}$ versus $\boldsymbol{x}$ curve is broken into strips.

In Figure 2.1 the area under the $F(x)$ versus x curve is broken into n strips, each of width Δx. Note that $F(x_i)\Delta x$ is the area of the cross-hatched ith strip shown in the figure. If the number of strips $n \to \infty$, then the width of each individual strip becomes vanishingly small and the sum of areas of all strips becomes equal to the area under the $F(x)$ versus x curve. The definite integral $\int_a^b F(x)\,dx$ thus represents the area under the $F(x)$ curve from $x = a$ to $x = b$.

Note that $\int_a^b F(x)\,dx$ is a number and its value does not depend on x. In fact,

$$\int_a^b F(x)\,dx = \int_a^b F(y)\,dy = \int_a^b F(z)dz\ .$$

Thus, in a definite integral, the symbol for variable of integration, x or y or z, is only a placeholder and it disappears after the integral is evaluated. So, in a definite integral the variable of integration is called the $dummy\ variable$. From the definition of a definite integral, it follows that

$$\int_a^b F(x)\,dx = f(b) - f(a) = -\,[f(a) - f(b)] = -\int_b^a F(x)\,dx$$

and

$$\int_a^a F(x)\,dx = 0\ .$$

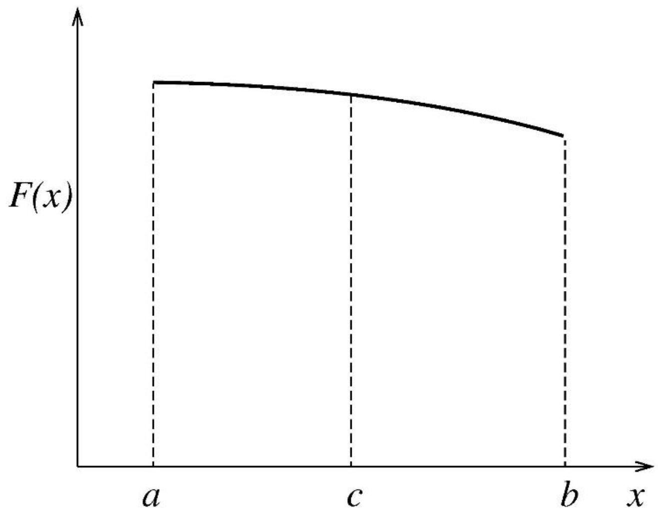


Figure 2.2. The point $\boldsymbol{x = c}$ lies inside the range, $\boldsymbol{a \le x \le b}$, of the integrand.

If the point $x = c$ lies somewhere in the middle of the range of x, then

$$\int_a^b F(x)\,dx = \text{area under the curve from } a \text{ to } b$$

$$= \text{area under the curve from } a \text{ to } c + \text{area under the curve from } c \text{ to } b$$

$$= \int_a^c F(x)\,dx + \int_c^b F(x)\,dx \ .$$

Also, if $F(x) \ge G(x)$ for $a \le x \le b$, then

$$\int_a^b F(x)\,dx \ \ge \int_a^b G(x)\,dx \ .$$

Similarly, if $F(x) \le G(x)$ for $a \le x \le b$, then

$$\int_a^b F(x)\,dx \ \le \int_a^b G(x)\,dx \ .$$

If $F(x)$ is positive for part of a range and negative for the remaining range of x, as shown in Figure 2.3, then

$$\int_a^b F(x)\,dx = \int_a^c F(x)\,dx + \int_c^b F(x)\,dx \equiv \ I_1 + I_2 \ ,$$

where $I_1 > 0$ and $I_2 < 0$. Note that even in this case the net area under the curve of a function $F(x)$ between $x = a$ and $x = b$ is $\int_a^b F(x)\, dx$. However, the magnitude of the area under the curve of $F(x)$ is $\int_a^b |F(x)| dx$.

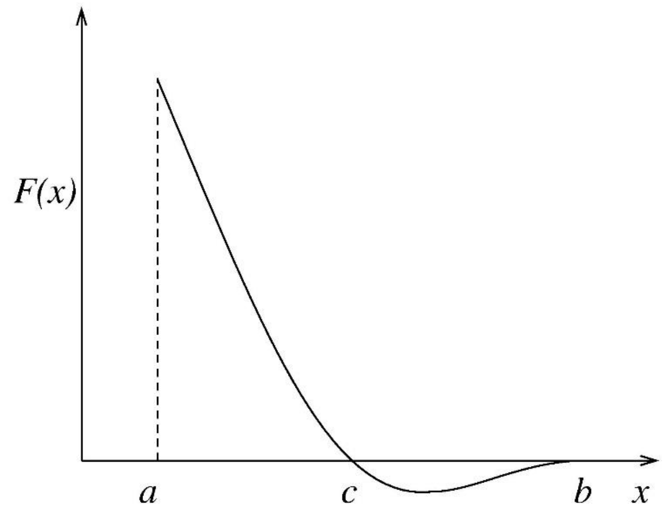


Figure 2.3. The function $\boldsymbol{F(x)}$ is positive as well as negative for parts of the range of $\boldsymbol{x}$.

Example: Find the net area and the magnitude of the area enclosed by function $\boldsymbol{F(x) = x^2 - 5x + 6 = (x-3)(x-2)}$ between $\boldsymbol{x = 0}$ and $\boldsymbol{x = 3}$ and the $\boldsymbol{x}$-axis.

Solution: In this case $F(x) \geq 0$ for $0 \leq x \leq 2$, and $F(x) \leq 0$ for $2 \leq x \leq 3$. The net area is

$$\int_0^3 F(x)\, dx = \int_0^3 (x^2 - 5x + 6)\, dx = \left(\frac{x^3}{3} - \frac{5x^2}{2} + 6x\right)\Bigg|_{x=0}^{x=3} = \frac{9}{2} - 0 = \frac{9}{2} .$$

To find the magnitude of the area under the curve, we break the integral from 0 to 2 and from 2 to 3. Then,

$$I_1 = \int_0^2 (x^2 - 5x + 6)\, dx = \left(\frac{x^3}{3} - \frac{5x^2}{2} + 6x\right)\Bigg|_0^2 = \frac{14}{3}$$

and

$$I_2 = \int_2^3 (x^2 - 5x + 6)\, dx = \left(\frac{x^3}{3} - \frac{5x^2}{2} + 6x\right)\Bigg|_2^3 = -\frac{1}{6} .$$

Clearly, $I_1 + I_2 = \frac{14}{3} - \frac{1}{6} = \frac{9}{2}$, same as the net area, but the magnitude of the area is

$$|I_1| + |I_1| = \left|\frac{14}{3}\right| + \left|-\frac{1}{6}\right| = \frac{29}{6} .$$

2.3 DIFFERENTIATION OF INTEGRALS: LEIBNIZ'S RULE

If one or both limits of a definite integral are functions of a variable, x, or if the integrand contains x as a parameter, then it is possible to differentiate the integral with respect to x. The example of such a definite integral is

$$I(x) = \int_{u(x)}^{v(x)} F(x,t)\, dt \ ,$$

where $u(x)$ and $v(x)$ are functions of x and t is the dummy variable. The derivative $\frac{dI}{dx}$ is evaluated using a method developed by Leibniz. To understand Leibniz's method of differentiating $I(x)$, we first look at three simpler cases of this integral, namely, $I_1(x), I_2(x), \text{and } I_3(x)$. In integral $I_1(x)$, the variable x appears only in the upper limit; in integral $I_2(x)$, the variable x appears only in the lower limit; and in integral $I_3(x)$, the variable x appears only in the integrand. If

$$I_1(x) = \int_a^x F(t)\, dt = \ f(x) - f(a) \ ,$$

then

$$\frac{dI_1}{dx} = \frac{df}{dx} = F(x) \ .$$

On generalizing, if

$$I_1(x) = \int_a^{v(x)} F(t)\, dt = \ f(v(x)) - f(a) \ ,$$

then

$$\frac{dI_1}{dx} = \frac{df(v(x))}{dx} = \frac{df(v(x))}{dv}\frac{dv(x)}{dx} = F(v(x))\ \frac{dv(x)}{dx} \ .$$

Next, if

$$I_2(x) = \int_x^b F(t)\, dt = \ f(b) - f(x) \ ,$$

then

$$\frac{dI_2}{dx} = -\frac{df}{dx} = -F(x) \ .$$

Again, on generalizing, if

$$I_2(x) = \int_{u(x)}^{b} F(t)\,dt = \ f(b) - f(u(x)) \ ,$$

then

$$\frac{dI_2}{dx} = -\frac{df\big(u(x)\big)}{dx} = -\frac{df\big(u(x)\big)}{du}\frac{du(x)}{dx} = -F\big(u(x)\big)\ \frac{du(x)}{dx} \ .$$

Finally, if

$$I_3(x) = \int_{a}^{b} F(x,t)\,dt \ ,$$

then

$$\frac{dI_3}{dx} = \int_{a}^{b} \frac{\partial F(x,t)}{\partial x}\ dt \ .$$

These three results of integrals $I_1(x), I_2(x),$ and $I_3(x)$ can be combined and written in a single form, known as Leibniz's rule,

$$\frac{dI(x)}{dx} = \frac{d}{dx}\left(\int_{u(x)}^{v(x)} F(x,t)\,dt\right) = F\big(x,v(x)\big)\frac{dv}{dx} - \ F\big(x,u(x)\big)\frac{du}{dx} + \int_{u(x)}^{v(x)} \frac{\partial F(x,t)}{\partial x}\ dt \ . \qquad Eq.(2.4)$$

Example: Determine $\frac{dI(x)}{dx}$ if

$$\boldsymbol{I(x) = \int_{x}^{2x} \frac{\exp(xt)}{t}\ dt} \ .$$

Solution: Using Leibniz's rule, we have

$$\frac{dI(x)}{dx} = \frac{\exp[x(2x)]}{2x}(2) - \frac{\exp[x(x)]}{x}(1) + \int_x^{2x} \frac{t\ \exp(xt)}{t}\ dt \ ,$$

or

$$\frac{dI(x)}{dx} = \frac{\exp[2x^2]}{x} - \frac{\exp[x^2]}{x} + \frac{\exp(xt)}{x}\bigg|_{t=x}^{t=2x} = \frac{2}{x}\left(\exp[2x^2] - \exp[x^2]\right) .$$

2.4 MULTIPLE INTEGRALS

Multiple integrals are generalizations of single integrals with one variable. For example, a double integral, I_2, with two variables, x and y, can be perceived as two single integrals.

$$I_2 = \int_{x=a_1}^{x=b_1} \int_{y=a_2}^{y=b_2} F(x,y)dy\,dx = \int_{x=a_1}^{x=b_1} f(x)dx \quad , \qquad Eq. (2.5)$$

where

$$f(x) = \int_{y=a_2}^{y=b_2} F(x,y)dy \ .$$

In evaluating the last integral for $f(x)$, the variable x inside the integrand is treated as a constant. Similarly, a triple integral, I_3, with three variables, x, y, and z, can be perceived as three single integrals.

$$I_3 = \int_{x=a_1}^{x=b_1} \int_{y=a_2}^{y=b_2} \int_{z=a_3}^{z=b_3} F(x,y,z)dz\,dy\,dx = \int_{a_1}^{b_1} f(x)dx \quad , \qquad Eq. (2.6)$$

where

$$f(x) = \int_{a_2}^{b_2} g(x,y)dy \ \ \text{and} \ \ g(x,y) = \int_{a_3}^{b_3} F(x,y,z)dz \ .$$

Again, in evaluating the integral $g(x, y)$, the variables x and y inside the integrand are treated as constants and in evaluating the integral for $f(x)$, the variable x inside its integrand is treated as a constant. Finally, if $F(x, y, z)$ is a separable function, it means that F can be expressed as a product of three separate functions with each function containing only one variable as $F(x, y, z) = h_1(x)h_2(y)h_3(z)$. So, for a separable integrand $F(x, y, z)$,

$$I_3 = \int_{a_1}^{b_1} h_1(x)dx \int_{a_2}^{b_2} h_2(y)dy \int_{a_3}^{b_3} h_3(z)dz$$

is also separable.

Example: Evaluate the double integral $\int_0^1 \int_0^1 (x-y)^2 dy\, dx$.

Solution: Consider the following integral,

$$\int_0^1 \int_0^1 (x-y)^2 dy\, dx = \int_0^1 \int_0^1 (x^2 - 2xy + y^2)\, dy\, dx$$

$$= \int_0^1 \left| x^2 y - 2x\frac{y^2}{2} + \frac{y^3}{3} \right|_{y=0}^{y=1} dx = \int_0^1 \left(x^2 - x + \frac{1}{3}\right) dx$$

$$= \left| \frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{3} \right|_0^1 = \frac{1}{6} \; .$$

2.5 BAG OF TRICKS

In this section we will introduce some tricks for evaluating somewhat complicated integrals by starting from a bit simpler integral which is commonly known. First, we look at some examples of indefinite integrals and follow it up by definite integrals (exponential and Gaussian integrals), which are useful in several areas of physics.

Trick Number 1: Sometimes a complicated integral, which may appear quite intractable, can be manipulated so that it is written as the derivative of a simpler $basic$ integral. In fact, it is possible to evaluate several new integrals by introducing a temporary parameter in the basic integral, and then differentiating both sides of the resulting integral with respect to this parameter to get the new integral.

Trick Number 1a: As the first example, indefinite integrals of the type

$$\int \frac{dx}{(x+a)^2}, \int \frac{dx}{(x+a)^3}, \ldots \qquad Eq.(2.7a)$$

are evaluated by starting from the $basic$ integral $\int \frac{dx}{x} = \ln(x)$. We introduce a parameter a in the basic integral by changing x to $x + a$ as

$$\int \frac{dx}{x+a} = \ln(x+a) \;\; . \qquad Eq.(2.7b)$$

After having introduced the parameter a, we can now differentiate both sides of the resulting integral with respect to a to evaluate the required integrals.

Example: Evaluate the indefinite integral $\int \frac{dx}{(x+a)^3}$.

Solution: Differentiate both sides of Eq. (2.7b) twice with respect to parameter a to get $(x+a)^3$ in the denominator. The first differentiation gives

$$-\int \frac{dx}{(x+a)^2} = \frac{1}{x+a} \ .$$

The second differentiation leads to

$$-(-2)\int \frac{dx}{(x+a)^3} = -\frac{1}{(x+a)^2} \ ,$$

which simplifies to the required integral

$$\int \frac{dx}{(x+a)^3} = -\frac{1}{2\,(x+a)^2} \ .$$

Trick Number 1b: Next, we can evaluate integrals of the type

$$\int \frac{1}{(a^2 - x^2)^{3/2}}\,dx\,, \int \frac{1}{(a^2 - x^2)^{5/2}}\,dx\,, \ldots \qquad Eq.(2.8a)$$

by starting with the $basic$ integral $\int \frac{1}{\sqrt{1-x^2}}\,dx = \arcsin x$. On replacing x by x/a in the basic integral, we get the integral with parameter a,

$$\int \frac{1}{\sqrt{a^2 - x^2}}\,dx = \arcsin\left(\frac{x}{a}\right) \ . \qquad Eq.(2.8b)$$

By differentiating both sides of this integral with respect to parameter a, we can evaluate integrals appearing in Eq. (2.8a).

Next, there are similar indefinite integrals of the type,

$$\int \frac{x}{(a^2 - x^2)^{3/2}}\,dx\,, \int \frac{x}{(a^2 - x^2)^{5/2}}\,dx\,, \ldots \qquad Eq.(2.9a)$$

In this case, the $basic$ integral is

$$\int \frac{x}{\sqrt{1 - x^2}}\,dx = -\sqrt{1 - x^2} \ ,$$

which can be evaluated easily using substitution $u^2 = 1 - x^2$. On replacing x by x/a in the basic integral, where a is a parameter, we get

$$\int \frac{x}{\sqrt{a^2 - x^2}}\, dx = -\sqrt{a^2 - x^2}\ . \qquad Eq. (2.9b)$$

Differentiating both sides of this integral with respect to parameter a leads to evaluation of integrals of Eq. (2.9a).

Example: Evaluate the indefinite integral $\int \frac{x}{(a^2 - x^2)^{3/2}}\, dx$ as well as the definite integral $\int_0^{a/2} \frac{x}{(a^2 - x^2)^{3/2}}\, dx$.

Solution: Differentiating both sides of Eq. (2.9b) with respect to a gives

$$\int \left[-\frac{1}{2}\frac{2a}{(a^2 - x^2)^{3/2}}\right] x\, dx = -\frac{1}{2}\frac{2a}{(a^2 - x^2)^{\frac{1}{2}}}\ ,$$

or

$$\int \frac{x}{(a^2 - x^2)^{3/2}}\, dx = \frac{1}{\sqrt{a^2 - x^2}}\ .$$

The definite integral is obtained by substituting the upper and the lower limits,

$$\int_0^{a/2} \frac{x}{(a^2 - x^2)^{3/2}}\, dx = \frac{2}{\sqrt{3}\, a} - \frac{1}{a}\ .$$

Trick Number 1c: Several indefinite integrals of the type

$$\int \frac{1}{(x^2 + a^2)^2}\, dx\, , \int \frac{1}{(x^2 + a^2)^3}\, dx\, , \ldots \qquad Eq. (2.10a)$$

can be evaluated by starting from the $basic$ integral $\int \frac{1}{x^2+1}\, dx = \arctan x$. Introduce a parameter a by replacing x by x/a in the basic integral to get

$$\int \frac{1}{x^2 + a^2}\, dx = \frac{1}{a}\arctan\left(\frac{x}{a}\right)\ . \qquad Eq. (2.10b)$$

Now, multiple differentiations of both sides of this integral with respect to parameter a lead to the required integrals of Eq. (2.10a).

Example: Using the integral $\int_0^\infty \frac{1}{1+x^2}\, dx = \frac{\pi}{2}$ as a guide, introduce a parameter and then differentiate both sides with respect to this parameter to evaluate the integral

$$\int_0^\infty \frac{\mathbf{1}}{(\boldsymbol{x}^2+\boldsymbol{y}^2)^3}\,\boldsymbol{dx} \ .$$

Solution: In the guiding integral $\int_0^\infty \frac{dx}{1+x^2} = \frac{\pi}{2}$, introduce a parameter y by changing x to x/y,

$$\int_0^\infty \frac{1}{1+\left(\frac{x}{y}\right)^2}\frac{dx}{y} = \frac{\pi}{2} \quad \text{or} \quad \int_0^\infty \frac{dx}{x^2+y^2} = \frac{\pi}{2y} \ .$$

Now, to get $(x^2+y^2)^3$ in the dominator, we differentiate both sides of the integral with respect to parameter y twice. The first differentiation gives

$$\int_0^\infty \left[-\frac{2y}{(x^2+y^2)^2}\right] dx = \frac{\pi}{2}\left[-\frac{1}{y^2}\right] ,$$

or

$$\int_0^\infty \frac{dx}{(x^2+y^2)^2} = \frac{\pi}{4}\frac{1}{y^3} \ .$$

The second differentiation gives

$$\int_0^\infty \left[-\frac{2(2y)}{(x^2+y^2)^3}\right] dx = \frac{\pi}{4}\left[-\frac{3}{y^4}\right] ,$$

or

$$\int_0^\infty \frac{dx}{(x^2+y^2)^3} = \frac{3\pi}{16}\frac{1}{y^5} ,$$

which is the value of the required integral.

Next, somewhat similar integrals of the type

$$\int \frac{x}{(x^2+\ a^2)^2}\,dx\,, \int \frac{x}{(x^2+\ a^2)^3}\,dx\,, \ldots \qquad Eq.(2.11a)$$

can be evaluated by starting from the *basic* integral

$$\int \frac{x}{x^2+\ 1}\,dx = \frac{1}{2}\ln(x^2+\ 1) \ .$$

This integral is easily evaluated by using the substitution $u = x^2+\ 1$. On replacing x by x/a, where a is a parameter, we get

$$\int \frac{x}{x^2 + a^2}\, dx = \frac{1}{2}\ln(x^2 + a^2) - \ln a \ . \qquad Eq. (2.11b)$$

Differentiating both sides of this integral with respect to parameter a provides the integrals of Eq. (2.11a).

Trick Number 2: Continuing with our bag of tricks to solve indefinite integrals, let us evaluate integrals of the form $\int \frac{P(x)}{Q(x)} dx$, where $P(x)$ and $Q(x)$ are polynomials of x.

Trick Number 2a: Our first case is the one in which degree of polynomial $P(x)$ is less than the degree of $Q(x)$. Examples of such integrals are

$$\int \frac{dx}{x^2 - x - 6} \qquad \text{or} \qquad \int \frac{x\, dx}{x^2 - 7x + 12} \ .$$

If the denominator can be factored, then use partial fractions to break the integrand of the first integral as

$$\frac{1}{x^2 - x - 6} = \frac{1}{(x-3)(x+2)} = \frac{A}{x-3} + \frac{B}{x+2} \ .$$

To determine A and B, multiply both sides by $(x^2 - x - 6)$ to get

$$1 = A(x+2) + B(x-3) \ .$$

Since this is true for any values of x, set $x = 3$ and $x = -2$ successively, to get $A = \frac{1}{5}$ and

$B = -\frac{1}{5}$. Thus,

$$\frac{1}{x^2 - x - 6} = \frac{1}{5}\left[\frac{1}{x-3} - \frac{1}{x+2}\right] \ .$$

The original integral then becomes

$$\int \frac{dx}{x^2 - x - 6} = \frac{1}{5}\int \frac{dx}{x-3} - \frac{1}{5}\int \frac{dx}{x+2} = \frac{1}{5}\ln(x-3) - \frac{1}{5}\ln(x+2) = \frac{1}{5}\ln\left(\frac{x-3}{x+2}\right) \ .$$

Similarly, for the second integral,

$$\frac{x}{x^2 - 7x + 12} = \frac{x}{(x-4)(x-3)} = \frac{A}{x-4} + \frac{B}{x-3} \ ,$$

or, on multiplying both sides by $(x^2 - 7x - 12)$, we get

$$x = A(x-3) + B(x-4) \ .$$

On setting $x = 3$ and $x = 4$ separately, we get $A = 4$ and $B = -3$. So,

$$\int \frac{x\,dx}{x^2 - 7x + 12} = 4\int \frac{dx}{x-4} - 3\int \frac{dx}{x-3} = 4\ln(x-4) - 3\ln(x-3) \ .$$

Trick Number 2b: There are situations in which the denominator of the integrand cannot be factored easily. In this case, one can try to complete the square and make a substitution that will turn the integral into a simpler well-known integral. For example,

$$\int \frac{dx}{3x^2 - 6x + 7} = \int \frac{dx}{3(x-1)^2 + 4} \ .$$

Define a new variable, $u = \sqrt{3}(x-1)/2$, or $dx = 2\,du/\sqrt{3}$. With this substitution,

$$\int \frac{dx}{3x^2 - 6x + 7} = \frac{1}{2\sqrt{3}}\int \frac{du}{u^2 + 1} = \frac{1}{2\sqrt{3}}\arctan u = \frac{1}{2\sqrt{3}}\arctan\left[\frac{\sqrt{3}(x-1)}{2}\right] \ .$$

Trick Number 2c: If the numerator contains a polynomial $P(x)$ of degree one less than the degree of polynomial $Q(x)$ in the denominator, then write the numerator as

$$numerator = a(derivative\ of\ denominator) + b \ ,$$

where a and b are some numbers. In this manner the integral becomes tractable. As an example, consider the integral

$$\int \frac{x+4}{x^2 + 2x + 5}\,dx \ .$$

The numerator can be expressed as

$$x + 4 = \frac{1}{2}\frac{d}{dx}[x^2 + 2x + 5] + 3 \ ,$$

so that

$$\int \frac{x+4}{x^2 + 2x + 5}\,dx = \int \frac{dx}{x^2 + 2x + 5}\left\{\frac{1}{2}\frac{d}{dx}[x^2 + 2x + 5] + 3\right\} = \frac{1}{2}\ln[x^2 + 2x + 5] + \int \frac{3\,dx}{x^2 + 2x + 5} \ .$$

The remaining integral can be determined using “complete the square” technique as described in *Trick Number 2b*.

Trick Number 2d: Next, we consider the case in which the degree of polynomial $P(x)$ in the numerator is more than the degree of polynomial $Q(x)$ in the denominator. In this situation, it is best to use long division to simplify the integrand.

Example: Evaluate the indefinite integral,

$$\int \frac{\mathbf{3x^3 - 7x^2 + 17x - 3}}{\mathbf{x^2 - 2x + 5}}\, \mathbf{dx} \ .$$

Solution: Using long division,

$$\frac{3x^3 - 7x^2 + 17x - 3}{x^2 - 2x + 5} = 3x - 1 + \frac{2}{x^2 - 2x + 5} \ .$$

Thus,

$$I = \int \frac{3x^3 - 7x^2 + 17x - 3}{x^2 - 2x + 5} dx = \int 3x\, dx - \int 1dx + \int \frac{2\, dx}{x^2 - 2x + 5}$$

$$= \frac{3x^2}{2} - x + 2I_1 \ ,$$

where I_1 can be evaluated using “complete the square” technique as described in *Trick Number 2b*.

Trick Number 3: Integrals involving trigonometric functions are simplified using standard identities such as,

$$\cos^2 x + \sin^2 x = 1 \ ,$$

$$\sin(2x) = 2 \sin x \cos x \ ,$$

$$\cos(2x) = 2\cos^2 x - 1 = 1 - 2\sin^2 x \text{, etc} \ .$$

Example: Evaluate $= \int \mathbf{\sin^3 x \ \cos^2 x \ dx}$**.**

Solution: The given integral is

$$I = \int \sin^2 x \ \cos^2 x \ [\sin x\, dx] = -\int (1 - \cos^2 x) \ \cos^2 x \ d(\cos x) \ .$$

On substituting $u = \cos x$, we get

$$I = -\int (1 - u^2)\, u^2\, du = -\frac{u^3}{3} + \frac{u^5}{5} == -\frac{\cos^3 x}{3} + \frac{\cos^5 x}{5} \ .$$

Example: Evaluate $I = \int \sin^2 x \ \cos^2 x \ dx$.

Solution: Since $\sin^2 x \ \cos^2 x = \frac{1}{4}\sin^2(2x) = \frac{1}{8}[1 - \cos(4x)]$, we get

$$I = \frac{1}{8}\int [1 - \cos(4x)] \ dx = \frac{1}{8}\left[x - \frac{\sin(4x)}{4}\right] .$$

2.6 EXPONENTIAL AND GAUSSIAN INTEGRALS

Continuing with our bag of tricks, now we look at some examples of definite integrals. Two types of integrals that we commonly encounter in physics are exponential integrals and Gaussian integrals. As the names suggest, the exponential integrals contain the exponential function, $\exp(-x)$, as a part of the integrand. The Gaussian integrals contain the Gaussian function, $\exp(-x^2)$, as a part of the integrand. The exponential integrals, I_e, are defined as

$$I_e^n = \int_0^\infty x^n \exp(-ax) \ dx \quad , \qquad Eq. (2.12)$$

and the Gaussian integrals, I_g, are,

$$I_g^n = \int_0^\infty x^n \exp(-ax^2) \ dx \quad . \qquad Eq. (2.13)$$

Sometimes we also encounter Gaussian integrals, I_{gg}, over an extended range of variable x,

$$I_{gg}^n = \int_{-\infty}^\infty x^n \exp(-ax^2) \ dx \quad . \qquad Eq. (2.14)$$

In these integrals n, a positive integer, is called the order of the integral and a is a positive constant. In each case we will try to set up a *reduction formula* that relates the integral I^n to similar integrals of lower order, such as I^{n-1} and/or I^{n-2}. Using this reduction formula successively, one can relate I^n to simpler integrals I^1 and/or I^0. In case of the exponential integral, we have, for $n = 0$,

$$I_e^0 = \int_0^\infty \exp(-ax) \ dx \ = \frac{1}{a} .$$

Also, for $n \geq 1$, we can integrate by parts to get the reduction formula

$$I_e^n = \int_0^\infty x^n \exp(-ax)\ dx = \left. x^n \frac{\exp(-ax)}{-a} \right|_0^\infty - \int_0^\infty \frac{\exp(-ax)}{-a} \cdot nx^{n-1}\,dx = 0 + \frac{n}{a} I_e^{n-1}\ .$$

The first integrated term vanishes at both limits. At the upper limit, as $x \to \infty$, the factor $\exp(-ax)$ approaches zero faster than any increase in x^n. At the lower limit, x^n goes to zero since $n \geq 1$. This reduction formula can be used successively to obtain

$$I_e^n = \frac{(n)}{a}\frac{(n-1)}{a} I_e^{n-2} = \frac{(n)}{a}\frac{(n-1)}{a}\frac{(n-2)}{a} I_e^{n-3} = \cdots = \frac{(n)}{a}\frac{(n-1)}{a} \cdots \frac{(2)}{a}\frac{(1)}{a} I_e^0\ ,$$

or, on substituting the value of I_e^0 here,

$$I_e^n = \frac{n!}{a^{n+1}}\ . \qquad Eq.(2.15)$$

In case of Gaussian integrals, I_g^n, we need I_g^0 and I_g^1 to use the reduction formula. We postpone the evaluation of I_g^0 until after we have discussed curvilinear coordinates and multiple integrals in Chapter 9. The value of this integral, however, is,

$$I_g^0 = \int_0^\infty \exp(-ax^2)\ dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\ .$$

The integral I_g^1, on the other hand, can be evaluated easily by substituting $u = ax^2$,

$$I_g^1 = \int_0^\infty x \exp(-ax^2)\ dx = \frac{1}{2a}\int_0^\infty \exp(-u)\ du = \frac{1}{2a}\ .$$

To set up the reduction formula for Gaussian integrals, we use the fact

$$\frac{d}{dx}\exp(-ax^2) = -2ax \exp(-ax^2) \quad \text{or } x \exp(-ax^2) = -\frac{1}{2a}\frac{d}{dx}\exp(-ax^2)\ .$$

Thus, for $n \geq 2$, using integration by parts we obtain

$$I_g^n = \int_0^\infty x^{n-1} \cdot x \exp(-ax^2)\ dx = -\frac{1}{2a}\int_0^\infty x^{n-1} \cdot \frac{d}{dx}\exp(-ax^2)\ dx$$

$$= -\frac{1}{2a} x^{n-1} \cdot \exp(-ax^2)|_0^\infty + \frac{1}{2a}\int_0^\infty (n-1)x^{n-2}\exp(-ax^2)\ dx \qquad Eq.(2.16)$$

$$= 0 + \frac{(n-1)}{2a}\int_0^{\infty} x^{n-2}\exp(-ax^2)\ dx = \frac{(n-1)}{2a} I_g^{n-2}\ .$$

Using this reduction formula, we can write $I_g^2, I_g^4, I_g^6\ \cdots$ in terms of I_g^0, and $I_g^3, I_g^5, I_g^7 \cdots$ in terms of I_g^1. For $n = 2m$, where m is an integer,

$$I_g^{2m} = \frac{(2m)!}{(4a)^m (m)!} I_g^0\ .$$

Similarly, for $n = 2m + 1$,

$$I_g^{2m+1} = \frac{(m)!}{a^m} I_g^1\ .$$

Finally, the Gaussian integral, I_{gg}, over an extended range of variables is

$$I_{gg}^n = \int_{-\infty}^{\infty} x^n \exp(-ax^2)\ dx = \int_0^{\infty} x^n \exp(-ax^2)\ dx + \int_{-\infty}^{0} x^n \exp(-ax^2)\ dx\ .$$

In the second integral, change the variable from x to y via $x = -y$. Then,

$$I_{gg}^n = \int_0^{\infty} x^n \exp(-ax^2)\ dx - \int_{\infty}^{0} (-y)^n \exp(-ay^2)\ dy = I_g^n + (-1)^n\ I_g^n\ .$$

Thus, if n is an odd integer, then $I_{gg}^n = 0$, and if n is an even integer, then $I_{gg}^n = 2\ I_g^n$.

PROBLEMS FOR CHAPTER 2

1. Evaluate using partial fractions

$$\int_2^4 \frac{2x+1}{x^2-1}\,dx\ .$$

2. Evaluate the indefinite integral

$$\int \frac{x^4-2x^2+4x+1}{x^3-x^2-x+1}\,dx\ ,$$

using the bag of tricks.

3. Use Leibniz's rule to find dI/dx for

$$I = \int_{a-x}^{x^2} (x-t)\,dt$$

with $a > 0$.

Next, evaluate the integral I explicitly, and then find dI/dx.

4. Use Leibniz's rule to find the value of θ that provides the extremum value of the integral

$$I(\theta) = \int_{\frac{\pi}{2}-\theta}^{\frac{\pi}{2}+\theta} \left(t-\frac{\pi}{4}\right)\frac{\sin(t)}{t}\,dt\ .$$

5. Define an integral $I_n = \int (\ln x)^n\ dx$. Using integration by parts, derive the following reduction formula:

$$I_n = x\,(\ln x)^n - n\,I_{n-1}\ .$$

Use this reduction formula to determine

$$I_3 = \int (\ln x)^3\ dx\ .$$

6. Consider the integral

$$I_n = \int_0^{\pi/2} \sin^{\mathrm{n}} x\,dx\ .$$

Explicitly evaluate the first two integrals I_0 and I_1.

Using integration by parts and the identity $\cos^2 x = 1 - \sin^2 x$, derive the reduction formula for $n \geq 2$:

$$I_n = \frac{n-1}{n} I_{n-2} \ .$$

Using the reduction formula, evaluate the integrals

$$I_5 = \int_0^{\pi/2} \sin^5 x \, dx \quad \text{and} \quad I_6 = \int_0^{\pi/2} \sin^6 x \, dx \ .$$

7. Consider the integral

$$I_n = \int_{-a}^{a} (x^2 - a^2)^n \, dx \ .$$

Explicitly evaluate the first integral I_0.

Using integration by parts, derive the reduction formula

$$I_n = -\frac{2na^2}{2n+1} I_{n-1}$$

for $n \geq 1$. Use the reduction formula to determine the value of

$$I_3 = \int_{-a}^{a} (x^2 - a^2)^3 \, dx \ .$$

8. Introduce a parameter in the standard integral

$$\int \frac{1}{\sqrt{1 - x^2}} \, dx = \arcsin x$$

and then differentiate both sides with respect to the parameter to determine the value of

$$I = \int_0^{a/2} \frac{dx}{(a^2 - x^2)^{3/2}} \ .$$

9. **Biomedical Physics Application.** A gunshot wound victim is bleeding at the rate given by

$$r(t) = r_0 t^2 \exp(bt) \ ,$$

where $r_0 = 0.000008 \ ml/s^3$, and $b = 0.004 \ s^{-1}$. The ambulance takes five minutes after the shooting to arrive at the crime scene. How much total blood has left victim's body before the medical attention is provided to the victim?

A healthy adult has about 5 liters of blood circulating in the human body. Most adults can lose up to 14% of their blood before their vital signs begin to deteriorate. What would be the outcome of the victim?

10. **Biomedical Physics Application.** The West Nile Virus is a disease that is contracted from infected mosquitoes. Scientists at the Centers for Disease Control and Prevention monitor the population growth of mosquitoes in a controlled environment. The mosquito population in this environment starts with 1000 mosquitoes at the start of the summer and grows at the rate of

$$r(t) = (1000)\exp(0.1\,t)$$

mosquitoes per day. What is the total mosquito population at the end of the seventh week of summer?

11. **Biomedical Physics Application.** In physiology we learn that a typical healthy adult human being breathes 12 to 15 times per minute. The volume of air inhaled in a single breath is about 500 mL. Assuming that breathing is a periodic process taking 5 seconds from the beginning of inhalation to the end of exhalation, the *rate* of air flow into the lungs can be represented by the function

$$f(t) = 100\,\pi\sin\left(\frac{2\pi t}{5}\right)\ \frac{mL}{s}\ .$$

Determine the total volume of the air inhaled at any time t starting at $t = 0$. Provide the volume of air inhaled by a healthy human being at $t = 1s, 2s, 3s, 4s, \text{and } 5\ s$.

Chapter 3: Infinite Series

In this chapter the convergent or divergent behavior of an infinite series is investigated. Several tests for checking this behavior are outlined. Taylor series and Maclaurin series are described and are used to obtain an infinite series expansion of several simple functions.

3.1 AN INFINITE SERIES

Consider an infinite series

$$a_1 + a_2 + a_3 \ldots \ldots + a_n + \cdots$$

where each element, a_n, of the series is a separate constant. For this series we define the partial sum, $S(N)$, as the sum of its first N terms

$$S(N) = a_1 + a_2 + \cdots + a_N = \sum_{n=1}^{N} a_n \ .$$

The infinite series is said to converge if the partial sum $S(N)$ has a finite limit as $N \to \infty$. This limit is the value of the converging series; that is, if

$$\lim_{N\to\infty} [S(N)] = S \ ,$$

then we write

$$S = a_1 + a_2 + \cdots + a_n + \cdots = \sum_{n=1}^{\infty} a_n \ . \qquad Eq. (3.1)$$

If the partial sum does not have a finite limit as $N \to \infty$, then the series is divergent. Note in passing that in the summation notation, the running variable n can be replaced by any other symbol as follows:

$$S = \sum_{n=1}^{\infty} a_n \ \text{ or } \ S = \sum_{m=1}^{\infty} a_m \ \text{ or } \ S = \sum_{p=1}^{\infty} a_p \ .$$

Therefore, the running variable n or m or p in the summation notation is referred to as the $dummy\ variable$ in the same sense as the dummy variable of a definite integral; that is, it serves as a placeholder that disappears after the sum is evaluated. To determine the convergence of a series, we compare it with another series whose convergence or divergence is known. For the comparison purpose we use either a geometric series or a harmonic series.

Geometric Series

The series in which the ratio of two successive terms is constant, namely

$$a + ax + ax^2 + ax^3 + \cdots ax^{n-1} + \cdots \qquad Eq. (3.2a)$$

is known as a geometric series. Here a is the *first term* and x is the $common\ ratio$. The partial sum of first N terms of this series is

$$S(N) = a + ax + ax^2 + \cdots ax^{(N-1)} \ . \qquad Eq. (3.2b)$$

On multiplying this partial sum by the common ratio x, we get

$$xS(N) = ax + ax^2 + \cdots ax^{(N-1)} + ax^N \ . \qquad Eq. (3.2c)$$

Subtracting $\mathrm{Eq.}\,(3.2c)$ from $\mathrm{Eq.}\,(3.2b)$ results in

$$(1 - x)S(N) = a - ax^N \ ,$$

or

$$S(N) = a\frac{1 - x^N}{1 - x} \ . \qquad Eq. (3.2d)$$

Now if $x \geq 1$, then $S(N) \rightarrow \infty$ as $N \rightarrow \infty$, so that geometric series is divergent for $x \geq 1$. For $x < 1$, $x^N \rightarrow 0$ for $N \rightarrow \infty$ so that

$$\lim_{N\rightarrow\infty}[S(N)] \rightarrow \frac{a}{1 - x} \ . \qquad Eq. (3.2e)$$

Therefore, the geometric series, $a + ax + ax^2 + \cdots$, converges to $\frac{a}{1-x}$ for $x < 1$ and diverges for $x \geq 1$. Note that in a converging geometric series, the common ratio x can take any value as long as this value is less than 1.

Harmonic Series

Next, we consider a harmonic series. Unlike a geometric series, which can take several values depending on the value of x, there is only one harmonic series. Successive terms of the harmonic series are reciprocals of the natural integers as follows:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots$$

Even though each successive term in this series is smaller than the previous term, the series is divergent. The divergent nature of the series can be seen by grouping the terms as

$$1 + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \cdots + \frac{1}{16}\right) + \cdots$$

A typical pair of parentheses will include terms like

$$\underbrace{\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{n+n}}_{n\ terms} > \underbrace{\frac{1}{n+n} + \frac{1}{n+n} + \cdots + \frac{1}{n+n}}_{n\ terms} = \frac{n}{2n} = \frac{1}{2}\ .$$

Thus, for harmonic series,

$$1 + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \cdots + \frac{1}{16}\right) + \cdots > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$$

Now, the series on the right-hand side of this inequality is clearly divergent since the partial sum of this series is

$$S(N) = 1 + \frac{1}{2}(N-1) = \frac{N+1}{2}\ ,$$

and

$$\lim_{N\to\infty}[S(N)] \to \infty\ .$$

Thus, the harmonic series is known to be a divergent series.

3.2 TESTS FOR CONVERGENCE

We can use several different tests to check the convergence of an unknown series. Three common tests are the comparison test, the ratio test, and the integral test.

Comparison Test

Now, to test the convergent or divergent nature of a series, we compare it term-by-term with another series whose convergence or divergence is known. For example, if the infinite series $a_1 + a_2 + a_3 + \cdots$ is known to be convergent, then consider an unknown series, $b_1 + b_2 + b_3$ .... If $b_n \leq a_n$ for all values of n from 1 to infinity, then $\sum_n b_n \leq \sum_n a_n$ and, therefore, $\sum_n b_n$ is also convergent.

On the other hand, suppose the infinite series $u_1 + u_2 + u_3$ ... is known to be divergent. Then, consider an unknown series $v_1 + v_2 + v_3$ ... If $v_n \geq u_n$ for all values of n from 1 to infinity, then $\sum_n v_n \geq \sum_n u_n$ and, therefore, $\sum_n v_n$is also divergent. For this comparison test, a standard geometric series will serve as a prototype convergent series, and the harmonic series will serve as a prototype divergent series.

Example: Check the convergence of the following series.

$$\boldsymbol{S = \sum_{n=1}^{\infty} \frac{1}{2 + 3^n}} .$$

Solution: Since $\frac{1}{2+3^n} < \frac{1}{3^n}$ for any n, and $\sum_{n=1}^{\infty} \frac{1}{3^n}$ is a convergent geometric series with the common ratio $1/3$, the unknown series S is also convergent.

Example: Check whether the following series converges or diverges.

$$\boldsymbol{S = \sum_{n=1}^{\infty} \frac{\ln(n)}{n}} .$$

Solution: We compare this series with the harmonic series. Since $\frac{\ln(n)}{n} > \frac{1}{n}$ for $n \geq 3$ and $\sum_{n=1}^{\infty} \frac{1}{n}$ is the divergent harmonic series, so S is also a divergent series.

Ratio test of Cauchy and D'Alembert

The Cauchy and D'Alembert ratio test provides the ability to compare an unknown infinite series with the geometric series $1 + x + x^2 + \cdots$. We know that the above geometric series converges only if x, the ratio of two successive terms, is less than 1. Thus, for the unknown series $a_1 + a_2 + a_3 \ldots + a_n + a_{n+1} + \cdots$, with all $a_i > 0$, we look for the ratio of two successive terms, $\frac{a_{n+1}}{a_n}$. If

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} < 1, \text{the series is convergent,}$$

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} > 1, \text{the series is divergent, and} \qquad Eq.(3.3)$$

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 1, \text{the nature of the series is indeterminate.}$$

Example: Determine the converging or diverging behavior of the infinite series,

$$\boldsymbol{\frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \cdots + \frac{n}{(n+1)!} + \cdots = \sum_{n=1}^{\infty} \frac{n}{(n+1)!}} .$$

Solution: In this case,

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=\frac{n+1}{(n+2)!}\frac{(n+1)!}{n}=\frac{n+1}{(n+2)\,n}\to 0\ .$$

Thus, according to the ratio test, the series converges. The value of this series is 1 [see Example 3 at the end of this chapter].

Example: Check the convergence of the following series,

$$\left(\frac{1}{3}\right)\mathbf{1}!+\left(\frac{1}{3}\right)^{2}\mathbf{2}!+\left(\frac{1}{3}\right)^{3}\mathbf{3}!\ldots+\left(\frac{1}{3}\right)^{n}\boldsymbol{n}!+\cdots=\sum_{n=1}^{\infty}\left(\frac{1}{3}\right)^{n}\boldsymbol{n}!\ .$$

Solution: In this case,

$$\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=\frac{\left(\frac{1}{3}\right)^{n+1}(n+1)!}{\left(\frac{1}{3}\right)^{n}n!}=\frac{1}{3}(n+1)\to\infty\ .$$

Thus, the unknown series is divergent.

Integral test of Cauchy and Maclaurin

Once again, consider an infinite series of constants,

$$a_1+a_2+a_3\ldots\ldots+a_n+\cdots$$

Imagine a function, $f(x)$, that is a continuous and monotonically decreasing function of x with $f(n)=a_n$. Consider its partial sum

$$S(N)=\sum_{n=1}^{N}a_n\ .$$

Now consider a series of unit-width rectangles, with heights a_1, a_2, a_3, ...etc, as shown in Figure 3.1. From the figure

$$a_1+a_2+\cdots+a_{N-1}>\int_1^N f(x)dx\qquad\text{[dashed line]}\qquad,$$

and

$$a_2 + a_3 + \cdots + a_N < \int_1^N f(x)dx \qquad \text{[solid line]} \qquad .$$

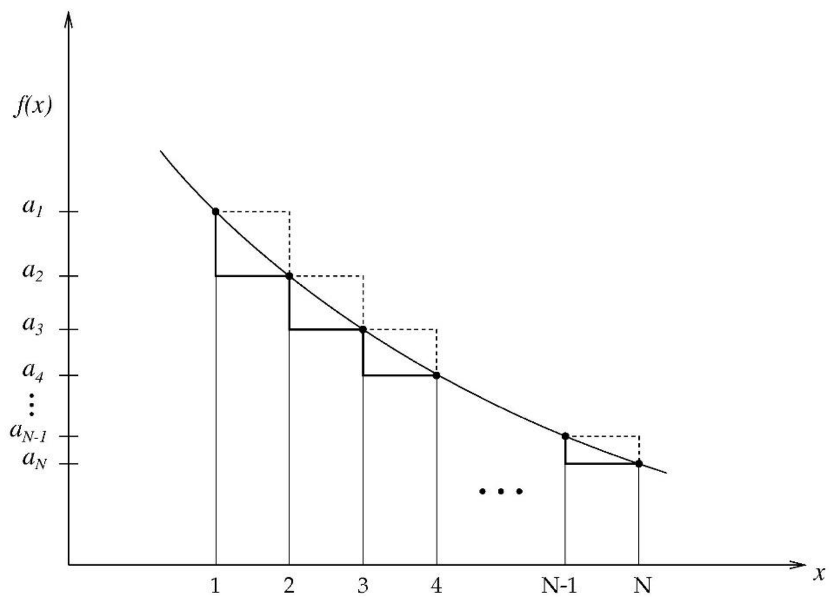


Figure 3.1. Integral test of Cauchy and Maclaurin.

That is,

$$S(N) > \int_1^{N+1} f(x)dx$$

and

$$S(N) < \int_1^N f(x)dx + a_1 \ .$$

On combining both the results, we get

$$\int_1^{N+1} f(x)dx < S(N) < \int_1^N f(x)dx + a_1 \ .$$

Now take the limit $N \to \infty$ to get,

$$\int_1^{\infty} f(x)dx < \sum_{n=1}^{\infty} a_n < \int_1^{\infty} f(x)dx + a_1 \ . \qquad Eq. (3.4)$$

Thus, the series $\sum_{n=1}^{\infty} a_n$ converges or diverges just as the integral $\int_1^{\infty} f(x)dx$ converges or diverges.

Example: Check the converging behavior of the Riemann Zeta function,

$$\boldsymbol{\zeta(p) = \sum_{n=1}^{\infty} \frac{1}{n^p}} \ .$$

Solution: We define $f(x) = \frac{1}{x^p}$ so that $f(n) = \frac{1}{n^p}$. Then,

$$\int_1^{\infty} f(x)dx = \int_1^{\infty} x^{-p} dx = \begin{cases} \left.\dfrac{x^{-p+1}}{-p+1}\right|_1^{\infty}, & \text{for } p \neq 1 \\ \ln(x)\,|_1^{\infty}, & \text{for } p = 1 \end{cases} \ .$$

The integral $\int_1^{\infty} f(x)dx$, therefore, diverges for $p \leq 1$ and converges for $p > 1$. The series for the Riemann Zeta function is thus convergent for $p > 1$. This is sometimes referred to as the p-series test.

Absolute Convergence

The series $\sum_{n=1}^{\infty} a_n$ is said to converge absolutely if the related series $\sum_{n=1}^{\infty} |a_n|$ also converges. An absolute convergence implies convergence but not vice versa. As an example, the series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$$

converges to $\ln 2$ [we will show it later in this chapter; see Eq. (3.12)]. But the corresponding absolute series,

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots \ ,$$

being the harmonic series, diverges. So, the original series is not absolutely convergent. An absolutely convergent series is important for two reasons: The product of two absolutely convergent series is another absolutely convergent series. The sum of the product of two absolutely convergent series is equal to the product of the sum of two individual series. That is,

$$S_a S_b + S_a S_c = S_a(S_b + S_c) \ .$$

If an infinite series is absolutely convergent, its sum is independent of the order in which the terms are added. Stated differently, if an infinite series is not absolutely convergent, the value of its sum will depend on the order in which the terms are added.

As an example, let us go back to the series

$$S = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \cdots \ ,$$

which is not absolutely convergent. Now we show explicitly that the value of this series depends on the order in which its terms are added. Separating out the positive and negative parts of this series, we get

$$S = \left(1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} \dots\right) - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \cdots\right) \ .$$

The terms in the second parenthesis can be further separated out as

$$S = \left(1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} \dots\right) - \left(\frac{1}{2} + \frac{1}{6} + \frac{1}{10} + \frac{1}{14} \dots\right) - \left(\frac{1}{4} + \frac{1}{8} + \frac{1}{12} + \frac{1}{16} \dots\right) \ .$$

Now, take out the common factor of $\frac{1}{2}$ in the second parenthesis and then easily combine the first and second parenthesis as

$$S = \left(\frac{1}{2} + \frac{1}{6} + \frac{1}{10} + \frac{1}{14} \dots\right) - \left(\frac{1}{4} + \frac{1}{8} + \frac{1}{12} + \frac{1}{16} \dots\right) \ .$$

On taking out an overall common factor of $\frac{1}{2}$, we get

$$S = \frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} \dots\right) \ ,$$

or

$$S = \frac{1}{2} S \ ,$$

which, of course, is an absurd result. The absurdity arises since this series is not absolutely convergent.

The second property of absolutely convergent series allows us to rearrange various series. Suppose $S_a = \sum_{n=1}^{\infty} a_n$ and $S_b = \sum_{n=1}^{\infty} b_n$ are two absolutely converging series, then $S = S_a S_b$ is the following double series:

$$S = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_m \, b_n = \sum_{m,n=1}^{\infty} c_{mn} \ .$$

In this double sum the summation is carried over rows (and columns) of the table shown in Figure 3.2. However, if we define

$$r = n + m - 1 \ ,$$

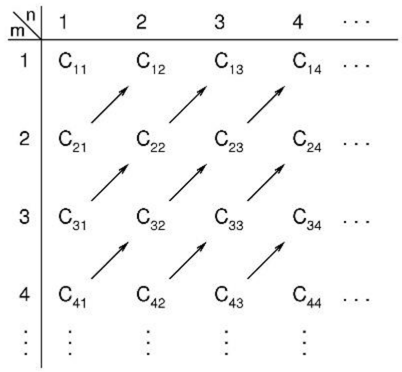


Figure 3.2. Evaluation of a double infinite sum.

then we note that r also takes all integer values over the range from 1 to ∞. So, we can replace the dummy variable m by variable r. Also, note $m = r - n + 1 \geq 1$ or $n \leq r$. Thus,

$$S = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} c_{mn} = \sum_{r=1}^{\infty}\sum_{n=1}^{r} c_{r-n+1,n} \quad . \qquad Eq. (3.5)$$

This summation is carried along the diagonal of the table in Figure 3.2. The advantage of this manipulation is that a double infinite sum is reduced to a single infinite sum plus a finite sum. This works only for absolutely convergent series.

3.3 SERIES OF FUNCTIONS

If each term of an infinite series is a function of a variable x, then the sum of the series, if it exists, is a function of x. Such a series looks like

$$a_0(x) + a_1(x) + a_2(x) + \cdots + a_n(x) + \cdots ,$$

with the sum of this infinite series being

$$S(x) = \sum_{n=0}^{\infty} a_n(x) = \lim_{N\to\infty} S(N; x) \; .$$

Here $S(N;x)$ is the partial sum, $S(N;x) = \sum_{n=0}^{N} a_n(x)$. In case of series of functions, we will start the sum from $n = 0$ instead of $n = 1$, as done previously for series of constant terms.

The dependence of $\lim\limits_{N\to\infty} S(N;x)$ on x is expressed through *uniform convergence* of the series. If for a certain range of values of x, $a \leq x \leq b$, and for any small number $\epsilon > 0$, it is possible to choose a number ν [which is independent of x] such that

$$S(x) - \epsilon < S(N; x) < S(x) + \epsilon \qquad \text{for all } N \geq \nu \ ,$$

or, alternatively,

$$|S(x) - S(N; x)| < \epsilon \qquad \text{for all } N \geq \nu \ ,$$

then the series converges uniformly to $S(x)$ in the range $a \leq x \leq b$. Thus, no matter how small ϵ is, it is always possible to choose a sufficiently large N such that the difference between the full sum $S(x)$ and the partial sum $S(N; x)$ is less than ϵ for all values of x between a and b. A good thing about a uniformly converging series is that it can be differentiated and integrated term by term. A series in x which converges uniformly in the interval $a \leq x \leq b$ can be integrated term by term provided the limits of the integration lie within $[a, b]$. Thus, if $S(x) = \sum_{n=0}^{\infty} a_n(x)$ is uniformly convergent, then

$$\int_p^q S(x)dx = \sum_{n=0}^{\infty} \int_p^q a_n(x)dx \ , \qquad Eq. (3.6a)$$

provided p and q lie within $[a, b]$. Similarly, a uniformly convergent series can be differentiated term by term within its range of convergence; that is, if $S(x) = \sum_{n=0}^{\infty} a_n(x)$, then

$$\frac{dS(x)}{dx} = \sum_{n=0}^{\infty} \frac{da_n(x)}{dx} \ . \qquad Eq. (3.6b)$$

Taylor and Maclaurin Series

The reverse process, in which a given function $S(x)$ is expanded into an infinite series, is called the power series expansion. Two power series expansions of interest to us are the Taylor and the Maclaurin series expansions. To gain more insight into the Taylor series expansion of a given function $S(x)$, consider expanding this function as

$$S(x) = c_0 + c_1(x - a) + c_2(x - a)^2 + \cdots + c_n(x - a)^n + \cdots \ , \qquad Eq. (3.7)$$

where the coefficients c_n are to be determined. This is the power series expansion of $S(x)$ about the point $x = a$. Define

$$S^{(n)}(x) = \frac{d^n S(x)}{dx^n} \text{ and } S^{(n)}(a) = S^n(x)|_{x=a} \ .$$

On setting $x = a$ in Eq. (3.7), we note that $c_0 = S(a)$. On taking the first derivative of the series in Eq. (3.7) and then setting $x = a$, we get

$$\frac{dS(x)}{dx} = 0 + c_1 + 2c_2(x - a) + \cdots$$

and

$$c_1 = \left.\frac{dS}{dx}\right|_{x=a} = S^{(1)}(a) \ .$$

Similarly, on taking the second derivative of the series in Eq. (3.7) and then setting $x = a$, we get

$$\frac{d^2S}{dx^2} = 2 \cdot 1c_2 + 3 \cdot 2c_3(x-a) + \cdots$$

and

$$c_2 = \left.\frac{1}{2!}\frac{d^2S}{dx^2}\right|_{x=a} = \frac{S^{(2)}(a)}{2!} \ .$$

Similarly, on taking the third derivative of the series in Eq. (3.7) and then setting $x = a$, we get

$$c_3 = \left.\frac{1}{3!}\frac{d^3S}{dx^3}\right|_{x=a} = \frac{S^{(3)}(a)}{3!} \ .$$

In general,

$$c_n = \left.\frac{1}{n!}\frac{d^nS}{dx^n}\right|_{x=a} = \frac{S^{(n)}(a)}{n!} \ .$$

Substituting the values of c_n in Eq. (3.7) provides the Taylor series expansion, namely

$$S(x) = \sum_{n=0}^{\infty} \frac{S^{(n)}(a)}{n!}(x-a)^n \ . \qquad\qquad Eq.(3.8a)$$

This is the Taylor series expansion of $S(x)$ about $x = a$. In general, the Taylor series can be truncated after N terms as

$$S(x) = \sum_{n=0}^{N-1} \frac{S^{(n)}(a)}{n!}(x-a)^n + R_N$$

where the remainder R_N is $\frac{S^{(N)}(b)}{N!}(x-a)^N$, for any b between a and x. The Taylor series expansion is convergent only if $R_N \xrightarrow[N\to\infty]{} 0$.

For $a = 0$,

$$S(x) = \sum_{n=0}^{\infty} \frac{S^{(n)}(0)}{n!} x^n \quad , \qquad Eq. (3.8b)$$

which is purely a power series expansion of $S(x)$ and is called the Maclaurin series for $S(x)$.

Example: Determine the Maclaurin series expansion of $S(x) = \exp x$.

Solution: In this case, we note that $S^{(n)}(x) = \exp x$ and $S^{(n)}(0) = 1$ for all n. The Maclaurin series becomes

$$\exp x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \qquad Eq. (3.9)$$

The convergence of the series can be checked by the ratio test since $\frac{a_{n+1}}{a_n} = \frac{x}{n+1} \to 0$ as $n \to \infty$ and x is finite. It is worth commenting that we can make expansion of $\exp x$ about any point (other than zero) using Taylor series.

Example: Determine Maclaurin series expansions of $\sin x$ and $\cos x$.

Solution: In this case,

$$f(x) = \sin x, \ f(0) = 0 \ ; \qquad g(x) = \cos x, \ g(0) = 1 \ ,$$

$$f^{(1)}(x) = \cos x, \ f^{(1)}(0) = 1 \ ; \qquad g^{(1)}(x) = -\sin x, \ g^{(1)}(0) = 0 \ ,$$

$$f^{(2)}(x) = -\sin x, \ f^{(2)}(0) = 0 \ ; \qquad g^{(2)}(x) = -\cos x, \ g^{(2)}(0) = -1 \ ,$$

$$f^{(3)}(x) = -\cos x, \ f^{(3)}(0) = -1 \ ; \qquad g^{(3)}(x) = \sin x, \ g^{(3)}(0) = 0 \ ,$$

$$f^{(4)}(x) = \sin x, \ f^{(4)}(0) = 0 \ ; \qquad g^{(4)}(x) = \cos x, g^{(4)}(0) = 1 \ .$$

Thus, in general, for $n = 0,1,2 \cdots$,

$$f^{(2n)}(0) = 0, \ f^{(2n+1)}(0) = (-1)^n \ ; \qquad g^{(2n)}(0) = (-1)^n, g^{(2n+1)}(0) = 0 \ .$$

Use these values in

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n \quad \text{and} \quad g(x) = \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{n!} x^n$$

to obtain

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \ , \qquad Eq.(3.10)$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \ . \qquad Eq.(3.11)$$

Example: Determine Maclaurin series expansion of $S(x) = \ln(1+x)$.

Solution: In this case, $S(0) = 0$ and

$$S^{(1)}(x) = (1+x)^{-1} \qquad S^{(1)}(0) = 1$$

$$S^{(2)}(x) = -(1+x)^{-2} \qquad S^{(2)}(0) = -1$$

$$S^{(3)}(x) = 2(1+x)^{-3} \qquad S^{(3)}(0) = 2!$$

$$S^{(4)}(x) = -3!\,(1+x)^{-4} \qquad S^{(4)}(0) = -3!$$

$$\vdots$$

$$S^{(n)}(x) = (-1)^{n-1}(n-1)!\,(1+x)^{-n} \quad n \geq 1 \qquad S^{(n)}(0) = (-1)^{n-1}(n-1)! \quad n \geq 1 \ .$$

Thus, the series for $\ln(1+x)$ is

$$S(x) = \ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \qquad Eq.(3.12)$$

Note in passing that $S(1) = \ln(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$ as mentioned earlier in this chapter.

Example: Determine Maclaurin series expansion of a binomial function (Binomial Theorem), namely

$$\boldsymbol{f(x) = (1+x)^m}$$

where m is a real number (positive or negative; integer or non-integer), and $|x| < 1$.

Solution: In this case, various derivatives of $f(x)$ are:

$$f^{(1)}(x) = m(1+x)^{m-1}$$

$$f^{(2)}(x) = m(m-1)(1+x)^{m-2}$$

$$f^{(3)}(x) = m(m-1)(m-2)(1+x)^{m-3} \ \vdots$$

$$f^{(n)}(x) = m(m-1)(m-2)\dots(m-n+1)(1+x)^{m-n} \ .$$

On setting $x = 0$, we get

$$f^{(n)}(0) = m(m-1)\dots(m-n+1) \ .$$

Thus, the Maclaurin series for the binomial function is

$$f(x) = (1+x)^m = 1 + mx + \tfrac{m(m-1)}{2!}x^2 + \tfrac{m(m-1)(m-2)}{3!}x^3 + \cdots \qquad Eq.(3.13)$$

As a particular case, if m is a positive integer, say N, then Nth derivative of $f(x)$ is

$$f^{(N)}(x) = N(N-1)(N-2)\dots(1) = N! \ ,$$

which is independent of x. So, all higher derivatives starting from $f^{(N+1)}(x)$ onwards are zero. In this case, the infinite series expansion of the binomial function $(1+x)^N$ reduces to a finite series. For $N = 2,3,4 \dots$ we have the well-known expansions

$$(1+x)^2 = 1 + 2x + x^2 \ ,$$

$$(1+x)^3 = 1 + 3x + 3x^2 + x^3 \ ,$$

$$(1+x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4 \ ,$$

etc.

Additional Examples

If a series of functions is a uniformly convergent series, then it can be differentiated and/or integrated to reduce an unfamiliar series to a familiar one. A few examples of this trick are discussed here.

Example 1: Consider the series $S(x)$,

$$S(x) = 1 + 2x + 3x^2 + 4x^3 + \cdots$$

where the range of convergence of this series is known to be $-1 < x < +1$. Integrate the series term-by-term to obtain a geometric series,

$$\int_0^x S(x)dx = x + x^2 + x^3 + x^4 + \cdots = \frac{x}{1-x} \ .$$

Now differentiate both sides of this expression to get

$$S(x) = \frac{d}{dx}\Big(\frac{x}{1-x}\Big) = \frac{1}{1-x} + \frac{x}{(1-x)^2} = \frac{1}{(1-x)^2} \ .$$

Example 2: Consider another series $S(x)$,

$$S(x) = \frac{1}{1.2} + \frac{x}{2.3} + \frac{x^2}{3.4} + \frac{x^3}{4.5} + \cdots$$

then,

$$x^2 S(x) = \frac{x^2}{1.2} + \frac{x^3}{2.3} + \frac{x^4}{3.4} + \frac{x^5}{4.5} + \cdots$$

On taking the second derivative of this series, it reduces to a geometric series, which can be easily summed as

$$\frac{d^2}{dx^2}[x^2 S(x)] = 1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x} \ .$$

Now, we integrate this expression twice to get $S(x)$. The first integration gives

$$\frac{d}{dx}\big(x^2 S(x)\big) = -\ln(1-x) + C_1 \ ,$$

and the second integration gives

$$x^2 S(x) = -\int \ln(1-x)\, dx + C_1 x + C_2 \ .$$

Here C_1 and C_2 are two constants of integration. The indefinite integral here can be evaluated using the substitution, $y = \ln(1-x)$, or equivalently, $x = 1 - \exp y$. Then,

$$\int \ln(1-x)\, dx = \int y[-\exp y] dy = -y \exp y + \int \exp y\, dy = (1-y)\exp y = (1-x)[1 - \ln(1-x)] \ .$$

So,

$$x^2 S(x) = -(1-x)[1 - \ln(1-x)] + C_1\, x + C_2 \ .$$

To evaluate the two constants of integration, we set $x = 0$ and recall that

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} \ldots \ ,$$

to obtain

$$C_2 = 1 \ .$$

Substitute this value of C_2 in the above expression for $S(x)$ and then divide both sides by x to get

$$xS(x) = 1 + \frac{(1-x)\ln(1-x)}{x} + C_1 \ .$$

Again, set $x = 0$ to obtain $C_1 = 0$. So, finally, the value of the original series $S(x)$ is

$$S(x) = \frac{1}{x} + \frac{1-x}{x^2}\ln(1-x) \ .$$

In some cases, even if the series does not contain a variable, one can still use the trick of differentiation and/or integration by introducing a variable judiciously.

Example 3: Consider the series

$$S = \tfrac{1}{2!} + \tfrac{2}{3!} + \tfrac{3}{4!} + \cdots + \tfrac{n-1}{n!} + \cdots \ .$$

We used this as an example of a converging series earlier; the value of this series was quoted as 1. Let us define

$$f(x) = \frac{x^2}{2!} + \frac{2x^3}{3!} + \frac{3x^4}{4!} + \frac{4x^5}{5!} + \cdots + \frac{(n-1)x^n}{n!} + \cdots$$

Clearly, $S = f(1) \text{ and } f(0) = 0$. Now,

$$\frac{df}{dx} = x + x^2 + \frac{x^3}{2!} + \frac{x^4}{3!} + \cdots + \frac{x^{n-1}}{(n-2)!} + \cdots$$

$$= x\left\{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \frac{x^{n-2}}{(n-2)!} + \cdots\right\} = x \exp x \ .$$

On integrating both sides, we get

$$\int_0^x \frac{df}{dx} dx = f(x) - f(0) = \int_0^x x \exp x \, dx \ .$$

Or

$$f(x) = x \exp x \,|_0^x - \int_0^x \exp x \, dx = x \exp x - \exp x + 1 = 1 + (x-1)\exp x \ .$$

Then, the value of the original series of constants is $S = f(1) = 1$.

PROBLEMS FOR CHAPTER 3

1. Find the value of α ($\alpha > 0$) if

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

2. Use the comparison test to check the convergence of following two series.

$$\sum_{n=1}^{\infty} \frac{5^n}{3^{2n}} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \ .$$

3. Use the Cauchy and D'Alembert ratio test to check the convergence of following two series [Useful information, $e = 2.72$].

$$\sum_{n=1}^{\infty} n \exp(-n) \qquad \text{and} \qquad \sum_{n=1}^{\infty} \frac{10^n}{(n!)^2} \ .$$

4. Use the Cauchy and Maclaurin integral test to check the convergence of following two series.

$$\sum_{n=1}^{\infty} \frac{1}{1+n^2} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{n}{1+n^2} \ .$$

5. Show that the first five terms in the Taylor series expansion of $\exp(-x)\cos x$ about $x = 0$ are

$$\exp(-x)\cos x = 1 - x + \frac{x^3}{3} - \frac{x^4}{6} + \frac{x^5}{30} \dots \ .$$

6. Expand the integrand below using binomial theorem

$$\frac{1}{2}\ln\left(\frac{1+x}{1-x}\right) = \int_0^x \frac{du}{1-u^2} \ ,$$

and integrate term by term to obtain the power series expansion of

$$\ln\left(\frac{1+x}{1-x}\right) \quad \text{for} \quad |x| < 1 \ .$$

7. By comparing with other well-known series, evaluate the sum of the following series exactly

$$1 + \ln 3 + \frac{(\ln 3)^2}{2!} + \frac{(\ln 3)^3}{3!} + \cdots$$

8. The classical expression for the kinetic energy of a particle is $KE = \frac{1}{2} m_o v^2$ where m_o is the mass of the particle when it is at rest and v is the speed of the particle. When the particle is moving with speed comparable

to c, where c is speed of light, the motion of the particle is correctly described by the theory of relativity. According to this theory, the mass of a particle moving with speed v is

$$m = \frac{m_o}{\sqrt{1 - \frac{v^2}{c^2}}} \ ,$$

and the kinetic energy of the particle is the difference between its total energy mc^2 and its energy at rest $m_o c^2$, that is, $KE = mc^2 - \ m_o c^2$.

(a) Using binomial expansion show that when v is very small compared to c, this expression for the kinetic energy of particle agrees with the classical expression.

(b) The leading relativistic correction term for the classical expression for the kinetic energy is of the form $\alpha \ m_o \frac{v^4}{c^2}$. What is the value of α?

9. **Biomedical Physics Application.** In the thirteenth century, Italian mathematician Fibonacci investigated the growth of rabbit population by assuming,

(a) rabbits never die,

(b) rabbits can mate at the age of one month with the gestation period of one month,

(c) the mating pair produces a new pair every month starting at the age of two months.

Starting with a newly born pair of rabbits, what is the number of rabbit pairs at the beginning of *n*th month? Here is the accounting:

At the beginning of first month, there is only one original pair, $a_1 = 1$.

At the beginning of second month, the pair mates, but there is still only one pair, $a_2 = 1$.

At the beginning of third month, there is original pair and a newly born pair, $a_3 = 2$.

At the beginning of fourth month, original pair produces a new pair and the one-month-old pair mates, $a_4 = 3$.

At the beginning of fifth month, original pair and two-month-old pair reproduce while one-month-old pair mates, $a_5 = 5$, etc.

At the beginning of *n*th month, number of pairs, a_n, equals the sum of newly born pairs [which is equal to the number of pairs at the beginning of (*n-2*)th month, a_{n-2}] plus the number of pairs living at the beginning of (*n-1*)th month, a_{n-1}. Explicitly, the first few terms in the Fibonacci series look like

1 + 1 + 2 + 3 + 5 + 8 + 13 + 21 +… ,

with the recursion relation $a_n = a_{n-1} + a_{n-2}$.

(i) If $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lambda$, then show that $\lambda = \frac{1+\sqrt{5}}{2}$.

(ii) Prove the following five relationships for elements of the Fibonacci series,

$$\frac{1}{a_{n-1}a_{n+1}} = \frac{1}{a_{n-1}a_n} - \frac{1}{a_n a_{n+1}} \ ,$$

$$\sum_{i=0}^{n-1} a_{2i+1} = a_{2n} \ ,$$

$$\sum_{i=1}^{n} a_{2i} = a_{2n+1} - 1 \ ,$$

$$\sum_{i=1}^{n} a_i = a_{n+2} - 1 \ ,$$

$$\sum_{i=1}^{n} a_i^2 = a_n a_{n+1} \ .$$

The Fibonacci series appears in several biological settings, such as number of leaves on a stem, number of branches in a tree as a function of height, designs of seashell or pinecone or artichoke or pineapple fruitlets, seeds in a sunflower, family tree of honeybees, etc.

Interlude

In this short interlude we introduce a few mathematical bits and pieces which will be helpful in our future journey of mathematics. We will introduce a unit imaginary number i, Kronecker Delta δ_{ij}, Dirac Delta function $\delta(x)$, Levi-Civita symbol ϵ_{ijk} and Euler's formula.

Unit Imaginary Number: $\boldsymbol{i = \sqrt{-1}}$

The square root of some positive numbers is easy to figure out and remember, such as $\sqrt{4} = 2, \sqrt{25} = 5$, etc. For other positive numbers, we may need a calculator, such as $\sqrt{3} = 1.732, \sqrt{7} = 2.646$, etc. But, our beloved calculator is not very helpful when we attempt to find out the square root of some negative numbers. For example, what is $\sqrt{-4}$ or $\sqrt{-7}$? For this purpose, we define the square root of -1 as a unit imaginary number, i. In terms of i we can write $\sqrt{-4} = \sqrt{(-1)(4)} = 2\,i$ or $\sqrt{-7} = \sqrt{(-1)(7)} = 2.646\,i$. Various powers of i are $i^2 = -1, i^3 = i^2 i = -i, i^4 = i^2 i^2 = 1$. Using the fact that $i^4 = 1$, we can reduce i^n, where n is a large number, to a much simpler form. As an example, $i^{65} = i^{16(4)+1} = i$ or $i^{135} = i^{33(4)+3} = i^3 = -i$ or $i^{206} = i^{51(4)+2} = i^2 = -1$, etc.

Kronecker Delta: $\boldsymbol{\delta_{ij}}$

Kronecker Delta is a symbol with two indices, i and j, and it can take two possible values, 0 and 1, depending on the values of the indices. The two indices, i and j, themselves can assume all possible discrete integer values from $-\infty$ to $+\infty$. Specifically,

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \qquad Eq.(I.1a)$$

Using this definition of Kronecker Delta, we can write

$$\sum_{i=-\infty}^{\infty} f(x_i)\,\delta_{ij} = f(x_j) \ . \qquad Eq.(I.1b)$$

Dirac Delta Function

In the definition of the Kronecker Delta above, its two indices, i and j, can assume all possible $discrete$ integer values from $-\infty$ to $+\infty$. Its analog in which these two indices can take all possible $continuous$ values is referred to as Dirac Delta function. To make this connection more understandable, we rewrite the Kronecker Delta in an alternate notation as

$$\delta(i-j) = 0 \;\; if \;\;\; i-j \neq 0 \qquad , \qquad [i,j = \cdots -2,-1,0,+1,+2 \cdots\ ] \qquad\qquad Eq.(I.2a)$$

along with

$$\sum_{i=-\infty}^{\infty} f(x_i)\,\delta(i-j) = f\big(x_j\big)\ . \qquad\qquad Eq.(I.2b)$$

Now, if δ is generalized from discrete variables i and j to all possible continuous values, x and y, then the Dirac Delta function should have the properties:

$$\delta(x-y) = 0 \;\; if \; x-y \neq 0 \qquad \text{or simply} \qquad \delta(x) = 0 \;\; if \; x \neq 0\ , \qquad Eq.(I.3a)$$

as well as

$$\int_{-\infty}^{\infty} f(x)\,\delta(x-y)\,dx = f(y) \qquad \text{or simply} \qquad \int_{-\infty}^{\infty} \delta(x)\,dx = 1\ . \qquad Eq.(I.3b)$$

The Eqs. ($I.3a$ and $I.3b$) imply that $\delta(x)$ cannot be zero for $x = 0$ because then the integral (which represents the area under the $\delta(x)$ versus x curve) will be zero, not 1. In order to understand the value of $\delta(0)$, consider a step function $F(x)$, of width 2ϵ around $x = 0$ and height $1/(2\epsilon)$,

$$F(x) = \begin{cases} \quad 0 \;\; for & |x| > \epsilon \\ 1/(2\epsilon) \quad for & \epsilon \geq x \geq -\epsilon \end{cases} \qquad .$$

The area under this step function (or, the integral of the function $F(x)$ over x from $-\infty$ to $+\infty$) is 1. On taking the limit $\epsilon \to 0$ in $F(x)$, the step function will have value 0 for $x \neq 0$, value ∞ for $x = 0$ and will preserve the area under the $F(x)$ versus x curve to be 1. Thus, our functional definition of the Dirac Delta function, $\delta(x)$, is any function which satisfies the following three conditions:

$$\delta(x) = 0 \;\; if \;\; x \neq 0,\;\; \delta(x) = \infty \;\; if \;\;\; x = 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} \delta(x)\,dx = 1\ . \qquad Eq.(I.4)$$

It is easy to show that the following limiting relationships, as $\epsilon \to 0$, satisfy all three defining conditions of the Dirac Delta function and, therefore, can be taken as definitions of the Delta function itself:

$$\delta(x) = \frac{1}{\pi}\lim_{\epsilon\to 0}\frac{\epsilon}{x^2+\epsilon^2} \quad , \qquad\qquad Eq.(I.5a)$$

$$\delta(x) = \frac{1}{\sqrt{\pi}}\lim_{\epsilon\to 0}\frac{1}{\epsilon}exp(-x^2/\epsilon^2) \quad , \qquad\qquad Eq.(I.5b)$$

$$\delta(x) = \frac{1}{\pi}\lim_{\epsilon\to 0}\ \frac{1}{x}sin\left(\frac{x}{\epsilon}\right)\ . \qquad\qquad Eq.(I.5c)$$

There are two comments worth mentioning here. First, unlike the Kronecker Delta which is a dimensionless number, the Dirac Delta function $\delta(x)$ has dimensions of $1/x$ for the integral of $Eq. (I.4)$ to be meaningful. Second, the Dirac Delta function defined in Eq. (I.4) is a one-dimensional Delta function which can be generalized to two-dimensional and three-dimensional Dirac Delta functions.

Levi-Civita symbol: $\boldsymbol{\epsilon_{ijk}}$

Analogous to a Kronecker delta, one can define the Levi-Civita symbol ϵ_{ijk} with three indices (ijk). The three indices (ijk) can take only three values $(1 \text{ or } 2 \text{ or } 3)$ each. The Levi-Civita symbol itself can have one of the three possible values, +1 or 0 or -1. Specifically,

$$\epsilon_{ijk} = \begin{cases} 0 \text{ if any two indices are equal} \\ +1 \text{ if an even exchange of } (ijk) \text{ gives } (123) \\ -1 \text{ if an odd exchange of } (ijk) \text{ gives } (123) \end{cases} \qquad Eq. (I.6)$$

Then, out of 27 (3 times 3 times 3) possible Levi-Civita symbols, only six are nonzero. These are $\epsilon_{123} = +1, \epsilon_{132} = -1, \epsilon_{213} = -1, \epsilon_{231} = +1, \epsilon_{312} = +1, \text{and } \epsilon_{321} = -1$. As we will see later, one can write the cross product of two vectors in a very compact form using the Levi-Civita symbol. Also, the value of a determinant of order 3 can be written using the Levi-Civita symbol.

The Levi-Civita symbol is related to Kronecker delta by

$$\sum_i \epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} \quad . \qquad Eq. (I.7)$$

A nice mnemonics device to remember the order of indices in this relationship is illustrated in the figure below so that, right-hand-side of the relation is $(first)(second) - (outer)(inner)$.

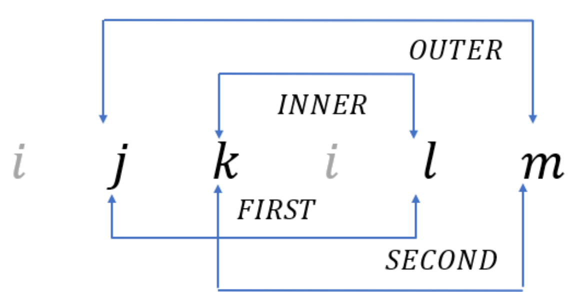


Figure I.1. Mnemonic for the order of indices in the Epsilon-Delta Identity

This relationship between Levi-Civita symbols and Kronecker deltas (called epsilon-delta identity) is very useful in deriving several identities of vector calculus as well as several properties of determinants. A proof of the epsilon-delta identity is given in Appendix C.

Euler's Formula

From Chapter 3, we recall the Maclaurin series expansion of some simple functions, such as:

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} ,$$

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

$$\exp x = \sum_{n=0}^{\infty} \frac{x^n}{n!} .$$

Using these expansions, it follows, with $i = \sqrt{-1}$, that

$$\exp(ix) = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} = \sum_{n=0,2,4\ldots} \frac{(ix)^n}{n!} + \sum_{n=1,3,5\ldots} \frac{(ix)^n}{n!} .$$

Now, on replacing the dummy index n by $2m$ in the first sum and n by $2m+1$ in the second sum, we get

$$\exp(ix) = \sum_{m=0}^{\infty} \frac{i^{2m} x^{2m}}{(2m)!} + \sum_{m=0}^{\infty} \frac{i^{2m+1} x^{2m+1}}{(2m+1)!}$$

$$= \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m}}{(2m)!} + i \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m+1}}{(2m+1)!} ,$$

or

$$\exp(ix) = \cos x + i \sin x \quad . \qquad Eq. (I.8)$$

This is known as Euler's formula. By many accounts, Euler's formula is the most beautiful equation of mathematics, or the jewel of mathematics. When written in the form $e^{i\pi} + 1 = 0$, it is made up of five different mathematical constants (namely, 0, 1, π, e, and i) each having its own independent value. Euler's formula will be very helpful in our further studies of Fourier series, Fourier transform, and complex variables.

PROBLEMS FOR INTERLUDE

1. Using properties of the Kronecker delta, show that

$$\sum_{i=1}^{3} \sum_{j=1}^{3} \delta_{ij}\, x_i x_j = \sum_{i=1}^{3} x_i^2 \ ,$$

and

$$\sum_{i=1}^{3} \sum_{j=1}^{3} \delta_{ij} \ = \sum_{i=1}^{3} \delta_{ii} = 3 \ .$$

2. Using properties of the Levi-Civita symbol, show that

$$\sum_{ij} \epsilon_{ijk}\, \epsilon_{ijm} = \ 2\delta_{km} \ ,$$

and

$$\sum_{ijk} \epsilon_{ijk}\, \epsilon_{ijk} = \ 6 \ .$$

3. Using Euler's formula, show that

$$\cos x = \frac{\exp(ix) + \exp(-ix)}{2} \ ,$$

$$\sin x = \frac{\exp(ix) - \exp(-ix)}{2i} \ .$$

Chapter 4: Fourier Series

Any periodic function can be expressed as an infinite series of sinusoidal (sine or cosine like) functions. This series is called the Fourier series of the periodic function. For a general nonperiodic function, the extension of the Fourier series leads to the Fourier transform. In this chapter, we will use orthogonality relationships of sinusoidal functions to derive coefficients in a Fourier series. We will also show how the Fourier series becomes the Fourier transform on increasing the periodicity of the periodic function.

4.1 PERIODIC FUNCTIONS

In preparation for our discussion of Fourier series and the Fourier transform, let us first get over some mathematical preliminaries. In our earlier discussion we derived an equation, the wave equation, whose solution describes any periodic function. In particular, we noted that a sinusoidal function

$$f(x,t) = \sin[kx \pm \omega t]$$

satisfies the wave equation with $k = 2\pi/\lambda,$ and $\omega = 2\pi/T$. We also noted that a periodic function that repeats itself both in time (t) as well as in a spatial coordinate (x) is not a function of variables x or t separately but is a function of dimensionless variables $kx \pm \omega t$. In general, a periodic function $f(\theta)$, of variable θ, repeats itself as the variable changes. The "periodicity" of such a repeating function is defined as the range of variable θ over which the function repeats itself. For example, if f is a function of time, it repeats itself after an amount of time called its period, T. If f repeats itself over space, then the length over which the function repeats itself is the wavelength, λ. So, we can define periodicity for these periodic functions as either T or λ. As another example, the periodicity of a simple sinusoidal function, $\sin\theta$ or $\cos\theta$, is 2π. The integral of these functions over a complete range of their periodicity is the same no matter what the starting (lower limit) or ending (upper limit) points of the integral are. Specifically, for $\alpha > 0$, we have

$$\int_{\alpha}^{\alpha+2\pi} \sin\theta\, d\theta = \int_{0}^{2\pi} \sin\theta\, d\theta + \int_{2\pi}^{\alpha+2\pi} \sin\theta\, d\theta - \int_{0}^{\alpha} \sin\theta\, d\theta = \int_{0}^{2\pi} \sin\theta\, d\theta \;,$$

$$\int_{\alpha}^{\alpha+2\pi} \cos\theta\, d\theta = \int_{0}^{2\pi} \cos\theta\, d\theta + \int_{2\pi}^{\alpha+2\pi} \cos\theta\, d\theta - \int_{0}^{\alpha} \cos\theta\, d\theta = \int_{0}^{2\pi} \cos\theta\, d\theta \;,$$

since in both cases a cancellation of integrals occurs in the middle step (on replacing the dummy variable θ by another variable ψ using $\theta = 2\pi + \psi$). Recall that a product of any two sinusoidal functions can be written as a linear combination of single sinusoidal functions, namely,

$$\sin\alpha\cos\beta = \frac{1}{2}[\sin(\alpha-\beta)+\sin(\alpha+\beta)] \ ,$$

$$\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta)-\cos(\alpha+\beta)] \ ,$$

$$\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha-\beta)+\cos(\alpha+\beta)] \ .$$

Thus, in general (with p and q integers),

$$\int_{\alpha}^{\alpha+2\pi} \sin(p\theta)\sin(q\theta)\,d\theta = \int_{0}^{2\pi} \sin(p\theta)\sin(q\theta)\,d\theta \ ,$$

$$\int_{\alpha}^{\alpha+2\pi} \cos(p\theta)\cos(q\theta)\,d\theta = \int_{0}^{2\pi} \cos(p\theta)\cos(q\theta)\,d\theta \ ,$$

$$\int_{\alpha}^{\alpha+2\pi} \cos(p\theta)\sin(q\theta)\,d\theta = \int_{0}^{2\pi} \cos(p\theta)\sin(q\theta)\,d\theta \ .$$

Orthogonality Relations

Recall that for an even function of θ, $f(-\theta)=f(\theta)$, and for an odd function, $f(-\theta)=-f(\theta)$. In our discussion of Fourier series and the Fourier transform, we will encounter integrals whose integrands are products of either two sine (ss), two cosine (cc), or a cosine and a sine (cs) functions. The complete range of periodicity of these functions, namely 2π, is taken to be from $-\pi$ to $+\pi$ so that $\sin\theta$ is an odd function and $\cos\theta$ is an even function of θ over this complete range. Assuming that p and q are *nonzero* integers, we have

$$I_{ss} = \int_{-\pi}^{\pi} \sin(p\theta)\sin(q\theta)\,d\theta$$

$$= -\frac{1}{2}\int_{-\pi}^{\pi} d\theta\{\cos[(p+q)\theta]-\cos[(p-q)\theta]\} = -\frac{1}{2}\left\{\frac{\sin(p+q)\theta}{p+q}-\frac{\sin(p-q)\theta}{p-q}\right\}\Bigg|_{-\pi}^{\pi}$$

$$= \begin{cases} 0 & \text{if} \quad p \neq q \\ \pi & \text{if} \quad p = q \end{cases}$$

$$\text{or} \qquad I_{ss} = \pi\delta_{pq} \ . \qquad\qquad Eq.(4.1a)$$

Similarly,

$$I_{cc} = \int_{-\pi}^{\pi} \cos(p\theta)\cos(q\theta)\, d\theta$$

$$= \frac{1}{2}\int_{-\pi}^{\pi} d\theta\{\cos[(p+q)\theta] + \cos[(p-q)\theta]\} = \frac{1}{2}\left\{\frac{\sin(p+q)\theta}{p+q} + \frac{\sin(p-q)\theta}{p-q}\right\}_{-\pi}^{\pi}$$

$$= \begin{cases} 0 & \text{if} \quad p \neq q \\ \pi & \text{if} \quad p = q \end{cases}$$

or $\quad I_{cc} = \pi\delta_{pq}$. $\qquad Eq. (4.1b)$

Finally,

$$I_{cs} = \int_{-\pi}^{\pi} \cos(p\theta)\sin(q\theta)\, d\theta$$

$$= \frac{1}{2}\int_{-\pi}^{\pi} d\theta\{\sin[(p+q)\theta] - \sin[(p-q)\theta]\} = \frac{1}{2}\left\{-\frac{\cos(p+q)\theta}{p+q} + \frac{\cos(p-q)\theta}{p-q}\right\}_{-\pi}^{\pi} \qquad Eq. (4.1c)$$

$= 0$ for all integer values of p and q .

We will refer to the three integrals I_{cc}, I_{ss} and I_{cs} as the orthogonality relationships for the sine and cosine functions. As an aside, we recall that Euler's formula relates sine and cosine functions to exponential functions. Thus, we can write an orthogonality relationship for exponential functions. More specifically,

$$I_{ee} = \int_{-\pi}^{\pi} \exp(ip\theta)\exp(-iq\theta)\, d\theta = \int_{-\pi}^{\pi} [\cos(p\theta) + i\sin(p\theta)][\cos(q\theta) - i\sin(q\theta)]\, d\theta$$

$$= \int_{-\pi}^{\pi} \cos(p\theta)\cos(q\theta)\, d\theta + \int_{-\pi}^{\pi} \sin(p\theta)\sin(q\theta)\, d\theta = \pi\,\delta_{pq} + \pi\,\delta_{pq} = 2\pi\delta_{pq} \qquad Eq. (4.2)$$

Periodic Functions of Spatial Coordinate, x, and Time, t.

If the periodic function under consideration is a function of a spatial coordinate x, with periodicity of λ, then we replace θ in the above examples by kx. Substituting $\theta = kx = 2\pi x/\lambda$, we get

$$I_{ss} = \int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} \sin(pkx)\sin(qkx)\, dx = \frac{\lambda}{2}\delta_{pq} \quad (p, q \neq 0) \ , \qquad Eq. (4.3a)$$

$$I_{cc} = \int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} \cos(pkx)\cos(qkx)\,dx = \frac{\lambda}{2}\delta_{pq} \;\; (p, q \neq 0) \;, \qquad Eq. (4.3b)$$

and

$$I_{cs} = \int_{-\frac{\lambda}{2}}^{\frac{\lambda}{2}} \cos(pkx)\sin(qkx)\,dx = 0 \;, \qquad Eq. (4.3c)$$

for all integer values of p and q.

If the periodic function under consideration is a function of time t, with periodicity of T, then we replace θ by ωt. Substituting $\theta = \omega t = 2\pi t/T$,

$$I_{ss} = \int_{-\frac{T}{2}}^{\frac{T}{2}} \sin(p\omega t)\sin(q\omega t)\,dt = \frac{T}{2}\delta_{pq} \quad (p, q \neq 0) \;, \qquad Eq. (4.4a)$$

$$I_{cc} = \int_{-\frac{T}{2}}^{\frac{T}{2}} \cos(p\omega t)\cos(q\omega t)\,dt = \frac{T}{2}\delta_{pq} \;\; (p, q \neq 0) \;, \qquad Eq. (4.4b)$$

and

$$I_{cs} = \int_{-\frac{T}{2}}^{\frac{T}{2}} \cos(p\omega t)\sin(q\omega t)\,dt = 0 \;\; \text{for all integer values of } p \text{ and } q \;. \qquad Eq. (4.4c)$$

4.2 FOURIER SERIES

Now, look at the three functions of variable θ shown in Figure 4.1. Clearly these functions are periodic functions with periodicity of 2π. The first function $f_1(\theta)$ is an even function of θ since $f_1(\theta) = f_1(-\theta)$. The second function $f_2(\theta)$ is an odd function of θ since $f_2(\theta) = -f_2(-\theta)$. Finally, the third function $f_3(\theta)$ is neither even nor odd since there is no direct relationship between $f_3(\theta)$ and $f_3(-\theta)$. According to Fourier, any arbitrary periodic function $f(\theta)$ can be represented as a linear combination of sinusoidal functions as:

$$f(\theta) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos(m\theta) + \sum_{m=1}^{\infty} b_m \sin(m\theta) \qquad Eq. (4.5)$$

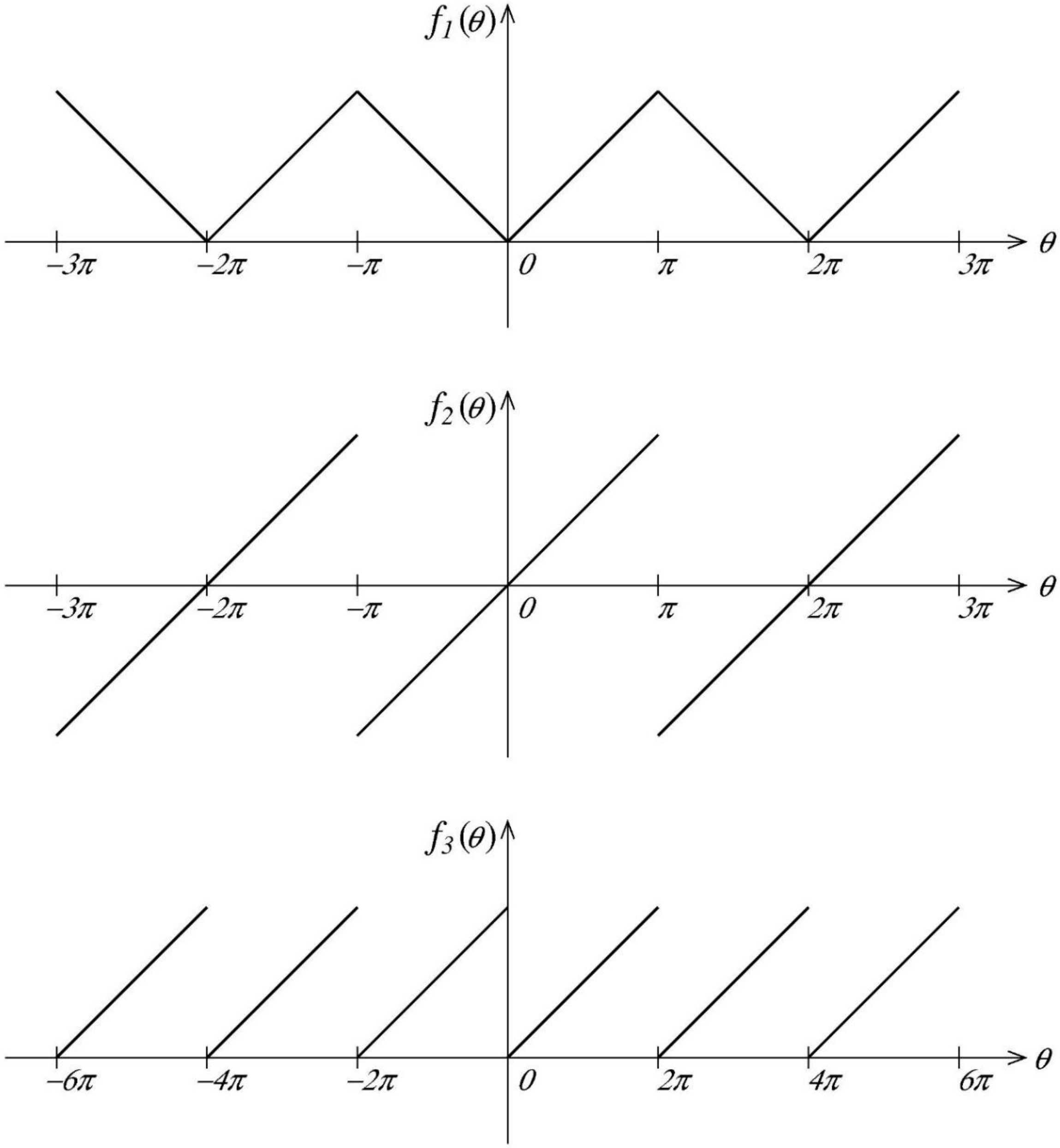


Figure 4.1. Three periodic functions of θ: $f_1(\theta)$ is an even function of θ, $f_2(\theta)$ is an odd function of θ, and $f_3(\theta)$ is neither even nor odd function of θ.

which is known as the Fourier series. Here $a_0, a_m,$ and b_m are Fourier coefficients. Note an additional factor of ½ with coefficient a_0. The mystery of this factor will become clear after we evaluate a_0. Given a periodic function $f(\theta)$, we can obtain its Fourier series by simply calculating its Fourier coefficients. To obtain these coefficients, we multiply Eq. (4.5) by either $\cos(p\theta)$ or $\sin(p\theta)$, where p is a nonzero integer, and integrate over θ from $-\pi$ to π to get

$$\int_{-\pi}^{\pi} \cos(p\theta)\, f(\theta) d\theta = \frac{a_0}{2} \int_{-\pi}^{\pi} \cos(p\theta)\, d\theta$$

$$+ \sum_{m=1}^{\infty} a_m \int_{-\pi}^{\pi} \cos(p\theta)\cos(m\theta)\, d\theta + \sum_{m=1}^{\infty} b_m \int_{-\pi}^{\pi} \cos(p\theta)\sin(m\theta)\, d\theta$$

$$= 0 + \sum_{m=1}^{\infty} a_m\, \pi\delta_{mp} \; + 0 = a_p\pi \;\; ,$$

or

$$a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(m\theta)\, f(\theta) d\theta \qquad m \geq 1 \; . \qquad\qquad Eq.\;\; (4.6)$$

Similarly, on multiplying Eq. (4.5) by $\sin(p\theta)$ and going through similar steps,

$$b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin(m\theta)\, f(\theta) d\theta \qquad m \geq 1 \; . \qquad\qquad Eq. (4.7)$$

Finally, integrating Eq. (4.5) directly,

$$\int_{-\pi}^{\pi} f(\theta) d\theta = \frac{a_0}{2} \int_{-\pi}^{\pi} d\theta + 0 + 0 = a_0\pi$$

or

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) d\theta \qquad , \qquad\qquad Eq. (4.8)$$

which is the same as Eq. (4.6) with $m = 0$. The reason for associating an extra factor of ½ with a_0 was to ensure that evaluation of this coefficient using Eq. (4.8) is similar to the evaluation of a_m $(m \geq 1)$ using Eq. (4.6). This completes the evaluation of Fourier coefficients $a_0, a_m,$ and b_m for a general periodic function $f(\theta)$.

Even and Odd Functions

Further simplification of Fourier coefficients occurs if the function $f(\theta)$ is either an even function or an odd function of θ. An even function of variable θ is symmetric about the origin, namely, $f(-\theta) = f(\theta)$, whereas for an odd function, $f(-\theta) = -f(\theta)$. First, consider the Fourier coefficients b_m. Split the integral in Eq. (4.7) into two parts as

$$b_m = \frac{1}{\pi}\int_0^{\pi} \sin(m\theta)\, f(\theta)d\theta + \frac{1}{\pi}\int_{-\pi}^{0} \sin(m\theta)\, f(\theta)d\theta \ = I_1 + I_2 \ .$$

In the second integral, substitute $y = -\theta$ or $\theta = -y$, and note that for an even function

$$I_2 = \frac{1}{\pi}\int_{\pi}^{0} [\sin(-my)] f(-y)d(-y) \ = \frac{1}{\pi}\int_{\pi}^{0} \sin(my)\, f(y)d(y) = -I_1 \ .$$

So, $b_m = 0$ for an even function $f(\theta)$. Similarly, for an odd function, $I_2 = I_1$ so that $b_m \neq 0$ and it is given by

$$b_m = \frac{2}{\pi}\int_0^{\pi} \sin(m\theta)\, f(\theta)d\theta \ .$$

Next, consider the Fourier coefficients a_m. In this case, splitting the integral of Eq. (4.6) into two parts gives

$$a_m = \frac{1}{\pi}\int_0^{\pi} \cos(m\theta)\, f(\theta)d\theta + \frac{1}{\pi}\int_{-\pi}^{0} \cos(m\theta)\, f(\theta)d\theta \ = J_1 + J_2 \ .$$

Again, in the second integral, substitute $y = -\theta$ or $\theta = -y$, and for an odd function, $f(-\theta) = -f(\theta)$, so that

$$J_2 = \frac{1}{\pi}\int_{\pi}^{0} [\cos(-my)] f(-y)d(-y) \ = \frac{1}{\pi}\int_{\pi}^{0} \cos(my)\, f(y)d(y) = -J_1 \ .$$

Thus, for an odd function, $a_m = 0$, including a_0. On the other hand, for an even function, $J_2 = J_1$ so that $a_m \neq 0$ and it is given by

$$a_m = \frac{2}{\pi}\int_0^{\pi} \cos(m\theta)\, f(\theta)d\theta \ .$$

This is really not a very surprising result since it implies that for an odd function $[f(-\theta) = -f(\theta)]$, the Fourier series consists of only odd sine functions:

$$f(\theta) = \sum_{m=1}^{\infty} b_m \sin(m\theta) \ ,$$

while for an even function $[f(-\theta) = f(\theta)]$, the Fourier series consists of only even cosine functions:

$$f(\theta) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos(m\theta) \ .$$

Here are examples of determining Fourier series for three different periodic functions. The first function is an odd function of θ, the second function is an even function of θ, and the third function is neither an even nor odd function of θ.

Example: Given a periodic function,

$$\boldsymbol{f(\theta) = \theta} \ \ \textbf{for} \ -\boldsymbol{\pi} \le \boldsymbol{\theta} \le \boldsymbol{\pi} \qquad\qquad [\textbf{Sawtooth function}],$$

determine its Fourier series.

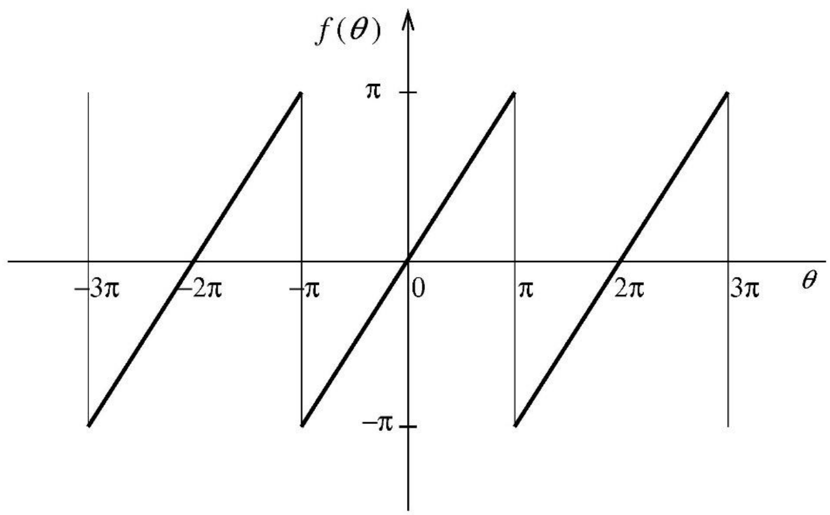


Figure 4.2. The sawtooth function, $f(\theta)$, as a function of θ.

Solution: As seen in Figure 4.2, $f(\theta)$ is an odd function of θ. So, all $a_m = 0$ (including a_0), and

$$f(\theta) = \sum_{m=1}^{\infty} b_m \sin(m\theta)$$

with

$$b_m = \frac{2}{\pi}\int_0^{\pi} \sin(m\theta)\, f(\theta) d\theta = \frac{2}{\pi}\int_0^{\pi} \theta \sin(m\theta)\, d\theta$$

$$= \frac{2}{\pi}\left\{ -\frac{\theta\cos(m\theta)}{m}\Bigg|_0^{\pi} + \int_0^{\pi}\frac{\cos(m\theta)}{m}\,d\theta \right\}$$

$$= \frac{2}{\pi}\left\{ -\frac{\pi}{m}\cos(m\pi) + \frac{\sin(m\theta)}{m^2}\Bigg|_0^{\pi} \right\}$$

$$= -\frac{2}{m}\cos(m\pi) = (-1)^{m+1}\frac{2}{m} \; .$$

Thus,

$$f(\theta) = \theta = 2\sum_{m=1}^{\infty}\frac{(-1)^{m+1}}{m}\sin(m\theta) \;\; ,$$

$$= 2\left\{ \sin(\theta) - \frac{1}{2}\sin(2\theta) + \frac{1}{3}\sin(3\theta) - \frac{1}{4}\sin(4\theta) + \cdots \right\} \; .$$

This is the Fourier series for the sawtooth function. Figure 4.3 shows the sawtooth function as well as the sum of first N terms of the Fourier series (red dotted line) for $N = 2, 5, 10,$ and 100.

Clearly, the depiction of the sawtooth function by the Fourier series becomes more accurate as N increases.

Example: Find the Fourier series for a periodic function $f(\theta)$ which is defined in the interval $-\pi \le \theta \le \pi$ as

$$\boldsymbol{f(\theta) = \theta^2} \; .$$

Solution: Since $f(\theta) = \theta^2$ is an even function of $\theta, b_m = 0$ for $m \geq 1$. Multiply both sides of the above function by $\cos n\theta$ and integrate over θ from $-\pi$ to π to get

$$\int_{-\pi}^{\pi} f(\theta)\cos(n\theta)\,d\theta = \frac{a_0}{2}\int_{-\pi}^{\pi}\cos(n\theta)\,d\theta + \sum_{m=1}^{\infty} a_m \int_{-\pi}^{\pi}\cos(m\theta)\cos(n\theta)\,d\theta$$

$$= 0 + \pi a_n$$

or

$$a_n = \frac{2}{\pi}\int_0^{\pi}\theta^2\cos(n\theta)\,d\theta = \frac{2\,\theta^2}{\pi}\cdot\frac{\sin(n\theta)}{n}\Bigg|_0^{\pi} - \frac{2}{\pi}\int_0^{\pi} 2\theta\cdot\frac{\sin(n\theta)}{n}\,d\theta$$

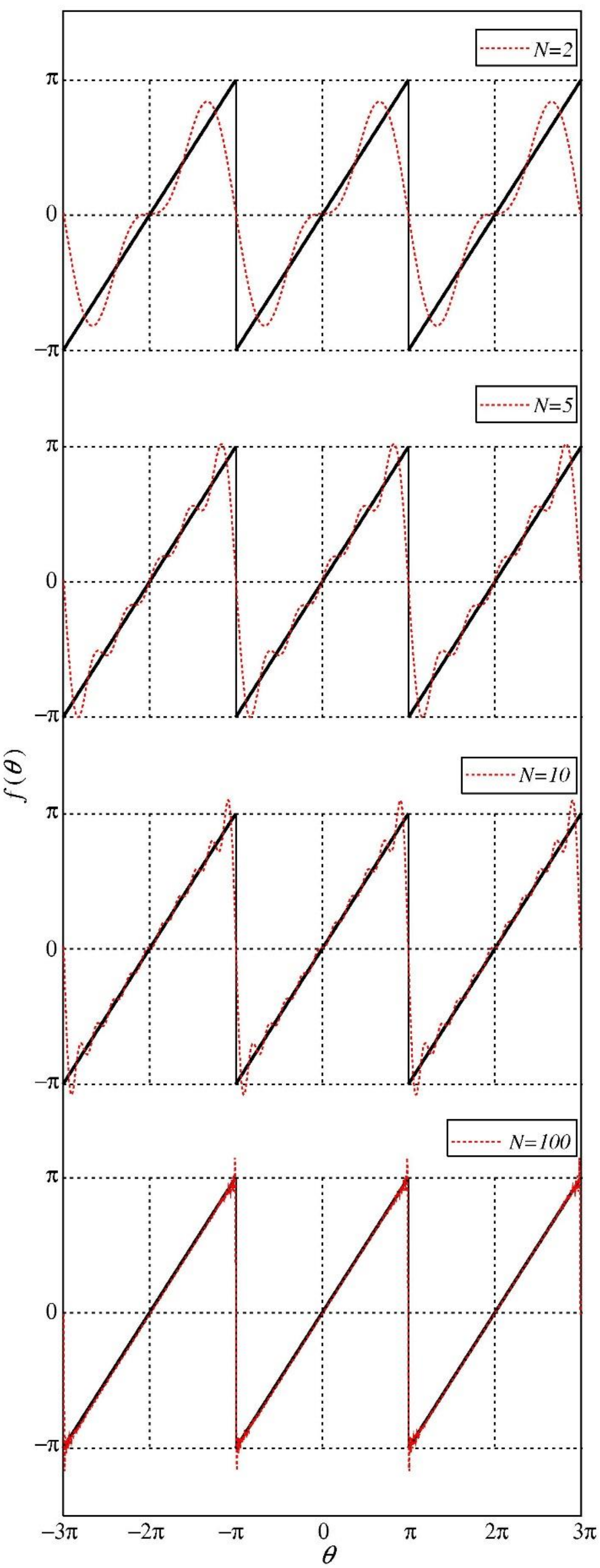


Figure 4.3. The sum of first N terms in the Fourier series of a sawtooth function.

$$= -\frac{4}{n\pi}\int_0^\pi \theta\sin(n\theta)\,d\theta = \frac{4}{n\pi}\theta\cdot\frac{\cos(n\theta)}{n}\bigg|_0^\pi - \frac{4}{n\pi}\int_0^\pi \frac{\cos(n\theta)}{n}d\theta$$

$$= \frac{4}{\pi n^2}[\pi\cos(n\pi)] - 0 = \frac{4}{n^2}(-1)^n \ .$$

Next, integrate the given function from $-\pi$ to π to get

$$\int_{-\pi}^{\pi}\theta^2 d\theta = \frac{a_0}{2}\int_{-\pi}^{\pi} d\theta + \sum_{m=1}^{\infty} a_m \int_{-\pi}^{\pi}\cos(m\theta)d\theta$$

or

$$\frac{\theta^3}{3}\bigg|_{-\pi}^{\pi} = \frac{a_0}{2}\theta\bigg|_{-\pi}^{\pi} + 0 \qquad \text{or} \qquad \frac{2}{3}\pi^3 = \frac{a_0}{2}\cdot 2\pi \qquad \text{or} \qquad a_0 = \frac{2}{3}\pi^2 \ .$$

Thus,

$$f(\theta) = \frac{\pi^2}{3} + \sum_{m=1}^{\infty}\frac{4(-1)^m}{m^2}\cos(m\theta) \ .$$

Example: Determine the Fourier series for a unit step function defined as

$$\boldsymbol{f(\theta) = \begin{cases} 0 & \textbf{for} \qquad -\pi < \theta < 0 \\ 1 & \textbf{for} \qquad\ \ 0 < \theta < \pi \end{cases}}$$

Solution: The function $f(\theta)$ is neither an odd function nor an even function of θ, so the Fourier series for $f(\theta)$ will consist of all Fourier coefficients, $a_0, a_m,$ and b_m, namely,

$$f(\theta) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_m\cos(m\theta) + \sum_{m=1}^{\infty} b_m\sin(m\theta) \quad .$$

We evaluate these coefficients using Eqs. (4.6) to Eq. (4.8),

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)d\theta = \frac{1}{\pi}\int_0^\pi (1)\,d\theta = \frac{\pi}{\pi} = 1 \ ,$$

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi}\cos(m\theta)\,f(\theta)d\theta = \frac{1}{\pi}\int_0^\pi \cos(m\theta)\,d\theta = \frac{1}{\pi}\cdot\frac{\sin(m\theta)}{m}\bigg|_0^\pi = 0 \qquad m \geq 1 \ ,$$

$$b_m = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin(m\theta)\, f(\theta) d\theta = \frac{1}{\pi}\int_{0}^{\pi} \sin(m\theta)\, d\theta = -\frac{1}{\pi}\cdot \frac{\cos(m\theta)}{m}\bigg|_0^{\pi} = \frac{1}{\pi m}[1-(-1)^m] \ ,$$

or,

$$b_m = \begin{cases} 0 & \text{for} \quad m = 2, 4, 6, 8 \ \ldots \\ \dfrac{2}{\pi m} & \text{for} \quad m = 1, 3, 5, 7 \ \ldots \end{cases} .$$

So, the Fourier series for $f(\theta)$ is

$$f(\theta) = \frac{1}{2} + \frac{2}{\pi}\left[\frac{\sin(\theta)}{1} + \frac{\sin(3\theta)}{3} + \frac{\sin(5\theta)}{5} + \frac{\sin(7\theta)}{7} + \cdots\right] \quad .$$

Complex Fourier Series

Since Euler's formula relates sine and cosine functions to exponential functions, it is possible to write the Fourier series as a series of exponential functions instead of sine and cosine functions,

$$f(\theta) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \left(\frac{\exp(im\theta) + \exp(-im\theta)}{2}\right) + \sum_{m=1}^{\infty} b_m \left(\frac{\exp(im\theta) - \exp(-im\theta)}{2i}\right)$$

$$= \frac{a_0}{2} + \sum_{m=1}^{\infty} \left(\frac{a_m - ib_m}{2}\right) \exp(im\theta) + \sum_{m=1}^{\infty} \left(\frac{a_m + ib_m}{2}\right) \exp(-im\theta) \ .$$

On changing m to $-m$ in the second sum, we have

$$f(\theta) = \frac{a_0}{2} + \sum_{m=1}^{\infty} \left(\frac{a_m - ib_m}{2}\right) \exp(im\theta) + \sum_{m=-1}^{-\infty} \left(\frac{a_{-m} + ib_{-m}}{2}\right) \exp(im\theta)$$

Now, we define new coefficients c_m as follows,

$$c_0 = \frac{a_0}{2} \ ,$$

$$c_m = \frac{a_m - ib_m}{2} \text{ for } m = 1,2, \cdots \infty \ ,$$

and

$$c_m = \frac{a_{-m} + ib_{-m}}{2} \text{ for } m = -1, -2, \cdots -\infty \ .$$

In terms of these new coefficients, the Fourier series becomes

$$f(\theta) = \sum_{m=-\infty}^{\infty} c_m \exp(im\theta) \quad . \qquad Eq.(4.9a)$$

This is the complex (or exponential) form of the Fourier series. The coefficients c_m are obtained by using the orthogonality relation [Eq. 4.2] for exponential functions,

$$c_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\psi)\exp(-im\psi)\,d\psi \quad . \qquad Eq.(4.9b)$$

Example: Let us repeat the sawtooth function,

$$\boldsymbol{f(\theta) = \theta \qquad \text{for} \quad -\pi \le \theta \le \pi} \ .$$

using the complex Fourier series.

Solution:

The Fourier coefficients for this complex series are,

$$c_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} \theta\exp(-im\theta)\,d\theta = \frac{1}{2\pi}\left\{\frac{\theta\exp(-im\theta)}{-im}\bigg|_{-\pi}^{\pi} - \int_{-\pi}^{\pi}\frac{\exp(-im\theta)}{-im}d\theta\right\} for\ m \neq 0 \ ,$$

$$= \frac{1}{2\pi}\left\{\frac{\theta\exp(-im\theta)}{-im}\bigg|_{-\pi}^{\pi} - \frac{\exp(-im\theta)}{(-im)^2}\bigg|_{-\pi}^{\pi}\right\}$$

$$= \frac{1}{2\pi}\left\{\frac{\pi\exp(-im\pi)}{-im} - \frac{(-\pi)\exp(im\pi)}{-im} - \frac{\exp(-im\pi)}{(-im)^2} + \frac{\exp(im\pi)}{(-im)^2}\right\}$$

$$= -\frac{1}{im}\cos(m\pi) - \frac{i}{\pi m^2}\sin(m\pi) = -\frac{(-1)^m}{im} - 0 = \frac{i(-1)^m}{m} \ for\ m \neq 0 \ .$$

For $m = 0$,

$$c_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\theta\,d\theta = 0 \ .$$

So, the complex Fourier series looks like

$$f(\theta) = \theta = 0 + \sum_{m=1}^{\infty} \frac{i(-1)^m}{m} \exp(im\theta) + \sum_{m=-1}^{-\infty} \frac{i(-1)^m}{m} \exp(im\theta)$$

On changing m to $-m$ in the second sum, we have

$$f(\theta) = \sum_{m=1}^{\infty} \frac{i(-1)^m}{m} \exp(im\theta) + \sum_{m=1}^{\infty} \frac{i(-1)^{-m}}{-m} \exp(-im\theta)$$

$$= \sum_{m=1}^{\infty} \frac{i(-1)^m}{m} [\exp(im\theta) - \exp(-im\theta)] = 2 \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \sin(m\theta) \;,$$

the same as before.

4.3 FOURIER TRANSFORMS

In our discussion of Fourier series, our motivation was to write any arbitrary periodic function as a series of some well-known periodic functions—namely, sine and cosine functions. A question arises: Can we write something analogous to a Fourier series for an arbitrary function that does not appear to be a periodic function? Let us explore this possibility. Look at an isolated step-like function in space as shown in the Figure 4.4a.

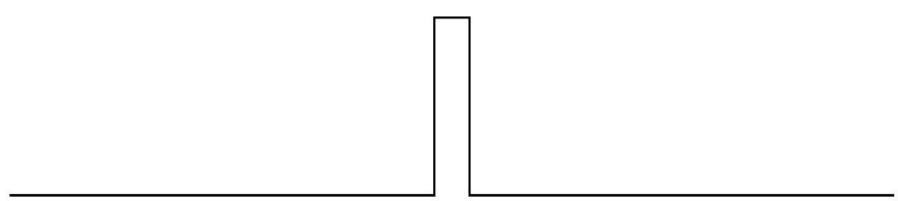


Figure 4.4a. An isolated step-like function in space.

At first sight this function appears as a non-periodic step function in space. Now, imagine starting with a periodic step function, as in part a of Figure 4.4b, of some definite wavelength as indicated. If we were to remove alternate spikes in part a of this figure, we will end up with part b, which still shows up as a periodic function, except with a longer wavelength. We repeat the process of removing alternate spikes in part b and end up with part c, which also shows a periodic function but with a still-longer wavelength. If we continue the process of removing alternate spikes from each new figure ad infinitum, we will end up with the isolated function of Figure 4.4a.

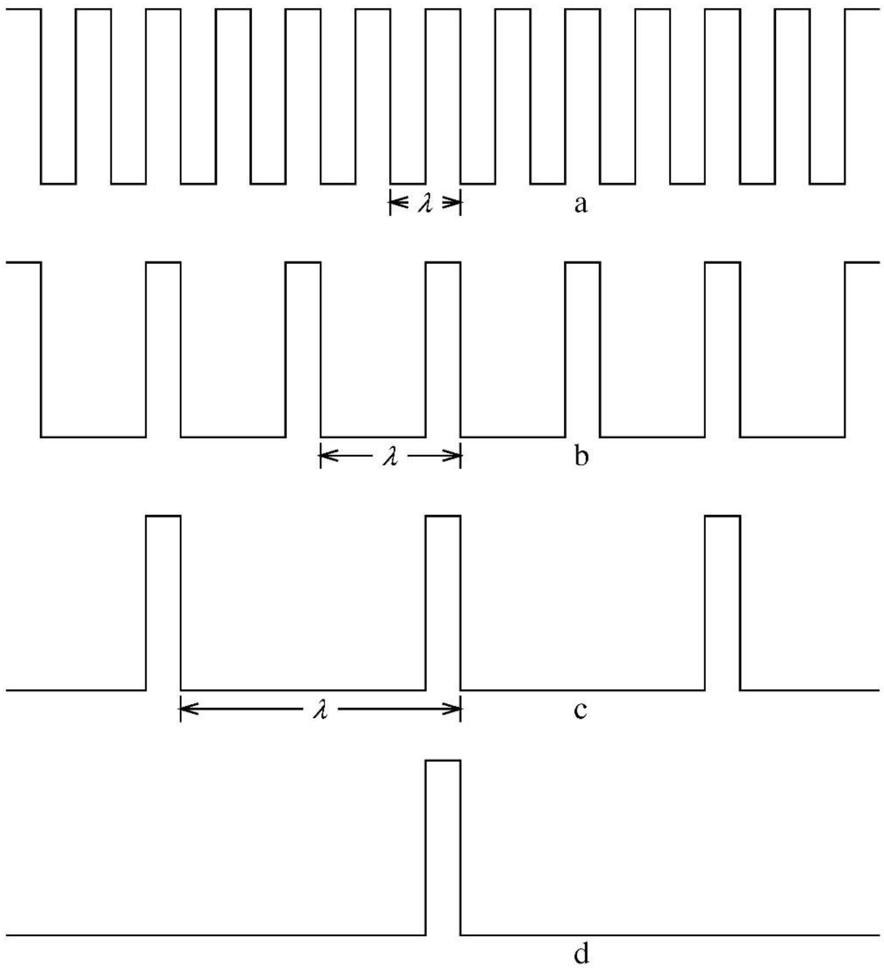


Figure 4.4b. A periodic function with an infinite wavelength appears as an isolated function.

This procedure suggests that an equivalent of the Fourier series of an isolated function, of spatial coordinate x, will be obtained by taking the limit of an infinite wavelength ($\lambda \to \infty$). Using complex Fourier series, with $\theta = kx$ we have

$$f(x) = \sum_{m=-\infty}^{\infty} c_m \exp(imkx) \ ,$$

where

$$c_m = \frac{1}{\lambda} \int_{-\lambda/2}^{\lambda/2} f(x') \exp(-imkx') \, dx' \ .$$

The dummy variable, m, in the infinite sum above goes from $-\infty$ to $+\infty$ in steps of 1, that is, $\Delta m = 1$. Now we make a change of dummy variable in this infinite sum from m to q, defined as $q = mk = m(2\pi/\lambda)$. Then,

$$\Delta q = \frac{2\pi}{\lambda}\Delta m = \frac{2\pi}{\lambda} \quad \text{or} \quad \frac{1}{\lambda} = \frac{\Delta q}{2\pi} \ .$$

Note that values of q are discrete with step size $2\pi/\lambda$. When $\lambda \to \infty$, the sum over m becomes an integral over q, and $\Delta q \to dq$. Then,

$$f(x) = \lim_{\lambda\to\infty} \int_{m \ or \ q=-\infty}^{\infty} \exp(iqx) \frac{\Delta q}{2\pi} \int_{-\lambda/2}^{\lambda/2} f(x') \exp(-iqx')\, dx' \ ,$$

or

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dq \int_{-\infty}^{\infty} f(x') \exp[iq(x - x')] dx' \ .$$

If we define

$$F(q) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x') \exp(-iqx')\, dx' \ ,$$

then

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dq\, F(q)\ \exp(iqx) \ .$$

$F(q)$ is called the Fourier transform of $f(x)$ and vice versa. The factor of $1/2\pi$ can be distributed in many ways between *F* and *f.* However, it is a common practice to distribute $1/\sqrt{2\pi}$ with function f and $1/\sqrt{2\pi}$ with function F to preserve symmetry between a function and its Fourier transform.

Let us then summarize the relationship between a function and its Fourier transform. Consider a pair of two physical quantities, u and v, whose product is dimensionless and, therefore, can be measured in radians. Such variables are called conjugate variables. We have seen an example of this in our discussion of the wave equation, namely the pair of wave number k and spatial coordinate x such that the product kx is dimensionless. Another pair consisting of angular frequency ω and time t is such that the product ωt is dimensionless. Since u and v are conjugate variables, the Fourier transform "transforms" the u-dependent function $f(u)$ into a completely equivalent representation $F(v)$, a v-dependent function, in the following way:

$$F(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(u) \exp(-iuv)\, du \ .$$

$F(v)$ is called the Fourier transform of $f(u)$. The inverse Fourier transform creates the opposite transformation, namely,

$$f(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(v) \exp(iuv)\, dv \ .$$

Thus, $f(u)$ is the Fourier transform of $F(v)$. In particular, a function of spatial coordinate x (position in space) is cast into an equivalent function of wavenumber k (momentum space) by the Fourier transformation,

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp(-ikx)\, dx \ , \qquad Eq. (4.10a)$$

and its inverse,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(k) \exp(ikx)\, dk \ . \qquad Eq. (4.11a)$$

As an aside, k-space is also called momentum space (or, p-space) since in quantum physics $p = \hbar k$, where $\hbar$ is the universal Planck's constant. Similarly, using the Fourier transformation, a temporal function of time t is cast into an equivalent function of frequency ω as

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \exp(-i\omega t)\, dt \ , \qquad Eq. (4.10b)$$

and its inverse,

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t)\, d\omega \ . \qquad Eq. (4.11b)$$

Example: Determine the Fourier transform of a Gaussian function $\boldsymbol{f(x)}$ given in the coordinate space as

$$\boldsymbol{f(x) = f(0) \exp\left(-\frac{x^2}{a^2}\right) = \frac{1}{a\sqrt{\pi}} exp\left(-\frac{x^2}{a^2}\right)} \ .$$

Solution: The constant in front, namely, $f(0) = 1/(a\sqrt{\pi})$, normalizes the Gaussian function as

$$\int_{-\infty}^{\infty} f(x)dx = 1 \ .$$

The integral here is same as integral I_{gg}^{0} of Eq. (2.14). The bell-shaped function $f(x)$ has its maximum value, $f(0)$, for $x = 0$ and it goes to zero as $x \to \pm\infty$. The extent of values of x, namely Δx, over which the function $f(x)$ is appreciable is given by the full width of the function at half of its maximum value (FWHM). Thus,

$$\Delta x = FWHM = 2a\sqrt{ln\, 2} \ . \qquad\qquad Eq. (4.12a)$$

In other words, the parameter a in the Gaussian function is a measure of the approximate width of the function; the larger the parameter a is, the wider the Gaussian function, and vice versa. Figure 4.5 shows the function $f(x)$ for $a = 1, 2, \text{and } 3$.

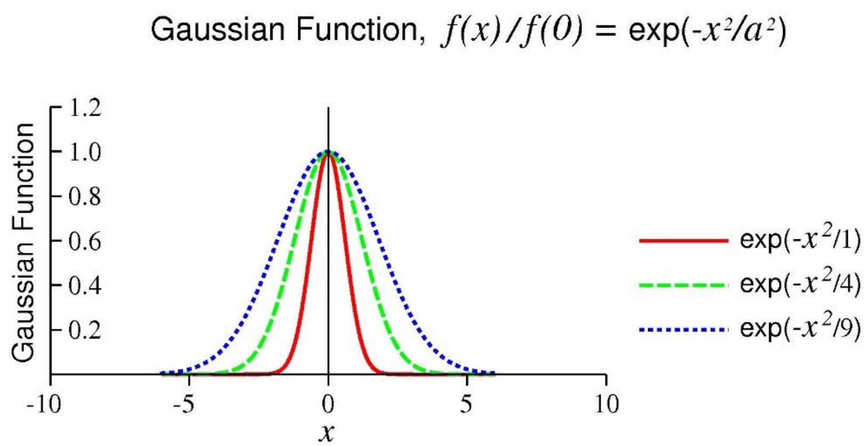


Figure 4.5. The Gaussian function $f(x)/f(0) = \ exp(-x^2/a^2)$ for various values of a.

The Fourier transform of $f(x)/f(0)$ is

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{a^2}\right] \exp[-ikx]\, dx \ .$$

On perfecting the square in the exponent, we get

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{a^2}\left(x^2 + ika^2x + \frac{(ika^2)^2}{4}\right) + \frac{(ika^2)^2}{4a^2}\right] dx$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{k^2a^2}{4}\right] \int_{-\infty}^{\infty} \exp\left[-\frac{(x + ika^2/2)^2}{a^2}\right] dx$$

On making a change of variable from x to u, using $u = \frac{(x+ika^2/2)}{a}$,

$$F(k) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{k^2a^2}{4}\right] \int_{-\infty}^{\infty} \exp(-u^2)\, a du \ ,$$

$$= \frac{a}{\sqrt{2}} \exp\left[-\frac{k^2a^2}{4}\right] ,$$

on using the integral I_{gg}^0 of Eq. (2.14). We note that the Fourier transform $F(k)$ of a Gaussian function in the coordinate space is also a Gaussian function, but in the k-space. The extent of values of k, namely Δk, over which the function $F(k)$ is appreciable is again given by the full width of this function at half of its maximum value (FWHM). Thus,

$$\Delta k = FWHM = \frac{4}{a}\sqrt{ln\, 2} \ . \qquad Eq. (4.12b)$$

Now the parameter a has an inverse relationship to the width of the function $F(k)$; the larger the parameter a is, the thinner the function $F(k)$. Note, the product of Δx and of Δk in Eq. (4.12) is independent of a. This essentially is the statement of the Uncertainty Principle in quantum physics.

Example: Determine the Fourier transform of a single rectangular pulse in the coordinate space looking like,

$$\boldsymbol{f(x)} = \begin{cases} \frac{\mathbf{1}}{(\mathbf{2a})} & -\boldsymbol{a} \le \boldsymbol{x} \le \boldsymbol{a} \\ \mathbf{0}, & |\boldsymbol{x}| \ge \boldsymbol{a} \end{cases} .$$

Solution: For this function,

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{2a} \int_{-a}^{a} dx = 1 \ .$$

Furthermore, the extent of the values of x over which the function is appreciable is,

$$\Delta x = 2a \ . \qquad Eq. (4.13a)$$

The Fourier transform of the single rectangular pulse is

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp(-ikx)\, dx = \frac{1}{\sqrt{2\pi}} \frac{1}{2a} \int_{-a}^{a} \exp(-ikx)\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{2a} \frac{\exp(-ikx)}{-ik}\bigg|_{-a}^{a} = -\frac{1}{2aik} \frac{1}{\sqrt{2\pi}} [\exp(-ika) - \exp(ika)]$$

$$= -\frac{1}{2aik} \frac{1}{\sqrt{2\pi}} [-2i \sin(ka)] = \frac{1}{\sqrt{2\pi}} \frac{\sin(ka)}{ka} \ .$$

Figure 4.6 shows $F(k)$ as a function of the variable k. It is a wiggly function with the central wiggle being the largest one, extending from $k = -\pi/a$ to $+\pi/a$.

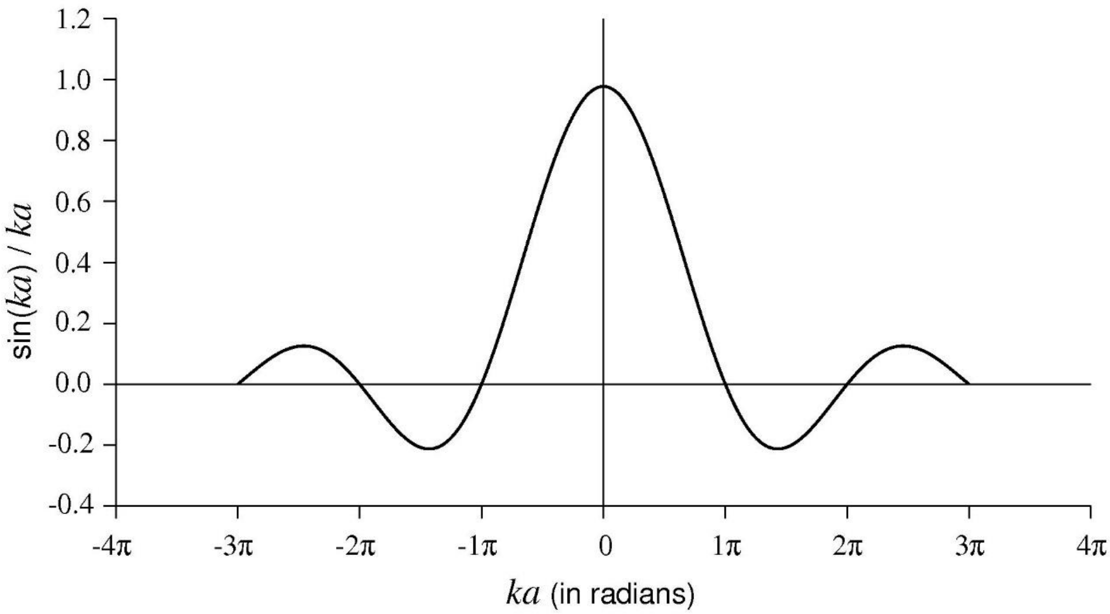


Figure 4.6. The function $\sin(ka)/(ka)$ as a function of ka.

Thus, the extent of the values of k over which $F(k)$ is appreciable is,

$$\Delta k = \frac{2\pi}{a} \ . \qquad Eq. (4.13b)$$

Once again, the product of Δx and of Δk, in Eqs. (4.13), is independent of a, which confirms the statement of the Uncertainty Principle in quantum physics.

From examples of Fourier transforms, it can be concluded that if the extent of a function is very wide, then the extent of its Fourier transform is very narrow and vice versa. This fact is very useful in defining the reciprocal lattices in condensed matter physics.

PROBLEMS FOR CHAPTER 4

1. Show explicitly, for *p* and *q* as positive integers and α as an arbitrary constant,

$$\int_{\alpha}^{\alpha+2\pi} \sin(p\theta)\sin(q\theta)\ d\theta = \int_{0}^{2\pi} \sin(p\theta)\sin(q\theta)\ d\theta \ ,$$

$$\int_{\alpha}^{\alpha+2\pi} \cos(p\theta)\cos(q\theta)\ d\theta = \int_{0}^{2\pi} \cos(p\theta)\cos(q\theta)\ d\theta \ ,$$

$$\int_{\alpha}^{\alpha+2\pi} \sin(p\theta)\cos(q\theta)\ d\theta = \int_{0}^{2\pi} \sin(p\theta)\cos(q\theta)\ d\theta \ .$$

2. Find the Fourier series for a periodic function $f(\theta)$ which is defined in the interval $-\pi \le \theta \le \pi$ as

$$f(\theta) = \begin{cases} \pi + \theta & \text{for } \ -\pi \le \theta \le 0 \\ \pi - \theta & \text{for } \ 0 \le \theta \le \pi \end{cases} \ .$$

3. Determine the Fourier series for the "square" wave defined as

$$f(x) = \begin{cases} -1 \ \text{ for } -\dfrac{\lambda}{2} < x < 0 \\ \ \ 1 \ \text{ for } 0 < x < \dfrac{\lambda}{2} \end{cases} \ .$$

This series appears in discussions of high frequency electronic circuits.

4. Determine the Fourier series for the function defined as

$$f(t) = |\sin \omega t| \ \text{ for } -\pi < \omega t < \pi \ .$$

This series appears in discussions of the full-wave rectifiers in electronics.

5. Determine the Fourier series for the function defined as

$$f(t) = \begin{cases} \sin \omega t & \text{for } \ 0 < \omega t < \pi \\ 0 & \text{for } -\pi < \omega t < 0 \end{cases}$$

This series appears in discussions of the half-wave rectifiers in electronics.

6. Consider a periodic function $f(\theta)$ which is defined in the interval $-\pi \le \theta \le \pi$ as

$$f(\theta) = \begin{cases} +\cos\theta & \text{for } \ 0 \le \theta \le \pi \\ -\cos\theta & \text{for } -\pi \le \theta \le 0 \end{cases}$$

(a) Is the function $f(\theta)$ an even function or an odd function or neither?

(b) Determine the Fourier series for $f(\theta)$.

7. Find the Fourier series for a periodic function $f(\theta)$ which is defined as

$$f(\theta) = |\theta| \quad \text{for} \quad -\pi \le \theta \le \pi \ .$$

8. Show that the Fourier transform of the function

$$f(x) = \begin{cases} \cos\left(\frac{\pi x}{2a}\right) & \text{for } -a \le x \le a \\ 0 & \text{otherwise} \end{cases}$$

(where a is a constant) is

$$F(k) = \frac{a}{\sqrt{2\pi}} \frac{\pi}{\left(\frac{\pi}{2}\right)^2 - (ka)^2} \cos(ka) \ .$$

Plot $f(x)$ as a function of x and $F(k)$ as a function of k. Using these plots, estimate the extents Δx and Δk of the functions $f(x)$ and $F(k)$, respectively. Show that Δx times Δk is independent of a.

{Hint: Use $\cos x = [\exp(ix) + \exp(-ix)]/2.$ }

9. Show that the Fourier transform of the parabolic function

$$f(x) = \begin{cases} a^2 - x^2 & \text{for } |x| \le a \\ 0 & \text{for } |x| > a \end{cases}$$

is

$$F(k) = \frac{a^3}{\sqrt{2\pi}} \left\{ \frac{4}{(ka)^3} [\sin(ka) - (ka)\cos(ka)] \right\} \ .$$

Plot $f(x)$ as a function of x and $F(k)$ as a function of k. Using these plots, estimate the extents Δx and Δk of the functions $f(x)$ and $F(k)$, respectively. Show that Δx times Δk is independent of a.

10. Show that the Fourier transform of the triangular pulse function,

$$f(x) = \begin{cases} 1 - |x|/a & \text{for } |x| \le a \\ 0 & \text{for } |x| > a \end{cases}$$

(where a is a positive constant) is

$$F(k) = \frac{a}{\sqrt{2\pi}} \left(\frac{\sin(ka/2)}{ka/2} \right)^2 \ .$$

Plot $f(x)$ as a function of x and $F(k)$ as a function of k. Using these plots, estimate the extents Δx and Δk of the functions $f(x)$ and $F(k)$, respectively. Show that Δx times Δk is independent of a.

Chapter 5: Complex Variables

In this chapter we will describe algebra related to complex numbers. Representations of complex numbers in two-dimensional planes will be explained with several examples. Algebraic properties of addition, subtraction, multiplication, and division of complex variables will be outlined. Finally, DeMoivre's formula will be derived and applied to obtain some useful trigonometric identities.

5.1 COMPLEX NUMBERS AND COMPLEX ALGEBRA

A complex number is defined as an *ordered* pair of two real numbers (x, y) in which the first number, x, is called the real part of the complex number and the second number, y, is called the imaginary part of the complex number. It is customary to reserve the letter z for complex numbers and write the complex number as

$$z = x + iy \ . \qquad Eq. (5.1a)$$

The real part of z, namely $Re(z)$, is x and the imaginary part of z, namely $Im(z)$, is y. Complex numbers can be represented as points in a two-dimensional plane, which is analogous to the common $x-y$ plane and is called the complex plane or the z-plane. Because of its similarity with Cartesian coordinates, the representation of a complex number as $z = x + iy$ is called the Cartesian representation of the complex number. In a plane one can switch from Cartesian coordinates (x, y) to plane polar coordinates (r, θ). From Figure 5.1,

$$x = r\cos\theta \, , y = r\sin\theta \ . \qquad Eq. (5.2a)$$

Thus, a complex number can be written as

$$z = r(\cos\theta + i\sin\theta) = r\exp(i\theta) \ . \qquad Eq. (5.1b)$$

Here r is the magnitude of z and θ is the argument or phase of z. The magnitude of a complex number is also commonly written as $|z|$ so that $r = |z|$. Note,

$$r = \sqrt{x^2 + y^2}, \theta = \arg(z) = \arctan\left(\frac{y}{x}\right) \ . \qquad Eq. (5.2b)$$

The representation of a complex number as $z = |z|\exp(i\theta) = r\exp(i\theta)$ is called the polar or exponential representation of the complex number.

The complex conjugate of a complex number is obtained by replacing i by $-i$ and is denoted by placing a bar on top of the number. Thus $\bar{z} = r\exp(-i\theta)$ or $\bar{z} = x - iy$. Note $|z| = r = |\bar{z}|$. Also $z\bar{z} = r^2 = |z|^2 = x^2 + y^2$. Thus,

$$|Re(z)| \leq |z| \text{ and } |Im(z)| \leq |z| \ .$$

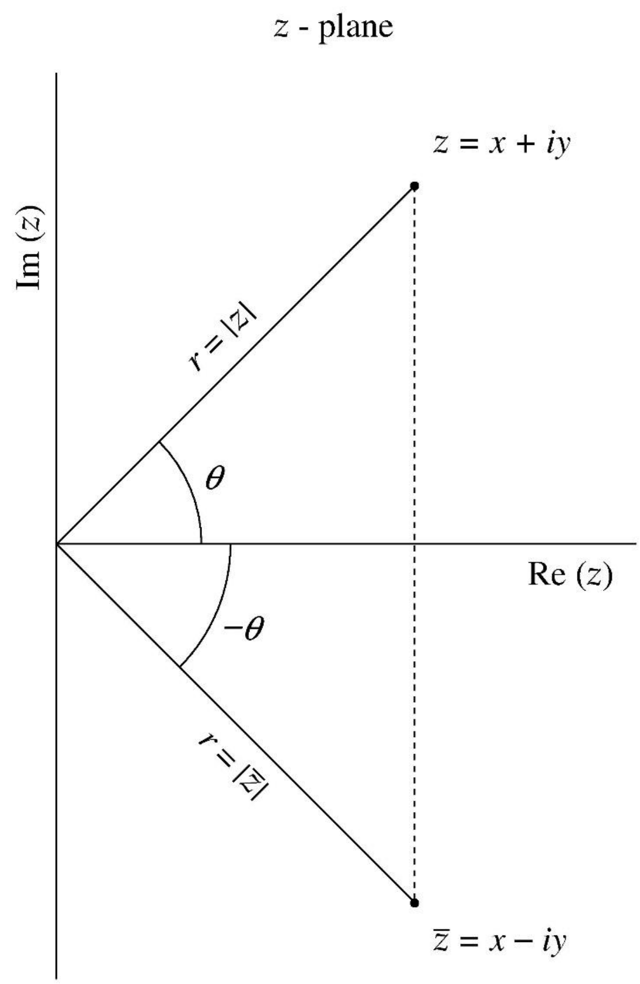


Figure 5.1. Cartesian and polar (or exponential) representation of a complex number.

Example: Given, $z_1 = 2 + 5i$, determine $Re(z_1)$, $Im(z_1)$, $|z_1|$, phase angle θ, complex conjugate $\overline{z_1}$ and exponential representation of z_1.

Solution: Looking at the complex number $z_1 = 2 + 5i$, we note $Re(z_1) = 2$, $Im(z_1) = 5$, $|z_1| = \sqrt{2^2 + 5^2} = \sqrt{29}$, $\tan\theta = \frac{5}{2} = 2.5$, $\theta = 68.2° = 1.19$ radians, $\overline{z_1} = 2 - 5i$, and $z_1 = \sqrt{29}\exp(1.19\ i)$.

Example: Given $z_2 = 4 + 3i$, determine $Re(z_2)$, $Im(z_2)$, $|z_2|$, phase angle θ, complex conjugate $\overline{z_2}$, and exponential representation of z_2.

Solution: Again, looking at the complex number $z_2 = 4 + 3i$, we note $Re(z_2) = 4$, $Im(z_2) = 3$, $|z_2| = \sqrt{4^2 + 3^2} = 5$, $\tan\theta = \frac{3}{4} = 0.75$, $\theta = 36.9° = 0.64$ radians, $\overline{z_2} = 4 - 3i$, and $z_2 = 5\exp(0.64\ i)$.

5.2 PROPERTIES OF COMPLEX NUMBERS

Here is a list of some general properties of complex numbers.

Property Number 1: Addition or subtraction of two complex numbers is achieved by adding or subtracting the real parts and the imaginary parts separately. If

$$z_1 = x_1 + iy_1 \text{ and } z_2 = x_2 + iy_2 \ ,$$

then

$$z_1 \pm z_2 = (x_1 \pm x_2) + i(y_1 \pm y_2) \ .$$

The addition of two complex numbers is both commutative as well as associative,

$$z_1 + z_2 = z_2 + z_1 \ ,$$

$$z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3 \ .$$

The product of two complex numbers is treated as simple polynomial multiplication,

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) \ .$$

The product of two complex numbers is commutative and associative as well as distributive,

$$z_1 z_2 = z_2 z_1 \ ,$$

$$z_1(z_2 z_3) = (z_1 z_2) z_3 \ ,$$

$$z_1(z_2 + z_3) = z_1 z_2 + z_1 z_3 \ .$$

Property Number 2: If two complex numbers, $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, are equal, then their real and imaginary parts are separately equal. In other words, $z_1 = z_2$ implies $x_1 = x_2$ and $y_1 = y_2$. If $z = x + iy = 0$, then it implies that both $x = 0$ and $y = 0$ simultaneously.

Property Number 3: If $z_1 z_2 = 0$ it implies that either $z_1 = 0$ or $z_2 = 0$ or both are zero. To prove this fact, we note

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) = 0 \ .$$

Thus $x_1 x_2 - y_1 y_2 = 0 \quad \text{and} \quad x_1 y_2 + x_2 y_1 = 0$. Therefore,

$$(x_1x_2 - y_1y_2)^2 + (x_1y_2 + x_2y_1)^2 = 0 \ ,$$

or

$$x_1^2(x_2^2 + y_2^2) + y_1^2(x_2^2 + y_2^2\ ) = 0 \ ,$$

or

$$(x_1^2 + y_1^2)(x_2^2 + y_2^2) = 0 \ .$$

Thus, either $x_1^2 + y_1^2 = 0,$ which implies both $x_1 = 0$ and $y_1 = 0,$ that is, $z_1 = 0$, or $x_2^2 + y_2^2 = 0$, which implies both $x_2 = 0$ and $y_2 = 0$, that is, $z_2 = 0$, or both z_1 and z_2 are zero. Note in passing,

$$|z_1z_2| = \{(x_2^2 + y_1^2)(x_2^2 + y_2^2)\}^{1/2} = |z_1||z_2| \ .$$

Property Number 4: Division of one complex number by another complex number works as follows. If $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2 \neq 0$, then

$$\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} = \frac{z_1\bar{z}_2}{z_2\bar{z}_2} = \frac{(x_1 + iy_1)(x_2 - iy_2)}{x_2^2 + y_2^2} = \frac{(x_1x_2 + y_1y_2)}{x_2^2 + y_2^2} + i\frac{(y_1x_2 - x_1y_2)}{x_2^2 + y_2^2} \ .$$

Property Number 5: The properties of complex conjugation imply

$$\overline{z_1 \pm z_2} = \bar{z}_1 \ \pm \bar{z}_2 \ ,$$

$$\overline{z_1z_2} = \bar{z}_1\ \bar{z}_2 \ ,$$

$$\overline{\left(\frac{z_1}{z_2}\right)} = \frac{\bar{z}_1}{\bar{z}_2} \qquad (z_2 \neq 0) \ ,$$

$$\overline{(\bar{z})} = z \ .$$

If $\bar{z} = z$ then z is pure real, that is, $Im(z) = 0$ and if $\bar{z} = -z$ then z is pure imaginary, that is, $Re(z) = 0$. In fact,

$$Re(z) = \frac{z + \bar{z}}{2}, \qquad Im(z) = \frac{z - \bar{z}}{2\,i} \ .$$

Property Number 6: If $z = r\exp(i\theta) = x + iy$ then $\theta = \arg(z) = \arctan(y/x)$, and

$$\arg(z_1z_2) = \arg(z_1) + \arg(z_2) \ ,$$

$$\arg\left(\frac{z_1}{z_2}\right) = \arg(z_1) - \arg(z_2) \ .$$

Example: In this example, we take the two complex numbers that we used in the previous two examples, namely, $z_1 = 2 + 5i$ and $z_2 = 4 + 3i$, and illustrate some of the properties that we discussed above.

Solution:

$$z_1 + z_2 = 6 + 8i, \qquad |z_1 + z_2| = \sqrt{6^2 + 8^2} = \sqrt{100} = 10 \ ,$$

$$z_1 - z_2 = -2 + 2\,i \qquad |z_1 - z_2| = \sqrt{(-2)^2 + 2^2} = \sqrt{8} \ ,$$

$$z_1 z_2 = (2 + 5i)(4 + 3i) = 8 + 20i + 6i + i^2 15 = -7 + 26i \ ,$$

$$|z_1 z_2| = \sqrt{(-7)^2 + 26^2} = \sqrt{49 + 676} = \sqrt{725} = \sqrt{25 * 29} = 5\sqrt{29} = |z_1||z_2| \ ,$$

$$\frac{1}{z_1} = \frac{1}{2 + 5i}\frac{2 - 5i}{2 - 5i} = \frac{2 - 5i}{29} = \frac{2}{29} - \frac{5}{29}i \ ,$$

$$\frac{1}{z_2} = \frac{1}{4 + 3i}\frac{4 - 3i}{4 - 3i} = \frac{4 - 3i}{25} = \frac{4}{25} - \frac{3}{25}i \ .$$

Here is a list of some items which are so important that one should commit them to memory. The useful things to remember, for *k* an integer (positive or negative) or zero, are:

$$\exp(i\pi/2) = i$$

$$\exp(-i\pi/2) = -i$$

$$\exp(ik\pi) = (-1)^k$$

$$\exp(i2k\pi) = +1$$

Based on these useful items to remember, Property Number 2 for two complex numbers, in exponential form, can be written as follows. If $z_1 = r_1 \exp(i\theta_1)$ and $z_2 = r_2 \exp(i\theta_2)$, then $z_1 = z_2$ implies $r_1 = r_2$ and $\theta_1 = \theta_2 + 2k\pi$, where k is an integer including zero.

In some situations, dividing by a complex number or finding the reciprocal of a complex number is best accomplished by using the exponential form of the complex number instead of using Property Number 4. The following example will illustrate this point.

Example: Write the complex number $z = 4i/(\sqrt{3}+i)$ in its Cartesian form.

Solution: We write the numerator and denominator of the complex number $\boldsymbol{z}$ in exponential form as

$$z = \frac{4i}{\sqrt{3}+i} = \frac{4\exp(i\pi/2)}{2\exp(i\pi/6)} = 2\exp\left(\frac{i\pi}{2} - \frac{i\pi}{6}\right) = 2\exp(i\pi/3) = 2\left(\cos\frac{\pi}{3} + i\sin\frac{\pi}{3}\right) = 1 + i\sqrt{3}\ .$$

5.3 POWERS OF COMPLEX NUMBERS

In order to calculate powers of a complex number, z, it is best to work with exponential notation, namely,

$$z^n = (x+iy)^n = r^n \exp(in\theta)\ .$$

Example: Given $z = \frac{1}{2} + i\frac{\sqrt{3}}{2}$, determine all higher powers of z.

Solution: First, we write z in its polar or exponential form. Since $r = \sqrt{\frac{1}{4}+\frac{3}{4}} = 1$ and $\theta = \arctan\sqrt{3} = \frac{\pi}{3}$, it follows that in exponential form

$$z = 1\exp\left(\frac{i\pi}{3}\right)\ .$$

Then,

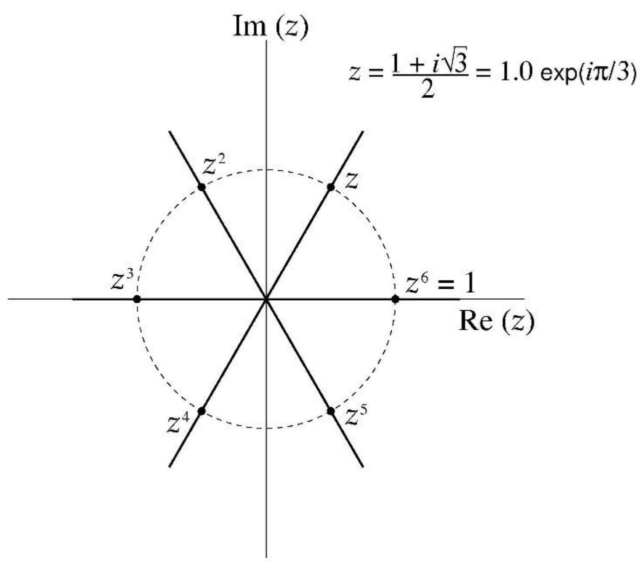


Figure 5.2. Powers of a complex number $z = 1\exp\left(\frac{i\pi}{3}\right)$.

$$z^2 = 1 \exp\left(i\frac{2\pi}{3}\right) \ ,$$

$$z^3 = 1 \exp(i\pi) \ ,$$

$$z^4 = 1 \exp\left(i\frac{4\pi}{3}\right) \ ,$$

$$z^5 = 1 \exp\left(i\frac{5\pi}{3}\right) \ ,$$

$$z^6 = 1 \exp\left(i\frac{6\pi}{3}\right) = 1 \exp(i2\pi) = 1 \ .$$

Since $z^6 = 1$, all higher powers of z, starting with z^7 onwards, take one of the six values above. These six values are shown in a complex plane in Figure 5.2. They lie on a circle of unit radius. If the magnitude r is different from 1, then we get a spiral instead of a circle. For example, for $r = 1.1$,

$$z = 1.1 \exp\left(i\frac{\pi}{2}\right) \ ,$$

$$z^2 = 1.21 \exp(i\pi) \ ,$$

$$z^3 = 1.331 \exp\left(i\frac{3\pi}{2}\right) \ ,$$

$$z^4 = 1.4641 \exp(2i\pi) \ , \ \ \text{etc} \ .$$

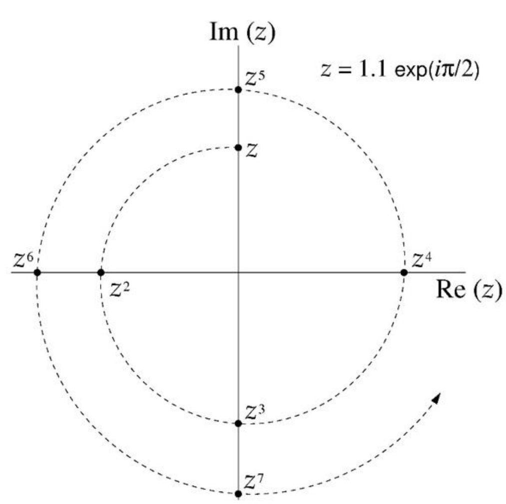


Figure 5.3. Powers of a complex number $z = 1.1 \exp\left(i\frac{\pi}{2}\right)$.

In the complex plane these values appear on an ever-expanding spiral, as seen in Figure 5.3. Similarly, if the magnitude of a complex number z is less than 1, that is $r < 1$, then all higher powers of z lie in the complex plane on an ever-contracting spiral. Also, for $r \neq 1$, all higher powers of z have a separate distinct value.

Example: Given $z = 1 + i$, what is z^8?

Solution: First, write the complex number z in polar or exponential form. Note $r = \sqrt{1+1} = \sqrt{2}$ and $\tan\theta = \frac{1}{1}$ so that $\theta = \pi/4$. So,

$$z = 1 + i = \sqrt{2}\exp\left(i\frac{\pi}{4}\right) \ .$$

Then, taking any higher power of z is straightforward,

$$z^8 = \left(\sqrt{2}\right)^8 \exp\left(8\frac{i\pi}{4}\right) = 2^4\exp(i2\pi) = 16 \ .$$

However, the long method would be to multiply $1 + i$ by itself eight times:

$$(1+i)^2 = 1 + 2i + i^2 = 1 + 2i - 1 = 2i \ ,$$

$$(1+i)^4 = (1+i)^2(1+i)^2 = (2i)(2i) = 4i^2 = -4 \ ,$$

$$(1+i)^8 = (1+i)^4(1+i)^4 = (-4)^2 = 16 \ .$$

5.4 ROOTS OF A COMPLEX NUMBER

Roots, such as square root or cube root, of a complex number are also evaluated most conveniently by using exponential form of the complex number. As a practice run, let us evaluate roots of $+1$ and -1. From algebra we know that a polynomial of order n,

$$1 + a_1 z + a_2 z^2 + \cdots + a_n z^n = 0 \ ,$$

where coefficients a_i can be real or complex, has exactly n roots. Some roots may be repeated. Thus, roots of a simple polynomial $1 - z^n = 0$ are the n roots of unity or +1. To obtain these roots, write

$$z^n = +1 = \exp(i2\pi k) \quad k = 0,1,2 \ldots$$

Then, $z = \exp\left(i\frac{2\pi k}{n}\right)$, which for different values of k provides n roots of unity. For $k = 0, 1, \ldots (n-1)$, these roots are, $1, \exp\left(\frac{i2\pi}{n}\right), \exp\left(\frac{i4\pi}{n}\right), \exp\left(\frac{i6\pi}{n}\right), \ldots \exp\left(\frac{i2\pi(n-1)}{n}\right)$. The next value of k, namely $k = n$, gives back the first root, that is, $z = 1$.

Similarly, from $z^n + 1 = 0$ or $z^n = -1 = \exp(i\pi)\ \exp(i2\pi k) = \exp[i\pi(2k+1)]$, the n roots of -1 are $\exp\left(\frac{i\pi}{n}\right), \exp\left(\frac{i3\pi}{n}\right), \exp\left(\frac{i5\pi}{n}\right), \ldots \exp\left(\frac{i(2n-1)\pi}{n}\right)$. The next value of k, namely $k = n$, gives back the first root, that is, $z = \exp\left(\frac{i\pi}{n}\right)$.

Example: Determine all possible values of square roots of +1.

Solution: In this case,

$$z^2 - 1 = 0 \quad \text{or} \quad z^2 = 1 = \exp(i2\pi k)$$

$$\text{or } z = \exp(i\pi k) \equiv z_k \text{ for } k = 0,1,2, \ldots.$$

The mathematical symbol ≡ means that it is an equality for any value of variable k. Thus, for $k = 0$ and 1,

$$z_0 = +1 \text{ and } z_1 = -1 \ .$$

Higher values of k simply keep repeating values of z_0 and z_1, that is,

$$z_k = +1 \ \text{ for } k = 2,4,6, \ldots \text{ and } \ z_k = -1 \text{ for } k = 3,5,7, \ldots$$

Thus, there are only two independent square roots of +1 and their values are ± 1. In the complex plane, these two roots of +1 lie on a unit circle as seen in Figure 5.4.

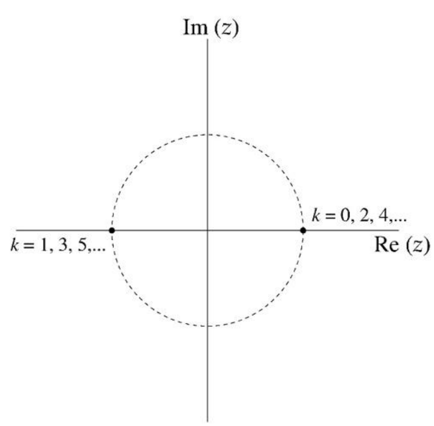


Figure 5.4. Square roots of +1.

Example: Determine all possible values of cube roots of +1.

Solution: In this case

$$z^3 - 1 = 0 \quad \text{or} \quad z^3 = 1 = \exp(i2\pi k)$$

$$\text{or,} \quad z = \exp\left(i\frac{2\pi}{3}k\right) \equiv z_k \text{ for } k = 0,1,2,\ldots$$

$$z_0 = +1,$$

$$z_1 = \exp\left(i\frac{2\pi}{3}\right) = \cos\left(\frac{2\pi}{3}\right) + i\sin\left(\frac{2\pi}{3}\right) = \frac{-1 + i\sqrt{3}}{2} ,$$

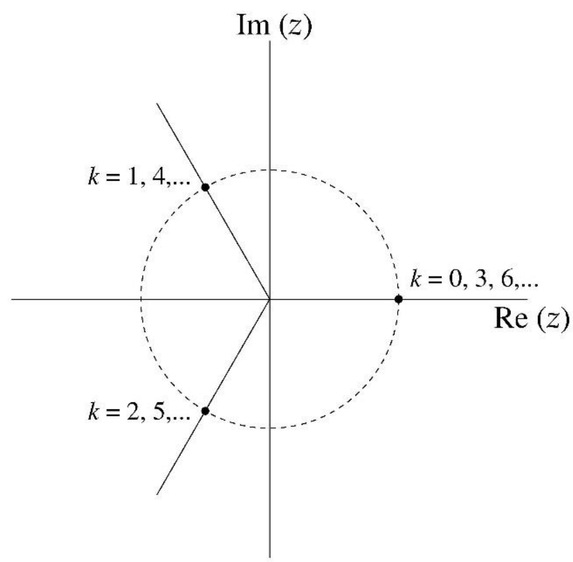


Figure 5.5. Cube roots of +1.

$$z_2 = \exp\left(i\frac{4\pi}{3}\right) = \cos\left(\frac{4\pi}{3}\right) + i\sin\left(\frac{4\pi}{3}\right) = \frac{-1 - i\sqrt{3}}{2} .$$

Since $z_3 = \exp(i2\pi) = +1$, which is same as z_0, the values of all z_k for $k \geq 3$ are repeated. So, there are only three independent cube roots of +1, namely, $z_0, z_1, \text{and } z_2$**.** Again, these three cube roots of +1 are seen to lie on a unit circle in the complex plane as in Figure 5.5.

Example: Given $\boldsymbol{z = 4i}$, what is $\boldsymbol{z^{1/2}}$?

Solution: First, convert z into its exponential form, $r = \sqrt{0 + 16} = 4$ and $\tan\theta = \frac{4}{0} = \infty$ or $\theta = \pi/2$, so that

$$z = 4\exp\left(i\frac{\pi}{2}\right)\ \exp(i2\pi k)$$

$$z^{1/2} \ = 2\exp(i\pi/4)\exp(i\pi k) \equiv z_k \text{ for } k = 0,1,2,\ldots.$$

For $k = 0$ and $k = 1$, we get

$$z_0 = 2\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right) = 2\left(\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}\right) = \sqrt{2}(1+i)\ \ ,$$

$$z_1 = 2\left(\cos\frac{5\pi}{4} + i\sin\frac{5\pi}{4}\right) = 2\left(-\frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2}\right) = -\sqrt{2}(1+i)\ \ .$$

For higher values of k, $k = 2, 3, 4, \ldots$, the roots z_0 and z_1 are repeated. Thus, $z = 4i$ has only two independent square roots, $\pm\sqrt{2}(1+i)$.

Example: Given $\boldsymbol{z = 1 - \sqrt{3}i}$, what is $\boldsymbol{\sqrt{z}}$?

Solution: Again, converting the complex number from its Cartesian form to its polar or exponential form, we get

$$z = 1 - \sqrt{3}i = 2\exp(-i\pi/3)\ \exp(i2\pi k)\ \ .$$

Then,

$$z^{1/2} = \sqrt{2}\exp(-i\pi/6)\exp(i\pi k) \equiv z_k \text{ for } k = 0,1,2,\ldots.$$

For $k = 0$ and $k = 1$, we have

$$z_0 = \sqrt{2}\left(\cos\frac{\pi}{6} - i\sin\frac{\pi}{6}\right) = \sqrt{2}\left(\frac{\sqrt{3}}{2} - i\frac{1}{2}\right) = \sqrt{\frac{3}{2}} - i\ \frac{1}{\sqrt{2}}\ ,$$

$$z_1 = \sqrt{2}\left(\cos\frac{5\pi}{6} + i\sin\frac{5\pi}{6}\right) = \sqrt{2}\left(-\frac{\sqrt{3}}{2} + i\frac{1}{2}\right) = -\sqrt{\frac{3}{2}} + i\ \frac{1}{\sqrt{2}}\ .$$

Again, for higher values of k, $k = 2, 3, 4, \ldots$, the roots z_0 and z_1 are repeated. Thus,

$z = 1 - \sqrt{3}i$ has only two independent square roots, $\pm\sqrt{\frac{3}{2}} \mp i\ \frac{1}{\sqrt{2}}$.

5.5 DEMOIVRE'S FORMULA

In the Interlude section we introduced the Euler's formula, namely,

$$\exp(ix) = \cos x + i \sin x \ .$$

If we put $x = n\theta$ in Euler's formula, we get

$$\exp(in\theta) = \cos(n\theta) + i \sin(n\theta) \ .$$

Also,

$$\exp(in\theta) = (\exp(i\theta))^n = (\cos\theta + i \sin\theta)^n \ .$$

So,

$$(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta) \qquad Eq. (5.3)$$

which is known as DeMoivre's formula.

Applications of DeMoivre's Formula

In some situations, it becomes useful to write powers of a sine or a cosine function, such as $\sin^{\mathrm{n}}\theta$ or $\cos^{\mathrm{n}}\theta$ [integer n], in terms of sine or cosine of various multiple angles like $\sin(m\theta)$ or $\cos(m\theta)$ [integer m]. This can be accomplished by a simple application of DeMoivre's formula. This formula also allows us to write $\sin(n\theta)$ or $\cos(n\theta)$ [integer n] in terms of multiple powers of sine or cosine, as $\sin^{\mathrm{m}}\theta$ or $\cos^{\mathrm{m}}\theta$ [integer m]. For convenience, write

$$z = \cos\theta + i\sin\theta = \exp(i\theta) \ ,$$

then

$$z^{-1} = [\exp(i\theta)]^{-1} = \cos\theta - i\sin\theta \ ,$$

and $z^n = \exp(in\theta)\,, z^{-n} = \exp(-in\theta)$.

Then,

$$z^n + z^{-n} = \exp(in\theta) + \exp(-in\theta) = 2\cos(n\theta) \ , \qquad Eq. (5.4a)$$

and

$$z^n - z^{-n} = \exp(in\theta) - \exp(-in\theta) = 2i\sin(n\theta) \quad . \qquad\qquad Eq. (5.4b)$$

These relationships along with DeMoivre's formula allow us to accomplish what we set out to do.

Example: Write $\cos^3\theta$ and $\sin^4\theta$ in terms of sine or cosine of angles which are multiples of θ.

Solution: Using Eqs. (5.4),

$$\cos^3\theta = \left(\frac{z+z^{-1}}{2}\right)^3 = \frac{1}{8}(z^3 + 3z + 3z^{-1} + z^{-3}) = \frac{1}{8}([z^3 + z^{-3}] + 3[z + z^{-1}])$$

$$= \frac{1}{8}[2\cos(3\theta) + 6\cos\theta] = \frac{1}{4}[\cos(3\theta) + 3\cos\theta]$$

and

$$\sin^4\theta = \left(\frac{z-z^{-1}}{2i}\right)^4 = \frac{1}{16}(z^4 - 4z^2 + 6 - 4z^{-2} + z^{-4})$$

$$= \frac{1}{16}([z^4 + z^{-4}] - 4[z^2 + z^{-2}] + 6)$$

$$= \frac{1}{16}[2\cos(4\theta) - 4\cdot 2\cos(2\theta) + 6]$$

$$= \frac{1}{8}[\cos(4\theta) - 4\cos(2\theta) + 3] \ .$$

Example: Write $\cos(4\theta)$, $\sin(4\theta)$, $\cos(2\theta)$, and $\sin(2\theta)$ in terms of multiples of $\sin\theta$ and $\cos\theta$.

Solution: Starting with DeMoivre's formula,

$$\cos(4\theta) + i\sin(4\theta) = (\cos\theta + i\sin\theta)^4$$

$$= \cos^4\theta + 4\cos^3\theta\,(i\sin\theta) + 6\cos^2\theta\,(i\sin\theta)^2 + 4\cos\theta\,(i\sin\theta)^3 + (i\sin\theta)^4$$

$$= (\cos^4\theta - 6\cos^2\theta\sin^2\theta + \sin^4\theta) + 4i\,(\cos^3\theta\sin\theta - \cos\theta\sin^3\theta)$$

Separating out the real and imaginary parts on both sides leads to

$$\cos(4\theta) = \cos^4\theta - 6\cos^2\theta\sin^2\theta + \sin^4\theta \ ,$$

and

$$\sin(4\theta) = 4\cos\theta\sin\theta\ (\cos^2\theta - \sin^2\theta)\ .$$

Similarly,

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

$$= (\cos^2\theta - \sin^2\theta) + i2\sin\theta\cos\theta\ .$$

Separating out real and imaginary parts, we get well-known relations

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

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

PROBLEMS FOR CHAPTER 5

1. Show that for any two complex numbers z_1 and z_2

$$|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2|z_1|^2 + 2|z_2|^2$$

2. Given a complex number,

$$z = \frac{5 + i\,10}{(1 - i\,2)(2 - i)},$$

determine the real and imaginary parts of z.

3. Determine the real part, imaginary part, and the absolute magnitude of the following three expressions,

$$\sin(i \ln i), \qquad \sqrt{i} \qquad \text{and} \qquad \frac{1}{2i}\ln\left(\frac{1 + ia}{1 - ia}\right).$$

4. If $z = x + iy$, determine the real part, imaginary part and the absolute magnitude of the following three functions,

(a) $\sin z$, (b) $\cos z$ and (c) $\exp(iz)$.

5. For two complex numbers $z_1 = \sqrt{3} + i$ and $z_2 = -\sqrt{2} + i\sqrt{2}$, determine

(a) $|z_1/z_2|^4$ and (b) $(z_1/z_2)^4$.

6. Find all roots of the equation $z^6 - 1 = 0$ and plot the results in a complex plane.

7. The three cube roots of +1 are of the form 1, ω and ω^2. Determine $Re\,[\omega]$ and $Im\,[\omega]$.

8. The three cube roots of -1 are of the form α, α^3 and α^5. Determine $Re\,[\alpha]$ and $Im\,[\alpha]$.

9. Determine the fourth root of $z = -8 + i\,8\sqrt{3}$.

10. Determine the third (or, cube) root of $z = -2 + i\,2$.

11. Use DeMoivre's formula to express $\sin^3\theta$ and $\cos^4\theta$ in terms of sine and/or cosine of multiples of θ.

12. Use DeMoivre's formula to express $\cos(3\theta)$ and $\sin(3\theta)$ in terms of powers of $\sin\theta$ and $\cos\theta$.

Chapter 6: Determinants

In this chapter we will first define a determinant as well as its minors and cofactors. Next, we will outline a procedure, due to Laplace, for determining the value of the determinant in terms of its cofactors. Several properties of determinants will be described which help in simplifying the determinant so that it can be reduced and evaluated easily. Two final examples will help in understanding the utility of properties in the simplification process for a determinant.

6.1 DEFINITION OF A DETERMINANT

A determinant is a *square* array of numbers (or elements) such as

$$\mathbb{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \cdots & a_{nn} \end{vmatrix} ,$$

which are combined together according to certain rules to provide the value of the determinant. The element a_{ij} is located at the intersection of i^{th} row and j^{th} column. The number of rows (or columns) is called the order of the determinant. The determinant shown above is of order n. We also define the *minor* and the *cofactor* of each element of a determinant. Starting with a determinant of order n, one can obtain a smaller determinant by omitting one row and one column. The determinant of order $(n-1)$ obtained by omitting the row and column containing a_{ij} (that is, omitting i^{th} row and j^{th} column) is called the minor M_{ij} of a_{ij} and the minor multiplied by a sign of $(-1)^{i+j}$ is called the cofactor C_{ij} of a_{ij}.

Let us start evaluating some determinants. A single number can be considered as a determinant of order 1 and the value of this determinant is the value of the single number.

Next higher order determinant is of order 2. Consider a determinant of order 2,

$$\mathbb{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} .$$

According to the rules of determining the values of a determinant, the value of a determinant of order 2 is the difference between the products of elements along two diagonals as

$$\det(\mathbb{A}) = a_{11}a_{22} - a_{21}a_{12} \ .$$

Note, if we interchange the rows and columns of this determinant to form a new determinant $\mathbb{B}$,

$$\mathbb{B} = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} ,$$

then

$$\det(\mathbb{B}) = \det(\mathbb{A}) \ .$$

Next higher order determinant is of order 3. A typical determinant of order 3 is

$$\mathbb{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \ .$$

Again, according to the rules of determining values of a determinant, the value of a determinant of order 3 can be written in terms of Levi-Civita symbols as

$$\det(\mathbb{A}) = \sum_{ijk} \epsilon_{ijk} a_{i1} a_{j2} a_{k3} = \epsilon_{123} a_{11} a_{22} a_{33} + \epsilon_{132} a_{11} a_{32} a_{23} + \epsilon_{213} a_{21} a_{12} a_{33} + \epsilon_{231} a_{21} a_{32} a_{13}$$
$$+ \epsilon_{312} a_{31} a_{12} a_{23} + \epsilon_{321} a_{31} a_{22} a_{13} = \sum_{ijk} \epsilon_{ijk} a_{1i} a_{2j} a_{3k} \ .$$

The last equality suggests that the value of a determinant of order 3 is not changed if its rows and columns are interchanged. In other words, if a determinant $\mathbb{B}$ is obtained from $\mathbb{A}$ by interchanging its rows and columns, then,

$$\det(\mathbb{B}) = \det(\mathbb{A}) \ .$$

Substituting explicitly the values of all Levi-Civita symbols, we get

$$\det(\mathbb{A}) = a_{11} a_{22} a_{33} - a_{11} a_{32} a_{23} - a_{21} a_{12} a_{33} + a_{21} a_{32} a_{13} + a_{31} a_{12} a_{23} - a_{31} a_{22} a_{13} \ .$$

Note that each product on the right-hand side contains only one element from each row and from each column of the determinant. Furthermore, factors on the right-hand side can be rearranged either as

$$\det(\mathbb{A}) = a_{11}(a_{22} a_{33} - a_{32} a_{23}) - a_{21}(a_{12} a_{33} - a_{13} a_{32}) + a_{31}(a_{12} a_{23} - a_{13} a_{22})$$

$$= a_{11}(-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{21}(-1)^{2+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}(-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

$$= a_{11}(-1)^{1+1} M_{11} + a_{21}(-1)^{2+1} M_{21} + a_{31}(-1)^{3+1} M_{31}$$

$$= a_{11}C_{11} + a_{21}C_{21} + a_{31}C_{31} \ ,$$

or as

$$\det(\mathbb{A}) = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$$

$$= a_{11}(-1)^{1+1}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1)^{1+2}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(-1)^{1+3}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}(-1)^{1+1}M_{11} + a_{12}(-1)^{1+2}M_{12} + a_{13}(-1)^{1+3}M_{13}$$

$$= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \ .$$

These latter forms suggest an alternate way, due to Laplace, of obtaining the value of a determinant of any order n as

$$\det(\mathbb{A}) = \sum_{i=1}^{n} a_{i1}C_{i1} \ ,$$

where the expansion of the determinant is across first column, or as

$$\det(\mathbb{A}) = \sum_{i=1}^{n} a_{1i}C_{1i} \ ,$$

where the expansion of the determinant is across first row. It suggests that rows and columns of a determinant can be interchanged without affecting its value. In fact, the value of a determinant can be obtained by expanding across *any* row or *any* column. We illustrate this fact in the example below.

Example: Determine the value of the following determinant of order 3, by expanding it across any row or any column.

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{0} & \mathbf{-1} & \mathbf{2} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} \end{vmatrix} \ .$$

Solution: Expanding across first row,

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix} -1 & 2 \\ 1 & 0 \end{vmatrix} + (2)(-1)^{1+2}\begin{vmatrix} 0 & 2 \\ 1 & 0 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 0 & -1 \\ 1 & 1 \end{vmatrix}$$

$$= 1(0-2) - 2(0-2) + 1(0+1) = +3 \ .$$

Expanding across second row,

$$\det(\mathbb{A}) = (0)(-1)^{2+1}\begin{vmatrix}2 & 1\\1 & 0\end{vmatrix} + (-1)(-1)^{2+2}\begin{vmatrix}1 & 1\\1 & 0\end{vmatrix} + (2)(-1)^{2+3}\begin{vmatrix}1 & 2\\1 & 1\end{vmatrix}$$

$$= -0(0-1) - 1(0-1) - 2(1-2) = +3 \ .$$

Expanding across third row,

$$\det(\mathbb{A}) = (1)(-1)^{3+1}\begin{vmatrix}2 & 1\\-1 & 2\end{vmatrix} + (1)(-1)^{3+2}\begin{vmatrix}1 & 1\\0 & 2\end{vmatrix} + (0)(-1)^{3+3}\begin{vmatrix}1 & 2\\0 & -1\end{vmatrix}$$

$$= 1(4+1) - 1(2-0) + 0(-1+0) = +3 \ .$$

Expanding across first column,

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix}-1 & 2\\1 & 0\end{vmatrix} + (0)(-1)^{2+1}\begin{vmatrix}2 & 1\\1 & 0\end{vmatrix} + (1)(-1)^{3+1}\begin{vmatrix}2 & 1\\-1 & 2\end{vmatrix}$$

$$= 1(0-2) - 0(0-1) + 1(4+1) = +3 \ .$$

Expanding across second column,

$$\det(\mathbb{A}) = (2)(-1)^{1+2}\begin{vmatrix}0 & 2\\1 & 0\end{vmatrix} + (-1)(-1)^{2+2}\begin{vmatrix}1 & 1\\1 & 0\end{vmatrix} + (1)(-1)^{3+2}\begin{vmatrix}1 & 1\\0 & 2\end{vmatrix}$$

$$= -2(0-2) - 1(0-1) - 1(2-0) = +3 \ .$$

Expanding across third column,

$$\det(\mathbb{A}) = (1)(-1)^{1+3}\begin{vmatrix}0 & -1\\1 & 1\end{vmatrix} + (2)(-1)^{2+3}\begin{vmatrix}1 & 2\\1 & 1\end{vmatrix} + (0)(-1)^{3+3}\begin{vmatrix}1 & 2\\0 & -1\end{vmatrix}$$

$$= 1(0+1) - 2(1-2) + 0(-1-0) = +3 \ .$$

Thus, in general,

$$\sum_k a_{ik}C_{jk} = \sum_k C_{ki}a_{kj} = \delta_{ij}\det(\mathbb{A}) \ . \qquad Eq.(6.1)$$

6.2 PROPERTIES OF DETERMINANT $\mathbb{A}$

There are several very useful properties of determinants which are helpful in determining the value of a large determinant. Since the value of a determinant does not change on interchanging the rows and the columns, all the remarks below for columns of a determinant will apply equally well to the rows of the determinant.

As a shorthand notation for writing the determinant, let us write A_l for the l^{th} column of $\mathbb{A}$. Then

$$\mathbb{A} = |A_1, A_2, \dots A_l, \dots A_n| \ .$$

Property Number 1: If any two columns of a determinant $\mathbb{A}$ are interchanged, then the value of the resulting determinant $\mathbb{B}$ is same as the value of the original determinant except for an overall change of sign. In other words, if

$$\mathbb{A} = |A_1, A_2 \dots A_l, \dots A_m, \dots A_n| \ ,$$

and

$$\mathbb{B} = |A_1, A_2 \dots A_m, \dots A_l, \dots A_n| \ ,$$

then

$$\det \mathbb{B} = -\det \mathbb{A} \ .$$

To prove this fact, we first assume that the l^{th} column and the m^{th} column of $\mathbb{A}$ are two adjacent columns, that is,

$$\mathbb{A} = |A_1, A_2 \dots A_l, A_m, \dots A_n|$$

and

$$\mathbb{B} = |A_1, A_2 \dots A_m, A_l, \dots A_n| \ .$$

Now, to determine the values of these two determinants, let us expand each determinant across the column that contains elements of the original l^{th} column. Each element of this column will have the same minor both in $\mathbb{A}$ and in $\mathbb{B}$; however, the cofactor of each element will differ in sign only since the l^{th} column is moved from left of m^{th} column to the right of m^{th} column. So, $\det \mathbb{B} = -\det \mathbb{A}$. This result is true even if the l^{th} column and the m^{th} column of $\mathbb{A}$ are not adjacent columns. If the l^{th} column and the m^{th} column are separated by j columns, such as

$$\mathbb{A} = |A_1, A_2 \dots A_l, \leftarrow j \text{ columns } \rightarrow, A_m, \dots A_n| \ ,$$

then a total of $(2j+1)$ interchanges of adjacent columns are required to exchange the l^{th} column and the m^{th} column and since $(-1)^{2j+1} = (-1)$ for any j, the fact that $\det \mathbb{B} = -\det \mathbb{A}$ is proved.

Example: A determinant $\mathbb{B}$ is obtained by interchanging the first two columns of another determinant $\mathbb{A}$ as shown below:

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix} \ \textbf{and}\ \mathbb{B} = \begin{vmatrix} \mathbf{-1} & \mathbf{1} & \mathbf{1} \\ \mathbf{3} & \mathbf{2} & \mathbf{-2} \\ \mathbf{1} & \mathbf{3} & \mathbf{-4} \end{vmatrix} .$$

Show that $\mathbf{det}\,\mathbb{B} = -\mathbf{det}\,\mathbb{A}$.

Solution: Expanding across first row, we get

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & -2 \\ 1 & -4 \end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix} 2 & -2 \\ 3 & -4 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= 1(-12+2) + 1(-8+6) + 1(2-9) = -19$$

and

$$\det(\mathbb{B}) = (-1)(-1)^{1+1}\begin{vmatrix} 2 & -2 \\ 3 & -4 \end{vmatrix} + (1)(-1)^{1+2}\begin{vmatrix} 3 & -2 \\ 1 & -4 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 3 & 2 \\ 1 & 3 \end{vmatrix}$$

$$= -1(-8+6) - 1(-12+2) + 1(9-2) = +19 \ .$$

Property Number 2: If a determinant has two identical columns, such as

$$\mathbb{A} = |A_1 A_2 \dots A_l \dots A_l \dots A_n| \quad ,$$

then $\det \mathbb{A} = 0$.

The proof of this statement easily follows from Property Number 1, since on interchanging the two columns we get a negative sign, while the determinant stays the same since the two exchanged columns are identical. Thus,

$$\det(\mathbb{A}) = -\det(\mathbb{A}) \ .$$

Or

$$\det(\mathbb{A}) = 0 \ .$$

Example: Two columns of a determinant $\mathbb{A}$ are identical. Determine the value of this determinant.

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{-1} \\ \mathbf{2} & \mathbf{3} & \mathbf{3} \\ \mathbf{3} & \mathbf{1} & \mathbf{1} \end{vmatrix} .$$

Solution: Expanding the determinant across first row, we get

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & 3 \\ 1 & 1 \end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix} + (-1)(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= 1(3-3) + 1(2-9) - 1(2-9) = 0 \ .$$

Property Number 3: If each element of a column of a determinant is multiplied by a constant k, then the value of the whole determinant is multiplied by the same constant k. Or, saying differently, multiplying a determinant by a constant k amounts to multiplying any one of the columns of the determinant by k.

Suppose determinant $\mathbb{B}$ is constructed by multiplying the l^{th} column of determinant $\mathbb{A}$ by a constant number k. Then

$$\mathbb{A} = \ |A_1, A_2 \dots A_l, \dots A_n \ |$$

and

$$\mathbb{B} = \ |A_1, A_2 \dots kA_l, \dots A_n \ | \ .$$

Also,

$$\det(\mathbb{A}) = \sum_{i=1}^{n} a_{il} C_{il}$$

and

$$\det(\mathbb{B}) = \sum_{i=1}^{n} (k a_{il}) C_{il} = k \sum_{i=1}^{n} a_{il} C_{il} = k \det(\mathbb{A}) \ .$$

Example: A new determinant $\mathbb{B}$ is formed by multiplying the third column of another determinant $\mathbb{A}$ by a constant number $k = 5$, as shown:

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix} \ \textbf{and} \ \mathbb{B} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{5} \\ \mathbf{2} & \mathbf{3} & \mathbf{-10} \\ \mathbf{3} & \mathbf{1} & \mathbf{-20} \end{vmatrix} \ .$$

Show that $\det \mathbb{B} = k \det \mathbb{A}$.

Solution: Determinant $\mathbb{A}$ is the same determinant as in the previous example. Its value, as determined previously, is -19.

Now, expanding the determinant $\mathbb{B}$ across the first column, we get

$$\det(\mathbb{B}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & -10 \\ 1 & -20 \end{vmatrix} + (2)(-1)^{2+1}\begin{vmatrix} -1 & 5 \\ 1 & -20 \end{vmatrix} + (3)(-1)^{3+1}\begin{vmatrix} -1 & 5 \\ 3 & -10 \end{vmatrix}$$

$$= 1(-60+10) - 2(20-5) + 3(10-15) = -95 = 5(-19) \ .$$

Property Number 4: If each element of the l^{th} column of a determinant $\mathbb{A}$ can be expressed as

$a_{jl} = \alpha p_{jl} + \beta q_{jl}$ with $j = 1,2 \cdots n$, then, columnwise,

$$A_l = \alpha P_l + \beta Q_l \ .$$

Now, expanding the determinant $\mathbb{A}$ across its l^{th} column

$$\det(\mathbb{A}) = \det|A_1 A_2 \cdots A_{l-1}\, \alpha P_l + \beta Q_l\, A_{l+1} \cdots A_n|$$

$$= \sum_{j=1}^{n} (\alpha p_{jl} + \beta q_{jl}) C_{jl} = \alpha \sum_{j=1}^{n} p_{jl} C_{jl} + \beta \sum_{j=1}^{n} q_{jl} C_{jl}$$

$$= \alpha \det|A_1 A_2 \cdots A_{l-1}\, P_l\, A_{l+1} \cdots A_n| + \beta \det|\, A_1 A_2 \cdots A_{l-1}\, Q_l\, A_{l+1} \cdots A_n|$$

$$= \alpha\ \det(\mathbb{P}) + \beta\ \det(\mathbb{Q}) \ .$$

Note, the determinants $\mathbb{A}$, $\mathbb{P}$ and $\mathbb{Q}$ share all columns except the l^{th} column.

Example: Two new determinants, $\mathbb{P}$ and $\mathbb{Q}$, are created by splitting the third column of another determinant $\mathbb{A}$, as shown:

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix} \ \textbf{with}\ \mathbb{P} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-1} \\ \mathbf{3} & \mathbf{1} & \mathbf{-2} \end{vmatrix} \ \textbf{and}\ \mathbb{Q} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{0} \\ \mathbf{2} & \mathbf{3} & \mathbf{-1} \\ \mathbf{3} & \mathbf{1} & \mathbf{-2} \end{vmatrix} \ .$$

Show that $\mathbf{det}(\mathbb{A}) = \mathbf{det}(\mathbb{P}) + \ \mathbf{det}(\mathbb{Q})$.

Solution: Determinant $\mathbb{A}$ is the determinant of previous few examples. Its value, as determined previously, is -19.

Now, expanding the determinants $\mathbb{P}$ and $\mathbb{Q}$ across the first row, we get

$$\det(\mathbb{P}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & -1 \\ 1 & -2 \end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix} 2 & -1 \\ 3 & -2 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= 1(-6+1) + 1(-4+3) + 1(2-9) = -5 - 1 - 7 = -13 \ ,$$

$$\det(\mathbb{Q}) = (1)(-1)^{1+1}\begin{vmatrix}3 & -1\\1 & -2\end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix}2 & -1\\3 & -2\end{vmatrix} + (0)(-1)^{1+3}\begin{vmatrix}2 & 3\\3 & 1\end{vmatrix}$$

$$= 1(-6+1) + 1(-4+3) - 0 = -5 - 1 = -6 \ .$$

Clearly, $\det(\mathbb{A}) = \det(\mathbb{P}) + \det(\mathbb{Q})$.

Property Number 5: If a determinant $\mathbb{B}$ is obtained from determinant $\mathbb{A}$ by adding to the l^{th} column of $\mathbb{A}$ a scalar multiple of any other column of $\mathbb{A}$, then $\det\mathbb{B} = \det\mathbb{A}$.

Explicitly,

$$\mathbb{B} = |A_1A_2 \cdots A_l + \alpha A_m \cdots A_m \cdots A_n| \ .$$

Using Property Number 4,

$$\det\mathbb{B} = \det|A_1A_2 \cdots A_l \cdots A_m \cdots A_n| + \alpha\det|A_1A_2 \cdots A_m \cdots A_m \cdots A_n| \ .$$

Or, using Property Number 2,

$$\det\mathbb{B} = \det|A_1A_2 \cdots A_l \cdots A_m \cdots A_n| + 0 = \det\mathbb{A} \ .$$

Example: A new determinant $\mathbb{B}$ is created by adding first column to the third column of another determinant $\mathbb{A}$ as shown:

$$\mathbb{A} = \begin{vmatrix}\mathbf{1} & \mathbf{-1} & \mathbf{1}\\ \mathbf{2} & \mathbf{3} & \mathbf{-2}\\ \mathbf{3} & \mathbf{1} & \mathbf{-4}\end{vmatrix} \ \textbf{and} \ \mathbb{B} = \begin{vmatrix}\mathbf{1} & \mathbf{-1} & \mathbf{2}\\ \mathbf{2} & \mathbf{3} & \mathbf{0}\\ \mathbf{3} & \mathbf{1} & \mathbf{-1}\end{vmatrix} \ .$$

Show that $\mathbf{det}\,\mathbb{B} = \mathbf{det}\,\mathbb{A}$.

Solution: Determinant $\mathbb{A}$ is the same determinant as in the previous examples. Its value, as determined previously, is -19.

Now, expanding the determinant $\mathbb{B}$ across the third column, we get

$$\det(\mathbb{B}) = (2)(-1)^{1+3}\begin{vmatrix}2 & 3\\3 & 1\end{vmatrix} + (0)(-1)^{2+3}\begin{vmatrix}1 & -1\\3 & 1\end{vmatrix} + (-1)(-1)^{3+3}\begin{vmatrix}1 & -1\\2 & 3\end{vmatrix}$$

$$= 2(2-9) - 0 - 1(3+2) = -14 - 5 = -19 \ .$$

Finally, we provide two examples in which we use properties of determinants to first simplify the determinant, then evaluate it.

Example: Using properties of determinants, find out the value of

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix} .$$

Solution: We have already seen that the value of $\mathbb{A}$ is -19. Now we will use various properties of determinants to create as many zeros in a single row or in a single column as possible. We will do so for the first row of determinant $\mathbb{A}$.

First, we add column 2 to column 3 and use property number 5 to get

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix} .$$

Next, we add column 1 to column 2 and use property number 5 again to get

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 5 & 1 \\ 3 & 4 & -3 \end{vmatrix} .$$

Now, we expand across row 1 to get,

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 5 & 1 \\ 3 & 4 & -3 \end{vmatrix} = 1 \begin{vmatrix} 5 & 1 \\ 4 & -3 \end{vmatrix} .$$

The determinant of order 2 is easy to evaluate as

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 5 & 1 \\ 3 & 4 & -3 \end{vmatrix} = 1 \begin{vmatrix} 5 & 1 \\ 4 & -3 \end{vmatrix} = -19 .$$

As seen in the above example, the trick of simplifying the evaluation of a determinant is to create as many zeros in a single row (or column) as possible and then use this row (or column) to evaluate the determinant.

Example: Given that $\begin{vmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{d} & \boldsymbol{e} & \boldsymbol{f} \\ \boldsymbol{g} & \boldsymbol{h} & \boldsymbol{i} \end{vmatrix} = \mathbf{5}$, use the properties of general determinants to find the value of

$$\mathbb{D} = \begin{vmatrix} \boldsymbol{-2a} & \boldsymbol{3d-g} & \boldsymbol{g+a} \\ \boldsymbol{-2b} & \boldsymbol{3e-h} & \boldsymbol{h+b} \\ \boldsymbol{-2c} & \boldsymbol{3f-i} & \boldsymbol{i+c} \end{vmatrix} .$$

Solution: The determinant to be evaluated is

$$\mathbb{D} = \begin{vmatrix} -2a & 3d-g & g+a \\ -2b & 3e-h & h+b \\ -2c & 3f-i & i+c \end{vmatrix} .$$

First, take -2 common out of column 1 to get

$$\mathbb{D} = (-2)\begin{vmatrix} a & 3d-g & g+a \\ b & 3e-h & h+b \\ c & 3f-i & i+c \end{vmatrix} .$$

Now, subtract column 1 from column 3 to get

$$\mathbb{D} = (-2)\begin{vmatrix} a & 3d-g & g \\ b & 3e-h & h \\ c & 3f-i & i \end{vmatrix} .$$

Adding column 3 to column 2 gives

$$\mathbb{D} = (-2)\begin{vmatrix} a & 3d & g \\ b & 3e & h \\ c & 3f & i \end{vmatrix} .$$

Taking a common factor of 3 out of column 2, we have

$$\mathbb{D} = (-2)(3)\begin{vmatrix} a & d & g \\ b & e & h \\ c & f & i \end{vmatrix} .$$

Finally, interchange rows and columns of the determinant,

$$\mathbb{D} = (-6)\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = (-6)(5) = -30 .$$

PROBLEMS FOR CHAPTER 6

1. Determine the values of the following two determinants

$$\begin{vmatrix} 5 & 2 & 3 \\ -1 & 3 & 4 \\ 0 & -2 & 3 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} -2 & 3 & 5 \\ 1 & 2 & 4 \\ 0 & 3 & -2 \end{vmatrix}.$$

2. Using the properties of determinants, first simplify and then determine the value of the following determinant,

$$\begin{vmatrix} a-b & a^2-b^2 & a^3-b^3 \\ b-c & b^2-c^2 & b^3-c^3 \\ c-a & c^2-a^2 & c^3-a^3 \end{vmatrix}.$$

3. Without expanding the determinant, and using the properties of determinants only, show that

$$\begin{vmatrix} bc & a^2 & a^2 \\ b^2 & ca & b^2 \\ c^2 & c^2 & ab \end{vmatrix} = \begin{vmatrix} bc & ab & ca \\ ab & ca & bc \\ ca & bc & ab \end{vmatrix}.$$

4. Without expanding the determinant, and using the properties of determinants only, show that

$$\begin{vmatrix} n\,a_1+b_1 & n\,a_2+b_2 & n\,a_3+b_3 \\ n\,b_1+c_1 & n\,b_2+c_2 & n\,b_3+c_3 \\ n\,c_1+a_1 & n\,c_2+a_2 & n\,c_3+a_3 \end{vmatrix} = (n+1)\,(n^2-n+1)\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

5. Without expanding the determinant, and using the properties of determinants only, show that

$$\begin{vmatrix} a^2 & (a+1)^2 & (a+2)^2 \\ b^2 & (b+1)^2 & (b+2)^2 \\ c^2 & (c+1)^2 & (c+2)^2 \end{vmatrix} = -4\,(a-b)\,(b-c)\,(c-a)\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Chapter 7: Matrices

In this chapter we will learn about matrices and their properties. Just like determinants, matrices are also arrays of numbers or elements, but that is where the similarity between a determinant and a matrix ends. A determinant, by design, is a square array of numbers while a matrix, in general, is a rectangular array of numbers. With determinants, the elements are combined together to provide a single number, which is the value of the determinant. On the other hand, in the case of a matrix there is no concept of a single value of the matrix available by any combination of its elements. However, only for a square matrix, one can mention a corresponding determinant whose elements are the same as the elements of the square matrix, or one can loosely talk about determinant of a square matrix. As an application of matrices, in this chapter we will explore their use in solving n coupled inhomogeneous algebraic equations.

7.1 A GENERAL MATRIX

A general matrix is a rectangular array of numbers or elements. For example,

$$\mathbb{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is a matrix with m rows and n columns. It is common to say that $\mathbb{A}$ is an $m \times n$ matrix. The element a_{ij} is located at the intersection of i^{th} row and j^{th} column of the matrix. If $\mathbb{A}$ = $\mathbb{B}$, then all elements of matrix $\mathbb{A}$ are individually equal to all elements of matrix $\mathbb{B}$; that is, $a_{ij} = b_{ij}$ for all i and j. Clearly, if $\mathbb{A}$ is an $m \times n$ matrix, then so is $\mathbb{B}$. Matrices consisting of only one row or one column are referred to as vectors. For example, an $m \times 1$ matrix is a column vector of dimension m, and a $1 \times n$ matrix is a row vector of dimension n.

Various matrices combine (add, subtract or multiply) according to some well-defined rules. First, in order to add or subtract two matrices, they must be of the same size; that is, must have same number of rows and same number of columns. Furthermore, matrices add or subtract element-by-element. For example, if $\mathbb{C} = \mathbb{A} \pm \mathbb{B}$, then $c_{ij} = a_{ij} \pm b_{ij}$. Matrix addition is commutative as well as associative, that is,

$$\mathbb{A} + \mathbb{B} = \mathbb{B} + \mathbb{A} \text{ and } \mathbb{A} + (\mathbb{B} + \mathbb{C}) = (\mathbb{A} + \mathbb{B}) + \mathbb{C} \ .$$

Multiplication of two matrices, $\mathbb{A}$ and $\mathbb{B}$, leads to a new matrix $\mathbb{C}$, that is, $\mathbb{C} = \mathbb{A}\mathbb{B}$. The element c_{ij}, located at the intersection of i^{th} row and j^{th} column of matrix $\mathbb{C}$, is given by

$$c_{ij} = \sum_k a_{ik} b_{kj} \quad . \qquad\qquad Eq. (7.1)$$

In other words, element c_{ij} of $\mathbb{C}$ is obtained by multiplying, element-by-element, the i^{th} row of $\mathbb{A}$ with the j^{th} column of $\mathbb{B}$, and then adding all the products. Note that for the multiplication of two general rectangular matrices to have any meaning, the number of columns of $\mathbb{A}$ (namely, dummy index k in the sum in Eq. (7.1)) must be equal to the number of rows of $\mathbb{B}$ (again, the same dummy index k in the same sum). So, if $\mathbb{A}$ is an $m \times p$ matrix and $\mathbb{B}$ is a $p \times n$ matrix, then the product $\mathbb{C}$ is an $m \times n$ matrix, while the product $\mathbb{BA}$ is undefined. However, as long as matrix multiplication is allowed, it is both distributive and associative, that is,

$$\mathbb{A}(\mathbb{B} + \mathbb{C}) = \mathbb{AB} + \mathbb{AC} \ ,$$

$$(\mathbb{AB})\mathbb{C} = \mathbb{A}(\mathbb{BC}) \ .$$

The use of the word $vector$ for matrices with only one row or one column is justified since multiplying a 1×3 row matrix containing elements $A_x, A_y,$ and A_z with a 3×1 column matrix containing elements $B_x, B_y, and\ B_z$ is equivalent to taking the scalar or dot product of two vectors. Explicitly,

$$(A_x \quad A_y \quad A_z)\begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} = A_x B_x + A_y B_y + A_z B_z = \mathbf{A} \bullet \mathbf{B} \ .$$

Multiplication of a matrix $\mathbb{A}$ by a constant *k* simply multiplies each and every element of $\mathbb{A}$ by k; that is,

$$k\mathbb{A} = \begin{pmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{pmatrix} . \qquad\qquad Eq. (7.2)$$

If the number of rows and the number of columns of a matrix are equal, say to n, then the matrix is a $square$ matrix of order n. For two square matrices $\mathbb{A}$ and $\mathbb{B}$ of the same order, one can form the products $\mathbb{AB}$ as well as $\mathbb{BA}$. However, in general, $\mathbb{AB} \neq \mathbb{BA}$. If $\mathbb{AB} = \mathbb{BA}$, then the matrices $\mathbb{A}$ and $\mathbb{B}$ are said to $commute$.

Example: Consider two square matrices $\mathbb{A}$ and $\mathbb{B}$ of order 2,

$$\mathbb{A} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \qquad \mathbb{B} = \begin{pmatrix} \mathbf{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} .$$

Check whether the products $\mathbb{AB}$ and $\mathbb{BA}$ are equal or not.

Solution: The products $\mathbb{AB}$ and $\mathbb{BA}$ are

$$\mathbb{AB} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\mathbb{BA} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} .$$

Clearly, in this case $\mathbb{AB} \neq \mathbb{BA}$.

Example: Given two square matrices, $\mathbb{A} = \begin{pmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\beta} & \boldsymbol{\alpha} \end{pmatrix}$ **and** $\mathbb{B} = \begin{pmatrix} \boldsymbol{\gamma} & \boldsymbol{\delta} \\ \boldsymbol{\delta} & \boldsymbol{\gamma} \end{pmatrix}$**, determine whether they commute or not.**

Solution: In this case,

$$\mathbb{AB} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} \gamma & \delta \\ \delta & \gamma \end{pmatrix} = \begin{pmatrix} \alpha\gamma + \beta\delta & \alpha\delta + \beta\gamma \\ \beta\gamma + \alpha\delta & \beta\delta + \alpha\gamma \end{pmatrix} ,$$

$$\mathbb{BA} = \begin{pmatrix} \gamma & \delta \\ \delta & \gamma \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} = \begin{pmatrix} \gamma\alpha + \delta\beta & \gamma\beta + \delta\alpha \\ \alpha\delta + \beta\gamma & \beta\delta + \gamma\alpha \end{pmatrix} .$$

Since $\mathbb{AB} = \mathbb{BA}$, then $\mathbb{A}$ and $\mathbb{B}$ commute.

A matrix with all of its elements as zero is called a null or a zero matrix and it is denoted by $\mathbb{O}$. We note in passing that, in general, if $\mathbb{AB} = \mathbb{O}$, it does not imply either $\mathbb{A} = \mathbb{O}$ or $\mathbb{B} = \mathbb{O}$. As an example, consider two matrices,

$$\mathbb{A} = \begin{pmatrix} 2 & 1 \\ -6 & -3 \end{pmatrix} \quad \mathbb{B} = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} .$$

Their product is

$$\mathbb{AB} = \begin{pmatrix} 2 & 1 \\ -6 & -3 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \mathbb{O} .$$

Note that the product $\mathbb{AB}$ is a zero matrix while neither $\mathbb{A}$ nor $\mathbb{B}$ is a zero matrix!

7.2 PROPERTIES OF MATRICES

From this point onwards, we will focus our attention only on square matrices. First, we define some special square matrices. A square matrix whose elements are all zero except the elements along the diagonal is called a $diagonal\ matrix$. For such a matrix $\mathbb{A}$, $a_{ij} = 0$ if $i \neq j$, and it looks like

$$\mathbb{A} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} . \qquad Eq. (7.3)$$

In particular, a square diagonal matrix which has only 1 as its diagonal element is a $unit\ or\ identity\ matrix\ \mathbb{1}$,

$$\mathbb{1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} . \qquad Eq. (7.4)$$

Thus, various elements of the unit matrix can be written in terms of a Kronecker delta as, $a_{ij} = \delta_{ij}$. Multiplying any general matrix $\mathbb{A}$, either on the left or on the right, by $\mathbb{1}$ gives back the matrix $\mathbb{A}$. So,

$$\mathbb{A}\mathbb{1} = \mathbb{1}\mathbb{A} = \mathbb{A} \ .$$

Transpose of a Matrix $(\widetilde{\mathbb{A}})$

A square matrix obtained by interchanging rows and columns of a matrix $\mathbb{A}$ is its transpose, $\widetilde{\mathbb{A}}$. (Another common notation for the transpose of a matrix which you may encounter is $\mathbb{A}^T$.) Thus,

$$\left(\widetilde{\mathbb{A}}\right)_{ij} = (\mathbb{A})_{ji} \text{ or } \tilde{a}_{ij} = a_{ji} \ .$$

Obviously, $(\widetilde{\widetilde{\mathbb{A}}}) = \mathbb{A}$. There is an interesting property of transpose of a product of two general square matrices. The transpose of a product of two matrices is equal to the product of transposes of the two matrices in reverse order, that is, if $\mathbb{C} = \mathbb{A}\mathbb{B}$ then $\widetilde{\mathbb{C}} = \widetilde{\mathbb{B}}\widetilde{\mathbb{A}}$. As a proof, note

$$c_{ij} = \sum_k a_{ik} b_{kj} = \sum_k \tilde{a}_{ki}\, \tilde{b}_{jk} = \sum_k \tilde{b}_{jk} \tilde{a}_{ki} \ = \left(\widetilde{\mathbb{B}}\widetilde{\mathbb{A}}\right)_{ji} \ .$$

Also, $\tilde{c}_{ji} = c_{ij}$ by definition of transpose. Thus $\tilde{c}_{ji} = \left(\widetilde{\mathbb{B}}\widetilde{\mathbb{A}}\right)_{ji}$ or $\widetilde{\mathbb{C}} = \widetilde{\mathbb{B}}\widetilde{\mathbb{A}}$.

Symmetric and Antisymmetric Matrices

A matrix $\mathbb{A}$ for which $\widetilde{\mathbb{A}} = \mathbb{A}$ is called a symmetric matrix since for such a matrix $a_{ij} = a_{ji}$. On the other hand, a matrix $\mathbb{A}$ for which $\widetilde{\mathbb{A}} = -\mathbb{A}$ is called an antisymmetric matrix since for such a matrix $a_{ij} = -a_{ji}$. It implies that the diagonal elements a_{ii} of an antisymmetric matrix are zero. Now, for any general square matrix $\mathbb{A}$, $\mathbb{A} + \widetilde{\mathbb{A}}$ is a symmetric matrix while $\mathbb{A} - \widetilde{\mathbb{A}}$ is an antisymmetric matrix. It is so because

$$\left(\mathbb{A} + \widetilde{\mathbb{A}}\right)_{ij} = a_{ij} + \tilde{a}_{ij} = \tilde{a}_{ji} + a_{ji} = \left(\widetilde{\mathbb{A}} + \mathbb{A}\right)_{ji}$$

$$\left(\mathbb{A} - \widetilde{\mathbb{A}}\right)_{ij} = a_{ij} - \tilde{a}_{ij} = \tilde{a}_{ji} - a_{ji} = -\left(\mathbb{A} - \widetilde{\mathbb{A}}\right)_{ji} \ .$$

Since, for any general matrix $\mathbb{A}$, we can write

$$\mathbb{A} = \tfrac{1}{2}\left(\mathbb{A} + \widetilde{\mathbb{A}}\right) + \tfrac{1}{2}\left(\mathbb{A} - \widetilde{\mathbb{A}}\right) ,$$

it means that any general matrix can be written as a sum of a symmetric and an antisymmetric matrix.

Inverse of a Matrix $\left(\mathbb{A}^{-1}\right)$

If $\mathbb{AB} = \mathbb{BA} = \mathbb{1}$ (unit matrix), then matrix $\mathbb{B}$ is the inverse of matrix $\mathbb{A}$ and matrix $\mathbb{A}$ is the inverse of matrix $\mathbb{B}$. Thus $\mathbb{B} = \mathbb{A}^{-1}$ and $\mathbb{A} = \mathbb{B}^{-1}$. Not every square matrix has an inverse; however, the inverse of a matrix, if it exists, is unique. As a proof of the uniqueness of the inverse of a matrix, let us suppose that there are two different matrices $\mathbb{B}$ and $\mathbb{C}$ which are inverse of the same matrix $\mathbb{A}$. Then $\mathbb{AB} = \mathbb{1}$ and $\mathbb{CA} = \mathbb{1}$. Now consider the triple product of matrices, $\mathbb{CAB}$. Using the associative nature of matrix multiplication, it can be rewritten as,

$$\mathbb{CAB} = (\mathbb{CA})\mathbb{B} = \mathbb{C}(\mathbb{AB}) \qquad \text{or} \qquad (\mathbb{1})\mathbb{B} = \mathbb{C}(\mathbb{1}) \qquad \text{or} \qquad \mathbb{B} = \mathbb{C} \ .$$

Thus, the inverse of $\mathbb{A}$ is unique.

There is an interesting property of the inverse of a product of two general square matrices. The inverse of a product of two matrices is equal to the product of inverses of the two matrices in reverse order; that is, if $\mathbb{C} = \mathbb{AB}$, then $\mathbb{C}^{-1} = \mathbb{B}^{-1}\mathbb{A}^{-1}$. As a proof, notice that

$$\mathbb{C}^{-1}\mathbb{C} = \mathbb{B}^{-1}\mathbb{A}^{-1}\mathbb{AB} = \mathbb{B}^{-1}\mathbb{1}\mathbb{B} = \mathbb{1} \qquad \text{and} \qquad \mathbb{CC}^{-1} = \mathbb{ABB}^{-1}\mathbb{A}^{-1} = \mathbb{A}\mathbb{1}\mathbb{A}^{-1} = \mathbb{1} \ .$$

Since the inverse of a matrix is unique, $\mathbb{C}^{-1}$ is the true inverse of $\mathbb{C}$.

Adjoint of a Matrix $(\mathbb{A}^{a})$

In preparation for describing a technique for obtaining the inverse of a square matrix, we first define the adjoint of a square matrix. Note that for a square matrix $\mathbb{A}$ we may think of its determinant as the corresponding determinant whose elements are identical to the elements of $\mathbb{A}$ and whose value is $\det(\mathbb{A})$. Thus, if $\mathbb{A}$ is an $n \times n$ square matrix, its adjoint matrix is obtained by first replacing each element a_{ij} by the cofactor C_{ij} of the corresponding element in the determinant $\det(\mathbb{A})$, with appropriate sign, and then transposing the matrix. Specifically, adjoint of $\mathbb{A}$ is

$$\mathbb{A}^{a} = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix} ,$$

where C_{ij} is the cofactor of a_{ij}. From its construction, we infer $(\mathbb{A}^{a})_{ij} = C_{ji}$. Using Eq. (6.1), then

$$(\mathbb{A}\mathbb{A}^a)_{ij} = \sum_k (\mathbb{A})_{ik}(\mathbb{A}^a)_{kj} = \sum_k a_{ik}C_{jk} = \delta_{ij}\det(\mathbb{A}) \ .$$

or,

$$\mathbb{A}\mathbb{A}^a = \det(\mathbb{A})\ \mathbb{1} \ ,$$

where $\mathbb{1}$ is the unit matrix.

Similarly, using Eq. (6.1) again,

$$(\mathbb{A}^a\mathbb{A})_{ij} = \sum_k (\mathbb{A}^a)_{ik}(\mathbb{A})_{kj} = \sum_k C_{ki}a_{kj} = \delta_{ij}\det(\mathbb{A}) \ .$$

Thus,

$$\mathbb{A}\mathbb{A}^a = \mathbb{A}^a\mathbb{A} = \det(\mathbb{A})\mathbb{1} \ .$$

This can be rewritten as

$$\left(\frac{\mathbb{A}^a}{\det(\mathbb{A})}\right)\mathbb{A} = \mathbb{A}\left(\frac{\mathbb{A}^a}{\det(\mathbb{A})}\right) = \mathbb{1} \ ,$$

so that

$$\mathbb{A}^{-1} = \frac{\mathbb{A}^a}{det(\mathbb{A})} \ . \qquad Eq.(7.5)$$

Note that if $\det(\mathbb{A}) = 0$ then $\mathbb{A}^{-1}$ is undefined. In other words, the condition for $\mathbb{A}^{-1}$ to exist is that $\det(\mathbb{A}) \neq 0$. A matrix $\mathbb{A}$ for which $\det(\mathbb{A}) \neq 0$ is referred to as a nonsingular matrix while a matrix for which $\det(\mathbb{A}) = 0$ is a singular matrix. Thus, the inverse of a singular matrix does not exist.

Example: Find the inverse of the matrix

$$\mathbb{A} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{2} \\ \mathbf{2} & \mathbf{-1} & \mathbf{3} \\ \mathbf{0} & \mathbf{2} & \mathbf{3} \end{pmatrix}$$

using the adjoint method.

Solution: First, we note that $\det\mathbb{A} = +1(-9) + 2(4) = -1 \neq 0$, so that the matrix $\mathbb{A}$ is nonsingular and its inverse exists.

The cofactors of various matrix elements are

$$C_{11} = -9, C_{12} = -6, C_{13} = +4 \ ,$$

$$C_{21} = +4, C_{22} = +3, C_{23} = -2 \ ,$$

$$C_{31} = +2, C_{32} = +1, C_{33} = -1 \ .$$

The adjoint matrix is

$$\mathbb{A}^a = \begin{pmatrix} -9 & 4 & 2 \\ -6 & 3 & 1 \\ 4 & -2 & -1 \end{pmatrix} .$$

Finally, the inverse of $\mathbb{A}$ is

$$\mathbb{A}^{-1} = \frac{\mathbb{A}^a}{\det(\mathbb{A})} = \begin{pmatrix} 9 & -4 & -2 \\ 6 & -3 & -1 \\ -4 & 2 & 1 \end{pmatrix} \ .$$

Just to verify that our inverse is correct, we multiply the original matrix $\mathbb{A}$ by its inverse matrix $\mathbb{A}^{-1}$ and make sure that we get the identity matrix, $\mathbb{1}$.

$$\mathbb{A}^{-1}\mathbb{A} = \begin{pmatrix} 9 & -4 & -2 \\ 6 & -3 & -1 \\ -4 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 0 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbb{1} \ .$$

Example: Given the matrix

$$\mathbb{A} = \begin{pmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{2} & \mathbf{3} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{2} \end{pmatrix} ,$$

find $\mathbb{A}^{-1}$ using the adjoint method.

Solution: The determinant of this matrix is $\det \mathbb{A} = +1(6) - 2(1) = 4 \neq 0$, hence the matrix is nonsingular. The cofactors of various elements a_{ij} are

$$C_{11} = +6, C_{12} = -4, C_{13} = +2 \ ,$$

$$C_{21} = -1, C_{22} = +2, C_{23} = -1 \ ,$$

$$C_{31} = -9, C_{32} = +6, C_{33} = -1 \ .$$

The adjoint matrix is

$$\mathbb{A}^a = \begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & 6 \\ 2 & -1 & -1 \end{pmatrix} ,$$

and the inverse matrix is

$$\mathbb{A}^{-1} = \frac{\mathbb{A}^a}{\det(\mathbb{A})} = \frac{1}{4}\begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & 6 \\ 2 & -1 & -1 \end{pmatrix} .$$

7.3 SOLVING INHOMOGENEOUS ALGEBRAIC EQUATIONS

The idea of matrix inversion is very useful in solving a set of algebraic equations. Consider the following set of n algebraic equations containing n unknown variables $x_1, x_2, \cdots, x_n$,

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = y_1 \ ,$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = y_2 \ ,$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2x_2} + \cdots + a_{nn}x_n = y_n \ .$$

Here all the coefficients a_{ij} as well as the numbers on the right-hand sides of these equations, namely y_n, are known. If all the numbers y_n are zero, then these algebraic equations are called $homogeneous$ algebraic equations. On the other hand, if all or some of the numbers y_n are nonzero, then these algebraic equations are called $inhomogeneous$ algebraic equations. The equations can be rewritten in a matrix form as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} ,$$

or

$$\mathbb{A}\mathbb{X} = \mathbb{Y} \ , \qquad Eq.(7.6a)$$

where $\mathbb{X}$ and $\mathbb{Y}$ are column matrices (or vectors), $\mathbb{Y}$ is known and $\mathbb{X}$ is unknown. Then, using the idea of matrix inversion,

$$\mathbb{X} = \mathbb{A}^{-1}\mathbb{Y} = \frac{\mathbb{A}^a}{\det(\mathbb{A})}\mathbb{Y} \ . \qquad Eq.(7.6b)$$

Example: Solve the following three inhomogeneous algebraic equations of three variables, $x, y,$ and z using the inverse of a matrix.

$$\boldsymbol{x+2y+3z=4} \ ,$$

$$\boldsymbol{2x+3y=8} \ ,$$

$$\boldsymbol{y+2z=12} \ .$$

Solution: Rewriting these equations in a matrix form, we get

$$\mathbb{A}\mathbb{X} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix} = \mathbb{Y} \ .$$

The inverse of this matrix $\mathbb{A}$ was determined in the previous example. Using it, we get

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbb{A}^{-1} \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & +6 \\ 2 & -1 & -1 \end{pmatrix} \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix}$$

$$= \begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & +6 \\ 2 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -23 \\ 18 \\ -3 \end{pmatrix} .$$

Thus, $x = -23, y = 18$ and $z = -3$.

In case of homogenous algebraic equations, $\mathbb{Y} = 0$. Then, if $\det(\mathbb{A}) \neq 0$, we get the solution $\mathbb{X} = 0$, which is called the trivial solution. A nontrivial solution of homogenous algebraic equations is available only when $\det(\mathbb{A}) = 0$.

7.4 SPECIAL MATRICES

Now, we introduce some well-known and useful matrices. An $orthogonal\ matrix$ $\mathbb{A}$ is the one that satisfies

$$\mathbb{A}\widetilde{\mathbb{A}} = \widetilde{\mathbb{A}}\mathbb{A} = \mathbb{1} \ .$$

Clearly, for the orthogonal matrix,

$$\widetilde{\mathbb{A}} = \mathbb{A}^{-1} \ .$$

As an example of an orthogonal matrix, consider

$$\mathbb{A} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} .$$

For this matrix

$$\widetilde{\mathbb{A}} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

and

$$\mathbb{A}\,\widetilde{\mathbb{A}} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \mathbb{1}$$

as well as

$$\widetilde{\mathbb{A}}\,\mathbb{A} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} = \mathbb{1} .$$

So, we conclude that $\widetilde{\mathbb{A}} = \mathbb{A}^{-1}$ and $\mathbb{A}$ is an orthogonal matrix.

If the elements of a matrix consist of some complex numbers, then we can obtain the $complex\ conjugate\ matrix$, $\mathbb{A}^*$, by simply replacing each element of matrix $\mathbb{A}$ by its complex conjugate. Thus, if $(\mathbb{A})_{ij} = a_{ij}$, then $(\mathbb{A}^*)_{ij} = a_{ij}^*$. As an example, consider

$$\mathbb{A} = \begin{pmatrix} 2 & 1+i \\ 2-i & 3 \end{pmatrix} .$$

Its complex conjugate matrix is

$$\mathbb{A}^* = \begin{pmatrix} 2 & 1-i \\ 2+i & 3 \end{pmatrix} .$$

Now, we will introduce $Hermitian$ and $unitary\ matrices$. First, we define the Hermitian adjoint of a general matrix $\mathbb{A}$. It is denoted by the symbol † (called dagger). The Hermitian adjoint is obtained by first taking the complex conjugate of the matrix, followed by taking the transpose of the matrix. Thus,

$$\mathbb{A}^\dagger = \widetilde{\mathbb{A}^*}$$

is the definition of the Hermitian adjoint of $\mathbb{A}$. Now, a matrix $\mathbb{A}$ is a Hermitian matrix if its Hermitian adjoint is equal to the matrix itself. In other words, if

$$\mathbb{A}^\dagger = \mathbb{A} ,$$

then $\mathbb{A}$ is a Hermitian matrix.

Next, a matrix $\mathbb{A}$ is a unitary matrix if its Hermitian adjoint is equal to the inverse, $\mathbb{A}^{-1}$, of the matrix. In other words, if $\mathbb{A}^{\dagger} = \mathbb{A}^{-1}$, or if

$$\mathbb{A}\mathbb{A}^{\dagger} = \mathbb{A}^{\dagger}\mathbb{A} = \mathbb{1} \ ,$$

then $\mathbb{A}$ is a unitary matrix.

As an example, consider the matrix

$$\mathbb{A} = \frac{1}{\sqrt{a^2+b^2+c^2}}\begin{pmatrix} a & b-ic \\ b+ic & -a \end{pmatrix} \ ,$$

with $a, b,$ and c real. One can even visually see that $\mathbb{A}^{\dagger} = \mathbb{A}$ for this matrix, so that it is a Hermitian matrix. Furthermore, one can easily check that this matrix satisfies

$$\mathbb{A}\mathbb{A}^{\dagger} = \mathbb{A}^{\dagger}\mathbb{A} = \mathbb{1} \ ,$$

so that this matrix is also a unitary matrix.

7.5 CHARACTERISTIC OR SECULAR EQUATION OF A MATRIX

If $\mathbb{A}$ is an $n \times n$ matrix and λ is a scalar parameter, then

$$\mathbb{K} = \mathbb{A} - \lambda\mathbb{1}$$

is called the characteristic matrix of $\mathbb{A}$. The equation

$$\det(\mathbb{K}) = \det(\mathbb{A} - \lambda\mathbb{1}) = 0$$

is called the characteristic or secular equation of $\mathbb{A}$. If $\mathbb{A}$ is an $n \times n$ matrix, then the secular equation is of the form

$$f(\lambda) = 1 + a_1\lambda + a_2\lambda^2 + \cdots + a_n\lambda^n = 0$$

where the coefficients $a_i(i = 1, \ldots n)$ depend on the elements of $\mathbb{A}$. The n roots of this algebraic equation $(\lambda_1, \lambda_2, \ldots\ldots, \lambda_n)$ are called the characteristic roots of the matrix $\mathbb{A}$.

Eigenvalues/Eigenvectors of Matrices

Given a $n \times n$ matrix $\mathbb{A}$, there exist certain column matrices (or vectors) $\mathbb{X}$ such that $\mathbb{A}\mathbb{X}$ is a simple multiple of $\mathbb{X}$. In other words,

$$\mathbb{A}\mathbb{X} = \lambda\mathbb{X} \qquad Eq. (7.7)$$

where λ is a number (real or complex). Then λ is said to be an eigenvalue of $\mathbb{A}$ belonging to the eigenvector $\mathbb{X}$. Here $\mathbb{X}$ is a column matrix with n rows,

$$\mathbb{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} .$$

Thus, Eq. (7.7) gives

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda x_i \ .$$

Or,

$$\sum_{j=1}^{n} (a_{ij} - \lambda \delta_{ij}) x_j = 0 \ .$$

These are n homogenous algebraic equations, for $i = 1, 2, \dots n$, in variables $x_1, x_2, \dots, x_n$. A nontrivial solution exists for these equations only if

$$\det(\mathbb{A} - \lambda \mathbb{1}) = 0 \ .$$

This, of course, is the secular equation of matrix $\mathbb{A}$. The characteristic roots of $\mathbb{A}$ are, then, the eigenvalues of $\mathbb{A}$, and the corresponding vectors $\mathbb{X}$ are the eigenvectors. We briefly note that eigenvalues and eigenvectors are quite useful in quantum physics and this fact will be mentioned again in Chapter 13.

Example: Find the eigenvalues and eigenvectors of the matrix

$$\mathbb{A} = \begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix} .$$

Solution: The characteristic equation is $\det(\mathbb{A} - \lambda \mathbb{1}) = 0$, or

$$\det \begin{pmatrix} -\lambda & 3 & 4 \\ 3 & -\lambda & 0 \\ 4 & 0 & -\lambda \end{pmatrix} = 0 \ .$$

On expanding the determinant, we get

$$-\lambda^3 + 9\lambda + 16\lambda = 0 \qquad \text{or} \qquad \lambda(\lambda+5)(\lambda-5) = 0 \ .$$

Or,

$$\lambda = 0, +5, -5 \ .$$

These are the three eigenvalues of matrix $\mathbb{A}$. Next, we determine the eigenvectors one-by-one. For $\lambda = 0$, $\mathbb{A}\mathbb{X} = \lambda\mathbb{X}$ gives,

$$\begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

Or,

$$3x_2 + 4x_3 = 0, \qquad 3x_1 = 0, \qquad 4x_1 = 0 \ .$$

Thus, within an arbitrary constant, the eigenvector is

$$\begin{pmatrix} 0 \\ 4 \\ -3 \end{pmatrix} .$$

A *normalized* eigenvector is analogous to a unit vector since they both have magnitudes of 1. To normalize any eigenvector, we simply divide it by the magnitude of the eigenvector. So, for $\lambda = 0$, the magnitude of the eigenvector is $\sqrt{0 + 4^2 + (-3)^2} = 5$, and, therefore, the normalized eigenvector is

$$\frac{1}{5}\begin{pmatrix} 0 \\ 4 \\ -3 \end{pmatrix} .$$

For $= +5$, $\qquad \mathbb{A}\mathbb{X} = \lambda\mathbb{X}$ gives

$$\begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = +5\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} .$$

Or,

$$-5x_1 + 3x_2 + 4x_3 = 0$$

$$+3x_1 - 5x_2 = 0 \qquad \text{or} \qquad x_2 = \frac{3}{5}\,x_1$$

$$+4x_1 - 5x_3 = 0 \qquad \text{or} \qquad x_3 = \frac{4}{5}x_1 \ .$$

Again, within an arbitrary constant, the eigenvector is

$$\begin{pmatrix} 5 \\ 3 \\ 4 \end{pmatrix} ,$$

and the normalized eigenvector is $\frac{1}{5\sqrt{2}}\begin{pmatrix} 5 \\ 3 \\ 4 \end{pmatrix}$.

Finally, for $\lambda = -5$, $\mathbb{AX} = \lambda\mathbb{X}$ gives

$$\begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = -5\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} .$$

Or

$$+5x_1 + 3x_2 + 4x_3 = 0$$

$$+3x_1 + 5x_2 = 0 \qquad \text{or} \qquad x_2 = -\frac{3}{5}\,x_1$$

$$+4x_1 + 5x_3 = 0 \qquad \text{or} \qquad x_3 = -\frac{4}{5}x_1 \ .$$

So, the eigenvector is, within an arbitrary constant,

$$\begin{pmatrix} 5 \\ -3 \\ -4 \end{pmatrix} ,$$

and the normalized eigenvector is $\frac{1}{5\sqrt{2}}\begin{pmatrix} 5 \\ -3 \\ -4 \end{pmatrix}$.

PROBLEMS FOR CHAPTER 7

1. Given the following two matrices $\mathbb{A}$ and $\mathbb{B}$, determine $\mathbb{AB}$, $\mathbb{BA}$, $\mathbb{A}^2$, and $\mathbb{B}^2$ if they exist. If a product does not exist, make a note of it.

$$\mathbb{A} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \\ 3 & 2 \end{pmatrix} \qquad \text{and} \qquad \mathbb{B} = \begin{pmatrix} 3 & -1 \\ 4 & -3 \end{pmatrix} .$$

2. Given the three matrices,

$$\mathbb{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbb{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbb{C} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} ,$$

find all possible products of pairs of matrices (either $\mathbb{A}$ and $\mathbb{B}$, or $\mathbb{B}$ and $\mathbb{C}$, or $\mathbb{C}$ and $\mathbb{A}$) to determine which pair, or pairs, of matrices commute.

3. For the following matrix $\mathbb{A}$, determine $\mathbb{A}^{-1}$,

$$\mathbb{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 0 \\ 2 & 0 & 3 \end{pmatrix} .$$

Explicitly verify that $\mathbb{A}\mathbb{A}^{-1} = \mathbb{A}^{-1}\mathbb{A} = \mathbb{1}$.

4. Determine the eigenvalues and the normalized eigenfunctions of matrix $\mathbb{A}$ of problem 3.

5. Solve the following inhomogeneous algebraic equations by finding the inverse of the matrix of coefficients,

$$x - y + z = 4$$

$$2x + y - z = -1$$

$$3x + 2y + 2z = 5 \ .$$

6. In the theory of spin $\frac{1}{2}$ in quantum physics, we encounter three 2×2 matrices (*Pauli spin matrices*),

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \ \ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Prove the following properties of these matrices:

(a) $\sigma_l \sigma_m = i \sum_{n=1}^{3} \epsilon_{lmn}\, \sigma_n \quad$ for $l \neq m$.

(b) $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, where $\mathbb{1}$ is the unit, or identity, matrix of size 2×2.

7. Determine the eigenvalues and the normalized eigenvectors of the three Pauli spin matrices of problem 6, separately.

Chapter 8: Vector Analysis

In biomedical physics we encounter several physical quantities. Some of these quantities are characterized by their magnitude only, such as temperature of a sick patient, density of a blood sample, pressure in the lung, etc. Such quantities are called scalars and they are treated as ordinary algebraic numbers. On the other hand, there are physical quantities which are characterized by both magnitude and direction, such as forces on human spine during heavy lifting, velocity of blood flowing in veins, electric and magnetic fields in medical equipment, etc. Such quantities are called vectors. In this chapter we will learn about mathematical properties of vectors.

8.1 VECTORS AND THEIR ADDITION/SUBTRACTION

A vector quantity is represented by a directed line whose length is proportional to the magnitude of the physical quantity and whose direction is the direction of the physical quantity. These directed lines, representing vectors, look like an arrow with one end as the head of the arrow and the other end as the tail of the arrow, as shown in Figure 8.1. Vectors are shown as bold-faced letters like **A** and their magnitudes are shown as ordinary letters

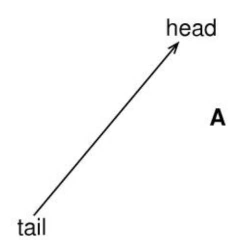


Figure 8.1. A vector, A, with its head and its tail.

like A. We can move a vector, or the directed line representing the vector, around in space and the vector will remain unchanged as long as both the length of the directed line as well as its direction are kept unchanged.

An alternate way of representing a vector is through its components. If the Cartesian coordinates of the head and of the tail of a vector **A** are (x_h, y_h, z_h) and (x_t, y_t, z_t), respectively, then, the components of **A** are

$$A_x = x_h - x_t \ ,$$

$$A_y = y_h - y_t \ ,$$

$$A_z = z_h - z_t \ .$$

The length of the vector is

$$\mathbf{A} = \left[A_x^2 + A_y^2 + A_z^2\right]^{1/2} \quad . \qquad Eq.(8.1a)$$

It is easier to see that if the origin of the coordinate system is chosen at the tail of the vector **A**, then $A_x, A_y,$ and A_z are merely the projections of **A** onto the three coordinate axes.

Addition and Subtraction of Vectors

Vectors do not add or subtract or multiply like ordinary numbers. First, we will learn the rules for adding and subtracting vectors, which lead to new vectors. Next, we will learn the rules for multiplying vectors, which can lead to either a scalar (scalar product) or a vector (vector product). The vectors can be added using a geometrical method, called head-to-tail method, in which we connect the head of the first vector, **A**, with the tail of the second vector, **B**. A new vector **X** whose head is the free unconnected head of **B** and whose tail is the free unconnected tail of **A** is the sum of vectors **A** and **B**, as shown in the Figure 8.2

$$\mathbf{X} = \mathbf{A} + \mathbf{B} \quad .$$

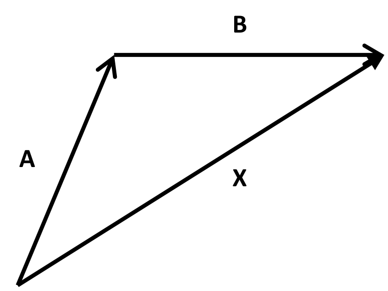


Figure 8.2. Vector **X** is the sum of vectors **A** and **B**.

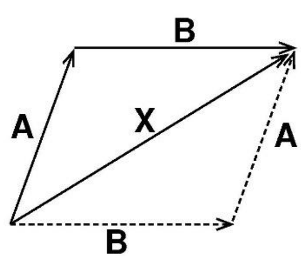


Figure 8.3. Commutative nature of vector addition.

By completing the parallelogram, as in Figure 8.3, it is seen that

$$\mathbf{X} = \mathbf{B} + \mathbf{A}$$

as well. Thus, vector addition is commutative. Addition of more than two vectors proceeds in the same manner. If **Z** = **A** + **B** + **C**, then from the Figure 8.4

$$\mathbf{X} = \mathbf{A} + \mathbf{B} \ ,$$

$$\mathbf{Y} = \mathbf{B} + \mathbf{C} \ ,$$

and

$$\mathbf{Z} = \mathbf{X} + \mathbf{C} = \mathbf{A} + \mathbf{Y} \ ,$$

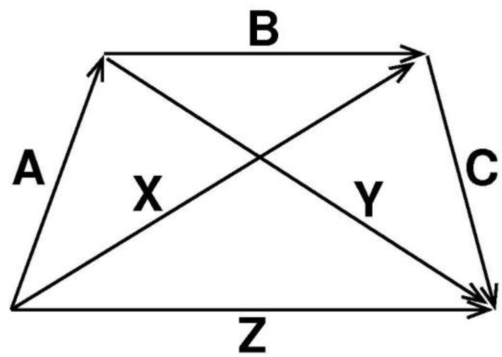


Figure 8.4. Sum of more than two vectors.

that is,

$$\mathbf{Z} = (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \ .$$

Thus, vector addition is associative as well. Now, let us talk about subtraction of one vector from another. Figure 8.5 shows two vectors, **A** and **B**, of equal length but with opposite directions. According to head-to-tail method of adding vectors, $\mathbf{A} + \mathbf{B} = \mathbf{0}$ or $\mathbf{B} = -\mathbf{A}$.

In other words, −**A** (or **B**) is a vector which has same magnitude as **A** but its direction is opposite to that of **A**. Subtraction of one vector from another can now be handled exactly in the same manner as vector addition except the sign (direction) of one of the vectors is reversed. For example,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) \ .$$

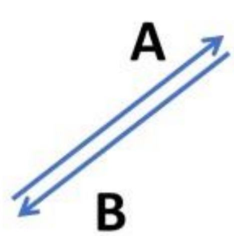


Figure 8.5. Two vectors **A** and **B** of equal length, but opposite directions.

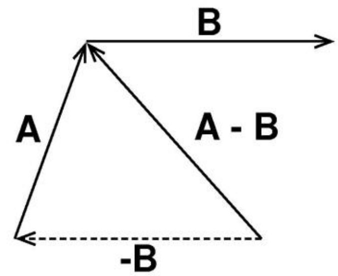


Figure 8.6. Subtraction of two vectors.

Also, graphically the "other" diagonal of the parallelogram in Figure 8.3 gives **A** $-$ **B** as shown in Figure 8.6.

An alternative way of adding vectors is by using the components of the vector. For this purpose, we choose a right-handed Cartesian coordinate system. The meaning of a right-handed coordinate system will be clearer after Eq. (8.6). For this coordinate system, we introduce unit vectors $\mathbf{e_x}$, $\mathbf{e_y}$, and $\mathbf{e_z}$. These three vectors are each of unit length and are directed along the x, y, and z axes, respectively. We note in passing that several other authors use the notation $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ or $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ for these Cartesian unit vectors. In our notation, then $A_x\mathbf{e_x}$ is a vector of length A_x along the x-axis, $A_y\mathbf{e_y}$ is a vector of length A_y along the y-axis, and $A_z\mathbf{e_z}$ is a vector of length A_z along the z-axis. Using vector addition by head-to-tail method, as seen in Figure 8.7,

$$\mathbf{A} = A_x\mathbf{e_x} + A_y\mathbf{e_y} + A_z\mathbf{e_z} \ . \qquad Eq. (8.1b)$$

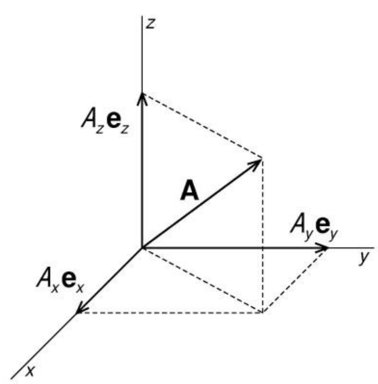


Figure 8.7. Components of a vector **A**.

Now, two vectors **A** and **B** can be added simply by adding their components, that is,

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \left(\mathbf{e_x}A_x + \mathbf{e_y}A_y + \mathbf{e_z}A_z\right) + \left(\mathbf{e_x}B_x + \mathbf{e_y}B_y + \mathbf{e_z}B_z\right) \\ &= \mathbf{e_x}(A_x + B_x) + \mathbf{e_y}(A_y + B_y) + \mathbf{e_z}(A_z + B_z) \ . \qquad Eq. (8.2a) \end{aligned}$$

Subtraction of one vector from another can now be handled exactly in the same manner as vector addition except sign (direction) of one of the vectors is reversed. For example,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) = \mathbf{e}_x(A_x - B_x) + \mathbf{e}_y(A_y - B_y) + \mathbf{e}_z(A_z - B_z) \ . \qquad Eq. (8.2b)$$

It follows from the discussion of addition and subtraction of vectors that if two vectors are equal to each other, then individually the x, y, and z components of those two vectors are also equal.

8.2 MULTIPLICATION OF VECTORS: SCALAR PRODUCT

Now, we talk about multiplying two vectors. There are two ways to multiply vectors. One way leads to a scalar and the other to a vector. The scalar (or inner or dot) product of two vectors **A** and **B** is a scalar number of magnitude AB $\cos\theta$, where A and B are the magnitudes of the two vectors and θ is the angle between the two vectors. This angle is obtained by joining the tails of both vectors as in Figure 8.8. The scalar product of two vectors is indicated by placing a large dot between the two vectors, **A•B** = AB $\cos\theta$ = **B•A**. Note that the scalar product of two vectors is commutative. The scalar product is also distributive, that is,

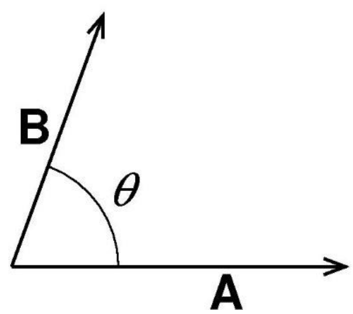


Figure 8.8. Scalar product of two vectors **A** and **B**.

A•(B + C) = A•B + A•C. Also, since unit vectors $\mathbf{e_x}, \mathbf{e_y}, \mathbf{e_z}$ are mutually orthogonal, it follows that.

$$\mathbf{e_x} \bullet \mathbf{e_x} = 1, \qquad \mathbf{e_y} \bullet \mathbf{e_y} = 1, \ \ \mathbf{e_z} \bullet \mathbf{e_z} = 1$$

and

$$\mathbf{e_x} \bullet \mathbf{e_y} = \mathbf{e_y} \bullet \mathbf{e_x} = 0, \mathbf{e_x} \bullet \mathbf{e_z} = \mathbf{e_z} \bullet \mathbf{e_x} = 0, \mathbf{e_y} \bullet \mathbf{e_z} = \mathbf{e_z} \bullet \mathbf{e_y} = 0 \ \ .$$

In particular, note

$$\mathbf{A}\bullet\mathbf{B} = (\mathbf{e_x}A_x + \mathbf{e_y}A_y + \mathbf{e_z}A_z) \bullet (\mathbf{e_x}B_x + \mathbf{e_y}B_y + \mathbf{e_z}B_z)$$

$$= +A_x B_x(\mathbf{e_x} \bullet \mathbf{e_x}) + A_y B_x(0) + A_z B_x(0)$$

$$+A_x B_y(0) + A_y B_y(\mathbf{e_y} \bullet \mathbf{e_y}) + A_z B_y(0)$$

$$+A_x B_z(0) + A_y B_z(0) + A_z B_z(\mathbf{e_z} \bullet \mathbf{e_z})$$

or

$$\mathbf{A} \bullet \mathbf{B} = A_x B_x + A_y B_y + A_z B_z \ . \qquad Eq.(8.3)$$

Also,

$$\mathbf{A \bullet A} = A_x^2 + A_y^2 + A_z^2$$

so that $\sqrt{\mathbf{A \bullet A}}$ gives the magnitude of vector **A**.

Now we introduce an alternate, but useful, new notation. We will use $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ for the three orthogonal unit vectors $\mathbf{e_x}, \mathbf{e_y}$, and $\mathbf{e_z}$, and A_1, A_2, and A_3 for the three components A_x, A_y, and A_z of a vector **A**. In this notation it follows that

$$\mathbf{e}_i \bullet \mathbf{e}_j = \delta_{ij} \ . \qquad Eq.(8.4)$$

Using this identity, we can write the scalar product of two vectors in terms of a Kronecker delta as

$$\mathbf{A \bullet B} = \left(\sum_i \boldsymbol{e_i}\, A_i\right) \bullet \left(\sum_j \boldsymbol{e_j} B_j\right) = \sum_{i,j} \delta_{ij}\, A_i B_j = \sum_i A_i\, B_i \ .$$

Finally, the law of cosines in trigonometry, which relates the lengths of three sides of a triangle, can be easily obtained using scalar product of two vectors. Three vectors **A, B,** and **C,** form the sides of a triangle as shown in the Figure 8.9. The angle θ between the two vectors **A** and **B** is obtained by joining the tails of both vectors. Using head-to-tail vector addition,

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \ ,$$

or

$$C^2 = \mathbf{C \bullet C} = (\mathbf{A} + \mathbf{B}) \bullet (\mathbf{A} + \mathbf{B}) = \mathbf{A \bullet A} + \mathbf{B \bullet B} + 2\, \mathbf{A \bullet B} = A^2 + B^2 + 2\, AB\cos\theta$$

or

$$C^2 = A^2 + B^2 - 2\, AB\cos\phi \qquad Eq.(8.5)$$

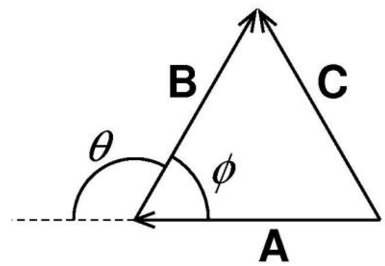


Figure 8.9. Law of cosines.

where ϕ is the angle inside the triangle opposite to side **C**, that is, the angle supplementary to θ. This is the law of cosines in trigonometry, and it is consistent with the Pythagorean theorem for a right-angled triangle (when $\phi = \pi/2$).

8.3 MULTIPLICATION OF VECTORS: VECTOR PRODUCT

Now, we talk about the second way of multiplying two vectors which leads to a vector. The vector (or cross) product of two vectors is indicated by placing a large cross between the two vectors **A** and **B**. It is defined to be a new vector **C**,

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} \ ,$$

such that the magnitude of **C** is $C = ABsin\theta$ and direction of **C** is perpendicular to the plane containing **A** and **B** in such a sense that **A**, **B**, and **C** form a right-handed system. In order to understand the direction of **C** better, let us imagine that the plane containing **A** and **B** is a big wall. We place a screwdriver, with a screw, against the wall such that the screwdriver can be rotated either clockwise or counterclockwise and the screw itself will move either into the wall or out of the wall. In the Figure 8.10 (a), the screwdriver is rotated counterclockwise, as shown by the arrow on angle θ, from **A** to **B,** and the direction of **A** $\times$ **B,** which is along the direction of motion of the screw itself, is coming out of the wall, indicated by symbol ⊙. In Figure 8.10 (b), the screwdriver is rotated clockwise along the arrow shown on angle θ, from **B** to **A,** and the direction of **B** $\times$ **A**, or the direction of motion of the screw itself, is going into the wall, indicated by symbol ⊗. This is referred to as *the right-hand rule*. With this definition, the vector **B** $\times$ **A** has the same magnitude as vector **A** $\times$ **B**; however, direction of **B** $\times$ **A** is opposite to that of **A** $\times$ **B**. Thus,

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \ ,$$

that is, the cross product is anticommuting. In the case of the three unit vectors $\mathbf{e_1}$, $\mathbf{e_2}$, and $\mathbf{e_3}$,

$$\mathbf{e_1} \times \mathbf{e_1} = 0 \;, \qquad \mathbf{e_2} \times \mathbf{e_2} = 0 \;, \qquad \mathbf{e_3} \times \mathbf{e_3} = 0 \;,$$

$$\mathbf{e_1} \times \mathbf{e_2} = \mathbf{e_3} = -\mathbf{e_2} \times \mathbf{e_1}, \mathbf{e_2} \times \mathbf{e_3} = \mathbf{e_1} = -\mathbf{e_3} \times \mathbf{e_2} \;, \qquad \mathbf{e_3} \times \mathbf{e_1} = \mathbf{e_2} = -\mathbf{e_1} \times \mathbf{e_3} \;.$$

In a more compact form, we can write all of these nine relationships as

$$\mathbf{e}_{\boldsymbol{i}} \times \mathbf{e}_{\boldsymbol{j}} = \sum_{k=1}^{3} \epsilon_{ijk}\, \mathbf{e}_{\boldsymbol{k}} \,. \qquad Eq.(8.6)$$

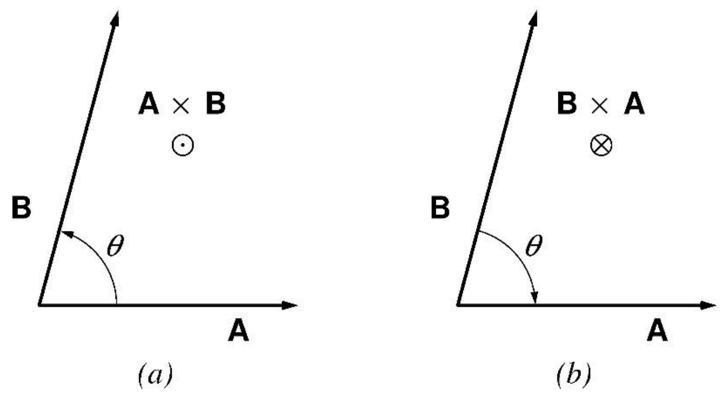


Figure 8.10. Vector product of two vectors **A** and **B**.

The three unit vectors, satisfying Eq. (8.6), describe a right-handed Cartesian coordinate system. Using these relations, the components of $\mathbf{C} = \mathbf{A} \times \mathbf{B}$ can be written as

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} = (\mathbf{e_1}A_1 + \mathbf{e_2}A_2 + \mathbf{e_3}A_3) \times (\mathbf{e_1}B_1 + \mathbf{e_2}B_2 + \mathbf{e_3}B_3)$$

$$= \mathbf{e_3}A_1B_2 - \mathbf{e_2}A_1B_3 - \mathbf{e_3}A_2B_1 + \mathbf{e_1}A_2B_3 + \mathbf{e_2}A_3B_1 - \mathbf{e_1}A_3B_2$$

$$= \mathbf{e_1}(A_2B_3 - A_3B_2) + \mathbf{e_2}(A_3B_1 - A_1B_3) + \mathbf{e_3}(A_1B_2 - A_2B_1) \;.$$

Thus,

$$C_1 = A_2B_3 - A_3B_2$$

$$C_2 = A_3B_1 - A_1B_3$$

$$C_3 = A_1B_2 - A_2B_1$$

or

$$C_i = A_jB_k - A_kB_j$$

with $i, j,$ and k all different and in cyclic permutation. In terms of Levi-Civita symbol, we can write the components of **C** as

$$C_i = \sum_{j,k} \epsilon_{ijk} A_j B_k \quad ,$$

and the vector **C** as

$$\mathbf{C} = \sum_i \mathbf{e_i} C_i \quad ,$$

or

$$\mathbf{A} \times \mathbf{B} = \sum_{i,j,k} \epsilon_{ijk} \mathbf{e}_i A_j B_k \quad . \qquad Eq. (8.7)$$

We note from this suggestive form that the vector product can be written in the form of a determinant as

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \quad . \qquad Eq. (8.8)$$

The cross product of two vectors has a very nice geometrical interpretation. Consider a parallelogram, in a plane, whose sides are made up by vectors **A** and **B**, as shown in Figure 8.11. Note that the area of this parallelogram is $AB \sin\theta$, which is the magnitude of **A** $\times$ **B**. However, the direction of **A** $\times$ **B** is perpendicular to the plane of the parallelogram. Thus, in general, the area of an element such as the parallelogram defined above can be treated as a vector with its direction perpendicular to the plane containing that element. Note that the direction of area vector, either into or out of the plane, is related to the sense of traversal of the periphery of the area by the right-hand rule. In Figure 8.11, the direction of the area element **A** $\times$ **B**, of magnitude AB$\sin\theta$, is obtained by traversing the periphery by going first over (solid line) vector **A** and then over (solid line) vector **B**, that is, traversing the periphery indicated by the thick counterclockwise arrow corresponding to direction coming out of the plane of parallelogram according to the right-hand rule. Similarly, the direction of the area element **B** $\times$ **A,** also of same magnitude AB$\sin\theta$, is obtained by traversing the periphery of the parallelogram by going first over (dashed line) vector **B** and then over (dashed line) vector **A**, that is, traversing the periphery indicated by the thick clockwise arrow corresponding to direction going into the plane of the parallelogram according to the right-hand rule.

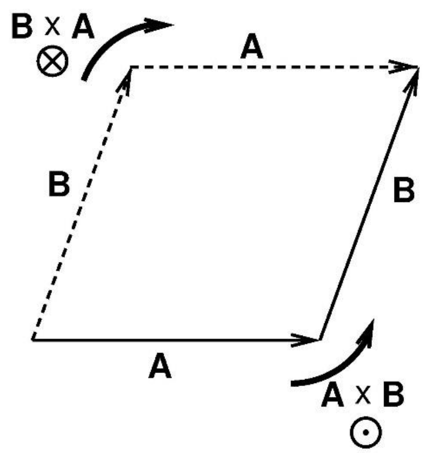


Figure 8.11. An area element as a vector.

8.4 TRIPLE SCALAR PRODUCT

Now let us explore how we can multiply three vectors **A**, **B**, and **C** to obtain either a scalar or a vector. The product **A**•(**B** $\times$ **C**) is a scalar and is referred to as the triple scalar product. In terms of components,

$$\mathbf{A}\bullet(\mathbf{B}\times\mathbf{C}) = A_1(\mathbf{B}\times\mathbf{C})_1 + A_2(\mathbf{B}\times\mathbf{C})_2 + A_3(\mathbf{B}\times\mathbf{C})_3$$

$$= A_1(B_2C_3 - B_3C_2) + A_2(B_3C_1 - B_1C_3) + A_3(B_1C_2 - B_2C_1) \ .$$

The expression on the right-hand side can be rewritten in two alternate forms:

$$\text{first form} = B_1(C_2A_3 - C_3A_2) + B_2(C_3A_1 - C_1A_3) + B_3(C_1A_2 - C_2A_1) = \mathbf{B}\bullet(\mathbf{C}\times\mathbf{A}) \ ,$$

or

$$\text{second form} = C_1(A_2B_3 - A_3B_2) + C_2(A_3B_1 - A_1B_3) + C_3(A_1B_2 - A_2B_1) = \mathbf{C}\bullet(\mathbf{A}\times\mathbf{B}) \ .$$

Thus, the triple scalar product can be written as,

$$\mathbf{A}\bullet(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\bullet(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\bullet(\mathbf{A}\times\mathbf{B}) = -\mathbf{A}\bullet(\mathbf{C}\times\mathbf{B}) = -\mathbf{B}\bullet(\mathbf{A}\times\mathbf{C}) = -\mathbf{C}\bullet(\mathbf{B}\times\mathbf{A}) \ . \qquad Eq.(8.9a)$$

In other words, dot and cross can be interchanged in a triple scalar product. The triple scalar product can also be written in the form of a determinant:

$$\mathbf{A}\bullet(\mathbf{B}\times\mathbf{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} \ , \qquad Eq.(8.9b)$$

or, in terms of Levi-Civita symbol,

$$\mathbf{A} \bullet (\mathbf{B} \times \mathbf{C}) = \sum_{i,j,k} \epsilon_{ijk} A_i B_j C_k \quad . \qquad Eq. (8.9c)$$

The triple scalar product has a nice geometrical interpretation. Consider the parallelepiped made up by vectors **A**, **B**, and **C** as shown in Figure 8.12. The area of the base of the parallelepiped is the parallelogram **A** $\times$ **B**. The direction of this area element is along $\boldsymbol{n}$, the normal to the base. Then,

$$(\mathbf{A} \times \mathbf{B}) \bullet \mathbf{C} = |\mathbf{A} \times \mathbf{B}|(\boldsymbol{n} \bullet \mathbf{C})$$

= (area of the base times height) = volume of the parallelepiped.

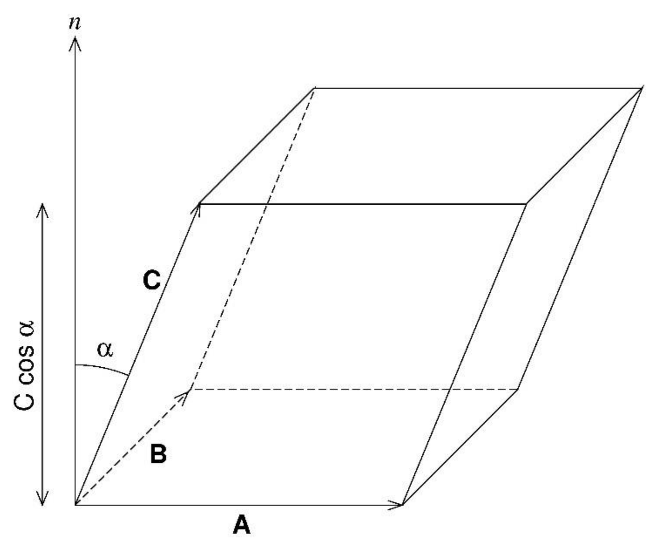


Figure 8.12. A parallelepiped made up by vectors **A**, **B**, and **C**.

8.5 TRIPLE VECTOR PRODUCT

Now we explore the possibility of obtaining a vector by multiplying three vectors **A**, **B**, and **C**. In the triple vector product **A** $\times$ (**B** $\times$ **C**), the order of multiplying vectors is important. For example,

$$\mathbf{e_1} \times (\mathbf{e_2} \times \mathbf{e_2}) \neq (\mathbf{e_1} \times \mathbf{e_2}) \times \mathbf{e_2} \ ,$$

since the left-hand side is zero because the vector product of a vector by itself is zero, but the right-hand side is nonzero, equal to $\mathbf{e_3} \times \mathbf{e_2} = -\mathbf{e_1}$. In general, the triple vector product can be written as

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{A} \times \mathbf{V} = \sum_{i,j,k} \epsilon_{ijk} \mathbf{e_i} A_j V_k \qquad Eq. (8.10a)$$

where

$$\mathbf{V} = \mathbf{B} \times \mathbf{C} \quad \text{and} \quad V_k = \sum_{l,m} \epsilon_{klm} B_l C_m \ . \qquad Eq.(8.10b)$$

On substituting for V_k from Eq. (8.10b) into Eq. (8.10a), we get

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \sum_{ijk} \epsilon_{ijk} \mathbf{e_i} A_j \sum_{lm} \epsilon_{klm} B_l C_m \ .$$

To carry out the sum over the repeated index k, we use the epsilon-delta identity (see Appendix C for proof),

$$\sum_k \epsilon_{ijk} \epsilon_{klm} = \sum_k \epsilon_{kij} \epsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \ ,$$

so that the triple vector product looks like:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \sum_{ij} \sum_{lm} \mathbf{e_i} A_j B_l C_m \left[\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right] \ .$$

Because of the Kronecker deltas, the sum over the repeated dummy indices l and m can be carried out easily,

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \sum_{ij} \mathbf{e}_i A_j \left[B_i C_j - B_j C_i\right] = \sum_i \mathbf{e}_i B_i \sum_j A_j C_j - \sum_i \mathbf{e}_i C_i \sum_j A_j B_j$$

$$= \mathbf{B}(\mathbf{A} \bullet \mathbf{C}) - \mathbf{C}(\mathbf{A} \bullet \mathbf{B}) \ . \qquad Eq.(8.11)$$

The order of vectors appearing on the right-hand side is remembered by using the mnemonic BAC-CAB. Another form of triple vector product is

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = -\mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = -\mathbf{A}(\mathbf{B} \bullet \mathbf{C}) + \mathbf{B}(\mathbf{C} \bullet \mathbf{A}) \ , \qquad Eq.(8.12)$$

after using the BAC-CAB rule in the last step. Now, the order of vectors in the two forms of the triple vector product, namely, Eqs. (8.11) and (8.12), can be determined by:

Vector product of three vectors = middle vector times scalar product of the remaining two vectors minus $(-)$ other vector in the parenthesis times scalar product of the remaining two vectors.

PROBLEMS FOR CHAPTER 8

1. Using the concept of scalar, or dot, product of two vectors, determine the angle between vectors $\mathbf{A} = \mathbf{e_x} + 2\mathbf{e_y} + 2\mathbf{e_z}$ and $\mathbf{B} = 2\mathbf{e_x} - 3\mathbf{e_y} + 6\mathbf{e_z}$.

2. Prove that the vectors $\mathbf{A} = 3\mathbf{e_x} + \mathbf{e_y} - 2\mathbf{e_z}$, $\mathbf{B} = -\mathbf{e_x} + 3\,\mathbf{e_y} + 4\,\mathbf{e_z}$, and $\mathbf{C} = 4\,\mathbf{e_x} - 2\,\mathbf{e_y} - 6\,\mathbf{e_z}$ can form the sides of a triangle. Find the lengths of the medians of this triangle.

3. Using the concept of vector, or cross, product of two vectors, determine the area of a triangle with vertices at points A(1,2,3), B(2,4,5), and C(1,-8,0). In our notation here, P(x, y, z) refers to the Cartesian coordinates of the point P.

4. Determine a unit vector perpendicular to the plane containing vectors $\mathbf{A} = 2\,\mathbf{e_x} - 6\mathbf{e_y} - 3\,\mathbf{e_z}$ and $\mathbf{B} = 4\,\mathbf{e_x} + 3\,\mathbf{e_y} - \mathbf{e_z}$.

5. The eight corners of a parallelepiped are labeled as C_1, C_2, C_3, C_4, C_5, C_6, C_7, and C_8. The Cartesian (x, y, z) coordinates of the corners are C_1 (5, 2, 0), C_2 (2, 5, 0), C_3 (4, 5, 0), C_4 (7, 2, 0), C_5 (6, 3, 3), C_6 (3, 6, 3), C_7 (5, 6, 3), and C_8 (8, 3, 3).

(a) Determine the three vectors **A**, **B**, and **C** that define this parallelepiped.

(b) Determine the volume of the parallelepiped.

[Hint: To visualize this parallelepiped, plot and label the first four and the last four corner points in two $x - y$ planes separately. The two $x - y$ planes are $z = 0$ plane and $z = 3$ plane.]

6. Prove the following identity for vector triple product:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) + \mathbf{B} \times (\mathbf{C} \times \mathbf{A}) + \mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = 0 \ .$$

7. Prove the following two identities:

$$(\mathbf{A} \times \mathbf{B}) \bullet (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \bullet \mathbf{C})\,(\mathbf{B} \bullet \mathbf{D}) - (\mathbf{A} \bullet \mathbf{D})\,(\mathbf{B} \bullet \mathbf{C}) \ ,$$

$$(\mathbf{A} \times \mathbf{B}) \times (\mathbf{C} \times \mathbf{D}) = [(\mathbf{A} \times \mathbf{B}) \bullet \mathbf{D}]\,\mathbf{C} - [(\mathbf{A} \times \mathbf{B}) \bullet \mathbf{C}]\,\mathbf{D} = [(\mathbf{C} \times \mathbf{D}) \bullet \mathbf{A}]\,\mathbf{B} - [(\mathbf{C} \times \mathbf{D}) \bullet \mathbf{B}]\,\mathbf{A} \ .$$

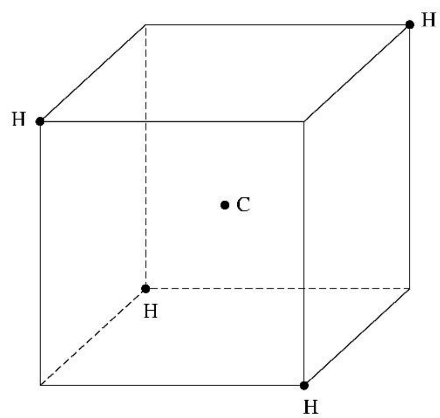
8. If **A** and **B** are any two vectors, then show that

$$(\mathbf{A} \times \mathbf{B}) \bullet (\mathbf{A} \times \mathbf{B}) = A^2 B^2 - (\mathbf{A} \bullet \mathbf{B})^2 \ .$$

9. A rhombus is a parallelogram with all four sides of equal length. Using the concept of scalar (or dot) product of vectors, show that the two diagonals of a rhombus are always perpendicular to each other.

10. **Biomedical Physics Application**. The "life", as we know it, is based on the chemistry of carbon-containing molecules. An understanding of the structure of these molecules is essential to understanding life. One of the simplest hydrocarbon molecules is methane (CH_4), in which four hydrogen atoms are placed symmetrically at the vertices of a tetrahedron around a single carbon atom. Alternatively, the four hydrogen atoms are located at the four corners of a cube, as shown in the figure below, with the carbon atom at the center of the cube. The bond angle is the angle between the lines that join the carbon atom to two of the hydrogen atoms. Using your knowledge of the scalar product of vectors, show that the bond angle in methane molecule is 109.5°.

[Hint: Take the origin of the coordinate system to be at the center of the cube (at carbon atom) of side length *2a* and determine the coordinates of each hydrogen atom.]



11. Consider an arbitrary triangle whose three sides are made up of three vectors **A**, **B** and **C**, then according to vector addition, **A** + **B** + **C** = 0. The three medians of this triangle, namely $\mathbf{M_A}$, $\mathbf{M_B}$, and $\mathbf{M_C}$, are also vectors. The median $\mathbf{M_A}$ has its tail at the vertex opposite **A** and its head at the midpoint of **A**. There is similar construction for other two medians $\mathbf{M_B}$ and $\mathbf{M_C}$. Using vector addition, show that

$$\mathbf{M_A} = -\frac{1}{2}(\mathbf{B} - \mathbf{C}) \ ,$$

$$\mathbf{M_B} = -\frac{1}{2}(\mathbf{C} - \mathbf{A}) \ ,$$

$$\mathbf{M_C} = -\frac{1}{2}(\mathbf{A} - \mathbf{B}) \ .$$

12. The three vertices of a triangle in the $x - y$ plane have Cartesian coordinates $(x_1, y_1), (x_2, y_2)$, and (x_3, y_3). Using the facts that three medians of a triangle intersect each other at a common point, called a centroid, and at this point each median is trisected, show, using vector addition, that the Cartesian coordinates of the centroid are

$$(x_c, y_c) = \left(\frac{x_1 + x_2 + x_3}{3}, \frac{y_1 + y_2 + y_3}{3}\right) .$$

13. The three vertices of a triangle in the $x - y$ plane have Cartesian coordinates $(x_1, y_1), (x_2, y_2), \text{and } (x_3, y_3)$. Using the cross product of two vectors, show that the area of this triangle is given by the following determinant:

$$\text{area } = \frac{1}{2}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} .$$

Chapter 9: Curvilinear Coordinates and Multiple Integrals

In this chapter, we will introduce two important curvilinear coordinate systems. These are the spherical coordinate system and the cylindrical coordinate system. They are very useful in carrying out multiple integrals encountered in problems with spherical and cylindrical symmetries.

9.1 CARTESIAN COORDINATES

First, we review the details of the well-known and more common rectangular Cartesian coordinate system. In this system, a point has three coordinates (x, y, z) which are the distances perpendicular to three mutually orthogonal surfaces. All three coordinates (x, y, z), vary independently from $-\infty$ to $+\infty$. Then we define some families of surfaces and families of curves for Cartesian coordinate system. A surface on which the value of the x

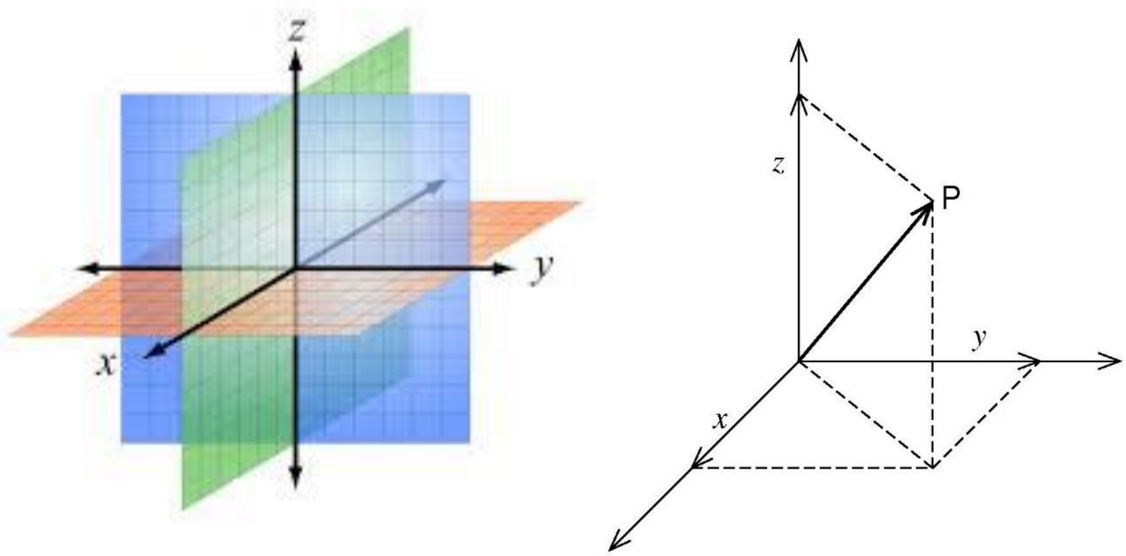


Figure 9.1. A point P in space with Cartesian coordinates x, y, z.

coordinate is constant, while the values of the y and z coordinates can vary, is called the $x =$ constant surface. This surface is a plane. Similarly, $y =$ constant and $z =$ constant surfaces are planes, as shown in Figure 9.1. A curve along which the x coordinate varies, while the y and z coordinates are constant and fixed is called the x–curve (or, x–axis). This curve is a straight line. Similarly, the y–curve and z– curve are straight lines. The unit vectors $\mathbf{e_x}$, $\mathbf{e_y}$, and $\mathbf{e_z}$ are tangentially along the x–curve, y–curve, and z–curve, respectively. Consider a point, P, in space with Cartesian coordinates (x, y, z). The vector, shown in Figure 9.1,

$$\mathbf{r} = x\,\mathbf{e_x} + y\,\mathbf{e_y} + z\,\mathbf{e_z} \ ,$$

locates the position of point P with respect to the origin. A length element **dl** in this coordinate system is

$$\mathbf{dl} = dx\,\mathbf{e}_x + dy\,\mathbf{e}_y + dz\,\mathbf{e}_z\ , \qquad Eq.(9.1a)$$

and the distance between two neighboring points is

$$(dl)^2 = (dx)^2 + (dy)^2 + (dz)^2\ . \qquad Eq.(9.1b)$$

A volume element in Cartesian coordinate system is

$$dV = dxdydz\ . \qquad Eq.(9.1c)$$

An area element in this system is

$$\mathbf{dS} = dxdy\,\mathbf{e}_z + dydz\,\mathbf{e}_x + dzdx\,\mathbf{e}_y\ . \qquad Eq.(9.1d)$$

9.2 SPHERICAL COORDINATES

Again, consider the same point, P, in space located at a position given by the vector $\mathbf{r}$ as shown in Figure 9.2. The spherical coordinates (r, θ, ϕ) of point P are defined as follows. The coordinate r is the length of vector $\mathbf{r}$ and its range is $0 \leq r \leq \infty$. The coordinate θ is the angle between vector $\mathbf{r}$ and the z-axis and its range is $0 \leq \theta \leq \pi$. It is similar to the angle of co-latitude in astronomy. In order to describe the third spherical coordinate, ϕ, we need to focus our attention on two planes, the $y = 0$ plane (or the $x-z$ plane) and the plane containing the vector $\mathbf{r}$ and the z-axis. Then, ϕ is the angle between these two planes. The coordinate ϕ is similar to the angle of longitude in astronomy and its range is $0 \leq \phi \leq 2\pi$. The surface $r =$ constant is a sphere of radius r. The surface

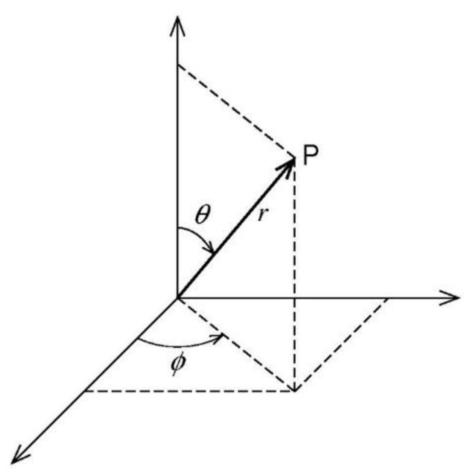


Figure 9.2. The point P of Figure 9.1 shown with its spherical coordinates r, θ, ϕ.

$\theta =$ constant is the curved surface of a cone having vertex at the origin. This cone becomes the $z = 0$ plane if $\theta = \pi/2$, becomes the $+z$-axis if $\theta = 0$, and becomes the $-z$-axis if $\theta = \pi$. The surface $\phi =$ constant is a semi-infinite plane. A curve along which the r coordinate varies, while the θ and ϕ coordinates are constant and fixed, is called the $r-$ curve. This curve is a straight line. Similarly, the $\theta-$ curve is a semicircle and the $\phi-$ curve is a full circle. The unit vectors $\mathbf{e_r}$, $\mathbf{e_\theta}$, and $\mathbf{e_\phi}$ are tangentially along the r–curve, θ–curve, and ϕ–curve, respectively. The relationships between spherical coordinates (r, θ, ϕ) and the Cartesian coordinates (x, y, z) are

$$x = r\sin\theta\cos\phi$$
$$y = r\sin\theta\sin\phi \qquad Eq. (9.2a)$$
$$z = r\cos\theta\ ,$$

or the inverse relationships,

$$r = (x^2 + y^2 + z^2)^{1/2}$$
$$\theta = \arccos\left(\frac{z}{(x^2+y^2+z^2)^{1/2}}\right) \qquad Eq. (9.2b)$$
$$\phi = \arctan(y/x)$$

Thus,

$$dx = \sin\theta\cos\phi\ dr + r\cos\theta\cos\phi\ d\theta - r\sin\theta\sin\phi\ d\phi\ ,$$
$$dy = \sin\theta\sin\phi\ dr + r\cos\theta\sin\phi\ d\theta + r\sin\theta\cos\phi\ d\phi\ , \qquad Eq. (9.3)$$
$$dz = \cos\theta\ dr - r\sin\theta\ d\theta\ .$$

The distance between two neighboring points is

$$(dl)^2 = (dx)^2 + (dy)^2 + (dz)^2\ ,$$

or, on substituting for dx, dy, and dz in terms of dr, $d\theta$, and $d\phi$ from Eq. (9.3),

$$(dl)^2 = (dr)^2 + (r\ d\theta)^2 + (r\sin\theta\ d\phi)^2\ . \qquad Eq. (9.4a)$$

Thus, we note that the roles of length elements dx, dy, and dz of Cartesian coordinates are taken over, in spherical coordinates, by length elements dr, $rd\theta$, and $r\sin\theta\ d\phi$. Thus, a volume element in spherical coordinates is

$$dV = r^2\sin\theta\ dr\ d\theta\ d\phi\ , \qquad Eq. (9.4b)$$

and an area element is

$$\mathbf{dS} = rdrd\theta\ \mathbf{e_\phi} + r^2\sin\theta\ d\theta d\phi\ \mathbf{e_r} + r\sin\theta\ drd\phi\ \mathbf{e_\theta}\ . \qquad Eq. (9.4c)$$

9.3 CYLINDRICAL COORDINATES

Once again, consider the same point, P, in space with Cartesian coordinates (x, y, z) and spherical coordinates (r, θ, ϕ). The cylindrical coordinates (ρ, ϕ, z) of this point are shown in Figure 9.3 and are defined as follows. The cylindrical z coordinate is same as the Cartesian z coordinate. Thus, the range of the z coordinate is $-\infty \leq z \leq \infty$. The cylindrical ϕ coordinate is same as the spherical ϕ coordinate. So, the range of this coordinate is $0 \leq \phi \leq 2\pi$. The cylindrical ρ coordinate is the projection of the position vector $\mathbf{r}$ onto the Cartesian $x-y$ plane. The range of this coordinate is $0 \leq \rho \leq \infty$. The surface $z =$ constant is a plane perpendicular to the z-axis. The surface $\phi =$ constant is a semi-infinite plane. The surface $\rho =$ constant is a cylinder coaxial with the z-axis. As in Cartesian coordinates, the z–curve is a straight line. As in spherical coordinates, the ϕ– curve is a full circle. The

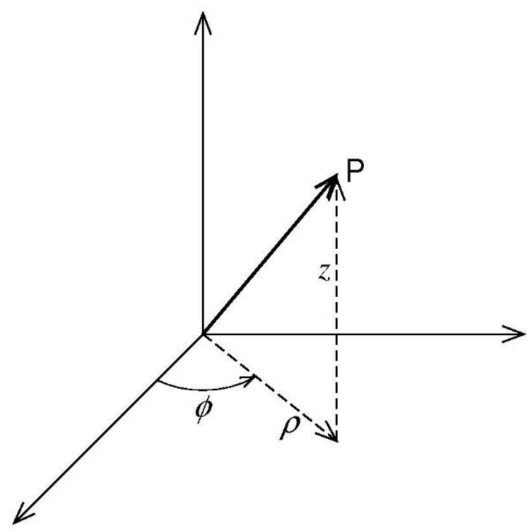


Figure 9.3. The point P of Figures 9.1 and 9.2 shown with its cylindrical coordinates ρ, ϕ, z.

ρ–curve is a curve along which the coordinate ρ varies while the other two coordinates ϕ and z are held constant. This curve is a straight line. The unit vectors $\mathbf{e}_\rho$, $\mathbf{e}_\phi$, and $\mathbf{e}_\mathbf{z}$ are tangentially along the ρ–curve, the ϕ–curve, and the z– curve, respectively. The relationships between cylindrical coordinates (ρ, ϕ, z) and the Cartesian coordinates (x, y, z) are

$$x = \rho \cos\phi \,, y = \rho \sin\phi \quad \text{and} \quad z = z \,, \qquad Eq. (9.5a)$$

and the inverse relationships are

$$\rho = \sqrt{x^2 + y^2}, \;\; \phi = \arctan(y/x) \;\; \text{and} \; z = z \;. \qquad Eq. (9.5b)$$

Using these relationships,

$$dx = \cos\phi \;\, d\rho - \rho \sin\phi \;\, d\phi \;\;,$$

$$dy = \sin\phi \; d\rho + \rho\cos\phi \; d\phi \;, \qquad Eq.(9.6)$$

$$dz = dz \;.$$

Now, the separation between two neighboring points, given by

$$(dl)^2 = (dx)^2 + (dy)^2 + (dz)^2,$$

can be rewritten on substituting for dx, dy, and dz in terms of $d\rho$, $d\phi$, and dz from Eq. (9.6), as

$$(dl)^2 = (d\rho)^2 + (\rho d\phi)^2 + (dz)^2 \;. \qquad Eq.(9.7a)$$

Thus, we note that the length elements dx, dy and dz of Cartesian coordinates are replaced, in cylindrical coordinates, by length elements $d\rho$, $\rho d\phi$ and dz. So, the volume element and the area element in cylindrical coordinates become

$$dV = \rho d\rho d\phi dz \;, \qquad Eq.(9.7b)$$

and,

$$\mathbf{dS} = \rho d\rho d\phi \; \mathbf{e_z} + \rho d\phi dz \; \mathbf{e_\rho} + dz d\rho \; \mathbf{e_\phi} \;, \qquad Eq.(9.7c)$$

respectively.

9.4 PLANE POLAR COORDINATES

We note in passing that because the z coordinate in Cartesian coordinates and in cylindrical coordinates is the same, it is possible to treat the $x-y$ plane as a $\rho-\phi$ plane. This relationship is a depiction of two-dimensional *plane polar coordinates*. In particular, an area element in this plane can be written as

$$dS = dxdy \;\text{ or }\; dS = \rho d\rho \; d\phi \;. \qquad Eq.(9.8)$$

As an example of plane polar coordinates, we attend to an unfinished integral that we encountered in our discussion of Gaussian integrals in Chapter 2. We stated there, without proof, that [see remark after Eq. (2.15)]

$$I_g^0 = \int_0^\infty \exp(-ax^2)\; dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} \;.$$

Now we evaluate this integral. But, first we evaluate the related integral,

$$I = \int_0^\infty \exp(-x^2)\, dx \;.$$

Then, the integral I_g^0 will be obtained from I by replacing x^2 by ax^2. We write the I integral twice, first with dummy variable x and next with dummy variable y, and multiply the two integrals to get

$$I^2 = \int_0^\infty \int_0^\infty \exp[-(x^2 + y^2)]\, dx\, dy \ .$$

Now, think of this integral as an integral over the first quadrant of the $x-y$ plane. Converting to plane polar coordinates in the first quadrant,

$$I^2 = \int_{\rho=0}^{\infty} \int_{\phi=0}^{\pi/2} \exp(-\rho^2)\, \rho d\rho\, d\phi = \frac{\pi}{2}\frac{1}{2}\int_0^\infty \exp(-u)\, du = \frac{\pi}{4} \ ,$$

with $u = \rho^2$. Thus,

$$I = \int_0^\infty \exp(-x^2)\, dx = \frac{\sqrt{\pi}}{2} \ . \qquad Eq.(9.9)$$

Now, to get the Gaussian integral I_g^0 simply replace x^2 by ax^2 in Eq.(9.9). This gives

$$I_g^0 = \int_0^\infty \exp(-ax^2)\ dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} \ .$$

9.5 MULTIPLE INTEGRALS

The curvilinear coordinates, discussed above, are very useful in carrying out integrals which involve areas or volumes of circles, spheres and cylinders. A few examples will impress the reader about the utility of curvilinear coordinates.

Example: Show that the area, A, of a circle of radius R is, $A = \pi R^2$.

Solution:

The statement of this problem is so well-known that it is almost taken as a fact. However, historically speaking, it was considered quite an intellectual feat by Eudoxus in the fifth century BC to realize that the area of a circle is proportional to the square of its radius. At the time of Eudoxus, the formula for finding out the area of a triangle was well-known. So, the area of a circle was determined by inscribing in it regular polygons, which can be broken into several triangles. If $\boldsymbol{A_n}$ is the area of an inscribed polygon of $\boldsymbol{n}$ sides, then $\boldsymbol{A_{circle}} = \lim\limits_{n\to\infty} \boldsymbol{A_n}$.

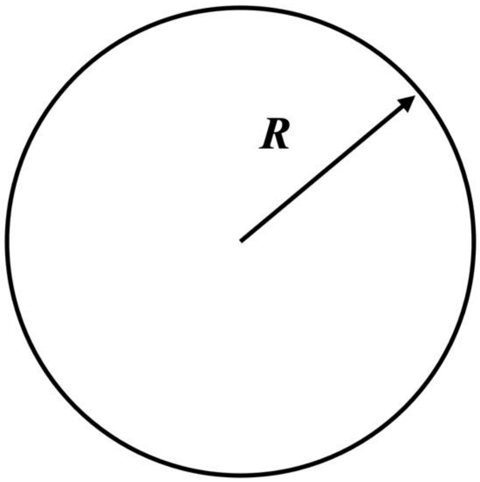


Figure 9.4. A circle of radius R.

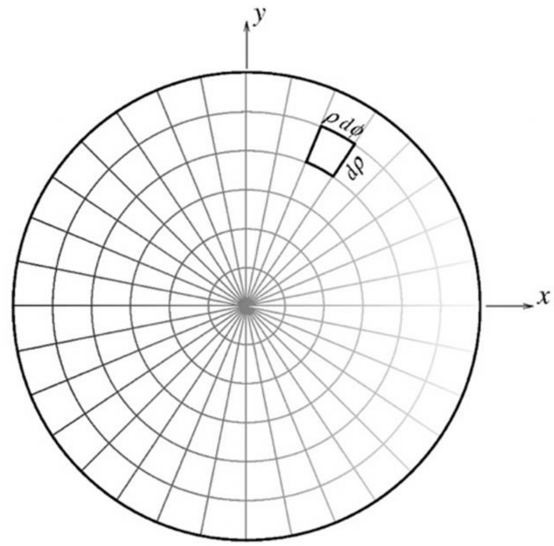


Figure 9.5. Calculating area of a circle.

Now, using multiple integrals, it is very easy to determine the area of a circle. Break the circle into large number of small pieces by drawing various concentric circles and radial spokes inside the circle, as in Figure 9.5. The area of the tiny area element shown in the figure is,

$$dA = (d\rho)(\rho d\phi) = \rho d\rho d\phi \ .$$

Then,

$$A_{circle} = \int_{circle} dA = \int_0^R \rho d\rho \int_0^{2\pi} d\phi = \frac{R^2}{2}(2\pi) = \pi R^2 \ . \qquad Eq. (9.10)$$

Example: Determine volume, V, of a sphere of radius R.

Solution: Since a small volume element in spherical coordinates, $\boldsymbol{dV}$, is given in Eq. (9.4b), the volume of a whole sphere can be determined by summing (or, integrating) over all the small volume elements that make up the sphere. Explicitly,

$$V = \int_{sphere} dV = \int_{r=0}^{R} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} r^2 \sin\theta \, d\theta dr d\phi \ ,$$

or,

$$V = \int_{\phi=0}^{2\pi} d\phi \int_{\theta=0}^{\pi} \sin\theta\, d\theta \int_{r=0}^{R} r^2 dr = (2\pi)(2)\left(\frac{R^3}{3}\right) = \frac{4\pi}{3} R^3 \ . \qquad Eq.(9.11)$$

9.6 SOLIDS OF REVOLUTION

Consider a small strip of length dx located in the $x-y$ plane at $(x, y) \equiv (x, R)$ as in Figure 9.6. If this strip is revolved about the x-axis, it creates a disk of radius R and thickness dx, and, therefore, of volume $V = \pi R^2 dx$. Generalizing this result, if a curve, representing a function $f(x)$ with $a \leq x \leq b$, is revolved about x-axis, as in Figure 9.6, the volume of the solid generated will be

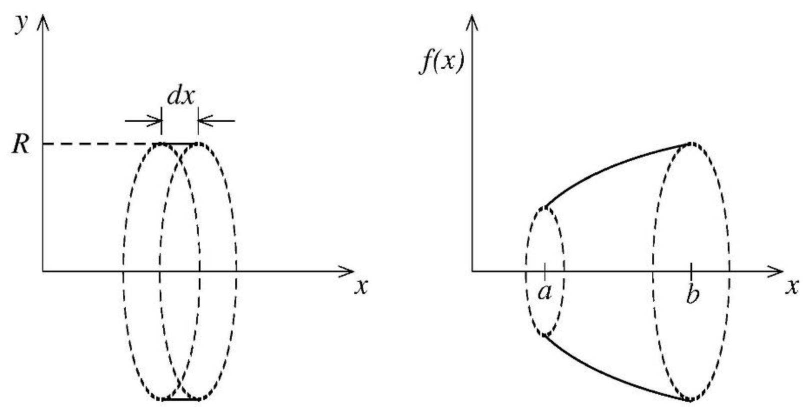


Figure 9.6. Solids of revolution.

$$V = \int_{a}^{b} \pi[f(x)]^2 dx \ . \qquad Eq.(9.12)$$

The solids obtained by revolving a curve about an axis are called the solids of revolution.

Example: **A straight line, given by the equation**

$$\boldsymbol{f(x) = \frac{R}{H} x},$$

is shown in the Figure 9.7. Revolution of this straight line about x-axis leads to a solid of revolution in the form of a cone of radius R and height H. The vertex of this cone is at the origin and its axis lies along the x-axis. The volume of this cone is

$$V = \int_{0}^{H} \pi \left[\frac{R}{H} x\right]^2 dx = \pi \frac{R^2}{H^2} \int_{0}^{H} x^2 dx = \pi \frac{R^2}{H^2} \frac{H^3}{3} = \frac{\pi}{3} R^2 H \ . \qquad Eq.(9.13)$$

Thus, in general, the volume of a cone of radius R and height H is $\pi R^2 H/3$.

Example: Consider a horizontal line of length H in the $x-y$ plane at $y=R$. Revolution of this line around x-axis generates a solid in the form of a cylinder of radius R and height H. Its volume is

$$\boldsymbol{V} = \int_{\boldsymbol{0}}^{\boldsymbol{H}} \boldsymbol{\pi}[\boldsymbol{R}]^2 \boldsymbol{dx} = \boldsymbol{\pi R^2 H} \ .$$

Thus, in general, the volume of a cylinder of radius R and height H is $\pi R^2 H$.

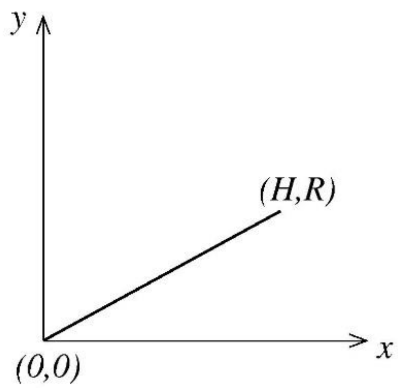


Figure 9.7. A cone generated as a solid of revolution.

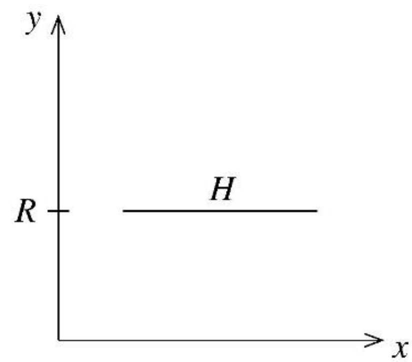


Figure 9.8. A cylinder generated as a solid of revolution.

9.7 CENTER OF MASS

The concept of center of mass is very useful in the study of mechanics of a distribution of masses. The center of mass of the distribution, under the influence of external forces, moves as if all the mass of the distribution were concentrated there and as if all the external forces were applied there.

Consider a mass distribution consisting of n discrete point-like masses, $\Delta m_1, \Delta m_2, \cdots \Delta m_i \cdots \Delta m_n$ located at $\boldsymbol{r_1}, \boldsymbol{r_2} \cdots \boldsymbol{r_i} \cdots \boldsymbol{r_n}$, respectively, as shown in Figure 9.9. The total mass M is the sum of all discrete masses. The center of mass of this distribution is located at

$$\boldsymbol{r}_{cm} = \frac{(\Delta m_1)\boldsymbol{r}_1 + (\Delta m_2)\,\boldsymbol{r}_2 + \ldots + (\Delta m_i)\,\boldsymbol{r}_i + \ldots + (\Delta m_n)\,\boldsymbol{r}_n}{(\Delta m_1) + (\Delta m_2) + \cdots + (\Delta m_n)}$$

$$= \frac{1}{M} \sum_{i=1}^{n} \boldsymbol{r}_i\, \Delta m_i \quad . \qquad Eq. (9.14a)$$

An extended object of total mass M can be imagined as a collection of a large number of point-like masses which are combined together to form a continuous distribution of mass. For such an object, the sum in Eq. (9.14a) can be replaced by an integral. Thus, for a continuous distribution of mass, M, the center of mass of the body is located at

$$\boldsymbol{r}_{cm} = \frac{1}{M} \int_{all\ body} \boldsymbol{r}\ dm \ , \qquad Eq. (9.14b)$$

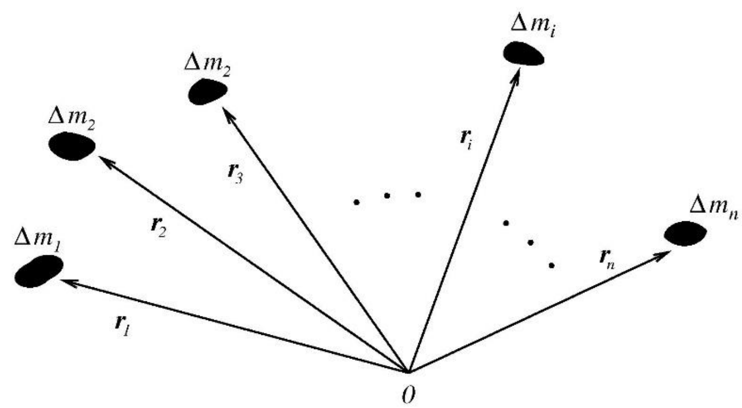


Figure 9.9. The center of mass of a distribution of n discrete point-like masses.

or, in Cartesian coordinates,

$$x_{cm} = \frac{1}{M} \int x\ dm \ ,$$

$$y_{cm} = \frac{1}{M} \int y\ dm \ ,$$

$$z_{cm} = \frac{1}{M} \int z\ dm \ .$$

For a mass distribution with some symmetry, the center of mass is normally at the center of symmetry or on the axis of symmetry. For example, for a full circular disk of radius R, the center of mass is at the center of symmetry, namely, the center of the disk.

Example: Determine the center of mass of a semicircular disk of radius *R* and mass *M*.

Solution:

In this case, the mass per unit area is $\mu = M/\left(\frac{1}{2}\pi R^2\right) = 2M/(\pi R^2)$. Using plane polar coordinates in the plane of the disk:

$$x_{cm} = \frac{1}{M}\int_0^R \int_0^\pi \mu\,(\rho\cos\phi)\, d\rho\, \rho\, d\phi \;= \frac{\mu}{M}\int_0^R \rho^2 d\rho \int_0^\pi \cos\phi\; d\phi = 0$$

$$y_{cm} = \frac{1}{M}\int_0^R \int_0^\pi \mu\,(\rho\sin\phi)\, d\rho\, \rho\, d\phi \;= \frac{\mu}{M}\int_0^R \rho^2 d\rho \int_0^\pi \sin\phi\; d\phi = \frac{\mu}{M}\frac{R^3}{3}2 = \frac{2}{\pi R^2}\frac{R^3}{3}2 = \frac{4R}{3\pi}\;.$$

Thus, center of mass of a semicircular disk is located on its axis of symmetry at $(0{,}4R/3\pi)$.

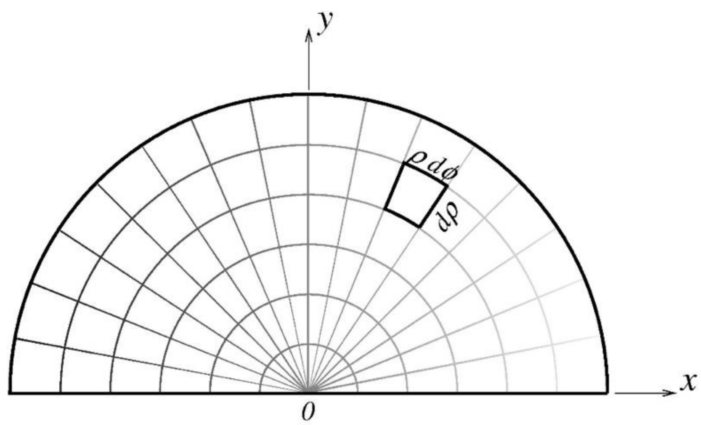


Figure 9.10. Center of mass of a uniform semicircular disk.

9.8 MOMENT OF INERTIA ABOUT AN AXIS

In elementary physics we learn that, during a translational motion, the kinetic energy of a body of mass m moving with a linear velocity v is $K = \frac{1}{2}mv^2$. Consider a person riding a stationary bicycle for exercise. Since the bicycle is stationary, there is no translational kinetic energy associated with the exercise machine. All the work done by pedaling the bicycle is converted into $rotational\ kinetic\ energy$ of the rotating wheel. If the wheel of the exercise bicycle is rotating with a constant angular velocity ω, then its kinetic energy is $K = \frac{1}{2}I\omega^2$, where I is the moment of inertia or the rotational inertia of the wheel. In order to define the moment of inertia of a mass distribution, imagine several discrete point-like masses $\Delta m_1, \Delta m_2, \cdots \Delta m_i \cdots \Delta m_n$ rotating about a certain common axis of rotation, as shown in Figure 9.11. Furthermore, $r_1, r_2 \cdots r_i \cdots r_n$, respectively, are the shortest, perpendicular distances of various masses from the axis of rotation. Then, the moment of inertia of this mass distribution about the axis of rotation is defined as

$$I = \sum_{i=1}^{n} r_i^2\,(\Delta m_i)\;\;. \qquad\qquad Eq.(9.15a)$$

An extended object can be imagined as a collection of a large number of point-like masses which are combined together to form a continuous distribution of mass. For such an object, the sum in Eq. (9.15a) can be replaced by an integral so that

$$I = \int_{all\ body} r^2\, dm \ . \qquad Eq.(9.15b)$$

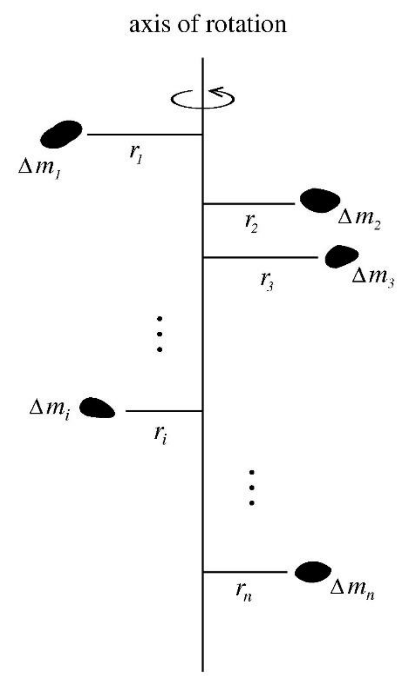


Figure 9.11. Moment of inertia of a distribution of n discrete point-like masses rotating about an axis.

If the mass is distributed uniformly in a volume V with density D_V (mass per unit volume), then

$$I = \int_V r^2\, D_V\, dV \ .$$

If the mass is distributed uniformly in an area A with density D_A (mass per unit area), then

$$I = \int_A r^2 D_A\, dA \ .$$

If the mass is distributed uniformly along a line L with density D_L (mass per unit length), then

$$I = \int_L r^2 D_L \, dL \ .$$

Example: Determine the moment of inertia of a thin spherical shell of mass M and radius R rotating about any diameter.

Solution: The shell has surface area of $4\pi R^2$. If D_A is mass per unit area, $D_A = M/(4\pi R^2)$, then mass of a small patch of area $dA = (Rd\theta)(R\sin\theta \, d\phi)$ is $dm = D_A \, dA$. The perpendicular distance of this patch of area from the axis of rotation is $r = R\sin\theta$. So,

$$I = \int_{shell} [R\sin\theta]^2 dm = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} [R\sin\theta]^2 D_A \, R^2 \sin\theta \ d\theta \, d\phi$$

or, with $u = \cos\theta$,

$$I = D_A \, R^4 \int_0^{\pi} \sin^3\theta \, d\theta \int_0^{2\pi} d\phi = D_A \, R^4 \, 2\pi \int_{-1}^{1} (1-u^2) \, du$$

$$= 2\pi \, D_A \, R^4 \left(\frac{4}{3}\right) = \frac{8\pi}{3}\left(\frac{M}{4\pi R^2}\right) R^4 = \frac{2}{3} M R^2 \ . \qquad Eq. (9.16)$$

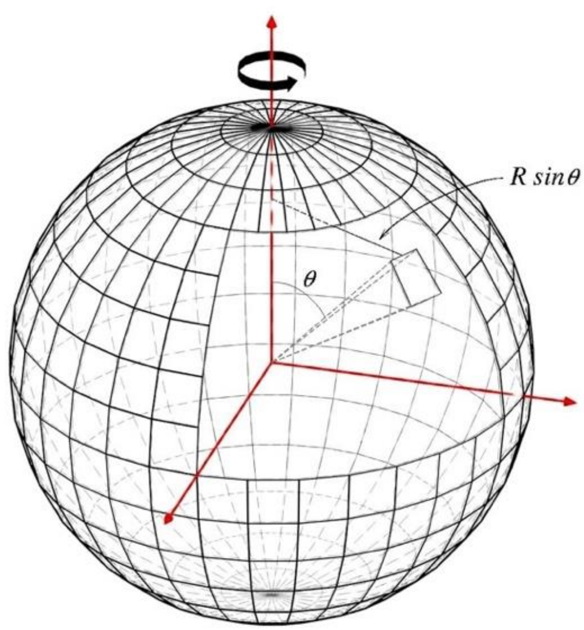


Figure 9.12. Moment of inertia of a thin spherical shell.

Example: A thin rod of length L and mass M is rotating about an axis. The axis is passing through the center of the rod and is perpendicular to the rod. Determine the moment of inertia of this rotating rod.

Solution:

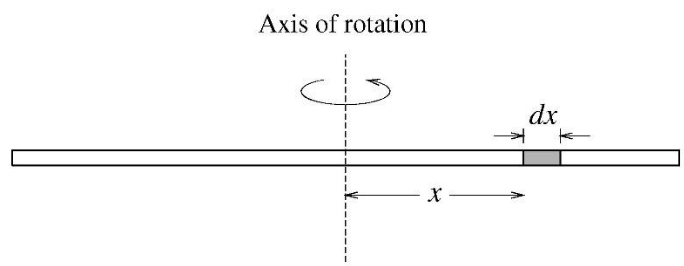


Figure 9.13. Moment of inertia of a rod rotating about an axis.

Since the mass is distributed along the length of the rod, we can define mass per unit length as $D_L = M/L$. Consider a small mass element of length dx located a distance x away from the axis of rotation. The mass of this element is $dm = D_L\, dx$. So, the moment of inertia is

$$I = \int_{rod} x^2 dm = \int_{-\frac{L}{2}}^{\frac{L}{2}} x^2 D_L\, dx = D_L \left.\frac{x^3}{3}\right|_{-\frac{L}{2}}^{\frac{L}{2}} = D_L \left[\frac{L^3}{24} - \left(-\frac{L^3}{24}\right)\right] ,$$

or,

$$I = \frac{M}{L}\left(\frac{L^3}{12}\right) = \frac{1}{12} M L^2 \ . \qquad Eq. (9.17)$$

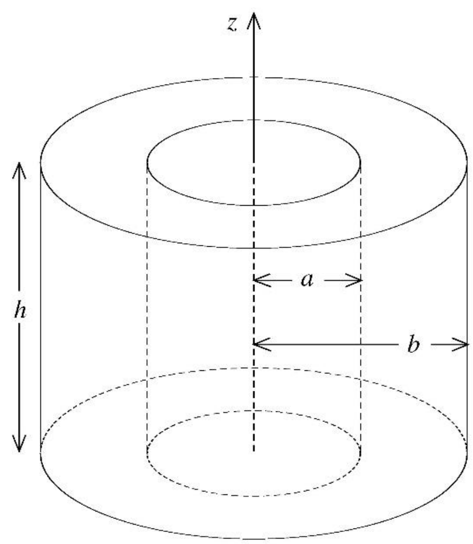
PROBLEMS FOR CHAPTER 9

1. Find the position of the center of mass of a solid cone of radius R and height H.

2. A solid sphere, of radius R, is cut into two identical hemispheres. Find the position of the center of mass of such a hemisphere.

3. Find the moment of inertia of a solid cone of radius R, uniform mass density D, height H, and mass M about its axis of symmetry. Give the answer in terms of M and R.

4. (a) Determine the moment of inertia of a thick-walled right circular cylindrical tube of mass M, inner radius a, outer radius b, and height h, about its central axis of symmetry.



(b) Using the result of part (a), determine the moment of inertia of a solid right circular cylinder of radius R and mass M.

(c) Using the result of part (a), determine the moment of inertia of a right circular cylindrical shell, open at both ends, of radius R and mass M. The thickness of the shell can be considered as almost zero.

5. (a) Determine the moment of inertia of a thick spherical shell, of mass M, about an axis passing through its center. The inner and outer radii of the shell are a and b, respectively.

(b) Using the result of part (a), determine the moment of inertia of a solid sphere of radius R and mass M.

(c) Using the result of part (a), determine the moment of inertia of a thin spherical shell of radius R and mass M. The thickness of the shell can be considered as almost zero.

6. **Biomedical Physics Application**. The human heart is a muscular organ that is responsible for pumping blood throughout the body. Without resorting to an actual surgery, the size of a healthy heart can be estimated by

using a medical imaging technique called CAT scan or simply CT scan (for computed tomography). The scan produces equally spaced cross-sectional views of the heart. Suppose that a CT scan of a human heart shows cross sections spaced 1.0 cm apart. The heart is about 10 cm long and the cross-sectional areas, in square centimeters, are 0, 18, 38, 57, 72, 84, 71, 60, 39, 20, and 0. Using this information estimate the volume of the heart.

7. **Biomedical Physics Application**. The velocity v of blood that flows in a blood vessel with radius R and length L at a distance r from the central axis is

$$v(r) = \frac{P}{4\eta L}\,(R^2 - r^2)\ ,$$

where P is the pressure difference between the ends of the vessel and η is the viscosity of the blood. The flux (volume of blood flowing per unit time) F of the blood can be calculated by approximating the cross-sectional area of the blood vessel by concentric rings and integrating over all such rings. Show that the flux is given by

$$F = \frac{\pi P R^4}{8\eta L}\ .$$

This is called Poiseuille’s Law.

Chapter 10: Vector Calculus

In this chapter we will combine our previous knowledge of calculus and of vectors to learn about vector functions and their derivatives. We will also gain knowledge of a couple of specialized theorems, the Gauss's theorem (or, divergence theorem) and Stoke's theorem, which are relevant for vector functions.

10.1 VECTOR FUNCTIONS

A vector function is a vector whose components are functions of coordinates. As an example, a vector function in $Cartesian\ coordinates$ is

$$\mathbf{A}(x_1, x_2, x_3) = \mathbf{e_1} A_1(x_1, x_2, x_3) + \mathbf{e_2} A_2(x_1, x_2, x_3) + \mathbf{e_3} A_3(x_1, x_2, x_3) \ . \qquad Eq. (10.1)$$

Here $\mathbf{e_1}, \mathbf{e_2}, \text{and } \mathbf{e_3}$ are the three orthogonal unit vectors and $A_1, A_2, \text{and } A_3$ are the three scalar functions of coordinates $x_1, x_2, \text{and } x_3$.

Now, in vector calculus, just as in multivariate calculus, all derivatives are related to the rate of change of a function. For a scalar function $f(x_1, x_2, x_3)$, the three derivatives, $\frac{\partial f}{\partial x_i}$ $(i = 1,2,3)$, represent the rate of variation of the function f with respect to x_i $(i = 1,2,3)$. These three derivatives of f behave like the three components of a vector, which is called the gradient of f. The gradient of f is written as

$$\boldsymbol{\nabla} f = \mathbf{e_1} \frac{\partial f}{\partial x_1} + \mathbf{e_2} \frac{\partial f}{\partial x_2} + \mathbf{e_3} \frac{\partial f}{\partial x_3} \ . \qquad Eq. (10.2)$$

Note that gradient of f is a vector function. The differential operator

$$\boldsymbol{\nabla} = \mathbf{e_1} \frac{\partial}{\partial x_1} + \mathbf{e_2} \frac{\partial}{\partial x_2} + \mathbf{e_3} \frac{\partial}{\partial x_3} = \sum_{i=1}^{3} \mathbf{e_i} \frac{\partial}{\partial x_i}$$

is called the del or $gradient$ operator. Let us now interpret the gradient of a scalar function. Consider the values of a scalar function f at two neighboring points in space. The first point is at $\mathbf{r} = (x_1, x_2, x_3)$ and the value of the function there is $f(x_1, x_2, x_3)$. The neighboring point is at $\mathbf{r} + \mathbf{dr} = (x_1 + dx_1, x_2 + dx_2, x_3 + dx_3)$ and the value of the function there is $f(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3)$, as shown in Figure 10.1. The change in the value of the function, f, on moving from one point to a neighboring point, is

$$df = f(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3) - f(x_1, x_2, x_3) \ ,$$

or, keeping only the first order terms in the Taylor-series expansion of the function,

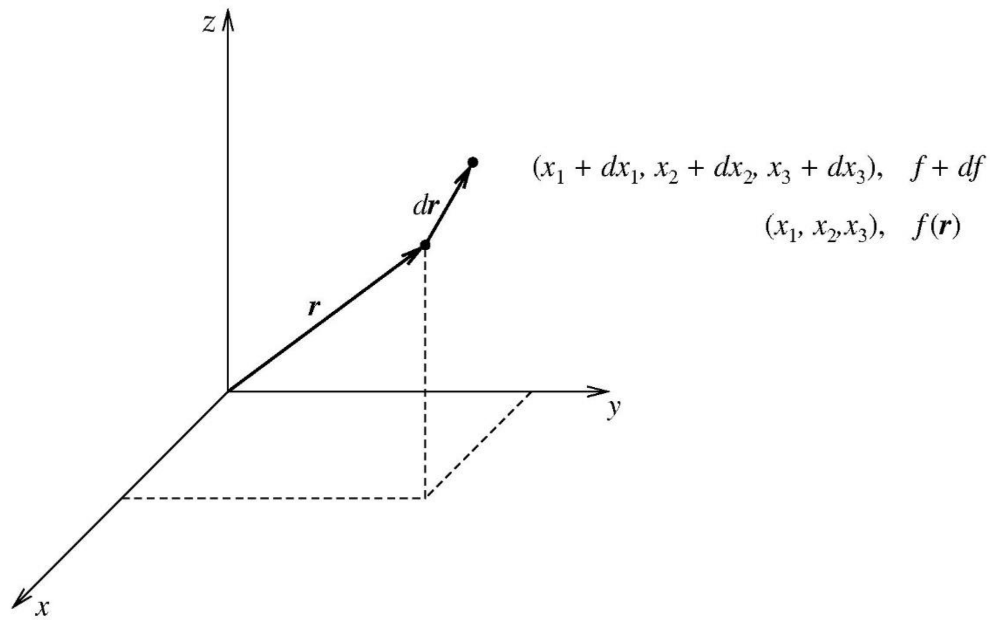


Figure 10.1. Values of a function at two neighboring points.

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \frac{\partial f}{\partial x_3} dx_3 = \boldsymbol{\nabla} f \bullet \mathbf{dr} \ . \qquad Eq. (10.3)$$

This expression leads to the following properties of the gradient of a scalar function f.

First, the change, df, in the value of the function is maximum when $\boldsymbol{\nabla} f$ is parallel to $\mathbf{dr}$. Thus $\boldsymbol{\nabla} f$ is a vector that represents the magnitude and direction of the greatest rate of change of f. Second, if $\mathbf{dr}$ is tangential to the surface $f(\mathbf{r}) = \text{constant}$, then both $\mathbf{r}$ and $\mathbf{r} + \mathbf{dr}$ lie on the surface. So,

$$df = f(\mathbf{r} + \mathbf{dr}) - f(\mathbf{r}) = 0 \ ,$$

or,

$$\boldsymbol{\nabla} f \bullet \mathbf{dr} = 0 \ .$$

Since $\mathbf{dr}$ is tangential to the surface, it means that $\boldsymbol{\nabla} f$ is normal (or, perpendicular) to the surface $f(\mathbf{r}) = \text{constant}$. We note in passing that in electrostatics the electric field, $\boldsymbol{E} = -\boldsymbol{\nabla}\phi$, is normal to an equipotential surface on which the electrostatic potential, ϕ, is constant. This fact is a direct consequence of the second property of the gradient of a function. The rate of variation of the scalar function f along any *arbitrary* direction $\boldsymbol{n}$ is given by $\boldsymbol{n} \bullet \boldsymbol{\nabla} f$, which is called the *directional derivative*.

Divergence and Curl of a Vector Function

Because of the vector nature of the del operator, $\boldsymbol{\nabla}$, we can form the following two combinations for a vector function **A**:

$$\boldsymbol{\nabla}\bullet \mathbf{A} = \left(\mathbf{e_1}\frac{\partial}{\partial x_1} + \mathbf{e_2}\frac{\partial}{\partial x_2} + \mathbf{e_3}\frac{\partial}{\partial x_3}\right)\bullet(\mathbf{e_1}A_1 + \mathbf{e_2}A_2 + \mathbf{e_3}A_3) = \frac{\partial A_1}{\partial x_1} + \frac{\partial A_2}{\partial x_2} + \frac{\partial A_3}{\partial x_3} \qquad Eq.(10.4)$$

which is called the divergence of function **A**, and

$$\boldsymbol{\nabla}\times \mathbf{A} = \left(\mathbf{e_1}\frac{\partial}{\partial x_1} + \mathbf{e_2}\frac{\partial}{\partial x_2} + \mathbf{e_3}\frac{\partial}{\partial x_3}\right)\times(\boldsymbol{e_1}\,A_1 + \mathbf{e_2}A_2 + \mathbf{e_3}A_3)$$

$$= \mathbf{e_1}\left(\frac{\partial A_3}{\partial x_2} - \frac{\partial A_2}{\partial x_3}\right) + \mathbf{e_2}\left(\frac{\partial A_1}{\partial x_3} - \frac{\partial A_3}{\partial x_1}\right) + \mathbf{e_3}\left(\frac{\partial A_2}{\partial x_1} - \frac{\partial A_1}{\partial x_2}\right) \qquad Eq.(10.5)$$

which is called the curl of function **A**. If, for a vector function **A**, $\boldsymbol{\nabla}\bullet\mathbf{A} = 0$, then **A** is called a solenoidal vector function. On the other hand, if $\boldsymbol{\nabla}\times\mathbf{A} = 0$, then **A** is called an irrotational or conservative vector function. Finally, we define $\boldsymbol{\nabla}\bullet\boldsymbol{\nabla}f$, which is the divergence of the gradient of a scalar function, as the Laplacian of f. It is written as $\nabla^2 f$. Explicitly,

$$\nabla^2 f = \left(\mathbf{e_1}\frac{\partial}{\partial x_1} + \mathbf{e_2}\frac{\partial}{\partial x_2} + \mathbf{e_3}\frac{\partial}{\partial x_3}\right)\bullet\left(\mathbf{e_1}\frac{\partial f}{\partial x_1} + \mathbf{e_2}\frac{\partial f}{\partial x_2} + \mathbf{e_3}\frac{\partial f}{\partial x_3}\right) = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_3^2}\ .$$

Or,

$$\nabla^2 f\ = \sum_{i=1}^{3}\frac{\partial^2 f}{\partial x_i^2}\ . \qquad Eq.(10.6)$$

The Laplacian operator, ∇^2, can also operate on a vector function $\mathbf{A}(x_1, x_2, x_3)$ as,

$$\nabla^2\mathbf{A} = \mathbf{e_1}\nabla^2 A_1 + \mathbf{e_2}\nabla^2 A_2 + \mathbf{e_3}\nabla^2 A_3\ . \qquad Eq.(10.7)$$

Note carefully that since the Cartesian unit vectors $\mathbf{e_1}, \mathbf{e_2},$ and $\mathbf{e_3}$ are constant in magnitude and direction, they can be taken out of the ∇^2 operator. Finally, we prove an important identity,

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\bullet\mathbf{A}) - \nabla^2\mathbf{A}\ . \qquad Eq.(10.8)$$

To prove it, we start with the left-hand side of the identity,

$$\text{Left Hand Side} = \sum_{ijk} \epsilon_{ijk}\ \mathbf{e_i}\ \frac{\partial}{\partial x_j}\{\nabla \times \mathbf{A}\}_k$$

$$= \sum_{ijk} \epsilon_{ijk}\ \mathbf{e_i}\ \frac{\partial}{\partial x_j}\left\{\sum_{lm} \epsilon_{klm} \frac{\partial}{\partial x_l}\ A_m\right\} = \sum_{ij}\sum_{lm} \mathbf{e_i}\ \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_l} A_m \sum_k \epsilon_{kij}\epsilon_{klm}$$

$$= \sum_{ij}\sum_{lm} \mathbf{e_i} \frac{\partial}{\partial x_j}\ \frac{\partial}{\partial x_l} A_m\{\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\} = \sum_{ij} \mathbf{e_i} \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} A_j - \sum_{ij} \mathbf{e_i} \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_j} A_i$$

$$= \sum_i \mathbf{e_i} \frac{\partial}{\partial x_i}\sum_j \frac{\partial A_j}{\partial x_j} - \sum_j \frac{\partial^2}{\partial x_j^2}\sum_i \mathbf{e_i} A_i = \nabla(\nabla \bullet \mathbf{A}) - \nabla^2 \mathbf{A}\ .$$

Here we used the epsilon-delta identity of Appendix C. The above vector identity is valid only for Cartesian components of the vector function $\mathbf{A}(x_1, x_2, x_3)$ since the unit vectors $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ are taken to be constant in magnitude and direction.

In the remainder of this chapter, including the following examples, it will be easier to use (x, y, z) notation instead of (x_1, x_2, x_3) notation. However, for the Cartesian unit vectors we will continue to use $\mathbf{e_1}, \mathbf{e_2}$, and $\mathbf{e_3}$.

Example: Consider the scalar function,

$$\boldsymbol{f(x, y, z) = x^2 + y^2 - z^2 + xy - xz - yz}\ .$$

Determine ∇f, unit vector along ∇f, directional derivative along $\mathbf{e_1} + \mathbf{e_2}$, and $\nabla^2 f$, all at $(1, 1, 1)$.

Solution: Starting with gradient of $\boldsymbol{f}$,

$$\nabla f = \mathbf{e_1}\frac{\partial f}{\partial x} + \mathbf{e_2}\frac{\partial f}{\partial y} + \mathbf{e_3}\frac{\partial f}{\partial z}$$

$$= \mathbf{e_1}(2x + y - z) + \mathbf{e_2}\ (2y + x - z) + \mathbf{e_3}(-2z - x - y)\ .$$

$$\nabla f \text{ at } (1,1,1) = \mathbf{e_1}(2) + \mathbf{e_2}(2) + \mathbf{e_3}(-4)\ .$$

$$\text{Unit vector along } \nabla f \text{ at } (1,1,1) = \frac{1}{\sqrt{24}}(2\mathbf{e_1} + 2\mathbf{e_2} - 4\mathbf{e_3}) = \frac{1}{\sqrt{6}}(\mathbf{e_1} + \mathbf{e_2} - 2\mathbf{e_3})\ .$$

In order to determine the directional derivative along $\mathbf{e_1} + \mathbf{e_2}$, we first determine the unit vector along $\mathbf{e_1} + \mathbf{e_2}$, which we will label as $\boldsymbol{n}$. It is

$$\boldsymbol{n} = \frac{1}{\sqrt{2}}(\mathbf{e_1} + \mathbf{e_2}) \ .$$

So, the directional derivative is

$$\boldsymbol{n}\bullet\boldsymbol{\nabla}f = \frac{1}{\sqrt{2}}(\mathbf{e_1} + \mathbf{e_2})\bullet[\mathbf{e_1}(2x + y - z) + \mathbf{e_2}(2y + x - z) + \mathbf{e_3}(-2z - x - y)]$$

$$= \frac{1}{\sqrt{2}}[2x + y - z + 2y + x - z] = \frac{3x + 3y - 2z}{\sqrt{2}} = \frac{4}{\sqrt{2}} \quad \text{at } (1,1,1) \ .$$

Finally,

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 2 + 2 - 2 = 2 \ .$$

Example: Given the vector function

$$\mathbf{A} = \boldsymbol{x^2 y}\,\mathbf{e_1} + \boldsymbol{y^2 z}\,\mathbf{e_2} + \boldsymbol{z^2 x}\,\mathbf{e_3} \ ,$$

determine its divergence, curl, and Laplacian. Then, explicitly verify that

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\bullet\mathbf{A}) - \boldsymbol{\nabla}^2\mathbf{A} \ .$$

Solution: The divergence of **A** is

$$\boldsymbol{\nabla}\bullet\mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} = 2xy + 2yz + 2zx \ .$$

The curl of **A** is

$$\boldsymbol{\nabla} \times \mathbf{A} = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ x^2 y & y^2 z & z^2 x \end{vmatrix} = \mathbf{e_1}(0 - y^2) + \mathbf{e_2}(0 - z^2) + \mathbf{e_3}(0 - x^2) = -y^2\mathbf{e_1} - z^2\mathbf{e_2} - x^2\mathbf{e_3} \ .$$

The Laplacian of **A** is

$$\nabla^2\mathbf{A} = \mathbf{e_1}(2y) + \mathbf{e_2}(2z) + \mathbf{e_3}(2x) \ .$$

Now we verify the identity,

$$\nabla \times (\nabla \times \mathbf{A}) = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ -y^2 & -z^2 & -x^2 \end{vmatrix} = \mathbf{e_1}(2z) + \mathbf{e_2}(2x) + \mathbf{e_3}(2y) = 2[z\mathbf{e_1} + x\mathbf{e_2} + y\mathbf{e_3}]$$

$$\nabla(\nabla \bullet \mathbf{A}) = \mathbf{e_1}(2y + 2z) + \mathbf{e_2}(2x + 2z) + \mathbf{e_3}(2y + 2x) \ .$$

On substituting explicitly,

$$\nabla(\nabla \bullet \mathbf{A}) - \nabla^2 \mathbf{A} = 2z\mathbf{e_1} + 2x\mathbf{e_2} + 2y\mathbf{e_3} = \nabla \times (\nabla \times \mathbf{A}) \ .$$

Example: Find $\nabla f(r)$ where r $= x\, \mathbf{e}_1 + y\, \mathbf{e}_2 + z\, \mathbf{e}_3$ with $r^2 = x^2 + y^2 + z^2$. Note $f(r)$ is a function of only the magnitude r of the vector r.

Solution: Starting with gradient of $f(r)$,

$$\nabla f(r) = \mathbf{e_1}\frac{\partial}{\partial x} f(r) + \mathbf{e_2}\frac{\partial}{\partial y} f(r) + \mathbf{e_3}\frac{\partial}{\partial z} f(r)$$

$$= \mathbf{e_1}\frac{\partial r}{\partial x}\frac{df(r)}{dr} + \mathbf{e_2}\frac{\partial r}{\partial y}\frac{df(r)}{dr} + \mathbf{e_3}\frac{\partial r}{\partial z}\frac{df(r)}{dr}$$

Use

$$\frac{\partial r}{\partial x} = \frac{x}{r}, \qquad \frac{\partial r}{\partial y} = \frac{y}{r}, \qquad \frac{\partial r}{\partial z} = \frac{z}{r} \ ,$$

to obtain

$$\nabla f(r) = \frac{x\,\mathbf{e_1} + y\,\mathbf{e_2} + z\,\mathbf{e_3}}{r}\frac{df}{dr} = \frac{\mathbf{r}}{r}\frac{df}{dr} \ .$$

10.2 LINE INTEGRALS

The line integral of a vector function **A** over an arbitrary path joining point 1 to point 2 is written as,

$$I = \int_1^2 \mathbf{A} \bullet \mathbf{dr} \ . \qquad Eq. (10.9)$$

It is evaluated by breaking the path, shown in Figure 10.2, into a large number of small segments of lengths $\mathbf{dr}_1, \mathbf{dr}_2, \cdots \mathbf{dr}_i \cdots \mathbf{dr}_n$. For each tiny segment, we calculate the value of the integrand, namely, $\mathbf{A}_i \bullet \mathbf{dr}_i$ and add all the values. The limiting value of this sum as the number of segments becomes very large or as the segment sizes become very small is the value of the line integral. In other words,

$$I = \int_1^2 \mathbf{A}\bullet\mathbf{dr} = \lim_{n\to\infty}\sum_{i=1}^{n} \mathbf{A}_i\bullet\mathbf{dr}_i$$

From this definition of the line integrals, we can derive an important result. Recalling from Eq. (10.3) that for any scalar function

$$df = \boldsymbol{\nabla} f\bullet\mathbf{dr} \ ,$$

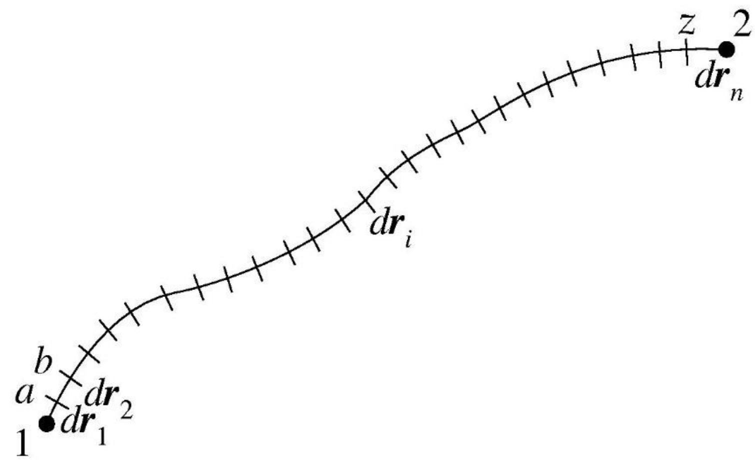


Figure 10.2. The path for a line integral.

we obtain

$$f(a) - f(1) = (\boldsymbol{\nabla} f)_1\bullet\ \mathbf{dr}_1 \ ,$$

$$f(b) - f(a) = (\boldsymbol{\nabla} f)_2\bullet\ \mathbf{dr}_2 \ ,$$

$$f(c) - f(b) = (\boldsymbol{\nabla} f)_3\bullet\ \mathbf{dr}_3 \ ,$$

$$\dots\dots\dots\dots$$

$$f(2) - f(z) = (\boldsymbol{\nabla} f)_n\bullet\ \mathbf{dr}_n \ .$$

Simply adding all these lines, we get

$$f(2) - f(1) = \sum_{i=1}^{n} (\boldsymbol{\nabla} f)_i\bullet\mathbf{dr}_i \to \int_1^2 \boldsymbol{\nabla} f\bullet\mathbf{dr} \ ,$$

independent of the path joining points 1 and 2. In particular, for a closed path,

$$\oint (\nabla f) \bullet \mathbf{dr} = 0 \ , \qquad Eq. (10.10)$$

for any scalar function $f(x, y, z)$. The symbol of a circle on an integral sign indicates that the path of integration is a closed path.

10.3 SURFACE INTEGRALS

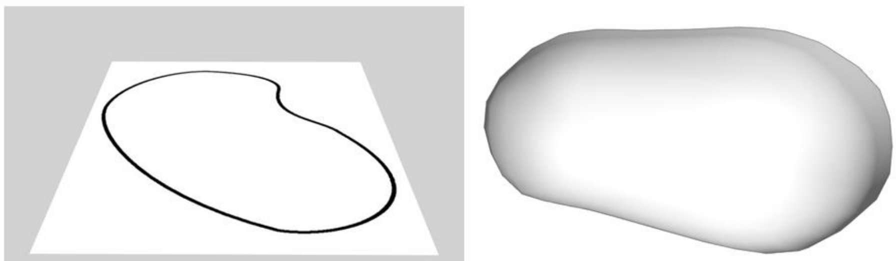


Figure 10.3a. Comparison of an open surface versus a closed surface.

In surface integrals, we encounter two different kinds of surfaces, open surfaces versus closed surfaces, as shown in Figure 10.3a. An open surface has two sides with a boundary C but no enclosed volume. An example of such a surface is a flat or a crumpled sheet of paper. In order to go from one side of an open surface to its other side, one has to cross the boundary C. A closed surface, on the other hand, has an enclosed volume but no boundary. Examples of closed surfaces include the surface of the Earth, the brown surface of an Idaho potato, etc. The surface integral of a vector function **A** over an arbitrary surface, open or closed, is written as

$$J = \int_S \mathbf{A} \bullet \mathbf{dS} = \int_S \mathbf{A} \bullet \boldsymbol{n} \, dS \quad . \qquad Eq. (10.11)$$

Here $\boldsymbol{n}$ is the normal to the area element $\mathbf{dS}$. The surface integral is evaluated by breaking the whole surface area, shown in Figure 10.3b, into a large number of small pieces of areas $\Delta\mathbf{S}_1, \Delta\mathbf{S}_2, \cdots \Delta\mathbf{S}_i \cdots \Delta\mathbf{S}_n$. For each small piece, we calculate the value of the integrand, namely, $\mathbf{A}_i \bullet \Delta\mathbf{S}_i$ and add all the values. The limiting value of this sum as the number of pieces becomes very large or as the size of all pieces becomes very small is the value of the surface integral. In other words,

$$J = \int_S \mathbf{A}\bullet\mathbf{dS} \; = \lim_{n\to\infty}\sum_{i=1}^{n}(\mathbf{A}\bullet\boldsymbol{n})_i\,\Delta S_i \;\; .$$

In the case of a closed surface, this integral is called $the\ flux\ of\ vector\ function$ ***A*** out of the closed volume V.

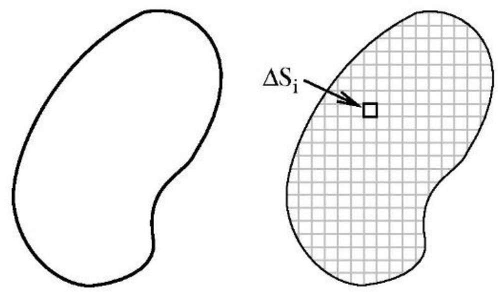


Figure 10.3b. Breaking a surface area into a large number of small area elements.

10.4 VOLUME INTEGRALS

The volume integral of a scalar function $f(x,y,z)$ over an arbitrary volume V is written as,

$$K = \int_V f(x,y,z)\,dV \quad . \qquad\qquad Eq.(10.12)$$

It is evaluated by breaking the volume, shown in Figure 10.4, into a large number of small elements of volume $\Delta V_1, \Delta V_2, \cdots \Delta V_i \cdots \Delta V_n$. For each little element, we calculate the value of the integrand, namely, $f(x_i,y_i,z_i)\Delta V_i$ and add all the values. The limiting value of this sum as the number of volume elements becomes very large

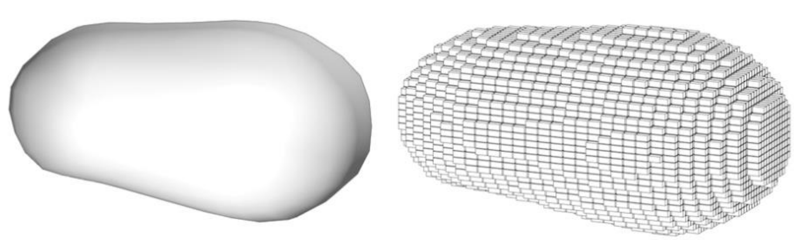


Figure 10.4. Breaking a volume into many small volume elements.

or as the element size becomes very small is the value of the volume integral. In other words,

$$K = \int_V f(x,y,z)\, dV = \lim_{n\to\infty} \sum_{\mathrm{i}=1}^{n} f(x_i, y_i, z_i)\Delta V_i \ .$$

10.5 FLUX OF A VECTOR FUNCTION: GAUSS'S (OR DIVERGENCE) THEOREM

The flux of a vector function **A** out of a closed volume V has a very interesting property. If we break the volume V into two smaller pieces, then the sum of fluxes of **A** out of two closed smaller volumes is equal to the flux of **A** out of the original large, closed volume V. Let us prove it.

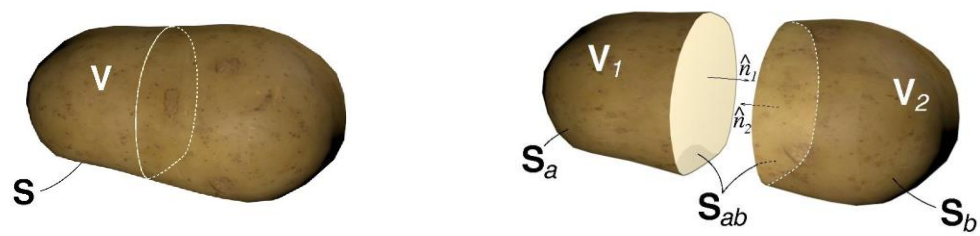


Figure 10.5. Breaking a volume V into two smaller volumes V_1 and V_2.

Consider a closed surface S enclosing a volume V, as in Figure 10.5. For clear visualization, imagine this surface to be the brown surface of an Idaho potato. In order to determine the total flux of a vector function **A** out of volume V, we first consider the flux of **A** through a small surface element $\mathbf{dS}$. It is $\mathbf{A} \bullet \boldsymbol{n}\, dS = \mathbf{A} \bullet \mathbf{dS}$. Total flux of **A** out of volume V is $\int_s \mathbf{A} \bullet \mathbf{dS}$. Now break the volume V into two pieces by cutting the potato with a knife. The two cut pieces have volumes V_1 and V_2 and the corresponding brown surface areas are S_a and S_b, respectively. Clearly, $V = V_1 + V_2$ and $S = S_a + S_b$. The cutting of the potato also exposes the interior white surface area, S_{ab}, that is common to both pieces of the potato. The volume V_1 is enclosed by surface $S_1 = S_a + S_{ab}$ and the volume V_2 is enclosed by surface $S_2 = S_b + S_{ab}$. Total flux of **A** out of volume V_1 is

$$\int_{S_a} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_1\, dS \ ,$$

where $\boldsymbol{n}_1$ is the outward normal on the S_{ab} part of S_1. Total flux of **A** out of volume V_2 is

$$\int_{S_b} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_2\, dS \ ,$$

where $\boldsymbol{n}_2$ is the outward normal on the S_{ab} part of S_2. Now, the sum of total flux of **A** out of volume V_1 and the total flux of **A** out of volume V_2 is

$$\int_{S_a} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_1\, dS + \int_{S_b} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_2\, dS \ .$$

Noting that $\boldsymbol{n}_1 = -\boldsymbol{n}_2$, the sum of fluxes out of volume V_1 and out of volume V_2 becomes,

$$\int_{S_a} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_b} \mathbf{A} \bullet \boldsymbol{n}\, dS = \int_{S_a+S_b} \mathbf{A} \bullet \boldsymbol{n}\ dS = \int_{S} \mathbf{A} \bullet \boldsymbol{n}\, dS\ ,$$

which is equal to the total flux of **A** out of volume V. We can generalize this result by subdividing a giant volume V into a large number of smaller component volumes, of any size and shape. Then the sum of fluxes of **A** out of smaller component volumes is equal to the flux of **A** out of giant volume V**.** For convenience, we take each small component volume to be a parallelepiped in shape.

Flux of A Out of a Parallelepiped

Now, we will determine the flux of a vector function **A** out of a parallelepiped of side lengths $(\Delta x, \Delta y, \Delta z)$. Total volume of the parallelepiped is $\Delta V = \Delta x \Delta y \Delta z$ and its surface area ΔS consists of six faces; namely, left face, right face, front face, back face, top face, and bottom face, as shown in Figure 10.6. Total flux of **A** out of the parallelepiped, $\int_{\Delta S} \mathbf{A} \bullet \mathbf{dS}$, is the sum of six terms each representing a flux out of a different face. We will determine the flux out of three pairs of faces, the left-right pair, the front-back pair and the top-bottom pair.

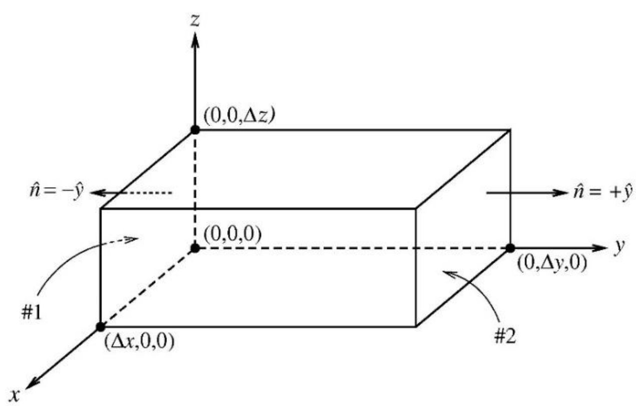


Figure 10.6. Flux of vector function **A** out of a parallelepiped.

Flux out of the left face 1 = $\int_{\#1} \mathbf{A} \bullet \boldsymbol{n}\, dS = -\int_{\#1} A_y\ dx\, dz = -A_y(1)\, \Delta x \Delta z$

Flux out of the right face 2= $\int_{\#2} \mathbf{A} \bullet \boldsymbol{n}\, dS = \int_{\#2} A_y\ dx\, dz = +A_y(2)\, \Delta x \Delta z$

Total flux out of the left-right pair of faces = $\left[A_y(2) - A_y(1)\right]\Delta x \Delta z \cong \left[\left(\frac{\partial A_y}{\partial y}\right)\ \Delta y\right] \Delta x \Delta z = \frac{\partial A_y}{\partial y} \Delta V.$

Similarly, total flux out of the front-back pair of faces = $\left[\left(\frac{\partial A_x}{\partial x}\right)\ \Delta x\right]\Delta y \Delta z = \frac{\partial A_x}{\partial x}\Delta V$ and total flux out of the top-bottom pair of faces = $\left[\left(\frac{\partial A_z}{\partial z}\right)\ \Delta z\right]\Delta x \Delta y = \frac{\partial A_z}{\partial z}\Delta V$.

Adding the contributions from all six faces, total flux of **A** out of volume ΔV is

$$\int_{\Delta S} \mathbf{A} \bullet \boldsymbol{n}\, dS = (\boldsymbol{\nabla} \bullet \mathbf{A})\Delta V \ .$$

Now, any arbitrary volume V can be broken into a large number of small parallelepipeds as shown in Figure 10.7, so that, in general, for a closed surface S,

Figure 10.7. Breaking an arbitrary volume into a large number of parallelepipeds.

$$\int_S \mathbf{A} \bullet \boldsymbol{n}\ dS = \int_{enclosed\ V} (\boldsymbol{\nabla} \bullet \mathbf{A}) dV \ . \qquad Eq. (10.13)$$

This is known as the Gauss's theorem or the divergence theorem for vector functions.

Example: Consider a vector function:

$$\mathbf{A} = xy\ \mathbf{e_1} + yz\ \mathbf{e_2} + zx\ \mathbf{e_3}\ ,$$

and a parallelepiped with side lengths a, b, c **and volume** $V = abc$. **Show that the divergence theorem is satisfied.**

Solution: The divergence of the vector function is

$$\boldsymbol{\nabla} \bullet \mathbf{A} = y + z + x\ .$$

The integral of this divergence over the volume of the parallelepiped of Figure 10.8 is

$$\int_V \boldsymbol{\nabla} \bullet \mathbf{A}\, dV = \int_V (x + y + z)\ dxdydz\ ,$$

$$\int_V \nabla\bullet \mathbf{A}\, dV = \int_0^a x\, dx \int_0^b dy \int_0^c dz + \int_0^a x \int_0^b y\, dy \int_0^c dz + \int_0^a dx \int_0^b dy \int_0^c z\, dz$$

$$= \frac{a^2}{2} bc + a\frac{b^2}{2} c + ab\frac{c^2}{2} = \frac{V}{2}(a+b+c) \ .$$

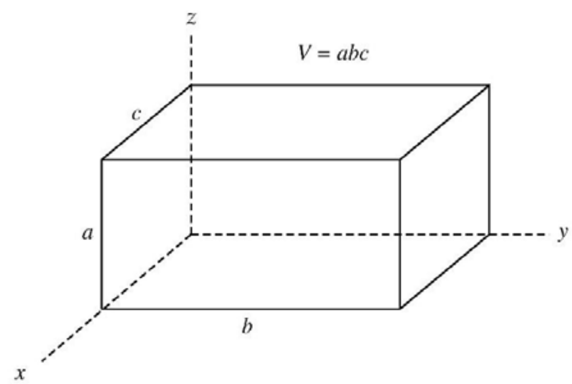


Figure 10.8. A parallelepiped with side lengths $a, b,$ and c.

The surface integral of the divergence theorem over the six faces (left-right pair, front-back pair and the top-bottom pair) of the parallelepiped is

$$\int_S \mathbf{A}\bullet \boldsymbol{n}\, dS = \int_{front} \mathbf{A}\bullet \boldsymbol{n}\, dS + \int_{back} \mathbf{A}\bullet \boldsymbol{n}\, dS + \int_{left} \mathbf{A}\bullet \boldsymbol{n}\, dS$$

$$+ \int_{right} \mathbf{A}\bullet \boldsymbol{n}\, dS + \int_{top} \mathbf{A}\bullet \boldsymbol{n}\, dS + \int_{bottom} \mathbf{A}\bullet \boldsymbol{n}\, dS$$

$$= \int_{front} \mathbf{A}\bullet(\mathbf{e_1}) dydz + \int_{back} \mathbf{A}\bullet(-\mathbf{e_1}) dydz + \int_{left} \mathbf{A}\bullet(-\mathbf{e_2}) dxdz \ .$$

$$+ \int_{right} \mathbf{A}\bullet(\mathbf{e_2})\, dxdz + \int_{top} \mathbf{A}\bullet(\mathbf{e_3})\, dxdy + \int_{bottom} \mathbf{A}\bullet(-\mathbf{e_3})\, dxdy$$

$$= \int_0^b \int_0^c a\, y\, dydz + \int\int (-0) dydz + \int\int (-0) dzdy$$

$$+ \int_0^a \int_0^c b\, z\, dxdz + \int_0^a \int_0^b c\, x\, dxdy + \int\int (-0) dxdy$$

$$= a\frac{b^2}{2} c + ab\frac{c^2}{2} + \frac{a^2}{2} bc = \frac{V}{2}(a+b+c) \ .$$

Thus, divergence theorem:

$$\int_{six\ faces} \mathbf{A} \bullet \boldsymbol{n}\, dS = \int_{parallelepped} (\boldsymbol{\nabla} \bullet \mathbf{A}) dV \ ,$$

is satisfied.

10.6 CIRCULATION OF A VECTOR FUNCTION: STOKE'S THEOREM

Consider an open surface S whose boundary is the closed loop L as in Figure 10.9. We define the $circulation$ of a vector function **A** in closed loop L as the line integral

$$\oint_L \mathbf{A} \bullet \mathbf{dr} \ . \qquad Eq. (10.14)$$

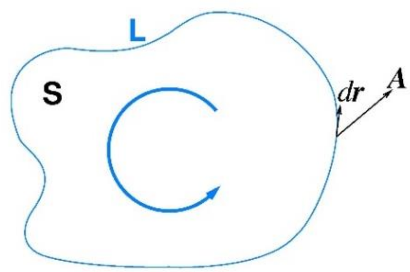


Figure 10.9. Circulation of a vector function **A** in loop L.

Here $\mathbf{dr}$ is everywhere tangential to the loop L. In general, the circulation of a vector function in a loop L has a very interesting property. Imagine that the area enclosed by loop L is broken into two smaller areas, with each area enclosed by a smaller loop. Then, the sum of circulations of a vector function **A** in two smaller loops (calculated in the same sense as in the original loop) is equal to the circulation of **A** in the large original loop L.

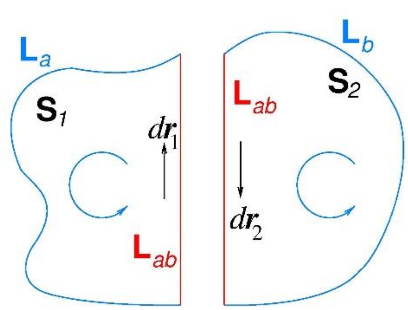


Figure 10.10. Breaking a loop L into two smaller pieces L_a and L_b.

In order to have a clear visualization, imagine that the large loop is made up of a blue color line as in Figure 10.10. This loop encloses an area S. Now, using a pair of scissors, we cut the area along the line L_{ab} to get two smaller areas S_1 and S_2. Also, during this cutting, the blue line is cut into two pieces L_a and L_b. Clearly, $S = S_1 + S_2$ and $L = L_a + L_b$. Area S_1 is enclosed by loop $L_1 = L_a + L_{ab}$ and area S_2 is enclosed by loop $L_2 = L_b + L_{ab}$. Now, circulation of **A** in loop L_1 is

$$\int_{L_a} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_1 \ ,$$

where $\mathbf{dr}_1$ is along the line of cut L_{ab}. The circulation of **A** in loop L_2 is

$$\int_{L_b} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_2 \ ,$$

where $\mathbf{dr}_2$ is along the line of cut L_{ab}. Thus, the sum of the circulations of **A** in loop L_1 and in loop L_2 becomes

$$\int_{L_a} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_1 + \int_{L_b} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_2 \ .$$

If the sense of circulations in the two smaller loops is the same, as shown in Figure 9.10, then $\mathbf{dr}_1 = -\mathbf{dr}_2$. So, the sum of circulations in loops L_1 and L_2 becomes

$$\int_{L_a} \mathbf{A} \bullet \mathbf{dr} + \int_{L_b} \mathbf{A} \bullet \mathbf{dr} = \int_{L_a + L_b} \mathbf{A} \bullet \mathbf{dr} = \int_{L} \mathbf{A} \bullet \mathbf{dr} \ ,$$

which is equal to the circulation of **A** in loop L. This result holds even when S_1 and S_2 are not *coplanar.*

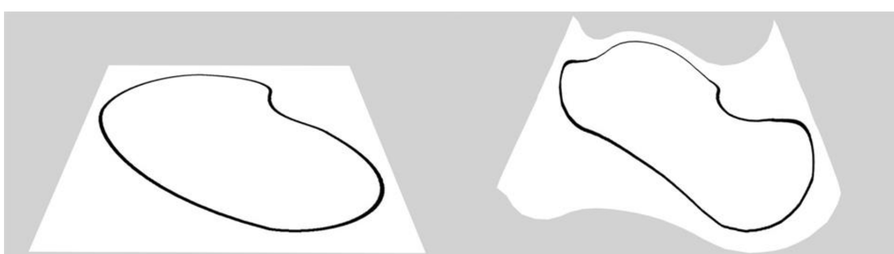


Figure 10.11. Planar versus nonplanar open surfaces.

Generalization is achieved by noting that loop L can enclose an infinite number of open surfaces. These surfaces need not be planar in shape and can be warped like a crumpled sheet of paper, as in Figure 10.11. Any one of these surfaces can be broken into a large number of smaller surface areas of any size and shape. Then the sum of circulations in loops surrounding these smaller surface areas is equal to the circulation in original loop L. For convenience, we take each of the smaller surface areas as a flat rectangle.

Circulation of A in the Loop Surrounding a Flat Rectangle

Now, we will determine the circulation of a vector function **A** in the loop surrounding a flat rectangle, in the $x-y$ plane, of side lengths Δx and Δy, as shown in Figure 10.12. The area of the loop is $\mathbf{dS} = \Delta x \Delta y\ \mathbf{e_3}$. the loop consists of four sides of the rectangle labeled 1, 2, 3, and 4. So, the circulation of **A** has four contributions,

$$\oint_{Rectangle} \mathbf{A}\bullet\mathbf{dr} = \int_1 A_x dx + \int_2 A_y dy - \int_3 A_x dx - \int_4 A_y dy \ ,$$

$$\cong A_x(1)\Delta x + A_y(2)\Delta y - A_x(3)\Delta x - A_y(4)\Delta x \ ,$$

$$= -[A_x(3) - A_x(1)]\Delta x + [A_y(2) - A_y(4)]\Delta y \ ,$$

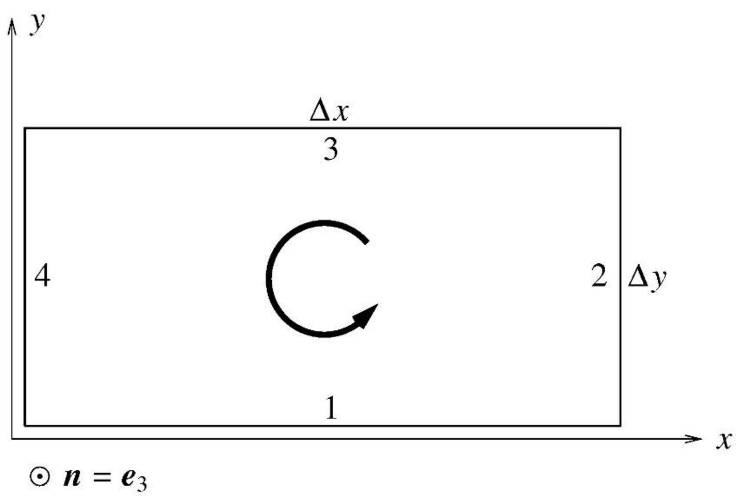


Figure 10.12. A rectangular loop in the $x-y$ plane.

or

$$\oint_{Rectangle} \mathbf{A}\bullet\mathbf{dr} = -\left[\frac{\partial A_x}{\partial y}\ \Delta y\right]\Delta x + \left[\frac{\partial A_y}{\partial x}\ \Delta x\right]\Delta y = (\boldsymbol{\nabla}\times\mathbf{A})_z\ \Delta x \Delta y = (\boldsymbol{\nabla}\times\mathbf{A})\bullet \boldsymbol{n}\ dS \ .$$

Now, going back to any general open surface, with loop L as its boundary, that is broken into a large number of smaller surface areas in the form of flat rectangles:

$$\oint_L \mathbf{A}\bullet\mathbf{dr} = \int_{Any\ open\ surface\ bounded\ by\ L} (\boldsymbol{\nabla}\times\mathbf{A}) \bullet \boldsymbol{n}\ dS \ . \qquad Eq.(10.15)$$

This is the statement of the Stoke’s theorem.

Example: Consider a vector function A $= xy\,\mathbf{e_1} + yz\,\mathbf{e_2} + zx\,\mathbf{e_3}$ and a square of side a in the y– z plane. Show that Stoke's theorem is satisfied by the vector function A and the square loop.

Solution:

For this vector function, the curl is

$$\nabla \times \mathbf{A} = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ xy & yz & zx \end{vmatrix}$$

$$= \mathbf{e_1}\left[\frac{\partial}{\partial y}(zx) - \frac{\partial}{\partial z}(yz)\right] + \mathbf{e_2}\left[\frac{\partial}{\partial z}(xy) - \frac{\partial}{\partial x}(zx)\right] + \mathbf{e_3}\left[\frac{\partial}{\partial x}(yz) - \frac{\partial}{\partial y}(xy)\right] ,$$

or

$$\nabla \times \mathbf{A} = -\mathbf{e_1}y - \mathbf{e_2}z - \mathbf{e_3}x \ .$$

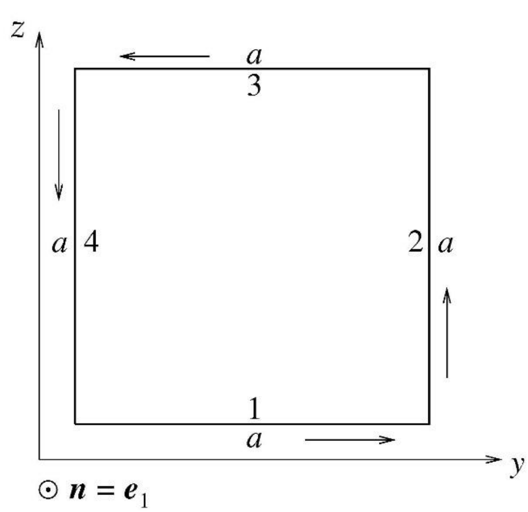


Figure 10.13. A square loop in the y– z plane.

Since the square lies in the y– z plane and we will traverse it in the counterclockwise sense, $\boldsymbol{n} = \mathbf{e_1}$. So,

$$\int_S (\nabla \times \mathbf{A}) \bullet \boldsymbol{n}\, dS = \int_0^a \int_0^a (-\mathbf{e_1}y - \mathbf{e_2}z - \mathbf{e_3}x) \bullet \mathbf{e_1}\, dy\, dz$$

$$= -\int_0^a y\, dy \int_0^a dz = -\frac{a^2}{2} \cdot a = -\frac{a^3}{2} \ .$$

Now, four sides of the square are labeled with numbers 1, 2, 3, and 4. Circulation of **A** in the square is

$$\int_{square} \mathbf{A} \bullet \mathbf{dr} = \int_{1+2+3+4} [xy\, dx + yz\, dy + zx\, dz] = 0 + \int_{1+2+3+4} y\, z\, dy + 0 \ ,$$

since $x = 0$ everywhere on the square in the $y\!-z$ plane. So, circulation becomes

$$\int_{square} \mathbf{A} \bullet \mathbf{dr} = \int_{1} (z = 0)\, ydy + \int_{2} yz\, (dy = 0) + \int_{3} (z = a)\, ydy + \int_{4} yz\, (dy = 0)$$

$$= 0 + 0 + a \int_{a}^{0} ydy + 0 = -\frac{a^3}{2} \ .$$

Thus, Stoke's theorem,

$$\oint_{Square\ loop\ L} \mathbf{A} \bullet \mathbf{dr} = \int_{Square\ surface\ bounded\ by\ L} (\boldsymbol{\nabla} \times \mathbf{A}) \bullet \boldsymbol{n}\, dS \ .$$

is satisfied.

PROBLEMS FOR CHAPTER 10

1. Determine gradient ($\boldsymbol{\nabla}$) and Laplacian (∇^2) of r and $1/r$ where $r = \sqrt{x^2 + y^2 + z^2}$.

2. Determine the divergence and curl of the vector function $\boldsymbol{F} = \frac{yz}{x}\,\mathbf{e_x} + \frac{zx}{y}\,\mathbf{e_y} + \frac{xy}{z}\,\mathbf{e_z}$.

3. If $\mathbf{r} = x\,\mathbf{e_x} + y\,\mathbf{e_y} + z\,\mathbf{e_z}$ is the position vector and $r = \sqrt{x^2 + y^2 + z^2}$ is its magnitude, then determine,

(a) $\boldsymbol{\nabla}\bullet(r\,\boldsymbol{r})$,

(b) $\boldsymbol{\nabla} \times (r\,\boldsymbol{r})$

(c) $\boldsymbol{\nabla}^2(r^2)$.

4. Consider a vector function **A** = $\ln(xyz)\,\mathbf{e_x}$ + $\ln(yz)\,\mathbf{e_y}$ + $\ln(z)\,\mathbf{e_z}$. For this vector determine

(a) $\boldsymbol{\nabla} \bullet \mathbf{A}$,

(b) $\boldsymbol{\nabla} \times \mathbf{A}$,

(c) $\nabla^2\mathbf{A}$,

(d) $\boldsymbol{\nabla}(\boldsymbol{\nabla} \bullet \mathbf{A})$,

(e) $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A})$.

(f) Using these results, verify $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \bullet \mathbf{A}) - \boldsymbol{\nabla}^2\mathbf{A}$.

5. Consider a vector function **A** = $(x^2 - y^2)\,\mathbf{e_x} + xyz\,\mathbf{e_y} - (x + y + z)\,\mathbf{e_z}$ and a cube bounded by the planes $x = 0, x = 1, y = 0, y = 1, z = 0$ and $z = 1$.

(a) Determine the volume integral $\int_V \boldsymbol{\nabla} \bullet \mathbf{A}\,\mathrm{dV}$, where V is the volume of the cube.

(b) Next determine the surface integral $\int_S \mathbf{A} \bullet \mathbf{n}\,dS$, where S is the surface area of the cube. Compare with the result of part (a)

6. Consider a vector function **A** = $(2x - y)\,\mathbf{e_x} + yz^2\,\mathbf{e_y} + y^2z\,\mathbf{e_z}$. Also, S is the flat surface area of a rectangle bounded by the lines $x = \pm 1$ and $y = \pm 2$ and C is its (rectangular) boundary in the $x - y$ plane.

(a) Determine the line integral $\int_C \mathbf{A} \cdot \mathbf{dr}$

(b) Next determine the surface integral $\int_S (\boldsymbol{\nabla} \times \mathbf{A}) \bullet \mathbf{n}\,dS$. Compare with the result of part (a).

7. **Biomedical Physics Application** Entomologists have known that fruit flies are attracted to sugary substances and feed on overripe fruit as well as on spilled sodas. The fruit flies approach the fruit along the direction in

which the smell of the decaying fruit is strongest. The concentration of the smell of the rotten fruit at a point (x, y, z), due to some overripe fruit at the origin (0, 0, 0), is,

$$C(x, y, z) = C_o \exp(-2\,x^2 - y^2 + \,z^2)\ .$$

Here x, y and z are in meters. If a fruit fly detects the presence of rotten food while it is at the point $(0.6\ m, 0.8\ m, 0.9\ m)$, determine the unit vector along the initial direction in which the fruit fly approaches the food.

Chapter 11: First Order Differential Equations

In our previous discussion of calculus, we were given a known function $F(x)$ and were asked to find either the derivative or the integral of this function. In some other situations, we are given the first, second or higher order derivative of an unknown function and are asked to figure out the function. That brings us to the domain of differential equations. Our aim in this chapter is to provide merely a flavor of differential equations, so we will confine our attention only to first order ordinary differential equations. The prolific topics related to differential equations, ordinary as well as partial, are covered in books which are entirely devoted to this subject.

11.1 A FIRST ORDER DIFFERENTIAL EQUATION

Consider the case where an unknown function $y(x)$, which is yet to be determined, satisfies,

$$\frac{dy}{dx} = F(x) \quad \text{or} \quad \frac{dy}{dx} = G(y) \quad \text{or} \quad \frac{dy}{dx} = H(x, y) \ , \qquad Eq. (11.1)$$

where the functional form of $F(x)$ or $G(y)$ or $H(x, y)$ is known. Such equations are called differential equations. The *order* of a differential equation is the order of highest derivative appearing in the equation. An $ordinary$ differential equation involves derivatives of a function of a single independent variable. All equations shown in Eq. (11.1) are first order ordinary differential equations. A $partial$ differential equation, on the other hand, involves derivative of a function of multiple variables. The wave equation, that we encountered in Chapter 1, serves as an example of a partial differential equation since it contains the following derivatives,

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \ .$$

The wave equation is a second order equation. When a differential equation for the function $y(x)$ is integrated, a constant of integration, C, will appear naturally as a part of integration. This constant will need to be evaluated using the constraints on the function y at some fixed value of variable x. Such constraining relationships are called the boundary conditions. If the variable in the differential equation is time, t, the boundary conditions are referred to as the initial conditions.

Several first order differential equations can be solved by direct integration. The following examples will illustrate this point.

Example: Solve the differential equation

$$\frac{dy}{dx} = \frac{x}{\sqrt{1 - x^2}} \qquad |x| \leq 1 \ , \qquad Eq. (11.2)$$

for the function $\boldsymbol{y(x)}$ with the boundary condition $\boldsymbol{y(1) = 0}$.

Solution: On integrating both sides of the differential equation (11.2) with respect to variable x, we get

$$y = \int \frac{xdx}{\sqrt{1-x^2}} + C \ .$$

To do the integral, substitute $1 - x^2 = u^2$, or $x\,dx = -u\,du$,

$$\int \frac{xdx}{\sqrt{1-x^2}} = -\int \frac{u\,du}{u} = -u = -\sqrt{1-x^2} \ .$$

So, $y = -\sqrt{1-x^2} + C$ is the solution of the differential equation and the boundary condition gives $C = 0$.

Example: Solve the differential equation for $\boldsymbol{y(x)}$,

$$\frac{dy}{dx} = 1 + y^2 \ , \qquad Eq. (11.3)$$

with boundary condition $\boldsymbol{y(0) = 0}$.

Solution: Since we have to solve for function y in terms of variable x, we separate out the x and y dependent parts in Eq. (11.3) and then integrate to get

$$\int \frac{dy}{1+y^2} = \int dx + C \ .$$

Or

$$\arctan y = x + C \ .$$

Or $y = \tan(x + C)$ is the solution of the differential equation. The boundary condition implies $C = 0$.

In some problems $\frac{dy}{dx}$ can be written in a separable form as

$$\frac{d}{dx} = g(x)h(y) \ .$$

Then,

$$\int \frac{dy}{h(y)} = \int g(x)dx + C \ .$$

Example: Solve the following differential equation for $y(x)$,

$$\frac{dy}{dx} = \frac{9\exp x}{y^2} \ , \qquad Eq.(11.4)$$

with $y(0) = 3$.

Solution: Write Eq. (11.4) in separable form and then integrate to get

$$\int y^2 dy = \int 9\exp x \, dx + C \ .$$

Or,

$$\frac{y^3}{3} = 9\exp x + C \ .$$

On substituting boundary condition, $C = 0$. Thus, the function $y(x)$ is,

$$y = 3\exp\left(\frac{x}{3}\right) \ .$$

11.2 INTEGRATING FACTOR

Some first order differential equations are of the form

$$a\frac{dy}{dx} + by = F(x) \ , \qquad Eq.(11.5a)$$

where a and b are constants. These equations are solved by introducing an $integrating\ factor$. The integrating factor for equations of this type is

$$\exp\left(\frac{b}{a}x\right) \ .$$

First, we rewrite Eq. (11.5a) after dividing by constant a,

$$\frac{dy}{dx} + \frac{b}{a}y = \frac{1}{a}F(x) \ . \qquad Eq.(11.5b)$$

On multiplying both sides of Eq. (11.5b) by the integrating factor, the left-hand side becomes an exact differential,

$$\frac{d}{dx}\left[y\exp\left(\frac{b}{a}x\right)\right] = \frac{\exp\left(\frac{b}{a}x\right)}{a}F(x) \ .$$

Integrate both sides of this equation with respect to x to obtain the solution

$$y\exp\left(\frac{b}{a}x\right) = \frac{1}{a}\int \exp\left(\frac{b}{a}x\right)\ F(x)dx + C\ , \qquad Eq.(11.6)$$

where C, the constant of integration, will be determined by the boundary conditions.

Example: $\boldsymbol{An\ R\text{–}L\ circuit}$. The circuit consisting of an inductor of inductance $\boldsymbol{L}$ and a resistor of resistance $\boldsymbol{R}$ is connected to a source of alternating current of frequency $\boldsymbol{\omega}$ and voltage $\boldsymbol{V_0\cos(\omega t)}$. Using Kirchhoff's rule in physics, the current $\boldsymbol{I(t)}$ in the circuit satisfies the first order differential equation

$$L\frac{dI}{dt} + RI = V_0\cos(\omega t) \ . \qquad Eq.(11.7a)$$

Determine the current $\boldsymbol{I(t)}$ as a function of time $\boldsymbol{t}$.

Solution: In this case the differential Eq. (11.7a) is of the same form as Eq. (11.5a). So, the integrating factor is $\exp\left(\frac{R}{L}t\right)$. On multiplying the differential Eq. (11.7a) by the integrating factor and dividing by L, it becomes

$$\frac{d}{dt}\left[\exp\left(\frac{R}{L}t\right)\ I(t)\right] = \frac{V_0}{L}\exp\left(\frac{R}{L}t\right)\cos(\omega t) \ .$$

On integrating both sides of this equation with respect to t using the standard integral,

$$\int \exp(at)\cos(bt)\ dt\ = \frac{\exp(at)}{[a^2+b^2]}\ [\mathrm{a}\cos(bt) + b\sin(bt)]\ ,$$

we obtain

$$\exp\left(\frac{R}{L}t\right)I(t) = \frac{V_0}{L}\frac{\exp\left(\frac{Rt}{L}\right)}{\left[\left(\frac{R}{L}\right)^2 + \omega^2\right]}\left[\frac{R}{L}\cos(\omega t) + \omega\sin(\omega t)\right] + C \ .$$

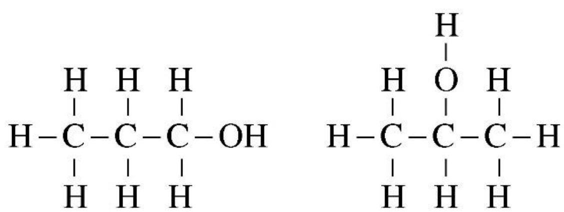
The constant of integration C is determined from the initial condition. If, for example, at $t = 0, I = 0$, then

$$C = -\frac{V_0}{L}\frac{R}{L}\frac{1}{\left[\left(\frac{R}{L}\right)^2 + \omega^2\right]} \ ,$$

so that, finally,

$$I(t) = \frac{V_0}{L}\frac{1}{\left[\left(\frac{R}{L}\right)^2 + \omega^2\right]}\left[\frac{R}{L}\cos(\omega t) + \omega\sin(\omega t)\right] - \frac{V_0}{L}\frac{R}{L}\frac{\exp\left(-\frac{Rt}{L}\right)}{\left[\left(\frac{R}{L}\right)^2 + \omega^2\right]} \quad . \qquad Eq.(11.7b)$$

Example: $\boldsymbol{Molecular\ Isomerization}$**. Some molecules with the same number of identical atoms can exist in several stable configurations or arrangements of their component atoms. Such configurations are called different isomers of the same molecule. For example, the molecule of propane-based alcohol, commonly called propanol, has two common isomers. The first isomer, 1-propanol, a primary alcohol, and the second isomer, 2-propanol, a secondary alcohol, have structural formulas as shown in the Figure 11.1.**



1-Propanol 2-Propanol

Figure 11.1. Two isomers of propanol.

The propanol molecule is of significant importance in biological physics. Determine the ratio of populations of two isomers of the propanol molecule in equilibrium.

Solution: A typical molecule isomerizes (or flips) between two configurations M_1 and M_2. The probability that the molecule in configuration M_2 flips into M_1 in time Δt is proportional to Δt—the longer the time period Δt, the higher the probability. Thus, the flipping probability is $k_+\Delta t$. The constant of proportionality k_+, which represents the probability per unit time, is called the reaction rate. Stated differently, the rate of change of a single molecule from configuration M_2 to configuration M_1 is k_+. If there are N_2 molecules in configuration M_2 then rate of change of all molecules from M_2 to M_1 is k_+N_2. Similarly, probability that the molecule in configuration M_1 flips into M_2 in time Δt is $k_-\Delta t$, where k_- is the reaction rate for the reverse isomerization. The complete isomerization process can be expressed as

$$M_2 \underset{k_-}{\overset{k_+}{\rightleftharpoons}} M_1$$

Note that in this isomerization process a single molecule of first configuration is converted into a single molecule of second configuration. Thus, the total number of molecules, $N_{tot} = N_1 + N_2$, stays the same at all times.

Assume that in the beginning, at $t = 0, N_2 = N_{tot}$ and $N_1 = 0$, which is the initial condition. The rates of change of N_1, the number of molecules with configuration M_1, and of N_2, the number of molecules with configuration M_2, are

$$\frac{dN_2}{dt} = -k_+ N_2 + k_- N_1 = -(k_+ + k_-)N_2 + k_-\, N_{tot} \ ,$$

$$\frac{dN_1}{dt} = +k_+ N_2 - k_- N_1 = -(k_+ + k_-)N_1 + k_+\, N_{tot} = -\frac{dN_2}{dt} \ .$$

Thus, the first-order differential equation satisfied by $N_2(t)$ is

$$\frac{dN_2}{dt} + (k_+ + k_-)N_2 = k_-\, N_{tot} \ , \qquad Eq.(11.8a)$$

which is of the same form as Eq. (11.5b). The integrating factor of this equation is $\exp[(k_+ + k_-)t]$. Therefore,

$$\frac{d}{dt}\{N_2 \exp[(k_+ + k_-)t]\} = k_- N_{tot} \exp[(k_+ + k_-)t] \ ,$$

which on integration gives

$$N_2(t) = \frac{k_- N_{tot}}{(k_+ + k_-)} + C_2 \exp[-(k_+ + k_-)t] \ . \qquad Eq.(11.8b)$$

Here C_2 is the constant of integration. Similarly,

$$N_1(t) = \frac{k_+ N_{tot}}{(k_+ + k_-)} + C_1 \exp[-(k_+ + k_-)t] \ , \qquad Eq.(11.8c)$$

where C_1 is the constant of integration. Using the initial condition, we get

$$C_1 = -\,C_2 = -\frac{k_+ N_{tot}}{(k_+ + k_-)} \ .$$

Thus

$$N_2(t) = \frac{k_- N_{tot}}{(k_+ + k_-)} + \frac{k_+ N_{tot}}{(k_+ + k_-)} \exp[-(k_+ + k_-)t]$$

and

$$N_1(t) = \frac{k_+ N_{tot}}{(k_+ + k_-)} [1 - \exp[-(k_+ + k_-)t]] \ .$$

Note $N_1(t) + N_2(t) = N_{tot}$ at all times. Also, $N_2(0) = N_{tot}$, $N_1(0) = 0$. Finally, after a sufficiently long time $[t \to \infty]$, when equilibrium is attained,

$$N_2(\infty) = \frac{k_- N_{tot}}{(k_+ + k_-)} \ , \qquad N_1(\infty) = \frac{k_+ N_{tot}}{(k_+ + k_-)} \ .$$

Thus, at equilibrium

$$N_{1,eq} = \frac{k_+}{k_+ + k_-} N_{tot}, \qquad N_{2,eq} = \frac{k_-}{k_+ + k_-} N_{tot} \text{ and } \frac{dN_2}{dt} = 0 \text{ as well as } \frac{dN_1}{dt} = 0 \ .$$

Also, note that at equilibrium,

$$\frac{N_{1,eq}}{N_{2,eq}} = \frac{k_+}{k_-} \ ,$$

that is, the ratio of population of two isomers is the same as the ratio of two rates.

Example: $\boldsymbol{Newton's\ Law\ of\ Cooling}$. It is a common experience that if a glass of water at room temperature is placed outdoors on a hot day in summer, the temperature of water starts rising quickly. Similarly, if the same glass of water at room temperature is placed outdoors during a cold wintry night when the temperature has plunged due to the polar vortex, the temperature of water in the glass starts falling rapidly. Newton formulated the Law of Cooling by stating that the rate of change of temperature of a body, at any time $\boldsymbol{t}$, is proportional to the difference between the constant temperature of the body's surroundings, $\boldsymbol{T_s}$, and the temperature of the body, $\boldsymbol{T(t)}$, at time $\boldsymbol{t}$. Mathematically,

$$\boldsymbol{\frac{dT(t)}{dt} = C\,[T_s - T(t)]} \ ,$$

where $\boldsymbol{C\ (C > 0)}$ is the constant of proportionality. This is a first-order differential equation whose solution provides the temperature of a body, versus the constant temperature of its surroundings, as a function of time. Determine an explicit expression for the temperature of the body, $T(t)$, **at any time $\boldsymbol{t}$.**

Solution: The differential equation to be solved is

$$\frac{dT(t)}{dt} + C\,T(t) = C\,T_s \ . \qquad Eq. (11.9a)$$

If we define a new function, $u(t)$, as $u(t) = T(t) - T_s$, then

$$\frac{du(t)}{dt} = -C\,u(t) \quad \text{or} \quad \frac{du(t)}{u} = -C\,dt \ . \qquad Eq. (11.9b)$$

A straightforward integration gives

$$\ln u(t) = -C\ t + k\ \ ,$$

where k is the constant of integration. This equation can be rewritten as

$$u(t) = T(t) - T_s = \exp(-C\ t + k)\ \ ,$$

and can be rearranged as

$$T(t) = T_s + \exp(-C\ t + k)\ \ .$$

The constant of integration, k, can be obtained by setting $t = 0$, to get

$$\exp(k) = T(0) - T_s\ \ ,$$

so that, finally,

$$T(t) = T_s + [T(0) - T_s]\exp(-C\ t)\ \ . \qquad Eq. (11.9c)$$

Note that at a much later time, that is, $t \to \infty$, $T(\infty) = T_s$ and the temperature of the body becomes equal to the temperature of its surroundings, as expected.

11.3 APPLICATIONS OF NUCLEAR PHYSICS

Without any doubt, nuclear physics has made tremendous contributions to modern medicine. Radiology and nuclear medicine involve the use of radioactive nuclei to diagnose, evaluate, and treat various diseases. The nucleus of any atom contains protons and neutrons. A single nuclear species, having unique values of the number of protons and number of neutrons, is called a nuclide. About 80% of the nuclides occurring on the Earth are stable nuclides; that is, they do not disintegrate, or decay, into other nuclides. On the other hand, about 20% of the nuclides are unstable, or radioactive, nuclides which can decay into other stable or unstable nuclides. In a sample of radioactive material, the number of radioactive nuclei continues to decrease with time as some of them disintegrate into other nuclides. In a radioactive transformation, the disintegrating nuclides are called $parent$ nuclides and they decay into so-called $daughter$ nuclides. Generally, the rate at which the parent nuclides decay varies from nuclide to nuclide. If $N(t)$ is the number of radioactive nuclides in a sample, then the number dN of nuclides that will undergo disintegration in time dt will be proportional both to dt as well as to $N(t)$. Saying this differently, the longer the time period dt, the more decays occur, and the larger the number N of nuclides available to decay, the more decays occur. Thus,

$$dN = -\lambda\ N\ dt\ \ ,$$

where the negative sign signifies a decrease in the number N. The constant of proportionality λ is called the decay constant and it is different for different nuclides. Therefore,

$$\frac{dN}{dt} + \lambda N = 0 \ . \qquad Eq.(11.10)$$

Example: $\boldsymbol{Radioactivity}$**. In a sample of radioactive material, the number of nuclides at $\boldsymbol{t = 0}$ is $\boldsymbol{N(0)}$. Determine $\boldsymbol{N(t)}$, the number of radioactive nuclides later at time $\boldsymbol{t}$. Also, determine the time at which the number of radioactive nuclides has decreased to half of the number at $\boldsymbol{t = 0}$. This time is called the half-life, $\boldsymbol{T_{1/2}}$, of the radioactive sample.**

Solution: The first order differential equation satisfied by $N(t)$ is

$$\frac{dN}{dt} = -\lambda N \ ,$$

which can be rewritten as

$$\frac{dN}{N} = -\lambda \, dt \ .$$

A simple integration of this equation gives

$$\ln N(t) = -\lambda t + C \ .$$

Here C is the constant of integration. Its value can be obtained using the initial condition that at $t = 0, N = N(0)$. So, $C = \ln N(0)$. Therefore,

$$\ln N = -\lambda t + \ln N(0) \ ,$$

or

$$N(t) = N(0) \exp(-\lambda t) \ . \qquad Eq.(11.11)$$

To determine the half-life, note that at $t = T_{1/2}$, we have $N = N(0)/2$, so that

$$\frac{1}{2} = \exp(-\lambda T_{1/2}) \ ,$$

or

$$T_{1/2} = \frac{ln\, 2}{\lambda} \ . \qquad Eq.(11.12)$$

Thus, starting from the original number $N(0)$, the number of remaining radioactive nuclides after each successive time interval of $T_{1/2}$ are $\frac{N(0)}{2}, \frac{N(0)}{4}, \frac{N(0)}{8}, \ldots$ and so on.

Example: $\boldsymbol{Successive\ Radioactivity}$. A radioactive nuclide $\boldsymbol{A}$ decays into a daughter nuclide $\boldsymbol{B}$ with a decay constant of $\boldsymbol{\lambda_A}$. The daughter nuclide $\boldsymbol{B}$ itself is radioactive and it decays into a stable nuclide $\boldsymbol{C}$ with a decay constant of $\boldsymbol{\lambda_B}$. Schematically, the nuclear reaction can be expressed as

$$\boldsymbol{A \xrightarrow{\lambda_A} B \xrightarrow{\lambda_B} C} \ .$$

If at the beginning, that is, at $\boldsymbol{t = 0}$, the number of nuclides is, $\boldsymbol{N_A(0) = N_0}$ and $\boldsymbol{N_B(0) = 0}$, determine the number of nuclides later at time $\boldsymbol{t}$.

Solution: Following the previous example, since the parent A decays only into the daughter B, the first order differential equation for $N_A(t)$ and its solution are

$$\frac{dN_A}{dt} = -\lambda_A\, N_A \ , \qquad Eq. (11.13a)$$

and

$$N_A(t) = N_0 \exp(-\lambda_A\, t) \ . \qquad Eq. (11.13b)$$

The daughter B is produced by decay of parent A and is lost by its own decay into C. So, the first order differential equation for $N_B(t)$ is

$$\frac{dN_B(t)}{dt} = \lambda_A\, N_A - \lambda_B\, N_B \ . \qquad Eq. (11.14)$$

The first term on the right-hand side represents the rate of production of B and the second term represents the rate of decay of B. Substituting $N_A(t)$ from Eq. (11.13b) into the right-hand side of Eq. (11.14), we get

$$\frac{dN_B(t)}{dt} + \lambda_B\, N_B = \lambda_A\, N_0 \exp(-\lambda_A\, t) \ . \qquad Eq. (11.15)$$

This differential equation is of the same form as Eq. (11.5b). In this case, the integrating factor is $\exp(\lambda_B\, t)$ and the solution of Eq. (11.15) is same as in Eq. (11.6), namely,

$$N_B(t)\ \exp(\lambda_B\, t) = \lambda_A\, N_0 \int \exp[(\lambda_B - \lambda_A)t]\ dt + C = \frac{\lambda_A\, N_0}{\lambda_B - \lambda_A} \exp[(\lambda_B - \lambda_A)t] + C \ ,$$

where C is the constant of integration. Using the initial condition for $N_B(t)$ at $t = 0$,

$$C = -\frac{\lambda_A \, N_0}{\lambda_B - \lambda_A} \; ,$$

so that

$$N_B(t) = \frac{\lambda_A \, N_0}{\lambda_B - \lambda_A} \, [\exp(-\lambda_A \, t) - \exp(-\lambda_B \, t)] \; . \qquad Eq. (11.16)$$

Eq. (11.16) provides the number of daughter nuclides as a function of time. On comparing the half-life of parent A with the half-life of daughter B, we arrive at two conclusions. First, if the half-life of the parent is much larger than the half-life of the daughter, it implies, using Eq. (11.12), that λ_A is much smaller than λ_B. Then, $\lambda_B - \lambda_A \approx \lambda_B$ and $\exp(-\lambda_A \, t)$ is much larger than $\exp(-\lambda_B \, t)$, so that

$$N_B(t) = \frac{\lambda_A \, N_0}{\lambda_B} \, \exp(-\lambda_A \, t) \; .$$

Using Eq. (11.13b), it means that

$$\lambda_A N_A(t) = \lambda_B N_B(t) \qquad \text{or} \qquad \frac{dN_A}{dt} = \frac{dN_B}{dt} \; . \qquad Eq. (11.17)$$

So, after a sufficiently long time, both the parent and daughter nuclides disintegrate at the same rate. Second, we conclude that if the half-life of the parent is much shorter than the half-life of the daughter, no equilibrium can be reached since the parent nuclide will disintegrate quickly, leaving behind a much longer-lived daughter nuclide.

PROBLEMS FOR CHAPTER 11

1. Determine the function $u(t)$ satisfying the differential equation

$$\frac{du}{dt} = \frac{t^2 + 1}{\cos u}$$

and $u(0) = \pi/2$.

2. Determine the function $y(x)$ satisfying

$$\frac{dy}{dx} = \frac{1}{2}\frac{y}{x}$$

with $y(4) = 1$.

3. On a calculus exam, students were asked to use the product rule to determine the derivative of a function $F(x) = f(x)\, g(x)$. A student forgot the product rule but mistakenly assumed that product rule was,

$$\frac{dF}{dx} = \frac{df(x)}{dx}\frac{dg(x)}{dx} \; .$$

Luckily, the student got the correct answer.

If one of the functions in the product was, $g(x) = \exp(3x)$, then set up a a first-order differential equation for the other function $f(x)$.

Solve the differential equation to obtain the second function $f(x)$.

4. **Biomedical Physics Application**. A glucose solution is administered intravenously into the bloodstream at a constant rate *r*. As the glucose is added, it is converted into other substances and removed from the bloodstream at a rate that is proportional to the concentration at that time. Thus, a model for the concentration *C*(*t*) of the glucose solution in the bloodstream is

$$\frac{dC}{dt} = r - kC$$

where *k* is a positive constant.

(a) Determine the concentration at any time *t* if the initial concentration at *t* = 0 is C_o.

(b) Determine the value of the concentration at sufficiently long time $(t \to \infty)$ after the administration begins. How will this value change if C_o is doubled?

5. **Biomedical Physics Application**. A device called a pacemaker can be implanted inside the human body to monitor the heart activity. The pacemaker is essentially a capacitor, with capacitance C, that stores some charge

at a certain voltage V. If the heart skips a beat, the pacemaker releases the stored energy through a lead wire of resistance R to get the heart back to beating normally. In a pacemaker circuit, the charge, $Q(t)$, on the plates of the capacitor varies with time as

$$R\frac{dQ}{dt} + \frac{Q}{C} = V\exp(-\lambda t)$$

where $\lambda = 2/(RC)$. If at $t = 0, Q(t = 0) = CV$, solve the above differential equation to obtain $Q(t)$ as a function of time, t.

6. **Biomedical Physics Application**. On a clear wintry night, at midnight, when the outdoor temperature is $0°C$, the body of an old man was discovered on a park bench. The temperature of the body at the time of discovery was determined to be $28°C$. The body temperature of a normal healthy adult is $37°C$. It took an hour to take the body from the park to the nearby hospital. At 1:00 AM, the pathologist doing the autopsy finds the temperature of the body to be $24°C$ and declares the cause of death to be hypothermia. Based on this information, determine the time of death for the old man.

7. **Biomedical Physics Application**. In medical research, radioactive materials are helpful in identifying and treating diseases in a noninvasive manner. In a chain of successive radioactivity, a radioactive nuclide A decays into a daughter nuclide B and the daughter nuclide B decays into a stable nuclide C. The decay constants (or half-lives) of both the parent nuclide and the daughter nuclide are equal. If at $t = 0$ the number of nuclides is $N_A(0) = N_0$ and $N_B(0) = 0$, then determine the time when the population of the daughter nuclide, B, is largest. Express this time in terms of the half-life of either nuclide.

Chapter 12: Diffusion Equation

Let us start our discussion of diffusion through some everyday observations. Imagine placing a sugar cube in a glass of water. After a few hours, you notice that sugar cube is no longer there; it has become part of the water and now the water tastes sweet. The sugar cube has dissolved due to diffusion. This process is similar to the process of gas exchange in multicellular organisms. Other examples of diffusion include noticing the smell of spray perfume or cigarette smoke even when the person using perfume or smoking is far away from the observer.

12.1 FICK'S LAW

The process of diffusion refers to the movement of a substance, or more accurately its atoms and molecules, from a region of higher concentration to a region of lower concentration. Diffusion is important in biology in transporting biomolecules from one location to another. Thermal energy can spontaneously move the molecules, and this spontaneous motion is governed by the diffusion equation. Diffusion plays a role in cellular transport and membrane permeability, and in determining conformation of certain large biomolecules, along with other applications. The rate of enzymatic reaction in many cases is also diffusion-limited as it determines the time it will take the reactant molecules to come closer and interact with each other.

To begin with, let us consider the flow of diffusing particles back and forth in one dimension (say, along the x-axis) across a surface S during a time t. The number of particles flowing across is directly proportional to time t and to area S. We define net flux, J_x, as the number of particles flowing per unit area per unit time. It was empirically noted by Fick that the net flux is proportional to the rate at which the concentration of particles varies with distance (that is, the gradient of the concentration). In other words, Fick's law states

$$J_x = -D\frac{\partial C}{\partial x} \ ,$$

where C is the concentration (that is, number of particles per unit volume). The negative sign is indicative of the direction of the flow of particles. The direction of flux, J_x, is the direction of increasing x and of decreasing concentration, C. The constant of proportionality D is called the diffusion constant and it has dimensions of $[\mathrm{L}^2/T]$. Similarly, the flow of particles along y and z directions satisfy

$$J_y = -D\frac{\partial C}{\partial y} \ ,$$

$$J_z = -D\frac{\partial C}{\partial z} \ .$$

So, in three dimensions, we can write Fick's law as

$$\boldsymbol{J} = -D\,\boldsymbol{\nabla} C \ . \qquad Eq. (12.1)$$

Equation of Continuity

In order to study the diffusion of particles in three-dimensional space, let us break the whole space into tiny boxes. Over time, as particles in the whole space diffuse from one box to another, the total number of particles in the whole space remains fixed. In other words,

$$N = \int_{whole\ space} C(x, y, z, t)dxdydz \qquad Eq. (12.2)$$

is constant, independent of time.

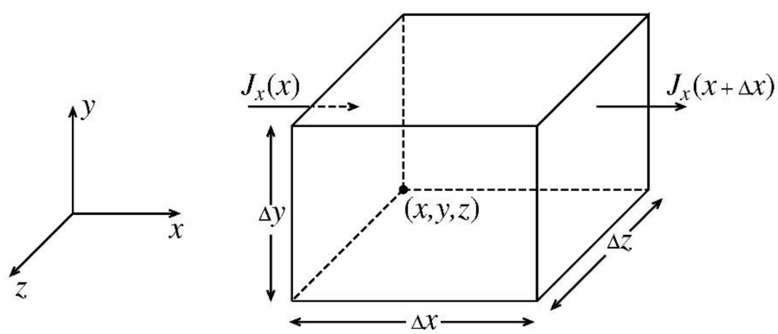


Figure 12.1. Particles diffusing into and out of a box.

Let us consider flow of particles through a box of sides $\Delta x, \Delta y$ and Δz between times t and $t + \Delta t$. The volume of this box is $\Delta V = \Delta x \Delta y \Delta z$. The concentration of particles in the box at time t is $C(x, y, z, t)$ and the number of particles in the box at time t is $n(t) = C(x, y, z, t)\Delta V$. Similarly, the concentration of particles in the box at a later time $t + \Delta t$ is $C(x, y, z, t + \Delta t)$ and the number of particles in the box at later time $t + \Delta t$ is $n(t + \Delta t) = C(x, y, z, t + \Delta t)\Delta V$. Now using the definition of flux as the particles flowing per unit area per unit time, we get

$$[J_x(x + \Delta x) - J_x(x)]\Delta t\Delta y\Delta z + [J_y(y + \Delta y) - J_y(y)]\Delta t\Delta x\Delta z + [J_z(z + \Delta z) - J_z(z)]\Delta t\Delta x\Delta y$$
$$= -\{n(t + \Delta t) - n(t)\} \ .$$

The negative sign on the right-hand side signifies that if more particles diffuse out of the box than diffuse into the box, then the number of particles in the box at the later time will be less than at the earlier time. Similarly, if more particles diffuse into the box than out of the box, then the number of particles in the box at the later time will be more than at the earlier time. Now, keeping only first-order terms in the difference of fluxes, we get

$$\left[\left(\frac{\partial J_x}{\partial x}\right)\Delta x\right]\Delta t\Delta y\Delta z + \left[\left(\frac{\partial J_y}{\partial y}\right)\Delta y\right]\Delta t\Delta x\Delta z + \left[\left(\frac{\partial J_z}{\partial z}\right)\Delta z\right]\Delta t\Delta x\Delta y = -\{C(x,y,z,t+\Delta t) - C(x,y,z,t)\}\Delta V \ .$$

On dividing both sides by Δt and ΔV, we get

$$\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} = -\frac{C(x,y,z,t+\Delta t) - C(x,y,z,t)}{\Delta t} = -\frac{\partial C}{\partial t} \ ,$$

or

$$\boldsymbol{\nabla}\cdot\boldsymbol{J} = -\frac{\partial C}{\partial t} \ . \qquad Eq.(12.3)$$

This is the equation of continuity. Physically, it represents the conservation of total number of particles, namely N, independent of time, as in Eq (12.2).

12.2 DIFFUSION EQUATION IN THREE DIMENSIONS

Combining the equation of continuity with Fick's law gives the diffusion equation,

$$D\,\boldsymbol{\nabla}^2 C = D\left[\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2}\right] = \frac{\partial C}{\partial t} \ . \qquad Eq.(12.4)$$

This is a time-dependent second order partial differential equation. Solving such an equation is beyond the scope of this book. The solution of the diffusion equation provides the time dependence of the concentration of the particles at any point in space.

Solution of Diffusion Equation

The solution of the diffusion equation, after solving the second order partial differential equation, is

$$C(x,y,z,t) = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3}\exp\left[-\frac{x^2+y^2+z^2}{4Dt}\right] \ . \qquad Eq.(12.5a)$$

Let us verify that this indeed is the solution of the diffusion equation. The first and second derivatives of C with respect to x are

$$\frac{\partial C}{\partial x} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3}\exp\left[-\frac{x^2+y^2+z^2}{4Dt}\right]\left(-\frac{x}{2Dt}\right) \ ,$$

$$\frac{\partial^2 C}{\partial x^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3}\left\{\exp\left[-\frac{x^2+y^2+z^2}{4Dt}\right]\left(-\frac{x}{2Dt}\right)^2 + \exp[\cdots]\left(-\frac{1}{2Dt}\right)\right\} \ ,$$

or, with $r^2 = x^2 + y^2 + z^2$,

$$\frac{\partial^2 C}{\partial x^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{x^2}{(2Dt)^2} - \frac{1}{2Dt}\right\} .$$

Similarly, the second derivatives with respect to variables y and z are

$$\frac{\partial^2 C}{\partial y^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{y^2}{(2Dt)^2} - \frac{1}{2Dt}\right\} ,$$

and

$$\frac{\partial^2 C}{\partial z^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{z^2}{(2Dt)^2} - \frac{1}{2Dt}\right\} .$$

Thus, combining all the derivatives together,

$$\boldsymbol{\nabla}^2 C = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{r^2}{(2Dt)^2} - \frac{3}{2Dt}\right\} .$$

Also, the derivative of C with respect to time, t, is

$$\frac{\partial C}{\partial t} = \exp\left[-\frac{r^2}{4Dt}\right]\frac{N}{\left(\sqrt{4\pi D}\right)^3}\left\{\frac{1}{\left(\sqrt{t}\right)^3}\frac{r^2}{4Dt^2} - \frac{3}{2}\frac{1}{\left(\sqrt{t}\right)^5}\right\}$$

$$= \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{r^2}{4Dt^2} - \frac{3}{2t}\right\} .$$

On placing all these derivatives in the diffusion equation, Eq. (12.4), we note that the equation is indeed satisfied. Notice that the concentration $C(x, y, z, t)$ of diffusing particles, Eq. (12.5a), is a spherically symmetric function, namely, it can be rewritten as

$$C(r,t) = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right] . \qquad Eq.(12.5b)$$

Here r is the distance from the point of release, or the source point, of diffusing particles. So, starting at the source, the diffusing particles spread out in a radial manner if there are no extenuating conditions. As time t increases, the spreading out of the particles increases the spatial extent of diffusion. However, the total number of particles in the whole space stays the same, namely, N. Explicitly,

$$I = \int_{whole\ space} C(x,y,z,t)\ dxdydz = \int_{whole\ space} C(r,t)\ 4\pi r^2 dr = \frac{4\ \pi\ N}{\left(\sqrt{4\pi Dt}\right)^3} \int_0^\infty \exp\left[-\frac{r^2}{4Dt}\right] r^2 dr\ .$$

Make a change of variables from r to $u = r/\sqrt{4Dt}$, so that

$$I = \frac{4\ \pi\ N}{\left(\sqrt{4\pi Dt}\right)^3}\ (4Dt)^{\frac{3}{2}} \int_{-\infty}^{\infty} \exp\left(-u^2\right) u^2\ du = \frac{4\ \pi\ N}{\pi^{\frac{3}{2}}} \frac{\sqrt{\pi}}{4} = N\ ,$$

independent of time, t. In the last step we used the integral I_g^2 [see, Eq. (2.16)]. The quantity $\sqrt{4Dt}$ is called the diffusion length, L_D.

12.3 DIFFUSION EQUATION IN ONE DIMENSION

For diffusion of particles in one dimension only (say, along the x-axis), the function describing the concentration of particles (that is, particles per unit length) at location x, at time t, is

$$C(x,t) = \frac{N}{\sqrt{\pi}\ L_D} \exp\left[-\frac{x^2}{L_D^2}\right] = C(0,t) \exp\left[-\frac{x^2}{L_D^2}\right]\ . \qquad Eq. (12.6)$$

This is a bell-shaped Gaussian function that we encountered previously in Figure 4.5 and its related discussion. The Gaussian function of Eq. (12.6) has its maximum, central, value of $C_{max}(t) = C(0,t) = \frac{N}{\sqrt{\pi}\ L_D}$. Just as in the three-dimensional case, the total number of diffusing particles N does not change with time for the one-dimensional case of Gaussian concentration. Explicitly,

$$\int_{-\infty}^{\infty} C(x,t)\ dx = \frac{N}{\sqrt{\pi}\ L_D} \int_{-\infty}^{\infty} \exp\left[-\frac{x^2}{L_D^2}\right] dx\ ,$$

or, on changing the variables from x to $u = x/L_D$,

$$\int_{-\infty}^{\infty} C(x,t)\ dx = \frac{N}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp[-u^2]\ d = \frac{N}{\sqrt{\pi}}\sqrt{\pi} = N\ ,$$

independent of time, t. The maximum value of the concentration function C_{max} is a function of time and it decreases with time. So, to keep N fixed the concentration spreads out in space. The extent of values of x, namely Δx, over which the function $C(x,t)$ is appreciable is given, at any time, by the full width of the function at half of its maximum value (FWHM). Thus,

$$\Delta x = FWHM = 2\sqrt{\ln 2}\ L_D\ . \qquad Eq. (12.7)$$

Note L_D increases with time; that is, the extent to which certain particles diffuse in time t is proportional to $\sqrt{t}$. Also, the time dependence of the concentration

$$C(x,t) = \frac{N}{\sqrt{\pi}\, L_D} \exp\left[-\frac{x^2}{L_D^2}\right] ,$$

comes through the time dependence of the diffusion length $L_D = \sqrt{4Dt}$. For $t \to 0$, the concentration of the diffusing particles is represented by a Dirac Delta function; that is, $C(x,0) = N\,\delta(x)$, using $Eq. (I.5b)$.

12.4 FINAL COMMENTS

We note in passing that the diffusion equation described here is same as the heat equation in thermal physics. That is not surprising since the heat equation describes diffusion of heat in a solid. In the thermal case, the concentration $C(\boldsymbol{r},t)$ is replaced by temperature, $T(\boldsymbol{r},t)$, at any point; thermal energy diffuses from a region of higher temperature to a region of lower temperature.

It is worth pointing out that the process of diffusion is also very important in industrial settings, such as in metallurgy. For example, to convert ordinary iron into steel we must add carbon to it, a process called carburizing, which hardens the surface of iron. Parts of iron in a high temperature furnace are placed in contact with carbon-rich gases and carbon diffuses into the parts, following the diffusion equation, turning iron into hardened steel. More generally, the process of diffusion coating, in which some elements are diffused onto the surface of metals to improve their hardness and corrosion properties, has been used in nonstick Teflon coatings of kitchen pots and pans.

PROBLEMS FOR CHAPTER 12

1. Consider a diffusing gas whose source is located at $r = 0$ and its diffusion constant D is $4 \times 10^{-4}\ cm^2/s$. At $t = 0$, the source spews out 1000 particles localized in the air. Plot the diffusion length L_D as a function of time t. What is the rate of growth of the horizontal area $(A = \pi L_D^2)$ filled by the diffusing particles?

2. Again, consider the diffusing gas of 1000 particles and the diffusion constant of $D = 4\ \times 10^{-4}\ cm^2/s$. For diffusion in one dimension and the source located at $x = 0$, plot the concentration $C(x,t)$ as a function of x for $t = 1\ s, 4\ s \text{ and } 9\ s$. What is the area under the $C(x,t)$ versus t curve for each value of time t?

3. **Biomedical Physics Application**. A truck carrying hydrogen fluoride, a toxic gas, gets into an accident on the road, causing leakage of about 5,000 molecules of the gas. If the diffusion constant of hydrogen fluoride is $1.7\ X\ 10^{-5}\ cm^2/s$, determine the concentrations of the toxic molecules at the site of the accident 6, 12, 18, and 24 hours after the accident.

4. **Biomedical Physics Application**. A researcher in a virology laboratory is working with a deadly virus in a sealed test tube. The laboratory is in the shape of a cube of side length $12\ m$. The researcher is at the center of the floor of the laboratory when the test tube is accidently dropped, spilling the deadly virus. As a result, the laboratory will need to be sealed and later clinically cleaned. If the diffusion constant of the virus is $9\ X\ 10^{-3}\ m^2/s$, how long will it take the virus to spread out to cover the whole floor of the laboratory?

Chapter 13: Probability Distribution Functions

In various branches of science, we often deal with $physical\ variables$, which are measurable and controllable quantities such as momentum and kinetic energy of a particle or rate of a chemical reaction, etc. The values of physical variables are determined by some physical or chemical property such as mass of the moving particle, temperature and concentration of chemicals, and others. On the other hand, a variable whose value is determined by some random phenomenon is called a $random\ variable$. Examples of random variables are the number of babies delivered in a hospital per day or the periods of rotation of various planets around the Sun, and so on. These quantities are measurable but not controllable. Typically, the outcome of an experiment is an event or number which could belong to either a set of discrete possibilities or to a continuous range of possible values. Flipping a coin or rolling a die are examples of experiments in which the outcome is discrete. On the other hand, observing the change in outdoor temperature in a day from morning to evening or measuring the speed of an accelerating car starting from rest leads to continuous outcomes. A probability distribution function is a mathematical way of associating the different outcomes in an experiment with the probability with which they occur. In this chapter we will learn about $discrete$ as well as $continuous$ probability distribution functions.

13.1 A DISCRETE PROBABILITY DISTRIBUTION FUNCTION

A pharmaceutical company assembles a group of twenty volunteers to test its new vaccine. Before starting the testing, the pharmaceutical company needs to get the vital signs of all volunteers. These include the height, weight, body temperature, and pulse rate of each volunteer. The heights of individual persons are, in inches, 60, 61, 62, 62, 64, 64, 64, 65, 65, 65, 65, 65, 65, 66, 66, 66, 68, 68, 69, and 70. To get the average or mean height, we simply add all the heights and divide by the number of persons, as

$$\bar{h} = \frac{60 + 61 + 62 + \cdots + 70}{20} = \frac{1300}{20} = 65 \text{ inches.}$$

[Note in passing that a variable with a bar on it is used to indicate the average or mean value of that variable.] Alternatively, we could organize the persons who have the same height into groups and write

$$\bar{h} = \frac{60 + 61 + 2(62) + 3(64) + 6(65) + 3(66) + 2(68) + 69 + 70}{20} = 65 \text{ inches.}$$

If there is a very large number of persons, say N, whose average height we need to find, we will first note that N_1 persons have height h_1, N_2 have height h_2, N_3 have height h_3… and N_n have height h_n, such that $N_1 + N_2 + \cdots + N_n = N$. Then,

$$\bar{h} = \frac{N_1h_1 + N_2h_2 + N_3h_3 + \cdots + N_nh_n}{N} = f_1h_1 + f_2h_2 + f_3h_3 + \cdots + f_nh_n \ , \quad Eq.(13.1a)$$

where,

$$f_i = \frac{N_i}{N} \text{ is the fraction of persons with height } h_i \ . \qquad Eq.(13.1b)$$

Thus,

$$\bar{h} = \sum_{i=1}^{n} f_i h_i \ , \qquad (Eq.13.1c)$$

where n is the number of different height groups. If instead of height, we are measuring a different vital sign, say, temperature of the volunteer, T, then,

$$\bar{T} = \sum_{i=1}^{n} f_i T_i \ ,$$

or pulse rate, P,

$$\bar{P} = \sum_{i=1}^{n} f_i P_i \ ,$$

or weight, w,

$$\bar{w} = \sum_{i=1}^{n} f_i w_i \ .$$

The number of groups n is different for different variables. In all these cases, the variable [height, temperature, weight, etc.] is a random variable and has discrete values (since the number of possible values is, at most, N). Also, note $\sum_{i=1}^{n} f_i$=1 in all cases, since all fractions of the sample (which, in this example, is the number of volunteers) must add up to 1. This is referred to as the $normalization\ condition$. Note that these fractions are measures of probabilities. For example, assume that in this cohort consisting of 10 males and 10 females, 14 volunteers are given the actual vaccine and the remaining six volunteers are given a placebo (a harmless medicine with no physiological effects). Males are given 4 placebos and 6 vaccines while the females are given 2 placebos and 8 vaccines. Then we can sort all the volunteers into four groups as shown in the following table.

	With placebo	With vaccine	Total
Male	4	6	10
Female	2	8	10
Total	6	14	20

The outcome of any measurement on this group is an event or a number which is discrete. Now, if a person is randomly chosen from this group, the probability of choosing a person who was given the actual vaccine will be 14/20 and probability of choosing a person given a placebo will be 6/20. The probability of choosing a male volunteer will be 10/20 and probability of choosing a female volunteer will also be 10/20. In general, if we are dealing with two or more events, then we need to find the ***joint probability*** that these events will occur simultaneously. If $P(X)$ is the probability of occurrence of event X and $P(Y)$ is the probability of occurrence of event Y, then joint probability of occurrence of both events X and Y simultaneously, if they are independent events, is

$$P(X \text{ and } Y) = P(X) \cdot P(Y) \ . \qquad\qquad Eq.(13.2a)$$

On the other hand, if the two events are not independent and their outcomes are related by some conditions, then we talk about ***conditional probability***. By definition, $P(X|Y)$ is the probability of occurrence of event X, given that event Y has already occurred. Similarly, $P(Y|X)$ is the probability of occurrence of event Y, given that event X has already occurred. The rules for finding the conditional probabilities are

$$P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)} \ \text{ and } \ P(Y|X) = \frac{P(X \text{ and } Y)}{P(X)} \ . \qquad\qquad Eq.(13.2b)$$

If X and Y are independent events, it means that occurrence of event X has no effect on the occurrence of event Y and vice versa and, therefore, $P(X|Y) = P(X)$ and $P(Y|X) = P(Y)$. Then, Eqs. (13.2b) reduce to Eq. (13.2a) for independent events. In case of conditional probabilities, it follows from Eq. (13.2b) that

$$P(X)\, P(Y|X) \ = \ P(Y)\, P(X|Y) \ , \qquad\qquad Eq.(13.2c)$$

which is called **Bayes' theorem** in statistics. This theorem provides the probability of an event, based on prior circumstances that might affect the event. Going back to the example of cohort with 20 volunteers, assume that event X is choosing a person with the vaccine and event Y is choosing a female person. Then $P(X)$ is $\frac{14}{20}$ and $P(Y)$ is $\frac{10}{20}$. Also from the table, $P(X \ and \ Y)$ is $\frac{8}{20}$. Since $P(X)$ multiplied by $P(Y)$ is not equal to $P(X \ and \ Y)$, it follows that events X and Y are not independent and, therefore, it is necessary to calculate the conditional probability. The probability of choosing a person with a vaccine, given that a female person has already been chosen, is

$P(X|Y) = \frac{8/20}{10/20} = 0.80$. Similarly, the probability of choosing a female person, given that a vaccinated person has already been chosen, is $P(Y|X) = \frac{8/20}{14/20} = \frac{4}{7} = 0.57$. It is easy to verify that Bayes' theorem is satisfied in this case.

If we have the average value $\bar{x}$ of a random variable x in a sample size of N, it does not provide any information about the spread of the values of x. To measure the spread of values, we define the $variance$ and $standard\ deviation$ of x. To quantify the spread of values of the variable, we measure the deviation of each data point from the mean value, namely, $x_i - \bar{x}$. The value of this deviation, for any individual data point, can be either positive or negative, and the simple sum of all of these deviations will be zero. That is so because, by definition of the average,

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \quad \text{or} \quad \sum_{i=1}^{N} x_i = N\,\bar{x} \quad \text{or} \quad \sum_{i=1}^{N} (x_i - \bar{x}) = 0 \ .$$

So, to define a variance and a standard deviation, σ, we first find the average value of the $squares$ of deviations over all data points. This is called variance and its square root is the standard deviation. Explicitly,

$$\begin{aligned} \text{variance}, \sigma^2 &= \textstyle\sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = \sum_{i=1}^{n} f_i x_i^2 - 2\,\bar{x} \sum_{i=1}^{n} f_i x_i + [\bar{x}]^2 \sum_{i=1}^{n} f_i \\ &= \overline{x^2} - 2\,\bar{x}\,\bar{x} + [\bar{x}]^2 = \overline{x^2} - [\bar{x}]^2 \ , \end{aligned} \qquad Eq.(13.3a)$$

and

$$\text{standard deviation}, \sigma = \sqrt{\overline{x^2} - [\bar{x}]^2} \ . \qquad Eq.(13.3b)$$

Roughly speaking, variance is an overall measure of how spread out the x_i values are from their mean value $\bar{x}$, while standard deviation is used when considering how an individual data point differs from the mean. In particular, the standard deviation is used when determining the statistical significance of an experimental result.

13.2 BINOMIAL DISTRIBUTION FUNCTION

Consider an experiment that has only two possible outcomes, which are mutually exclusive and can be thought of as a success or a failure. Examples of such experiments are tossing a coin or rolling a die. In case of coin-tossing, if getting a head is a success, then the probability of success is ½ and probability of failure is also ½. While in rolling a die, if getting a 4 is a success, then probability of success is 1/6 and probability of failure is 5/6. The binomial distribution provides the probability of observing n successes in N independent trials if the probability of success in a single trial is p. The success probability p is the same for all trials. The binomial distribution is

$$P(n, N, p) = \frac{N!}{n!\,(N-n)!} p^n (1-p)^{N-n} \quad . \qquad\qquad Eq.(13.4)$$

Example: In the maternity ward of a hospital, young moms can deliver either a boy or a girl with equal probability. On a particular day, six pregnant women are admitted in the hospital. What is the probability that only one boy will be delivered on this day? Or, two boys will be delivered? Or, three boys will be delivered? Or, four boys will be delivered? Or, five boys will be delivered? Finally, what is the probability that the six babies delivered on this day will be either all boys or all girls?

Solution: On this particular day six experiments are being done in the maternity ward and each experiment has two possible outcomes, either a boy or a girl. In this case $N = 6$. Assuming that delivering a boy is a success for the experiment (no sexism intended!), then $p = ½$. Using the binomial distribution, the probability of delivering only one boy (that is, $n = 1$) is

$$P(1,6,0.5) = \frac{6!}{1!\,(6-1)!} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{6-1} = \frac{6}{64} = \frac{3}{32} \; .$$

The probability of delivering two boys (that is, $n = 2$) is

$$P(2,6,0.5) = \frac{6!}{2!\,(6-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{6-2} = \frac{15}{64} \; .$$

Similarly, probabilities of delivering three or four or five boys are

$$P(3,6,0.5) = \frac{20}{64} = \frac{5}{16}, \qquad P(4,6,0.5) = \frac{15}{64} \qquad \text{and} \quad P(5,6,0.5) = \frac{6}{64} = \frac{3}{32} \; .$$

Finally, the probabilities of delivering all six boys $(n = 6)$ or all six girls $(n = 0)$ are

$$P(6,6,0.5) = \frac{1}{64} \quad \text{and} \quad P(0,6,0.5) = \frac{1}{64} \; .$$

Note that the sum of probabilities of all possibilities adds up to 1, that is

$$\sum_{n=0}^{6} P(n, 6, 0.5) = 1 \; ,$$

which is the normalization condition.

13.3 POISSON DISTRIBUTION FUNCTION

Let us, once again, consider an experiment which has two possible outcomes, but now we do not know what the probability of success, p, is in a single trial. If we do this experiment for a large number of times $(N \to \infty)$, then we can find an average rate of success. For example, if we toss a coin millions of times, we will infer that average rate of getting a head is ½. So, instead of dealing with p, we deal with average rate of success, namely, $Np \equiv \lambda$. So, we replace p by λ/N and take the limit $N \to \infty$ in the binomial distribution to get

$$P(n, \lambda) = \lim_{N\to\infty} \frac{N!}{n!\,(N-n)!}\left(\frac{\lambda}{N}\right)^n \left(1 - \frac{\lambda}{N}\right)^{N-n} \quad .$$

To evaluate this limiting expression, we look at it part-by-part. For example,

$$\lim_{N\to\infty} \frac{N!}{(N-n)!}\frac{1}{N^n} = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n} = \lim_{N\to\infty}\left(\frac{N}{N}\right)\left(\frac{N-1}{N}\right)\cdots\left(\frac{N-n+1}{N}\right) = 1,$$

and

$$\lim_{N\to\infty}\left(1 - \frac{\lambda}{N}\right)^{-n} = 1 \quad .$$

Also, in precalculus, Euler's number, e, is introduced as the limit of $(1 + 1/N)^N$ as N approaches infinity. The value of this constant is $e = 2.71828$. So,

$$\lim_{N\to\infty}\left(1 - \frac{\lambda}{N}\right)^{N} = e^{-\lambda} \quad .$$

Thus,

$$P(n, \lambda) = \frac{\lambda^n}{n!}\exp(-\lambda) \quad , \qquad\qquad Eq.(13.5)$$

which is known as the Poisson distribution. The Poisson distribution is appropriate if the average rate at which certain identical events occur during a time period is known. Furthermore, all events occur independently; that is, occurrence of one event does not affect the probability of occurrence of the next event. The Poisson distribution provides the probability that n events occur during a certain time period where $n = 0, 1, 2, \cdots$. Note in passing,

$$\sum_{n=0}^{\infty} P(n, \lambda) = \exp(+\lambda)\exp(-\lambda) = 1 \quad ,$$

which is the normalization condition.

Example: In several parts of the world, sightings of tornadoes are a common occurrence. According to the historical records of the National Weather Service of the USA, in the state of Michigan the average number of tornadoes is 15 per year, with the largest monthly average of three tornadoes occurring in June. With this information, what is the probability that there will be exactly one or two or three or four tornadoes in Michigan next year in June?

Solution: Since the average rate of tornadoes in June is three, $\lambda = 3$. The probability of n tornadoes in June is

$$P(n,3) = \frac{3^n}{n!}\exp(-3) \ .$$

Thus, probabilities of having one or two or three or four tornadoes next June are

$$P(1,3) = \frac{3^1}{1!}\exp(-3) = 0.149 \ ,$$

$$P(2,3) = \frac{3^2}{2!}\exp(-3) = 0.224 \ ,$$

$$P(3,3) = \frac{3^3}{3!}\exp(-3) = 0.224 \ ,$$

$$P(4,3) = \frac{3^4}{4!}\exp(-3) = 0.168 \ .$$

13.4 MAXWELL-BOLTZMANN (CONTINUOUS) DISTRIBUTION FUNCTION

Now suppose we wish to find the average speed of molecules in a room. It will be extremely difficult to do since the number of molecules is immeasurably large and difficult to count. Even if we only wish to find the fraction of molecules with a particular speed v_i, their number will be enormously large. In this case, we break the whole possible range of speeds into small groups or intervals Δv and concentrate on fraction of molecules with speeds between v_i and $v_i + \Delta v$. Then, the fraction of molecules in this interval is $f(v_i)\Delta v$. The function $f(v_i)$ is called a distribution function or probability density function (PDF). In this case,

$$\bar{v} = \sum_{i=1}^{n} v_i f(v_i)\Delta v \ ,$$

and, if $\Delta v \to 0$, the variable v becomes a continuous random variable,

$$\bar{v} = \int_{whole\ range\ of\ v} v f(v)\, dv \quad . \qquad Eq.(13.6a)$$

Also, since all fractions must add up to 1,

$$\int_{whole\ range\ of\ v} f(v)\, dv = 1 \quad . \qquad Eq.(13.6b)$$

This is the normalization condition for the distribution function. This is equivalent to the notion that we can be 100% certain that an observation will fall into the range of all possible outcomes.

The velocity distribution of molecules of mass m in a gas at temperature T is the Maxwell-Boltzmann distribution. It is a continuous distribution function given by

$$f(v) = 4\pi v^2 \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) \quad , \qquad Eq.(13.7)$$

where k is the Boltzmann constant. We note

$$\int_0^\infty f(v) dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \exp\left(-\frac{mv^2}{2T}\right) v^2\, dv \quad .$$

To evaluate this integral, we make a change of variables. Let $u^2 = \frac{m}{2kT} v^2$ or $u = \sqrt{\frac{m}{2kT}}\, v$ and $du = \sqrt{\frac{m}{2kT}} dv$.

Then,

$$\int_0^\infty f(v) dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \left(\frac{2kT}{m}\right)^{3/2} \int_0^\infty \exp(-u^2)\, u^2 du = 4\pi \frac{1}{\pi^{3/2}} \frac{\sqrt{\pi}}{4} = 1 \quad , \qquad Eq.(13.8a)$$

as expected, since this is the normalization condition. Here we used the integral I_g^2 from Eq. (2.13). We can also find average values of v and of v^2 as

$$\bar{v} = \int_0^\infty v f(v) dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \exp\left(-\frac{mv^2}{2kT}\right) v^3 dv \quad .$$

Again, we make a change of variables. Let $t = \frac{mv^2}{2kT}$ and $dt = \frac{m}{2kT} 2v dv$, so that

$$\bar{v} = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \frac{1}{2} \left(\frac{2kT}{m}\right)^2 \int_0^\infty \exp(-t)\, t\, dt = 2\left(\frac{2kT}{m\pi}\right)^{1/2} \int_0^\infty \exp(-t)\, t\, dt = \left(\frac{8kT}{\pi m}\right)^{1/2} \quad , \qquad Eq.(13.8b)$$

and

$$\overline{v^2} = \int_0^\infty v^2 f(v) dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \exp\left(-\frac{mv^2}{2kT}\right) v^4 dv \ .$$

With the change of variables, $u^2 = \frac{mv^2}{2kT}$,

$$\overline{v^2} = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \left(\frac{2kT}{m}\right)^{5/2} \int_0^\infty \exp(-u^2)\, u^4 du = \frac{4\pi}{\pi^{3/2}} \frac{2kT}{m} \frac{3}{8} \sqrt{\pi} = \frac{3kT}{m} \ . \qquad Eq. (13.8c)$$

Note in passing that $\frac{1}{2} m\overline{v^2} \ = \frac{3kT}{2}$, that is, average kinetic energy per molecule in a gas depends only on the temperature of the gas.

13.5 NORMAL OR GAUSSIAN (CONTINUOUS) DISTRIBUTION FUNCTION

Another useful continuous probability distribution function is the Gaussian, or normal, distribution function. A random variable x has normal distribution if $-\infty \leq x \leq \infty$ and the corresponding probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \ . \qquad Eq. (13.9)$$

By definition, $f(x)\ dx$ is the probability that value of variable x lies between x and $x + dx$. First, we verify the normalization condition, namely, $\int_{-\infty}^{\infty} f(x)\ dx = 1$. The integral is

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \ .$$

To evaluate this integral, we make a change of variable from x to u using $u = \frac{x-\mu}{\sigma}$, or $x = \sigma\, u + \mu$, and $dx = \sigma\, du$,

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{u^2}{2}\right] du \ = \frac{1}{\sqrt{2\pi}}\ \sqrt{2\pi} = 1 \ . \qquad Eq. (13.10a)$$

Here we used the integral I_{gg}^0 from Eq. (2.14). Next, we find average values of x and of x^2 using the same change of variable from x to u,

$$\bar{x} \ = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (u\sigma + \mu) \exp\left[-\frac{u^2}{2}\right] du \; = 0 + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{u^2}{2}\right] du \; = \mu \;\;, \qquad Eq.(13.10b)$$

and

$$\overline{x^2} \; = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (u^2\sigma^2 + 2\mu\sigma u \; + \mu^2) \exp\left[-\frac{u^2}{2}\right] du \; = \frac{1}{\sqrt{2\pi}} \{\sigma^2\sqrt{2\pi} + 0 + \mu^2\sqrt{2\pi}\,\}$$

or,

$$\overline{x^2} \; = \sigma^2 + \mu^2 \;\; . \qquad Eq.(13.10c)$$

Then,

$$\text{variance} = \; \overline{x^2} - [\bar{x}]^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$$

and,

$$\text{standard deviation} = \sigma \;\; . \qquad Eq.(13.11)$$

So, now we can interpret the normalized Gaussian distribution function. If a variable x is normally distributed, then $f(x)\,dx$ is the probability that x lies between x and $x + dx$. The function $f(x)$ has a mean value of μ and standard deviation of σ. If we change variable from x to $u = \frac{x-\mu}{\sigma}$, then, for the normal distribution,

$$f(x)\,dx = F(u)\,du \;\; ,$$

where

$$F(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \;\; .$$

We note that $F(u)$ is the same Gaussian function that we encountered in chapters 4 and 12. It is worth observing that $F(u)$ is a symmetric function of u, that is, $F(-u) = F(u)$. Also, $F(u)$ is a normalized function, that is, $\int_{-\infty}^{\infty} F(u)\,du = 1$. The quantity $F(u)\,du$ is the probability that u lies between u and $u + du$. The probability that u lies between a and b is

$$\int_a^b F(u)\,du = \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{u^2}{2}\right)du \ .$$

For given a and b, the numerical value of this integral is obtained from the tabulated values of the Gaussian function, $\Phi(x)$, defined as

$$\Phi(x) = \int_{-\infty}^{x} F(u)\,du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right)du \ . \qquad Eq. (13.12)$$

The tabulated numerical values of $\Phi(x)$, as a function of x, are given in Appendix D, in which the shaded area under the Gaussian curve is the value of $\Phi(x)$. Note $\Phi(\infty) = 1, \Phi(0) = \frac{1}{2}, \Phi(-\infty) = 0$. Also, due to the symmetric nature of $F(u)$,

$$\Phi(-x) = \int_{-\infty}^{-x} F(u)\,du = -\int_{\infty}^{x} F(-v)\,dv = \int_{x}^{\infty} F(v)\,dv = \int_{-\infty}^{\infty} F(v)\,dv - \int_{-\infty}^{x} F(v)\,dv$$

$$\Phi(-x) = 1 - \Phi(x) \ . \qquad Eq. (13.13)$$

This relationship allows us to find Φ for negative values of x in terms of Φ for positive values of x. So, in the standard table of Appendix D, the function $\Phi(x)$ is tabulated for positive values of x only. Also, if $b > a$, then

$$\int_a^b F(u)\,du = \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{u^2}{2}\right)du$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{b} \exp\left(-\frac{u^2}{2}\right)du - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \exp\left(-\frac{u^2}{2}\right)du = \Phi(b) - \Phi(a) \ . \qquad Eq. (13.14)$$

If the range of value of x is from x_L to x_H ($x_L \le x \le x_H$) instead of from $-\infty$ to ∞, then

$$f(x) = N\frac{1}{\sigma\sqrt{2\pi}}\begin{cases}\exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) & \text{for} \quad x_L \le x \le x_H \\ 0 & \text{otherwise}\end{cases} \ . \qquad Eq. (13.15a)$$

This is referred to as the $truncated\ normal\ distribution$. The scaling constant N is determined by the fact that $\int_{-\infty}^{\infty} f(x)dx = 1$, or,

$$N\frac{1}{\sigma\sqrt{2\pi}}\int_{x_L}^{x_H} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)dx = 1 \ .$$

Making a change of variable from x to $u = \frac{x-\mu}{\sigma}$,

$$1 = N\frac{1}{\sqrt{2\pi}}\int_{u_L}^{u_H} \exp\left(-\frac{u^2}{2}\right) \mathrm{du} = N[\Phi(u_H) - \Phi(u_L)] \ ,$$

or

$$N = \frac{1}{[\Phi(u_H) - \Phi(u_L)]} \ . \qquad Eq. (13.15b)$$

Example: The grade point averages (GPA) of a large population of students at a university are normally distributed with a mean of 2.4 and a standard deviation of 0.8. The students with GPAs higher than 3.3 receive honors credit while students with GPAs less than 2.0 do not receive the degree.

What percentage of students receive honors credit?

What percentage of students can never graduate?

Solution: Let x be GPA of students, $0 \le x \le 4$.

$$x_H = 4.0, u_H = \frac{x_H - \mu}{\sigma} = \frac{4 - 2.4}{0.8} = 2.0 \ ,$$

$$x_L = 0.0, u_L = \frac{x_L - \mu}{\sigma} = \frac{0 - 2.4}{0.8} = -3.0 \ ,$$

$$N = \frac{1}{\Phi(2.0) - \Phi(-3.0)} = \frac{1}{\Phi(2) - [1 - \Phi(3.0)]} = \frac{1}{\Phi(2) + \Phi(3.0) - 1}$$

$$= \frac{1}{0.9772 + 0.9987 - 1} = \frac{1}{1.9759 - 1} = \frac{1}{0.9759} \ .$$

(a)

$$\text{Percentage of honors students} = \int_{3.3}^{4.0} f(x)dx = N\frac{1}{\sigma\sqrt{2\pi}}\int_{3.3}^{4.0} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \ .$$

Let $u = \frac{x-\mu}{\sigma}$, then lower limit of $u = \frac{3.3-2.4}{0.8} = \frac{9}{8}$, upper limit of $u = \frac{4.0-2.4}{0.8} = 2$.

$$\text{Percentage} = N\frac{1}{\sqrt{2\pi}}\int_{9/8}^{2} \exp\left[-\frac{u^2}{2}\right] du$$

$$= \frac{\Phi(2) - \Phi\left(\frac{9}{8}\right)}{\Phi(2) - \Phi(-3)} = \frac{0.9772 - 0.8686}{0.9759} = \frac{0.1086}{0.9759} = 0.11 \quad \text{or} \quad 11\% \ .$$

(b)

$$\text{Percentage of failing students} = \int_{0}^{2.0} f(x)dx = N\frac{1}{\sigma\sqrt{2\pi}}\int_{0}^{2.0} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx \ .$$

Let $u = \frac{x-\mu}{\sigma}$, lower limit $= \frac{0-2.4}{0.8} = -3$, upper limit $= \frac{2.0-2.4}{0.8} = -0.5$.

$$\text{Percentage } = N\frac{1}{\sqrt{2\pi}}\int_{-3}^{-0.5} \exp\left[-\frac{u^2}{2}\right]du$$

$$= N[\Phi(-0.5) - \Phi(-3)] = N[1 - \Phi(0.5) - 1 + \Phi(3)] = N[\Phi(3) - \Phi(0.5)]$$

$$= \frac{0.9987 - 0.6915}{0.9759} = \frac{0.3072}{0.9759} = 0.31 \quad \text{or} \quad 31\% \ .$$

Example: The U.S. Centers for Disease Control and Prevention has reported several cases of Salmonella outbreak linked to guinea pigs. A veterinary research physician sets up a trap to catch a large number of guinea pigs. If the mean width of guinea pig's shoulder is $\mu = 3.8$ inches with a standard deviation of 0.6 inch, and if the door of the trap is 5 inches wide, what percentage of guinea pigs will pass through the door?

Solution: Let x be the shoulder width of a guinea pig, so that $0 \leq x \leq \infty$. The percentage of guinea pigs with shoulder length less than 5 inches is

$$\text{Percentage} = \frac{\frac{1}{\sigma\sqrt{2\pi}}\int_0^{5.0} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx}{\frac{1}{\sigma\sqrt{2\pi}}\int_0^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx} \ .$$

Make a change of variables from x to $u = \frac{x-\mu}{\sigma}$.

For $x = \infty, u = \infty$. For $x = 0, u = \frac{0-3.8}{0.6} = -\frac{19}{3}$. For $x = 5.0, u = \frac{5-3.8}{0.6} = 2.0$.

$$\text{Percentage} = \frac{\frac{1}{\sqrt{2\pi}}\int_{-19/3}^{2.0} \exp\left[-\frac{u^2}{2}\right]du}{\frac{1}{\sqrt{2\pi}}\int_{-19/3}^{\infty} \exp\left[-\frac{u^2}{2}\right]du} = \frac{\Phi(2) - \Phi(-\frac{19}{3})}{\Phi(\infty) - \Phi(-\frac{19}{3})}$$

$$= \frac{\Phi(2) - [1 - \Phi(6.33)]}{\Phi(\infty) - [1 - \Phi(6.33)]} = \frac{\Phi(2) + \Phi(6.33) - 1}{\Phi(\infty) + \Phi(6.33) - 1} = \frac{0.9772 + 1 - 1}{1 + 1 - 1} = 0.9772 \ ,$$

or 97.7% guinea pigs will fit through the door.

13.6 A FINAL REMARK

As a final remark, it is worth pointing out that concepts related to probability distribution functions are central to quantum physics. The initial formulation of quantum physics using matrix methods led to the understanding of eigenvalues and eigenfunctions of matrices as representing physical observables. The gist of quantum physics is to talk about *possibilities* and *probabilities* in the measurement of physical properties, such as energy, momentum, angular momentum, and so on, at the scale of atomic and subatomic particles. The eigenvalues represent the possibilities and the eigenfunctions (or, their modulus square) represent the corresponding probability distribution functions.

PROBLEMS FOR CHAPTER 13

1. To explain the conduction properties of metals, Paul Drude provided a simple model of a metal as a reservoir of free-electron gas. Even though Drude's model was successful in explaining the classical Ohm's law, it failed to account for the quantum behavior of electrons. The Exclusion Principle in quantum physics asserts that no more than two electrons can occupy the same atomic orbital. Classically, particles of a gas at absolute zero temperature will all have zero kinetic energy. In quantum physics, all electrons in an electron gas cannot be in ground state since that would contradict the Exclusion Principle. At absolute zero temperature, the fraction of electrons in a metal with energies between E and $E + dE$ is given by

$$f(E)\,dE = \begin{cases} C\sqrt{E}\,dE & \text{for } 0 \le E \le E_F \\ 0 & \text{for } E > E_F \end{cases} ,$$

where E_F is a constant (called the Fermi energy).

Determine C if this energy distribution is normalized to 1.

Find the average energy of electrons in terms of E_F.

2. A one-dimensional classical harmonic oscillator of mass m and frequency ω is oscillating along the x–axis with its equilibrium point at $x = 0$. The turning points of the classical oscillator at $x = \pm x_c$ refer to the values of x at which the oscillator momentarily comes to rest and reverses its direction of motion. In other words, x_c is the largest displacement or the amplitude of the classical oscillator. The total energy E of the classical oscillator is conserved and its value at the classical turning points is $E = \frac{1}{2} m\omega^2 x_c^2$. In quantum physics, the one-dimensional harmonic oscillator can be found anywhere along the x–axis from $x = -\infty$ to $x = +\infty$. The probability of finding the quantum oscillator between x and $x + dx$ is

$$P(x)\,dx = C \exp(-x^2/x_c^2)\ dx\ .$$

Determine C if this probability distribution function is normalized to 1.

Find the probability of locating the quantum oscillator outside the classical turning points.

3. The simplest of all atoms is the hydrogen atom consisting of a proton and an electron. The proton is in the nucleus and the electron can be anywhere in the space surrounding the nucleus, with its position given by a distribution function. The probability of finding the electron at a distance between r and $r + dr$ from the nucleus is

$$P(r)\,dr = C\,r^2 \exp(-2r/a_o)\,dr\ ,$$

where a_o is a constant (called the Bohr radius). Here, r can take any value between 0 and ∞.

Determine C if this probability distribution function is normalized to 1.

Using this probability distribution function, determine the average value $\bar{r}$ of r.

4. **Biomedical Physics Application.** The gestation period of human pregnancies has a normal distribution with a mean μ of 268 days and standard deviation σ of 16 days.

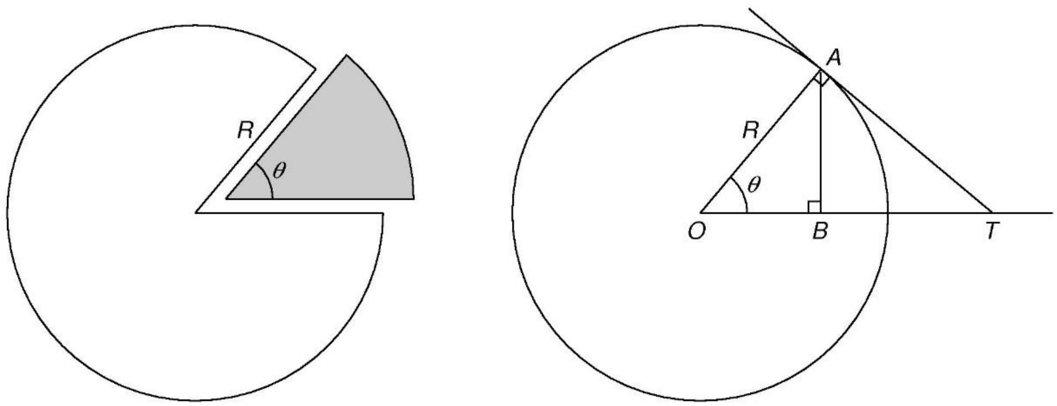
What percentage of pregnancies last between 260 and 284 days?

The colloquial term "preemies" refers to infants who have preterm birth with gestational age of less than 37 weeks, or 259 days. What percentage of infants are preemies?

Post-term babies are born after 284 days of pregnancy. What percentage of pregnancies leads to post-term babies?

Appendix A: Limiting Value of $\sin\theta/\theta$ as $\theta \to 0$.

Using simple facts from geometry and trigonometry, we can determine the limiting value of $\sin\theta/\theta$ as θ approaches 0.



Consider a circular pie, of radius R, which is cut into slices. One slice of angular size θ (shown in the left figure) is of area A_s. Using the concept of proportionalities, ratio of area of slice (A_s) and area of full pie (πR^2) is same as the ratio of angular size of slice (θ) and angular size of full circular pie, namely, 2π. Thus, $\frac{A_s}{\pi R^2} = \frac{\theta}{2\pi}$ or $A_s = \frac{\theta R^2}{2}$. In the adjoining right figure, we show the same slice of angular size θ and label one of its corners as A. From this corner A we drop a perpendicular line that meets the opposite side of the slice at point B. Also, at corner A, we draw a tangent to the circular pie (of center O) that meets the extension of line OB at point T. From this second figure, we note

$$\text{Area of triangle } OAB \leq \text{Area of slice} \leq \text{Area of triangle } OAT \;.$$

Or,

$$\frac{1}{2}\,(OB)(AB) \leq A_s \leq \frac{1}{2}\,(OA)(AT)$$

$$\frac{1}{2}\,(R\cos\theta)(R\sin\theta) \leq \frac{\theta\, R^2}{2} \leq \frac{1}{2}\,(R)(R\tan\theta)$$

$$\cos\theta \leq \frac{\theta}{\sin\theta} \leq \frac{1}{\cos\theta}$$

Since $\cos\theta \to 1$ as $\theta \to 0$, it follows that

$$\lim_{\theta\to 0}\frac{\sin(\theta)}{\theta} = 1 \quad . \qquad Eq.(A.1)$$

In addition, we note that for small values of θ, $\sin\theta \approx \theta$.

Appendix B: Mnemonic for Maxwell's Relations of Thermodynamics

Laws of thermodynamics play fundamental roles in physics, biology, chemistry, chemical engineering, mechanical engineering, and other scientific disciplines. In their mathematical descriptions, these laws relate four physical variables, pressure (P), volume (V), temperature (T) and entropy (S). Derivatives of these variables are related via four Maxwell's relations of thermodynamics, which are

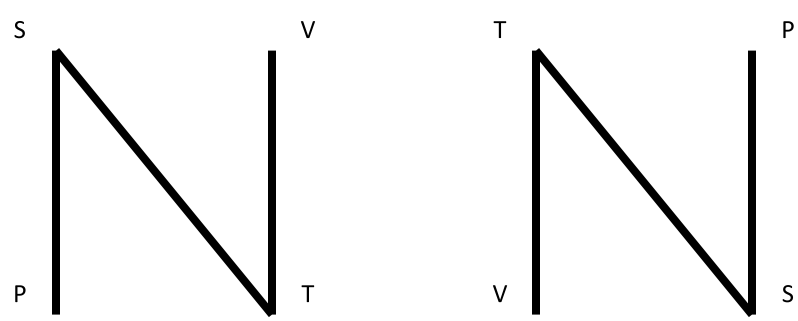
$$+\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P \ ,$$

$$+\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V \ ,$$

$$+\left(\frac{\partial P}{\partial T}\right)_V = +\left(\frac{\partial S}{\partial V}\right)_T \ ,$$

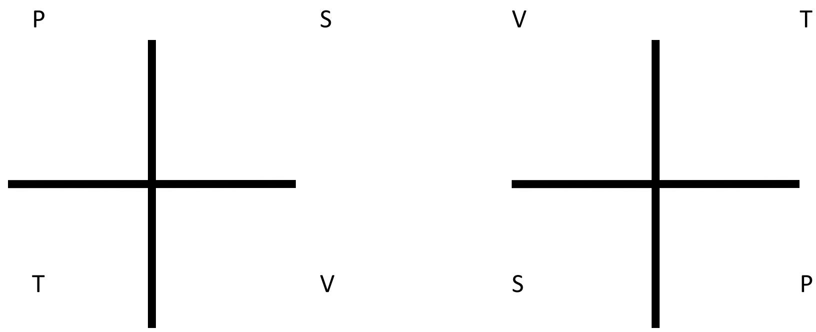
$$+\left(\frac{\partial V}{\partial S}\right)_P = +\left(\frac{\partial T}{\partial P}\right)_S \ .$$

This appendix presents a simple mnemonic device for recalling these relationships without any complications for remembering the correct signs. Note that two of these relations have a positive (+) sign on the right-hand side and the other two have a negative (N) sign on the right-hand side. In each relationship, the variable with respect to which a partial derivative is taken on one side is kept fixed on the other side. The first two relations with the negative sign on the right-hand side can be memorized as follows. Recalling the letter N for the negative relations, wrap the four letters PSTV for the thermodynamic variables, *in alphabetical order*, along this N in two possible ways as shown here:



In these figures, the placement of letters is in the same location/way as in the first two Maxwell's relations, which contain the negative sign on the right-hand side.

To get the mnemonic for the other two Maxwell relations—namely, the ones with the positive sign on the right-hand side—rotate these figures by $\pi/2$ (does not matter whether the rotation is clockwise or counterclockwise) and replace N (negative) by + (positive), to get



In these figures, the placement of letters is in the same location/way as in the last two Maxwell's relations, which contain the positive sign on the right-hand side.

Appendix C: Proof of the Epsilon-Delta Identity

In the introduction of Levi-Civita symbol, ϵ_{ijk}, in the Interlude chapter, a relationship between ϵ_{ijk} and Kronecker delta δ_{ij} was presented as

$$\sum_i \epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} \ .$$

This epsilon-delta identity can be proved by starting with a determinant of order 3 whose element, a_{ij}, located at the intersection of i^{th} row and j^{th} column, is δ_{ij}. This determinant is

$$\mathbb{A} = \begin{vmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 \ .$$

In the Interlude chapter, the value of Levi-Civita symbol, ϵ_{ijk}, was given as $+1$ (or -1) if an even (or odd) permutation of (ijk) gives (123). Thus, on replacing the row numbers (123) of $\mathbb{A}$ by (ijk), the value of the determinant will become ϵ_{ijk} instead of 1, that is

$$\begin{vmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{vmatrix} = \epsilon_{ijk} \ .$$

Similarly, on replacing the column numbers (123) by (nlm), the value of the determinant will become

$$\begin{vmatrix} \delta_{in} & \delta_{il} & \delta_{im} \\ \delta_{jn} & \delta_{jl} & \delta_{jm} \\ \delta_{kn} & \delta_{kl} & \delta_{km} \end{vmatrix} = \epsilon_{ijk}\epsilon_{nlm} \ .$$

Now, replace n by i, and expand the determinant across first column explicitly, to get

$$\epsilon_{ijk}\epsilon_{ilm} = \delta_{ii}\left(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\right) - \delta_{ji}(\delta_{il}\delta_{km} - \delta_{im}\delta_{kl}) + \delta_{ki}\left(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right) \ .$$

Finally, sum over i and use $\sum_{i=1}^{3} \delta_{ii} = 3$ to obtain

$$\sum_i \epsilon_{ijk}\epsilon_{ilm} = 3\left(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\right) - \left(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\right) + \left(\delta_{jm}\delta_{kl} - \delta_{jl}\delta_{km}\right) \ ,$$

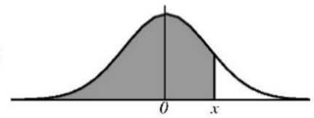
or,

$$\textstyle\sum_i \epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} \ .$$

Appendix D: Areas under the Standard Normal Curve

The table below provides the values of the Gaussian function, $\Phi(x)$, as a function of *x*.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2)\, du$$



x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5238	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6330	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7156	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7641	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 13.1. Areas under the standard curve from $-\infty$ to *x*.