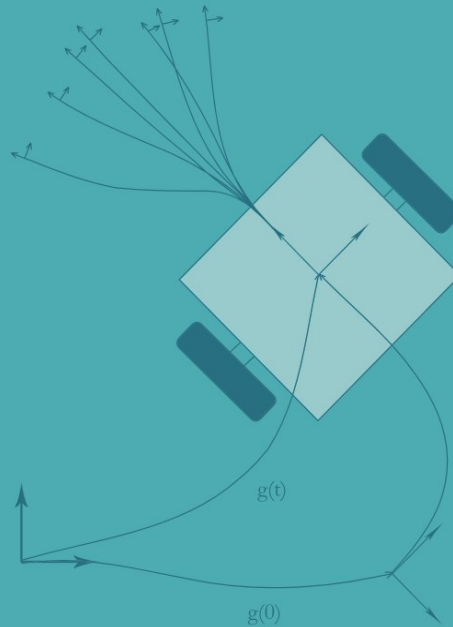


# Stochastic Models, Information Theory, and Lie Groups

Volume

1

*Classical Results and Geometric Methods***Gregory S. Chirikjian**



# Applied and Numerical Harmonic Analysis

## ***Series Editor***

**John J. Benedetto**

University of Maryland

## **Editorial Advisory Board**

***Akram Aldroubi***

Vanderbilt University

***Douglas Cochran***

Arizona State University

***Ingrid Daubechies***

Princeton University

***Hans G. Feichtinger***

University of Vienna

***Christopher Heil***

Georgia Institute of Technology

***Murat Kunt***

Swiss Federal Institute of Technology, Lausanne

***James McClellan***

Georgia Institute of Technology

***Wim Sweldens***

Lucent Technologies, Bell Laboratories

***Michael Unser***

Swiss Federal Institute  
of Technology, Lausanne

***Martin Vetterli***

Swiss Federal Institute  
of Technology, Lausanne

***M. Victor Wickerhauser***

Washington University

Gregory S. Chirikjian

Stochastic Models,  
Information Theory,  
and Lie Groups,  
Volume 1

Classical Results and Geometric Methods

Birkhäuser  
Boston • Basel • Berlin

Gregory S. Chirikjian  
Department of Mechanical Engineering  
The Johns Hopkins University  
Baltimore, MD 21218-2682  
USA  
gregc@jhu.edu

ISBN 978-0-8176-4802-2 e-ISBN 978-0-8176-4803-9  
DOI 10.1007/978-0-8176-4803-9

Library of Congress Control Number: 2009933211

Mathematics Subject Classification (2000): 22E60, 53Bxx, 58A15, 60H10, 70G45, 82C31, 94A15, 94A17

© Birkhäuser Boston, a part of Springer Science+Business Media, LLC 2009

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Birkhäuser Boston, c/o Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

*Cover design:* Alex Gerasev

Printed on acid-free paper

Birkhäuser Boston is part of Springer Science+Business Media ([www.birkhauser.com](http://www.birkhauser.com))

To my family

---

## ANHA Series Preface

The *Applied and Numerical Harmonic Analysis (ANHA)* book series aims to provide the engineering, mathematical, and scientific communities with significant developments in harmonic analysis, ranging from abstract harmonic analysis to basic applications. The title of the series reflects the importance of applications and numerical implementation, but richness and relevance of applications and implementation depend fundamentally on the structure and depth of theoretical underpinnings. Thus, from our point of view, the interleaving of theory and applications and their creative symbiotic evolution is axiomatic.

Harmonic analysis is a wellspring of ideas and applicability that has flourished, developed, and deepened over time within many disciplines and by means of creative cross-fertilization with diverse areas. The intricate and fundamental relationship between harmonic analysis and fields such as signal processing, partial differential equations (PDEs), and image processing is reflected in our state-of-the-art *ANHA* series.

Our vision of modern harmonic analysis includes mathematical areas such as wavelet theory, Banach algebras, classical Fourier analysis, time-frequency analysis, and fractal geometry, as well as the diverse topics that impinge on them.

For example, wavelet theory can be considered an appropriate tool to deal with some basic problems in digital signal processing, speech and image processing, geophysics, pattern recognition, biomedical engineering, and turbulence. These areas implement the latest technology from sampling methods on surfaces to fast algorithms and computer vision methods. The underlying mathematics of wavelet theory depends not only on classical Fourier analysis, but also on ideas from abstract harmonic analysis, including von Neumann algebras and the affine group. This leads to a study of the Heisenberg group and its relationship to Gabor systems, and of the metaplectic group for a meaningful interaction of signal decomposition methods. The unifying influence of wavelet theory in the aforementioned topics illustrates the justification for providing a means for centralizing and disseminating information from the broader, but still focused, area of harmonic analysis. This will be a key role of *ANHA*. We intend to publish with the scope and interaction that such a host of issues demands.

Along with our commitment to publish mathematically significant works at the frontiers of harmonic analysis, we have a comparably strong commitment to publish major advances in the following applicable topics in which harmonic analysis plays a substantial role:

<i>Antenna theory</i>	<i>Prediction theory</i>
<i>Biomedical signal processing</i>	<i>Radar applications</i>
<i>Digital signal processing</i>	<i>Sampling theory</i>
<i>Fast algorithms</i>	<i>Spectral estimation</i>
<i>Gabor theory and applications</i>	<i>Speech processing</i>
<i>Image processing</i>	<i>Time-frequency and</i>
<i>Numerical partial differential equations</i>	<i>time-scale analysis</i>
	<i>Wavelet theory</i>

The above point of view for the *ANHA* book series is inspired by the history of Fourier analysis itself, whose tentacles reach into so many fields.

In the last two centuries Fourier analysis has had a major impact on the development of mathematics, on the understanding of many engineering and scientific phenomena, and on the solution of some of the most important problems in mathematics and the sciences. Historically, Fourier series were developed in the analysis of some of the classical PDEs of mathematical physics; these series were used to solve such equations. In order to understand Fourier series and the kinds of solutions they could represent, some of the most basic notions of analysis were defined, e.g., the concept of “function.” Since the coefficients of Fourier series are integrals, it is no surprise that Riemann integrals were conceived to deal with uniqueness properties of trigonometric series. Cantor’s set theory was also developed because of such uniqueness questions.

A basic problem in Fourier analysis is to show how complicated phenomena, such as sound waves, can be described in terms of elementary harmonics. There are two aspects of this problem: first, to find, or even define properly, the harmonics or spectrum of a given phenomenon, e.g., the spectroscopy problem in optics; second, to determine which phenomena can be constructed from given classes of harmonics, as done, for example, by the mechanical synthesizers in tidal analysis.

Fourier analysis is also the natural setting for many other problems in engineering, mathematics, and the sciences. For example, Wiener’s Tauberian theorem in Fourier analysis not only characterizes the behavior of the prime numbers, but also provides the proper notion of spectrum for phenomena such as white light; this latter process leads to the Fourier analysis associated with correlation functions in filtering and prediction problems, and these problems, in turn, deal naturally with Hardy spaces in the theory of complex variables.

Nowadays, some of the theory of PDEs has given way to the study of Fourier integral operators. Problems in antenna theory are studied in terms of unimodular trigonometric polynomials. Applications of Fourier analysis abound in signal processing, whether with the fast Fourier transform (FFT), or filter design, or the adaptive modeling inherent in time-frequency-scale methods such as wavelet theory. The coherent states of mathematical physics are translated and modulated Fourier transforms, and these are used, in conjunction with the uncertainty principle, for dealing with signal reconstruction in communications theory. We are back to the *raison d’être* of the *ANHA* series!

*John J. Benedetto*  
Series Editor  
University of Maryland  
College Park



---

## Preface

As an undergraduate student at a good engineering school, I had never heard of stochastic processes or Lie groups (even though I double majored in Mathematics). As a faculty member in engineering I encountered many problems where the recurring themes were “noise” and “geometry.” When I went to read up on both topics I found fairly little at this intersection. Now, to be certain, there are many wonderful texts on one of these subjects or the other. And to be fair, there are several advanced treatments on their intersection. However, for the engineer or scientist who has the modest goal of modeling a stochastic (i.e., time-evolving and random) mechanical system with equations with an eye towards numerically simulating the system’s behavior rather than proving theorems, very few books are out there. This is because mechanical systems (such as robots, biological macromolecules, spinning tops, satellites, automobiles, etc.) move in multiple spatial dimensions, and the configuration space that describes allowable motions of objects made up of rigid components does not fit into the usual framework of linear systems theory. Rather, the configuration space manifold is usually either a Lie group or a homogeneous space<sup>1</sup>.

My mission then became clear: write a book on stochastic modeling of (possibly complicated) mechanical systems that a well-motivated first-year graduate student or undergraduate at the senior level in engineering or the physical sciences could pick up and read cover-to-cover without having to carry around twenty other books. The key point that I tried to keep in mind when writing this book was that the art of mathematical modeling is very different than the art of proving theorems. The emphasis here is on “how to calculate” quantities (mostly analytically by hand and occasionally numerically by computer) rather than “how to prove.” Therefore, some topics that are treated at great detail in mathematics books are covered at a superficial level here, and some concrete analytical calculations that are glossed over in mathematics books are explained in detail here. In other words the goal here is not to expand the frontiers of mathematics, but rather to translate known results to a broader audience.

The following quotes from Felix Klein<sup>2</sup> in regard to the modern mathematics of his day came to mind often during the writing process:

The exposition, intended for a few specialized colleagues, refrains from indicating any connection with more general questions. Hence it is barely accessible to colleagues in neighboring fields and totally inaccessible to a larger circle. . .

---

<sup>1</sup>The reader is not expected to know what these concepts mean at this point.

<sup>2</sup>F. Klein, *Development of Mathematics in the 19th Century*, translated by M. Ackerman as part of *Lie Groups: History, Frontiers and Applications*, Vol. IX, Math Sci Press, 1979.

In fact, the physicist can use little, and the engineer none at all, of these theories in his tasks.

The later of these was also referenced in Arnol'd's classic book<sup>3</sup> as an example of how work that is initially viewed as esoteric can become central to applied fields.

In order to emphasize the point that this book is for practitioners, as I present results they generally are not in "definition–proof–theorem" format. Rather, results and derivations are presented in a flowing style. Section headings punctuate results so that the presentation (hopefully) does not ramble on too much.

Another difference between this book and one on pure mathematics is that while pathological examples can be viewed as the fundamental motivation for many mathematical concepts (e.g., the behavior of  $\sin \frac{1}{x}$  as  $x \rightarrow 0$ ), in most applications most functions and the domains on which they are defined do not exhibit pathologies. And so practitioners can afford to be less precise than pure mathematicians.

A final major difference between this presentation and those written by mathematicians is that rather than the usual "top-down" approach in which examples follow definitions and theorems, the approach here is "bottom-up" in the sense that examples are used to motivate concepts throughout this book and the companion volume. Then after the reader gains familiarity with the concepts, definitions are provided to capture the essence of the examples.

To help with the issue of motivation and to illustrate the art of mathematical modeling, case studies from a variety of different engineering and scientific fields are presented. In fact, so much material is covered that this book has been split into two volumes. Volume 1 (which is what you are reading now) focuses on basic stochastic theory and geometric methods. The usefulness of some of these methods may not be clear until the second volume. For example, some results pertaining to differential forms and differential geometry that are presented in Volume 1 are not applied to stochastic models until they find applications in Volume 2 in the form of integral geometry (also called geometric probability) and in multivariate statistical analysis. Volume 2 serves as an in-depth (but accessible) treatment of Lie groups, and the extension of statistical and information-theoretic techniques to that domain.

I have organized Volume 1 into the following 9 chapters and an appendix: Chapter 1 provides an introduction and overview of the kinds of the problems that can be addressed using the mathematical modeling methods of this book. Chapter 2 reviews every aspect of the Gaussian distribution, and uses this as the quintessential example of a probability density function. Chapter 3 discusses probability and information theory and introduces notation that will be used throughout these volumes. Chapter 4 is an overview of white noise, stochastic differential equations (SDEs), and Fokker–Planck equations on the real line and in Euclidean space. The relationship between Itô and Stratonovich SDEs is explained and examples illustrate the conversions between these forms on multi-dimensional examples in Cartesian and curvilinear coordinate systems. Chapter 5 provides an introduction to geometry including elementary projective, algebraic, and differential geometry of curves and surfaces. That chapter begins with some concrete examples that are described in detail. Chapter 6 introduces differential forms and the generalized Stokes' theorem. Chapter 7 generalizes the treatment of surfaces and polyhedra to manifolds and polytopes. Geometry is first described using a coordinate-dependent presentation that some differential geometers may find old fashioned, but it

---

<sup>3</sup>See Arnol'd, VI, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, Berlin, 1978.

is nonetheless fully rigorous and general, and far more accessible to the engineer and scientist than the elegant and powerful (but cryptic) coordinate-free descriptions. Chapter 8 discusses stochastic processes in manifolds and related probability flows. Chapter 9 summarizes the current volume and introduces Volume 2. The appendix provides a comprehensive review of concepts from linear algebra, multivariate calculus, and systems of first-order ordinary differential equations. To the engineering or physical science student at the senior level or higher, some of this material will be known already. But for those who have not seen it before, it is presented in a self-contained manner. In addition, exercises at the end of each chapter in Volume 1 reinforce the main points. There are more than 150 exercises in Volume 1. Volume 2 also has many exercises. Over time I plan to build up a full solution set that will be uploaded to the publisher's webpage, and will be accessible to instructors. This will provide many more worked examples than space limits allow within the volumes.

Volume 1 can be used as a textbook in several ways. Chapters 2–4 together with the appendix can serve as a one-semester course on continuous-time stochastic processes. Chapters 5–8 can serve as a one-semester course on elementary differential geometry. Or, if chapters are read sequentially, the whole book can be used for self-study. Each chapter is meant to be relatively self-contained, with its own references to the literature. Altogether there are approximately 250 references that can be used to facilitate further study.

The stochastic models addressed here are equations of motion for physical systems that are forced by noise. The time-evolving statistical properties of these models are studied extensively. Information theory is concerned with communicating and extracting content in the presence of noise. Lie groups either can be thought of as continuous sets of symmetry operations, or as smooth high-dimensional surfaces which have an associated operator. That is, the same mathematical object can be viewed from either a more algebraic or more geometric perspective.

Whereas the emphasis of Volume 1 is on basic theory of continuous-time stochastic processes and differential geometric methods, Volume 2 provides an in-depth introduction to matrix Lie groups, stochastic processes that evolve on Lie groups, and information-theoretic inequalities involving groups. Volume 1 only has a smattering of information theory and Lie groups. Volume 2 emphasizes information theory and Lie groups to a much larger degree.

Information theory consists of several branches. The branch originating from Shannon's mathematical theory of communication is covered in numerous engineering textbooks with minor variants on the titles "Information Theory" or "Communications Theory." A second branch of information theory, due to Wiener, is concerned with filtering of noisy data and extracting a signal (such as in radar detection of flying objects). The third branch originated from the field of mathematical statistics in which people like Fisher, de Bruijn, Cramér, and Rao developed concepts in statistical estimation. It is primarily this third branch that is addressed in Volume 1, and so very little of the classical engineering information theory is found here. However, Shannon's theory is reviewed in detail in Volume 2, where connections between many aspects of information and group theory are explored. And Wiener's filtering ideas (which have a strong connection with Fourier analysis) find natural applications in the context of deconvolving functions on Lie groups (an advanced topic that is also deferred to Volume 2).

Volume 2 is a more formal and more advanced presentation that builds on the basics covered in Volume 1. It is composed of three parts. Part 1 begins with a detailed treatment of Lie groups including elementary algebraic, differential geometric, and func-

tional analytic properties. Classical variational calculus techniques are reviewed, and the coordinate-free extension of these concepts to Lie groups (in the form of the Euler–Poincaré equation) are derived and used in examples. In addition, the basic concepts of group representation theory are reviewed along with the concepts of convolution of functions and Fourier expansions on Lie groups. Connections with multivariate statistical analysis and integral geometry are also explored. Part 2 of Volume 2 is concerned with the connections between information theory and group theory. An extension of the de Bruijn inequality to the context of Lie groups is examined. Classical communication-theory problems are reviewed, and information inequalities that have parallels in group theory are explained. Geometric and algebraic problems in coding theory are also examined. A number of connections to problems in engineering and biology are provided. For example, it is shown how a spherical optical encoder developed by the author and coworkers<sup>4</sup> can be viewed as a decoding problem on the rotation group,  $SO(3)$ . Also, the problem of noise in coherent optical communication systems is formulated and the resulting Fokker–Planck equation is shown to be quite similar to that of the stochastic Kinematic cart that is described in the introductory chapter of Volume 1. This leads to Part 3 of Volume 2, which brings the discussion back to issues close to those in Volume 1. Namely, stochastic differential equations and Fokker–Planck equations are revisited. In Volume 2 all of these equations evolve on Lie groups (particularly the rotation and rigid-body-motion groups). The differential geometric techniques that are presented in Volume 1 are applied heavily in this setting. Several closely related (though not identical) concepts of “mean” and “covariance” of probability densities on Lie groups are reviewed, and their propagation under iterated convolutions is studied. As far as the descriptions of probability densities on Lie groups are concerned, closed-form Gaussian-like approximations are possible in some contexts, and Fourier-based solutions are more convenient in others. The coordinate-based tools needed for realizing these expressions as concrete quantities (which can in principle be implemented numerically) are provided in Volume 2.

During a lecture I attended while writing this book, an executive from a famous computer manufacturer said that traditionally technical people have been trained to be “I-shaped,” meaning an education that is very deep in one area, but not broad. The executive went on to say that he now hires people who are “T-shaped,” meaning that they have a broad but generally shallow background that allows them to communicate with others, but in addition have depth in one area. From this viewpoint, the present book and its companion volume are “III-shaped,” with a broad discussion of geometry that is used to investigate three areas of knowledge relatively deeply: stochastic models, information theory, and Lie groups.

It has been a joy to write these books. It has clarified many issues in my own mind. And I hope that you find them both interesting and useful. And while I have worked hard to eliminate errors, there will no doubt be some that escaped my attention. Therefore I welcome any comments/corrections and plan to keep an updated online erratum page which can be found by searching for my name on the Web.

There are so many people without whom this book would not have been completed. First, I must thank John J. Benedetto for inviting me to contribute to this series that he is editing, and Tom Grasso at Birkhäuser for making the process flow smoothly.

A debt of gratitude is owed to a number of people who have worked (and maybe suffered) through early drafts of this book. These include my students Kevin Wolfe,

---

<sup>4</sup>Stein, D., Scheinerman, E.R., Chirikjian, G.S., “Mathematical models of binary spherical-motion encoders,” *IEEE-ASME Trans. Mechatron.*, 8, pp. 234–244, 2003.

Michael Kutzer, and Matt Moses who received very rough drafts, and whose comments and questions were very useful in improving the presentation and content. I would also like to thank all of my current and former students and colleagues for providing a stimulating environment in which to work.

Mathematicians Ernie Kalnins, Peter T. Kim, Willard Miller, Jr., and Julie Mitchell provided comments that helped significantly in identifying mathematical errors, fine-tuning definitions, and organizing topics. I am thankful to Tamás Kalmár-Nagy, Jennifer Losaw, Tilak Ratnanather, and Jon Selig for finding several important typographical errors. John Oprea went way above and beyond the call of duty to read and provide detailed comments on two drafts that led to a significant reorganization of the material. Andrew D. Lewis provided some very useful comments and the picture of a torus that appears in Chapter 5. Andrew Douglas, Tak Igusa, and Frank C. Park each provided some useful and/or encouraging comments. Wooram Park helped with some of the figures.

I would like to thank William N. Sharpe, Jr. for hiring me many years ago straight out of graduate school (even after knowing me as an undergraduate), and Nick Jones, the Benjamin T. Rome Dean of the JHU Whiting School of Engineering, for allowing me to have the sabbatical during the 2008 calendar year that was used to write this book after my service as department chair finished.

I would also like to thank the faculty and staff of the Institute for Mathematics and Its Applications (IMA) at the University of Minnesota for the three week-long workshops that I attended there during part of the time while I was writing this book. Some of the topics discussed here percolated through my mind during that time.

Last but not least, I would like to thank my family. Writing a single-author book can be a solitary experience. And so it is important to have surroundings that are “fuuuun.”

Baltimore, Maryland

*Gregory Chirikjian*  
May 2009

---

# Contents

<b>ANHA Series Preface</b> .....	vii
<b>Preface</b> .....	ix
<b>1 Introduction</b> .....	1
1.1 What this Book is About .....	1
1.2 Different Meanings of Equality .....	4
1.2.1 Defining Equalities .....	4
1.2.2 Equality in the Sense of Zero Mean-Squared Error .....	5
1.2.3 Big-O Notation .....	6
1.2.4 A Philosophical View of Equality and Inequality .....	7
1.3 Other Useful Shortcuts .....	8
1.3.1 Simplifying Notation .....	8
1.3.2 Nice, Well-Behaved, Non-Pathological Functions .....	9
1.4 Modern Mathematical Notation and Terminology .....	10
1.4.1 What is Modern Mathematics? .....	10
1.4.2 Stating Mathematical Results .....	11
1.4.3 Sets and Mappings .....	11
1.4.4 Transformation Groups .....	17
1.4.5 Understanding Commutative Diagrams .....	17
1.5 Transport Phenomena and Probability Flow .....	19
1.5.1 Continuum Mechanics .....	20
1.5.2 Heat Flow and Entropy .....	25
1.6 Organization of this Book .....	27
1.7 Chapter Summary .....	28
1.8 Exercises .....	28
References .....	29
<b>2 Gaussian Distributions and the Heat Equation</b> .....	31
2.1 The Gaussian Distribution on the Real Line .....	31
2.1.1 Defining Parameters .....	31
2.1.2 The Maximum Entropy Property .....	34
2.1.3 The Convolution of Gaussians .....	36
2.1.4 The Fourier Transform of the Gaussian Distribution .....	37
2.1.5 Diffusion Equations .....	38
2.1.6 Stirling's Formula .....	39

2.2	The Multivariate Gaussian Distribution . . . . .	40
2.2.1	Conditional and Marginal Densities . . . . .	40
2.2.2	Multi-Dimensional Integrals Involving Gaussians . . . . .	42
2.3	The Volume of Spheres and Balls in $\mathbb{R}^n$ . . . . .	43
2.4	Clipped Gaussian Distributions . . . . .	45
2.4.1	One-Dimensional Clipped Gaussian Distributions . . . . .	45
2.4.2	Multi-Dimensional Clipped Gaussian Distributions . . . . .	46
2.5	Folded, or Wrapped, Gaussians . . . . .	47
2.6	The Heat Equation . . . . .	48
2.6.1	The One-Dimensional Case . . . . .	48
2.6.2	The Multi-Dimensional Case . . . . .	50
2.6.3	The Heat Equation on the Unit Circle . . . . .	51
2.7	Gaussians and Multi-Dimensional Diffusions . . . . .	51
2.7.1	The Constant Diffusion Case . . . . .	51
2.7.2	The Time-Varying Case . . . . .	52
2.8	Symmetry Analysis of Evolution Equations . . . . .	53
2.8.1	Symmetries in Parameters . . . . .	54
2.8.2	Infinitesimal Symmetry Operators of the Heat Equation . . . . .	54
2.8.3	Non-Linear Transformations of Coordinates . . . . .	57
2.9	Chapter Summary . . . . .	58
2.10	Exercises . . . . .	58
	References . . . . .	60
<b>3</b>	<b>Probability and Information Theory . . . . .</b>	<b>63</b>
3.1	Probability Theory in Euclidean Space . . . . .	64
3.1.1	Basic Definitions and Properties of Probability Density Functions . . . . .	64
3.1.2	Change of Variables . . . . .	65
3.1.3	Marginalization, Conditioning, and Convolution . . . . .	65
3.1.4	Mean and Covariance . . . . .	66
3.1.5	Parametric Distributions . . . . .	67
3.2	Conditional Expectation . . . . .	68
3.2.1	Jensen’s Inequality and Conditional Expectation . . . . .	71
3.2.2	Convolution and Conditional Expectation . . . . .	72
3.3	Information Theory . . . . .	73
3.3.1	Entropy of Conditional and Marginal Density Functions . . . . .	74
3.3.2	Entropy and Gaussian Distributions . . . . .	75
3.3.3	Mutual Information . . . . .	76
3.3.4	Information-Theoretic Measures of Divergence . . . . .	76
3.3.5	Fisher Information . . . . .	77
3.3.6	Information and Convolution . . . . .	78
3.3.7	Shift and Scaling Properties . . . . .	80
3.4	Parameter Estimation . . . . .	81
3.4.1	Unbiased Estimators . . . . .	82
3.4.2	The Cramér–Rao Bound . . . . .	82
3.4.3	Demonstration with Gaussian Distributions . . . . .	84
3.4.4	The de Bruijn Identity . . . . .	84
3.4.5	The Entropy Power Inequality . . . . .	85
3.4.6	Entropy of a Weighted Sum of Disjoint PDFs . . . . .	87
3.4.7	Change of Coordinates . . . . .	88
3.4.8	Computation of Entropy via Discretization . . . . .	89

3.5	The Classical Central Limit Theorem . . . . .	91
3.5.1	The Central Limit Theorem (Fourier Version) . . . . .	91
3.5.2	The Central Limit Theorem (RMSD Error Version) . . . . .	92
3.5.3	The Central Limit Theorem (Information-Theoretic Version) . . . . .	94
3.5.4	Limitations of the Central Limit Theorem . . . . .	94
3.6	An Alternative Measure of Dispersion . . . . .	95
3.7	Chapter Summary . . . . .	95
3.8	Exercises . . . . .	96
	References . . . . .	97
<b>4</b>	<b>Stochastic Differential Equations . . . . .</b>	<b>101</b>
4.1	Motivating Examples . . . . .	101
4.1.1	The Discrete Random Walker . . . . .	102
4.1.2	Continuous-Time Brownian Motion in Continuous Space . . . . .	103
4.2	Stationary and Non-Stationary Random Processes . . . . .	105
4.2.1	Weak and Strong Stationarity . . . . .	105
4.2.2	Non-Stationary Processes . . . . .	106
4.3	Gaussian and Markov Processes . . . . .	106
4.4	Wiener Processes and Stochastic Differential Equations . . . . .	108
4.4.1	An Informal Introduction . . . . .	108
4.4.2	Abstracted Definitions . . . . .	111
4.5	The Itô Stochastic Calculus . . . . .	112
4.5.1	Itô Stochastic Differential Equations in $\mathbb{R}^d$ . . . . .	114
4.5.2	Numerical Approximations . . . . .	115
4.5.3	Mathematical Properties of the Itô Integral . . . . .	116
4.5.4	Evaluating Expectations is Convenient for Itô Equations . . . . .	118
4.5.5	Itô's Rule . . . . .	119
4.5.6	The Fokker–Planck Equation (Itô Version) . . . . .	120
4.6	The Stratonovich Stochastic Calculus . . . . .	121
4.7	Multi-Dimensional Ornstein–Uhlenbeck Processes . . . . .	123
4.7.1	Steady-State Conditions . . . . .	124
4.7.2	Steady-State Solution . . . . .	126
4.7.3	Detailed Balance and the Onsager Relations . . . . .	128
4.8	SDEs and Fokker–Planck Equations Under Coordinate Changes . . . . .	130
4.8.1	Brownian Motion in the Plane . . . . .	130
4.8.2	General Conversion Rules . . . . .	134
4.8.3	Coordinate Changes and Fokker–Planck Equations . . . . .	134
4.9	Chapter Summary . . . . .	136
4.10	Exercises . . . . .	136
	References . . . . .	138
<b>5</b>	<b>Geometry of Curves and Surfaces . . . . .</b>	<b>141</b>
5.1	An Introduction to Geometry Through Robotic Manipulator Kinematics . . . . .	142
5.1.1	Forward (or Direct) Kinematics . . . . .	142
5.1.2	Reverse (or Inverse) Kinematics . . . . .	144
5.2	A Case Study in Medical Imaging . . . . .	147
5.2.1	A Parametric Approach . . . . .	148
5.2.2	An Implicit Approach . . . . .	151
5.3	Differential Geometry of Curves . . . . .	154
5.3.1	Local Theory of Curves . . . . .	155



5.3.2	Global Theory of Curves . . . . .	157
5.4	Differential Geometry of Surfaces in $\mathbb{R}^3$ . . . . .	159
5.4.1	The First and Second Fundamental Forms . . . . .	161
5.4.2	Curvature . . . . .	163
5.4.3	The Sphere . . . . .	166
5.4.4	The Ellipsoid of Revolution . . . . .	167
5.4.5	The Torus . . . . .	169
5.4.6	The Gauss–Bonnet Theorem and Related Inequalities . . . . .	170
5.5	Tubes . . . . .	171
5.5.1	Offset Curves in $\mathbb{R}^2$ . . . . .	171
5.5.2	Parallel Fibers, Ribbons, and Tubes of Curves in $\mathbb{R}^3$ . . . . .	172
5.5.3	Tubes of Surfaces in $\mathbb{R}^3$ . . . . .	175
5.6	The Euler Characteristic: From One Dimension to $N$ Dimensions . . . . .	176
5.6.1	The Euler Characteristic of Zero-, One-, and Three-Dimensional Bodies . . . . .	176
5.6.2	Relationship Between the Euler Characteristic of a Body and Its Boundary . . . . .	178
5.6.3	The Euler Characteristic of Cartesian Products of Objects . . . . .	179
5.7	Implicit Surfaces, Level Set Methods, and Curvature Flows . . . . .	179
5.7.1	Implicit Surfaces . . . . .	179
5.7.2	Integration on Implicitly Defined Surfaces and Curves in $\mathbb{R}^3$ . . . . .	181
5.7.3	Integral Theorems for Implicit Surfaces . . . . .	184
5.7.4	Level Sets and Curvature Flows . . . . .	184
5.8	Chapter Summary . . . . .	186
5.9	Exercises . . . . .	186
	References . . . . .	190
<b>6</b>	<b>Differential Forms . . . . .</b>	<b>193</b>
6.1	An Informal Introduction to Differential Forms on $\mathbb{R}^n$ . . . . .	194
6.1.1	Definitions and Properties of $n$ -Forms and Exterior Derivatives . . . . .	194
6.1.2	Exterior Derivatives of $(n - 1)$ -Forms on $\mathbb{R}^n$ for $n = 2, 3$ . . . . .	197
6.1.3	Products of Differential Forms . . . . .	199
6.1.4	Concise Notation for Differential Forms and Exterior Derivatives . . . . .	200
6.2	Permutations . . . . .	201
6.2.1	Examples of Permutations and Their Products . . . . .	201
6.2.2	The Sign of a Permutation . . . . .	202
6.2.3	Multi-Dimensional Version of the Levi–Civita Symbol . . . . .	202
6.3	The Hodge Star Operator . . . . .	204
6.4	Tensor Products and Dual Vectors . . . . .	206
6.5	Exterior Products . . . . .	207
6.5.1	A Concrete Introduction to Exterior Products . . . . .	207
6.5.2	Abstract Definition of the Exterior Product of Two Vectors . . . . .	208
6.5.3	The Exterior Product of Several Vectors . . . . .	209
6.5.4	Computing with Exterior Products . . . . .	210
6.5.5	The Exterior Product of Two Exterior Products . . . . .	210
6.5.6	The Inner Product of Two Exterior Products . . . . .	211
6.5.7	The Dual of an Exterior Product . . . . .	211
6.6	Invariant Description of Vector Fields . . . . .	211
6.7	Push-Forwards and Pull-Backs in $\mathbb{R}^n$ . . . . .	213
6.7.1	General Theory . . . . .	213

6.7.2	Example Calculations . . . . .	214
6.8	Generalizing Integral Theorems from Vector Calculus . . . . .	221
6.8.1	Integration of Differential Forms . . . . .	221
6.8.2	The Inner Product of Forms . . . . .	222
6.8.3	Green’s Theorem for a Square Region in $\mathbb{R}^2$ . . . . .	223
6.8.4	Stokes’ Theorem for a Cube in $\mathbb{R}^3$ . . . . .	223
6.8.5	The Divergence Theorem for a Cube in $\mathbb{R}^3$ . . . . .	224
6.8.6	Detailed Examples . . . . .	224
6.8.7	Closed Forms and Diffusion Equations . . . . .	226
6.9	Differential Forms and Coordinate Changes . . . . .	227
6.10	Chapter Summary . . . . .	228
6.11	Exercises . . . . .	229
	References . . . . .	232
<b>7</b>	<b>Polytopes and Manifolds</b> . . . . .	<b>233</b>
7.1	Properties and Operations on Convex Polytopes in $\mathbb{R}^n$ . . . . .	234
7.1.1	Computing the Volume and Surface Area of Polyhedra . . . . .	235
7.1.2	Properties of Minkowski Sums . . . . .	235
7.1.3	Convolution of Bodies . . . . .	237
7.2	Examples of Manifolds . . . . .	238
7.3	Embedded Manifolds, Part I: Using Vector Calculus . . . . .	246
7.3.1	The Inner Product of Vector Fields on Manifolds Embedded in $\mathbb{R}^n$ . . . . .	246
7.3.2	An Example: A Hyper-Spherical Cap . . . . .	247
7.3.3	Computing Normals Extrinsically Without the Cross Product . . . . .	250
7.3.4	The Divergence Theorem in Coordinates . . . . .	253
7.3.5	Integration by Parts on an Embedded Manifold . . . . .	254
7.3.6	Curvature . . . . .	257
7.4	Covariant vs. Contravariant . . . . .	258
7.4.1	Tensors . . . . .	258
7.4.2	Derivatives and Differentials . . . . .	259
7.5	Embedded Manifolds, Part II: Using Differential Forms . . . . .	261
7.5.1	Push-Forwards and Pull-Backs (Revisited) . . . . .	261
7.5.2	Expressing Pull-Backs of Forms in Coordinates . . . . .	262
7.5.3	Volume Element of an Embedded Manifold . . . . .	263
7.5.4	Conversion to Vector Notation . . . . .	264
7.5.5	General Properties of Differential Forms on Embedded Manifolds . . . . .	265
7.6	Intrinsic Description of Riemannian Manifolds . . . . .	266
7.6.1	Computing Tangent Vectors and Boundary Normals in Local Coordinates . . . . .	269
7.6.2	Stokes’ Theorem for Manifolds . . . . .	270
7.6.3	The Gauss–Bonnet–Chern Theorem . . . . .	277
7.7	Fiber Bundles and Connections . . . . .	278
7.7.1	Fiber Bundles . . . . .	279
7.7.2	Connections . . . . .	281
7.8	The Heat Equation on a Riemannian Manifold . . . . .	283
7.9	Chapter Summary . . . . .	283
7.10	Exercises . . . . .	284
	References . . . . .	286

<b>8</b>	<b>Stochastic Processes on Manifolds</b> . . . . .	289
8.1	The Fokker–Planck Equation for an Itô SDE on a Manifold: A Parametric Approach . . . . .	289
8.2	Itô Stochastic Differential Equations on an Embedded Manifold: An Implicit Approach . . . . .	291
8.2.1	The General Itô Case . . . . .	291
8.2.2	Bilinear Itô Equations that Evolve on a Quadratic Hyper-Surface in $\mathbb{R}^n$ . . . . .	292
8.3	Stratonovich SDEs and Fokker–Planck Equations on Manifolds . . . . .	293
8.3.1	Stratonovich SDEs on Manifolds: Parametric Approach . . . . .	293
8.3.2	Stratonovich SDEs on Manifolds: Implicit Approach . . . . .	294
8.4	Entropy and Fokker–Planck Equations on Manifolds . . . . .	295
8.5	Examples . . . . .	296
8.5.1	Stochastic Motion on the Unit Circle . . . . .	296
8.5.2	The Unit Sphere in $\mathbb{R}^3$ : Parametric Formulation . . . . .	297
8.5.3	SDEs on Spheres and Rotations: Extrinsic Formulation . . . . .	299
8.5.4	The SDE and Fokker–Planck Equation for the Kinematic Cart . . . . .	299
8.6	Solution Techniques . . . . .	300
8.6.1	Finite Difference and Finite Elements . . . . .	300
8.6.2	Non-Parametric Density Estimation . . . . .	301
8.6.3	Separation of Variables: Diffusion on $SE(2)$ as a Case Study . . . . .	301
8.7	Chapter Summary . . . . .	308
8.8	Exercises . . . . .	309
	References . . . . .	309
<b>9</b>	<b>Summary</b> . . . . .	313
	References . . . . .	314
<b>A</b>	<b>Review of Linear Algebra, Vector Calculus, and Systems Theory</b> . . . . .	315
A.1	Vectors . . . . .	315
A.1.1	Vector Spaces . . . . .	315
A.1.2	Linear Mappings and Isomorphisms . . . . .	317
A.1.3	The Scalar Product and Vector Norm . . . . .	317
A.1.4	The Gram–Schmidt Orthogonalization Process . . . . .	319
A.1.5	Dual Spaces . . . . .	319
A.1.6	The Vector Product in $\mathbb{R}^3$ . . . . .	320
A.2	Matrices . . . . .	321
A.2.1	Matrix Multiplication and the Trace . . . . .	322
A.2.2	The Determinant . . . . .	323
A.2.3	The Inverse of a Matrix . . . . .	325
A.2.4	Pseudo-Inverses and Null Spaces . . . . .	326
A.2.5	Special Kinds of Matrices . . . . .	327
A.2.6	Matrix Norms . . . . .	328
A.2.7	Matrix Inequalities . . . . .	331
A.3	Eigenvalues and Eigenvectors . . . . .	331
A.4	Matrix Decompositions . . . . .	334
A.4.1	Jordan Blocks and the Jordan Decomposition . . . . .	334
A.4.2	Decompositions into Products of Special Matrices . . . . .	335
A.4.3	Decompositions into Blocks . . . . .	336
A.5	Matrix Perturbation Theory . . . . .	337

A.6	The Matrix Exponential . . . . .	338
A.7	Kronecker Products and Kronecker Sums . . . . .	340
A.8	Complex Numbers and Fourier Analysis . . . . .	342
A.9	Important Inequalities from the Theory of Linear Systems . . . . .	344
A.10	The State-Transition Matrix and the Product Integral . . . . .	346
A.11	Vector Calculus . . . . .	349
	A.11.1 Optimization in $\mathbb{R}^n$ . . . . .	349
	A.11.2 Differential Operators in $\mathbb{R}^n$ . . . . .	351
	A.11.3 Integral Theorems in $\mathbb{R}^2$ and $\mathbb{R}^3$ . . . . .	351
	A.11.4 Integration by Parts in $\mathbb{R}^n$ . . . . .	353
	A.11.5 The Chain Rule . . . . .	353
	A.11.6 Matrix Differential Calculus . . . . .	354
A.12	Exercises . . . . .	356
	References . . . . .	360
<b>Index</b>	. . . . .	<b>363</b>

## Introduction

This chapter is an overview of the sorts of problems that can be addressed using the methods from this book. It also discusses the major differences between mathematical modeling and mathematics, and reviews some basic terminology that is used throughout the book. The appendix provides a much more in-depth review of engineering mathematics. This book is meant to be self-contained in the sense that only prior knowledge of college-level calculus, linear algebra, and differential equations is assumed. Therefore, if it is read sequentially and something does not make sense, then the appendix most likely contains the missing piece of knowledge. Standard references on classical mathematics used in engineering and physics include [2, 5], which also can be consulted to fill in any missing background.

Even after consulting the appendix and the cited references, some of the concepts presented toward the end of each chapter may be difficult to grasp on the first reading. That is okay. To a large extent, it should be possible to skip over some of the more difficult concepts in any given chapter, and still understand the fundamental ideas in subsequent chapters. In order to focus the reader on the most important ideas in each chapter, the equations that are necessary to successfully navigate through later chapters are circumscribed with a box. This also makes it easier to refer back to key equations.

The main things to take away from this chapter are:

- To become accustomed to the style and notation used in this book, including the concepts of sets, mappings, commutative diagrams, etc.;
- To understand that there are several different meanings of “equality” and “inequality”;
- To review topics in advanced calculus and its applications in mechanics, including the application of the divergence theorem and localization arguments;
- To be able to compose mappings and do calculations with Jacobian matrices;
- To understand the layout of the rest of the book and to get a feeling for the topics that can be addressed with the tools developed here.

### 1.1 What this Book is About

Practitioners (such as the author) are motivated to make the investment to learn new mathematics when the potential payoff of that investment is clear up front. Therefore, before delving into the intricate details of stochastic calculus, information theory, Lie groups, etc., consider the following simply stated problems:

**Problem 1:** A random walker on a sphere starts at the north pole. After some period of time, what will the probability be that the walker is at a particular location? And how long will it take before the walker’s location on the sphere is completely randomized (i.e., how long will it be before the initial location of the walker becomes irrelevant)?

**Problem 2:** The cart-like robot shown in Figure 1.1 moves around in the plane by turning each of its two wheels. Relative to a frame of reference fixed in the plane, the frame of reference fixed in the robot moves as a function of the torque inputs imparted by the motors to the wheels. This reference frame can be thought of as the time-dependent rigid-body motion

$$g = \begin{pmatrix} \cos \theta & -\sin \theta & x \\ \sin \theta & \cos \theta & y \\ 0 & 0 & 1 \end{pmatrix} \quad (1.1)$$

where  $\theta$  is the angle that the axle makes with the  $x$ -axis of the world frame, and  $x$  and  $y$  are the components of the translation of the center of the cart-like robot relative to the frame of reference fixed in the plane. If the robot’s motion has been observed, then  $g(t)$  is known for all times from  $t = 0$  up to the present time. However, the exact location of the future location of the robot is uncertain until it actually happens since, for example, the wheels might slip. Given models describing these uncertainties, what will the most likely position and orientation of the robot be at a given future time?

Let the two wheels each have radii  $r$ , and let the distance between the wheels (called the wheelbase) be denoted as  $L$ . Imagine that the angles through which the wheels turn around their axes are governed by “stochastic differential equations” of the form<sup>1</sup>

$$d\phi_1 = \omega(t)dt + \sqrt{D}dw_1 \quad (1.2)$$

$$d\phi_2 = \omega(t)dt + \sqrt{D}dw_2 \quad (1.3)$$

where  $dw_i$  each represent “uncorrelated unit white noise,”  $D$  scales the strength of the noise, and  $r\omega(t)$  is what the speed of the robot would be if  $D$  were zero. Then a “stochastic trajectory” for  $g(t)$  in (1.1) is defined by stochastic differential equations of the form [13]

$$\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix} = \begin{pmatrix} r\omega \cos \theta \\ r\omega \sin \theta \\ 0 \end{pmatrix} dt + \sqrt{D} \begin{pmatrix} \frac{r}{2} \cos \theta & \frac{r}{2} \cos \theta \\ \frac{r}{2} \sin \theta & \frac{r}{2} \sin \theta \\ \frac{r}{L} & -\frac{r}{L} \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}. \quad (1.4)$$

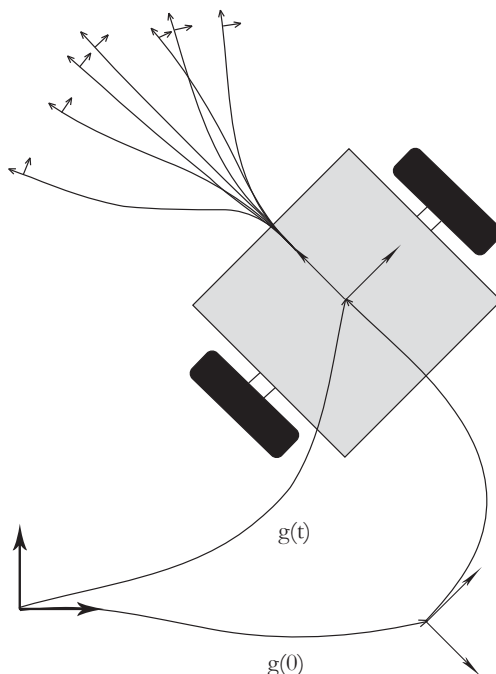
Stochastic trajectories, by definition, are not repeatable. However, if such an equation is simulated many times, each time starting from the same initial conditions (say,  $x = y = \theta = 0$ ), then a function,  $f(x, y, \theta; t)$  that records the distribution of positions and orientations of the cart at the same value of time,  $t$ , in each trajectory can be defined. As will be seen in Chapter 4, a well-developed theory for linking stochastic differential equations such as (1.4) to functions such as  $f(x, y, \theta; t)$  exists. This theory produces a partial differential equation (called a *Fokker–Planck equation*) for  $f(x, y, \theta; t)$ . In the present context, this equation is of the form:

$$\begin{aligned} \frac{\partial f}{\partial t} = & -r\omega \cos \theta \frac{\partial f}{\partial x} - r\omega \sin \theta \frac{\partial f}{\partial y} \\ & + \frac{D}{2} \left( \frac{r^2}{2} \cos^2 \theta \frac{\partial^2 f}{\partial x^2} + \frac{r^2}{2} \sin^2 \theta \frac{\partial^2 f}{\partial x \partial y} + \frac{r^2}{2} \sin^2 \theta \frac{\partial^2 f}{\partial y^2} + \frac{2r^2}{L^2} \frac{\partial^2 f}{\partial \theta^2} \right). \end{aligned} \quad (1.5)$$

---

<sup>1</sup>The terms in quotes will be defined in Chapter 4. It is not expected that the reader will understand these concepts at this point, but rather only get a taste of things to come.

By the end of this volume the reader will know how to derive such equations and generate sample paths from the corresponding stochastic differential equation.



**Fig. 1.1.** A Kinematic Cart with an Uncertain Future Position and Orientation

**Problem 3:** A long and slender semi-flexible biological macromolecule, such as double-helical DNA composed of 300 stacked base pairs, is subjected to random Brownian motion bombardment by the surrounding solvent molecules. If reference frames are attached to both ends of the DNA, what will the distributions of rigid-body motions between these reference frames look like as a function of temperature and the stiffness of the molecule?

**Problem 4:** An isolated *E. coli* bacterium swims in a medium and, based on sensory information, randomly reorients. For a given starting position, nutrient environment, and temperature, what will the probability be that it reaches a particular position at a particular time?

**Problem 5:** (a) One rigid body is set at a fixed pose (or position and orientation) in a box, and a second rigid body is allowed to move uniformly at random in the box under the constraint that it cannot intersect the first body. How much free space is there for the second body to move? (b) If the opposing faces of the box are identified with each other (by “gluing them together”), then the boundaries are removed, but the volume in this toroidal world will be the same as that of the original box. How much free space is there for the second body to move in this scenario?

All of these problems (and many more) are discussed in detail in Volume 2. The current volume establishes the methodological foundations that are required in order

to approach the applications. In the process of laying these foundations, numerous “toy examples” will be provided. The essence of the problems and methods addressed is that there is some “geometric” feature that is intertwined with a “probabilistic” one.

Lie groups are a natural tool to study geometry, and a mature theory for random processes on Lie groups has been developed by mathematicians over the past one hundred years. This theory is closely connected with applications. For example, in Problems 1–5 mentioned above, the group of rotations and full rigid-body motions (rotations + translations) figure prominently. Stochastic processes that evolve on Lie groups can be treated *more concretely* than those that evolve on abstract manifolds, because Lie groups have structure that is “close to” that of the vector space  $\mathbb{R}^n$ , with the group operation taking the place of regular addition. Volume 2 focuses on Lie groups.

Many excellent and precise mathematics books exist on stochastic calculus, manifolds, and stochastic calculus on manifolds. However, for the practitioner, the barrier to understanding the content of such books can be quite high. In part, this is because when mathematical statements are made, the level of precision can be greater than the practitioner needs. The approach taken in the current book to lower the bar to understanding these phenomena is to use a weak sense of the equality sign in equations. This is explained in detail in the next section.

## 1.2 Different Meanings of Equality

This section reviews several different meanings associated with the equality sign, “=.” At the outset, this seems like an absurd thing to discuss, since we all know from early educational experiences that two things are either equal or they are not. But sometimes it is convenient to claim that two quantities are equal when, strictly speaking, they are not. This avoids having to append to every approximate equality words like “almost surely” or “up to a set of measure zero” or “in the mean squared sense.” While such statements add to precision, they also distract from the main points that a modeler seeks to glean from the mathematics literature.

### 1.2.1 Defining Equalities

A first point of confusion concerning the equality sign is when an equality is derived (i.e., when it is the statement of a result), versus when it is part of a definition. The usual “=” will be used to denote any derived equality.

Let  $x$  be a real number, and let  $\mathbb{R}$  denote the set of all real numbers<sup>2</sup>. For a function, such as  $f(x) = x^2$ , it is possible for  $f(x) = 0$  (in this specific case, when  $x = 0$ ). In that context, making a statement like  $f(x) = 0$  might be the starting point for a root-finding problem. However, in other problems it may be desirable to say that  $f(x) = 0$  for all of the values of  $x$ . In this case, it is said that  $f(x)$  is identically equal to zero, which is denoted as  $f(x) \equiv 0$ . This is a kind of “temporary definition” of  $f(x)$ , but as the symbol  $f(x)$  is used elsewhere in the text, the restriction that it is set to zero for the time being is not carried forward. In contrast, “ $\stackrel{\text{def}}{=}$ ” will be used to denote a defining equality, and definitions made in this way will persist throughout the book. Sometimes defining equalities are composed of several conditional equalities. For example, a family of functions  $\phi_{\alpha,n}(x)$  can be defined as

---

<sup>2</sup>It is also convenient to denote  $\mathbb{R}^n$  to be the space of all  $n$ -dimensional column vectors with real entries.



$$\phi_{\alpha,n}(x) \doteq \begin{cases} e^{-\alpha/x^{2n}} & \text{if } x, \alpha > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (1.6)$$

This means that  $\phi_{\alpha,n}(x)$  is defined to be equal to one thing when certain conditions are met, and something else under different conditions.

In calculus, the derivative of a function  $f(x)$  on the open interval<sup>3</sup>  $(a, b) \subset \mathbb{R}$  is defined as the new function

$$\frac{df}{dx} \doteq \lim_{\epsilon \rightarrow 0} \frac{f(x + \epsilon) - f(x)}{\epsilon} \quad (1.7)$$

on the same interval. Here  $\epsilon$  can be thought of as  $1/N$  where  $N \rightarrow \infty$ . If  $f'(x) \doteq df/dx$  is a continuous function, then  $f(x)$  is called continuously differentiable. Likewise, if  $d^2f/dx^2 \doteq df'/dx$  is a continuous function, then  $f(x)$  is called twice continuously differentiable. A function for which  $n$  derivatives can be taken, each resulting in a continuous function, is called *n-times continuously differentiable*, and is denoted as  $C^n(a, b)$ .

The shorthand  $df/dx$  to describe the calculation in (1.7) thereby circumvents the complexity that would arise in writing higher derivatives such as

$$\frac{d^2f}{dx^2} = \lim_{\epsilon_2 \rightarrow 0} \lim_{\epsilon_1 \rightarrow 0} \frac{[f(x + \epsilon_1 + \epsilon_2) - f(x + \epsilon_1)] - [f(x + \epsilon_1) - f(x)]}{\epsilon_1 \epsilon_2}.$$

Indeed, such an expression would be quite confusing, which emphasizes the importance of simplifying notation.

As a result of the *Fundamental Theorem of Calculus*, if  $f'(x)$  is continuous then

$$\int_a^b f'(x)dx = f(b) - f(a) \quad \text{and} \quad \frac{d}{dt} \int_0^t f(x)dx = f(t) \quad (1.8)$$

where

$$\int_a^b f'(x)dx \doteq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^N f'(a + (b-a)n/N). \quad (1.9)$$

Rather than carrying around the explicit limits, in practice the now-standard notations of differential and integral calculus are well known to engineers and scientists. After the initial investment has been made to learn the notations of calculus, there is rarely any need to go back to the defining limits. But for the student who has never seen the notation, it would be impossible to understand the usefulness of calculus. Likewise, the task of assisting the engineer or applied scientist in understanding basic results of modern mathematics can be made much easier by keeping the presentation as explicit as possible.

### 1.2.2 Equality in the Sense of Zero Mean-Squared Error

Another issue worth discussing is the meaning given to a derived equality. For example, classical *Fourier analysis* is a useful tool for describing periodic functions. Such functions can be viewed as functions on the *unit circle*,  $S^1$ . The circle is a special case of a hypersphere,

---

<sup>3</sup>The open interval  $(a, b)$  is defined to be the set of values of  $x$  such that the strict inequalities  $a < x < b$  hold. In contrast, the closed interval  $[a, b]$  is defined by  $a \leq x \leq b$ . The half-open intervals  $(a, b]$  and  $[a, b)$  are defined by  $a < x \leq b$  and  $a \leq x < b$ , respectively.

$$S^{n-1} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}, \quad (1.10)$$

for the case when  $n = 2$ . (The above notation is read as “ $S^{n-1}$  is the set of all  $\mathbf{x} \in \mathbb{R}^n$  such that  $\|\mathbf{x}\| = 1$ .”) In the case when a sphere of radius  $r$  in  $n$ -dimensional space is of interest rather than a unit sphere, this is denoted as  $S_r^{n-1}$ .

The set of all real (or complex-valued) functions on  $S^1$  with an absolute value (or modulus) which when raised to the  $p$ th power and integrated yields a finite number is called  $L^p(S^1)$ :

$$\int_0^{2\pi} |f(\theta)|^p d\theta < \infty \quad \iff \quad f \in L^p(S^1). \quad (1.11)$$

The class of continuous functions on the unit circle is denoted as  $C^0(S^1)$ . Given a function that is both continuous and square-integrable on the unit circle (i.e.,  $f \in C^0(S^1) \cap L^2(S^1)$ ), classical Fourier analysis defines a *band limited* version of  $f$  with *band limit*  $N$  as

$$f_N(\theta) = \frac{1}{2\pi} \sum_{n=-N}^N \hat{f}(n) e^{in\theta} \quad \text{where} \quad \hat{f}(n) = \int_0^{2\pi} f(\theta) e^{-in\theta} d\theta \quad (1.12)$$

where  $i = \sqrt{-1}$ . The function  $f_N(\theta)$  should approach  $f(\theta)$  as  $N \rightarrow \infty$ . However, the sense of equality used here is that

$$\lim_{N \rightarrow \infty} \int_0^{2\pi} |f(\theta) - f_N(\theta)|^2 d\theta = 0. \quad (1.13)$$

When this condition holds, the Fourier series is said to converge, and the original function and its Fourier series are said to be equal *in the mean squared sense*.

Strictly speaking this is not the same thing as saying that  $f(\theta) = \lim_{N \rightarrow \infty} f_N(\theta)$ . For example, using a Fourier series to expand a step function that is equal to zero over the range  $0 \leq \theta \leq \pi$  and equal to unity otherwise will result in a Fourier expansion that satisfies (1.13) but exhibits Gibbs peaks around  $\theta = 0, \pi$ . At those points  $f(\theta) \neq \lim_{N \rightarrow \infty} f_N(\theta)$ .

The reason why this is relevant to the topic of this book is that many statements regarding *stochastic systems* (i.e., systems with some degree of noise) are statements about average behavior that are not strictly true on a pointwise basis. To make these statements absolutely precise requires a degree of mathematical rigor that places many texts on stochastic processes well outside the reach of the practitioner who seeks to model an engineering system or biological process. For this reason, the shortcuts taken in notation throughout this book are justified, but it needs to be understood up front that many of the stated equalities only hold in a limited sense.

### 1.2.3 Big-O Notation

Another way that the meaning of “equality” is distorted for the sake of convenience is by the use of “Big-O” notation. There are two variations to this notation. First, if  $f(x)$  and  $g(x)$  are real-valued functions on the real line, then [11]

$$f(x) = O(g(x)) \quad \text{as} \quad x \rightarrow \infty \quad (1.14)$$

if there exist positive constants  $M$  and  $x_0$  such that  $|f(x)| \leq M|g(x)|$  for all  $x > x_0$ . For example, the number of arithmetic operations used to multiply two  $n \times n$  matrices

$A = [a_{ij}]$  and  $B = [b_{jk}]$  using the formula  $c_{ik} = \sum_{j=1}^n a_{ij}b_{jk}$  is  $n^2 \cdot (2n - 1)$  since there are  $n^2$  values of  $i$  and  $k$  and the evaluation of this formula for each fixed  $i$  and  $k$  uses  $n$  multiplications and  $n - 1$  additions. Therefore

$$n^2 \cdot (2n - 1) = O(n^3).$$

It is implicit in writing this statement that  $n$  is a large number because for small values of  $n$  the computation would be fast anyway, in which case there would be no need to examine how it scales. Note that this “equality” is not equating two things that are exactly the same. Rather the right side is being used to summarize the essential features of the quantity on the left.

The second typical way that Big-O notation is used is [10]

$$f(x) = O(g(x)) \quad \text{as } x \rightarrow 0 \quad \iff \quad \lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = C \quad (1.15)$$

where  $0 < |C| < \infty$ .

For example, for a function that has a convergent *Taylor series expansion* about  $x = 0$ , it is common to write

$$f(x) = f(0) + xf'(0) + \frac{1}{2}x^2f''(0) + O(x^3).$$

The use of Big-O makes approximations such as  $f(x) \approx f(0) + xf'(0) + \frac{1}{2}x^2f''(0)$  more precise, since the amount of error in the approximation is quantified. On the other hand, if an equality sign is meant to convey the exactness of expressions on both sides, then the Big-O actually is not the right tool, since it destroys some information in return for the convenience of simpler expressions.

#### 1.2.4 A Philosophical View of Equality and Inequality

In addition to the way that mathematical “equality” can mean several different things, the inequality sign can also be used in nonstandard ways. For example, the elementary school metaphor in which the alligator’s open mouth,  $<$ , goes for the bigger number is not so easily applied when there are matrices on both sides of the inequality. In some treatments of matrix analysis, control systems, and information theory, the notation  $A < B$  is used when  $A$  and  $B$  are symmetric matrices of the same dimension. This is used to mean that all of the eigenvalues of the matrix  $B - A$  are greater than zero, or equivalently,  $\mathbf{x}^T A \mathbf{x} < \mathbf{x}^T B \mathbf{x}$  for any real vector  $\mathbf{x}$  of dimension compatible with  $A$  and  $B$ .

On a philosophical note, the concept of mathematical equality is not physically realizable anyway. Two real objects are never exactly equal. For example, two coins may have equal value (which is an artificial idealization that we impart on them, originally based on the observed weight of precious metals contained in them) but they will never have *exactly* the same weight, surface finish, etc. And so, mathematical equality is always either an approximation that holds within our ability to measure, or it is shorthand for an equivalence relation used as a proxy for true equality. For example, U.S. coins can be divided into pennies, nickels, etc., and the concept that every penny is equal to every other is really a statement that they are members of the same equivalence class, based on their monetary value. But actually, no two pennies are exactly the same.

Now in everyday life, a sphere is the surface of a ball. In some areas of pure mathematics, if one point is removed from a sphere, then it is considered to be a completely

different object. This is because, as *stereographic projection*<sup>4</sup> tells us, a so-called *punctured sphere* can be mapped to the plane, and vice versa. However, a full sphere cannot be mapped to the plane in this way. On the other hand, from the point of view of  $L^2$  equality of smooth functions, a function on the sphere and the same function restricted to the punctured sphere are indistinguishable, and therefore are in some sense equal.

The bottom line is that the sense in which an equality is stated depends on the context. In this book, when equalities such as Itô's formula are presented, it is important to note the sense of equality that is used. Likewise, in the study of stochastic differential equations, it is important to note that individual solutions (sample paths) do not have meaning. Only large ensembles of stochastic trajectories do. And the same ensemble characteristics can be achieved with different-looking stochastic differential equations. Therefore, exactly what it means for two such equations to be the "same" must be asked, because the resulting ensemble behavior reduces the discussion of equality of these equations to one of equivalence.

### 1.3 Other Useful Shortcuts

A number of shortcuts can be employed to make some of the basic mathematical ideas presented in this book more accessible.

#### 1.3.1 Simplifying Notation

It will often be the case that a family of functions is used to describe some phenomenon, and it will be convenient to hold one parameter fixed and examine the properties of an individual function in the family. In this context,  $f(x; a)$  means  $x$  is the variable and  $a$  is viewed as a fixed parameter that defines the particular function. For example, given a function of two variables  $f(x, y)$ , if  $y$  is fixed at the value  $a$ , then  $f(x; a) \doteq f(x, y = a)$ . This is more convenient than writing  $f_a(x)$  when the number of fixed variables becomes large. It also is useful to avoid notational clashes when there is a family of functions  $f_i(x; a)$ , such as when the subscript denotes the  $i$ th entry in a vector function.

Vectors, when viewed as a column array, will be denoted in bold. For example,  $\mathbf{x} \in \mathbb{R}^n$  (the  $n$ -dimensional space of all vectors with real entries). Whereas some books on mathematics denote Euclidean space as  $E^n$ , here no distinction is made between the vector space  $\mathbb{R}^n$  and  $n$ -dimensional Euclidean space. The use of vectors has been so thoroughly engrained into engineering education that the difference between the vector space  $\mathbb{R}^n$  and the underlying geometric object,  $E^n$ , is hardly worth mentioning.

Other shorthand notation is used throughout mathematics (and this book) to avoid needless complexity. However, whenever a shorthand is introduced, a new notation must be learned, and the possibility for misunderstanding or misusing the new notation exists. For this reason, the presentation throughout this book is very explicit. For example, whereas a pure mathematician would write Stokes' law for a manifold<sup>5</sup>  $M$  with boundary  $\partial M$  in the "light" (coordinate-free) form as

---

<sup>4</sup>This is a projection that identifies every point on a punctured sphere (with the point at the north pole removed), by having the south pole of the sphere sit on the plane, and connecting each point on the punctured sphere with a corresponding point on the plane. This is accomplished by constructing straight lines passing through the north pole, each of which intersects the punctured sphere and plane each exactly once.

<sup>5</sup>For now, think of this as a surface. A precise definition will come in Chapter 7.

$$\boxed{\int_M d\omega = \int_{\partial M} \omega} \quad (1.16)$$

the functional version of this expression that we will use will be expressed in parameters and take on a “heavy” (coordinate-dependent) appearance. (In fact, it is so heavy that it will not be presented here, for fear of scaring the reader away!) An expression of “intermediate weight” that is less general than (1.16), but can be used more easily with little training is the *divergence theorem* in  $\mathbb{R}^3$ :

$$\boxed{\int_V \left( \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3} \right) dV = \int_S (f_1 n_1 + f_2 n_2 + f_3 n_3) dS} \quad (1.17)$$

where  $dS$  is an element of surface area for the surface  $S$  that bounds the finite volume (or body)  $V$  with volume element  $dV$ ,  $n_i$  are the components of the outward-pointing unit normal for  $S$ , and  $f_i(\mathbf{x})$  are the components of a vector field defined over the body and the surface. All quantities are expressed in Cartesian coordinates  $x_1, x_2, x_3$ . Equation (1.17) is often abbreviated as

$$\int_V \nabla_{\mathbf{x}} \cdot \mathbf{f} dV = \int_S \mathbf{f} \cdot \mathbf{n} dS. \quad (1.18)$$

For the purpose of proving theorems, and gaining deep understanding, the light form may be preferred. For the purpose of performing computations with a minimal amount of new terminology and notation, the heavy form has value. Equations such as (1.18) are a convenient compromise.

### 1.3.2 Nice, Well-Behaved, Non-Pathological Functions

Many classes of functions exist. For example, the set of functions that are continuously differentiable  $n$  times on the interval  $(a, b)$  was denoted earlier as  $C^n(a, b)$ . And the functions on the unit circle with  $p$ th power of absolute value that integrates to a finite number form a set called  $L^p(S^1)$ .

More generally, for a continuous domain  $D$  contained in  $\mathbb{R}^n$ , a huge variety of different classes of functions can be defined that assign real values to each point in  $D$ . However, in order to avoid delving into the very precise terminology that is required to distinguish one class of functions from another, the scope here will be limited to “nice” functions. These functions can also be called “well-behaved,” or “non-pathological.” This means that the Taylor series of such functions will be convergent.<sup>6</sup> As a consequence, these nice functions are in  $C^\infty(D)$  (where the extension of the definition of  $C^n$  from the interval to a domain follows by simply requiring that partial derivatives of all orders result in continuous functions at every point in the domain). In addition, nice functions are defined here to be in  $L^p(D)$  for  $p = 1$  and  $p = 2$  (where the concept of  $L^p$  generalizes from the specific case of the circle to any continuous domain).

While this limits the scope of the presentation somewhat, functions that are encountered in many practical applications can be treated as being “nice” in the sense defined above. Mathematically, the class of nice functions on a domain  $D$  can be expressed as  $\mathcal{N}(D) = \mathcal{A}(D) \cap L^1(D) \cap L^2(D)$  or

$$\boxed{\mathcal{N}(D) = (\mathcal{A} \cap L^1 \cap L^2)(D)} \quad (1.19)$$

<sup>6</sup>Such functions are called *analytic*.

where  $\mathcal{A}(D)$  denotes the class of analytic functions on  $D$ . The symbols  $\cap$  denotes the intersection of these classes of functions, resulting in a more restricted class. The following section discusses mathematical terminology and symbols in greater detail.

## 1.4 Modern Mathematical Notation and Terminology

In undergraduate programs in engineering and the sciences, students are introduced to basic courses in Calculus, Linear Algebra, Differential Equations, and perhaps Fourier Analysis. These powerful areas of mathematics literally are now hundreds of years old. Very little of modern mathematics has worked its way into the modern undergraduate training of students in fields other than mathematics and physics. For this reason, some review of concepts and terminology from modern mathematics is in order, since one of the goals of this book is to introduce some of the useful concepts of modern mathematics to the practitioner.

Throughout this text many mathematical statements such as “A is true if and only if B is true,” “ $x \sim y \in X/Y \subset S$ ,” and “Z is topologically equivalent to W” will be made. The meaning of these and other ubiquitous mathematical terms are defined here. More specialized concepts are defined in the particular chapters where they are introduced and used. But first, a brief review of the different areas of modern mathematics is in order.

### 1.4.1 What is Modern Mathematics?

Roughly speaking, modern mathematics consists of a number of areas including (but not limited to) algebra, geometry, analysis, topology, probability and statistics, and number theory. The topics in this book draw from ideas in geometry, probability and statistics, and analysis, though some very basic ideas from algebra and topology are also used.

Algebra is concerned with sets that have associated with them specialized operations that recombine elements of the set in particular ways. Geometry is concerned with the shape of objects. In high dimensions such objects cannot be visualized, but it is still possible to define concepts of curvature to describe their shape at each point. Geometry and algebra have ancient roots. Analysis is concerned with the properties of functions on Euclidean space, such as continuity, differentiability and the approximation of functions using series expansions, such as the Fourier series discussed in Section 1.2. Topology is concerned with issues such as whether or not abstract spaces are *connected* (i.e., if it is possible to define a path within the space connecting any two points in the space), and if so how so. And given a closed path within the space, can it be shrunk to a point, or is there an inherent barrier to doing so imposed by the space? For example, any closed curve drawn on the surface of a sphere can be shrunk to a point. But in the case of closed curves drawn on a torus (surface of a donut), some can be shrunk to a point, and others cannot be (e.g., the circles resulting from the transverse intersection of a plane and the torus cannot be shrunk to a point without shrinking the torus itself). Topology is also concerned with the issue of whether or not spaces are *compact* (closed and bounded). The sphere and torus are examples of compact spaces. That is, they can be completely covered by a finite number of finite-area overlapping patches. The same cannot be said for the real line or the plane. Two spaces can be considered to be topologically equivalent if they have all topological features in common.

Topology of abstract spaces is studied by establishing relationships between a space under investigation and a known one. Topology and geometry also have some overlaps, especially when attempting to relate local and global features. Often techniques from analysis and algebra are used in geometry (resulting in differential geometry and algebraic geometry) as well as in topology (resulting in differential topology and algebraic topology). Some researchers have studied the interface between geometry and probability and statistics (see, e.g., [1]).

### 1.4.2 Stating Mathematical Results

In (1.11) the arrow  $A \iff B$  was used to denote that the statement on the left *implies* the one on the right, and vice versa. This is the same as saying that  $A$  is true *if and only if*  $B$  is true. Another shorthand for the same statement is to say “iff.” This statement consists of two parts: (1)  $A \implies B$ , which is read “ $A$  implies  $B$ ,” or “ $B$  is true if  $A$  is true,” or “if  $A$  is true then  $B$  is true”; and (2)  $A \impliedby B$ , which is read “ $A$  is implied by  $B$ ,” or “ $B$  is true *only if*  $A$  is true,” or “if  $B$  is true then  $A$  is true.” The words “ $B$  only if  $A$ ” mean that  $B$  cannot be true if  $A$  is not true. Since  $B$  being true automatically implies that  $A$  is, this makes sense.

Another way to articulate in words the same mathematical statements is by the use of the words *necessary* and *sufficient*. The statement  $A \implies B$  means that  $A$  being true is a sufficient condition for  $B$  to be true, since if  $A$  is true it implies that  $B$  is true also. On the other hand,  $B$  might be true independent of whether or not  $A$  is true. But if  $A$  is true, it is “enough” to guarantee that  $B$  is true. On the other hand,  $A \impliedby B$  (which is exactly the same as  $B \implies A$ ) means that  $B$  cannot be true without  $A$  also being true. Therefore  $A$  being true is “required” for  $B$  to be true since the truth of  $A$  results from the truth of  $B$ . That makes  $B$  a necessary condition.

Given any two statements,  $A$  and  $B$ , establishing that  $A \iff B$  (or equivalently  $A$  iff  $B$ ) is to say that  $A$  is necessary and sufficient for  $B$  (and  $B$  is necessary and sufficient for  $A$ ). Such a statement establishes that  $A$  and  $B$  are, in some sense, merely two different descriptions of the same underlying phenomenon. A nice summary of these concepts is provided in the appendix of [4].

When it comes to presenting mathematical results, there are several sorts of sub-headings. *Axioms* (also called *postulates*) are the minimal set of starting conditions that are obvious to all without proof. *Theorems* are the main mathematical points that are to be proven. *Lemmas* are lesser points that are proved on the way to proving theorems. And *corollaries* are interesting results that follow easily from the statement of a theorem. Also highlighted in mathematics books are definitions and remarks.

The style here will be to minimize the use of these. The section and subsection headings in each chapter will be subdivided finely enough that there is little need for further subdivision for the presentation of results. The emphasis will not be on proving theorems, but the illustration of how to use the results. However, in special cases when a particularly impressive theorem from the literature is reviewed it will be stated as such, and sometimes a sketch of the proof will be reproduced. This is particularly important when the proof is constructive, and therefore instructive regarding how to apply the result.

### 1.4.3 Sets and Mappings

As an elementary introduction, consider the scenario of a graduate course at an elite east-coast university in which there are five students. The collection (or *set*) of five

students in the course can be denoted as

$$S = \{s_1, s_2, s_3, s_4, s_5\}$$

where  $S$  is shorthand for “set” and “ $s_i$ ” is shorthand for “student  $i$ .” Membership in a set is denoted with the symbol  $\in$ , as  $s_i \in S$  for  $i = 1, \dots, 5$ . This is read as “ $s_i$  is *in*  $S$  for each value  $i = 1, i = 2$ , up to  $i = 5$ .” In set theory the particular ordering of elements within a set is not important. Viewed in the context of this example, this means that the students can sit anywhere in the classroom, and it is still the same class.

Consider the following set:  $G = \{m, f\}$  where  $G$  is shorthand for “gender,”  $m$  is shorthand for “male,” and  $f$  is shorthand for “female.” A *mapping* is an assignment of each element of one set to one element from a second set. For example, in the current context the mapping  $g : S \rightarrow G$  (which is read as “ $g$  takes elements of  $S$  and assigns one element of  $G$  to each”) is simply the evaluation  $g(s_i)$  that assesses the gender of each student.<sup>7</sup> In other words, if students 1 and 2 are female, and 3, 4, 5 are male, then  $g(s_1) = g(s_2) = f$  and  $g(s_3) = g(s_4) = g(s_5) = m$ .

The set of all students can be divided into two *subsets*, one consisting of female students, and the other males:

$$S_f = \{s_1, s_2\} \quad \text{and} \quad S_m = \{s_3, s_4, s_5\}.$$

Each of these subsets is *contained* in the original set. This is written as  $S_m \subseteq S$  and  $S_f \subseteq S$ . Since  $S_m \neq S$  and  $S_f \neq S$ , these two subsets are *strictly contained* in  $S$ , meaning that they are not equal to  $S$ . Such subsets are called *proper*. In this case the notation  $S_m \subset S$  and  $S_f \subset S$  is used. In cases where the possibility exists that a subset  $A$  might be equal to the whole set  $S$ , then the symbol  $A \subseteq S$  will be used. If  $A \subseteq S$  and  $S \subseteq A$ , then  $S = A$ . In contrast, if  $A \subseteq S$  and  $A \neq S$ , then this is when the notation  $A \subset S$  is used.<sup>8</sup>

In the particular example above, the original set can be reconstructed by pooling all of the elements of these two subsets. This pooling of elements of two subsets is called the *union*, and in this example  $S_m \cup S_f = S$ . In contrast, the two subsets in this example have no members in common. Therefore their *intersection* is  $S_m \cap S_f = \emptyset$  where the *empty set*  $\emptyset = \{\}$  is the set containing no elements. Since ordering of elements does not matter in the definition of a set, neither does the order in which the union or intersection of two subsets is computed. For example,  $S_m \cap S_f = S_f \cap S_m = \emptyset$  and  $S_m \cup S_f = S_f \cup S_m = S$ .

## Equivalence Relations and Equivalence Classes

The mapping  $g : S \rightarrow G$  can be viewed as having established two *equivalence classes*  $S_m$  and  $S_f$ , where members of  $S_m$  all share a common feature, as do members of  $S_f$ . This is denoted as  $s_1 \sim s_2$  since  $g(s_1) = g(s_2)$  and likewise  $s_3 \sim s_4 \sim s_5$  because  $g(s_3) = g(s_4) = g(s_5)$ . This is *not* the same as saying  $s_1 = s_2$ , etc. The symbol  $\sim$  is read as *similar to* or *equivalent to*, and is called an *equivalence relation*. Let  $i, j, k \in \{1, 2, 3, 4, 5\}$ . Then in this example,  $\sim$  has the following properties (which are true in general for equivalence relations): (1)  $s_i \sim s_i$  (the reflexive property); (2)  $s_i \sim s_k$  implies

<sup>7</sup>Note that the arrow,  $\rightarrow$ , of a mapping (which means “goes to”) does not have the same meaning as the logical  $\implies$  (which means “implies”).

<sup>8</sup>This is not standard in the literature. Sometimes  $\subsetneq$  is used for what is being denoted as  $\subset$  here, and sometimes  $\subset$  is used to denote what is being denoted  $\subseteq$  here.



$s_k \sim s_i$  (the symmetric property); (3)  $s_i \sim s_j$  and  $s_j \sim s_k$  implies  $s_i \sim s_k$  (the transitive property). In a sense, the original set is “broken up” (or *partitioned*) into two subsets by the equivalence relation induced by the mapping  $g$ . In the current context this can be written as  $\{S_m, S_f\} = S/G$  or  $\{S_m, S_f\} = S/g$  or  $\{S_m, S_f\} = S/\sim$ . In other words, a set of subsets of the original set is produced, the union of which is the original set, and the intersection of which is the empty set.

Now suppose that there is another set, the set of names of all students in the graduate program (not only the students in this class):

$$N = \{Abigail, Kathy, Matt, Kevin, Mary, Mike, Susan, \dots\}.$$

(The “...” here means that there are too many to write down explicitly.) Out of all of these names, five correspond to the names of students in the class. The mapping  $n : S \rightarrow N$  can be defined explicitly as  $n(s_1) = Abigail$ ,  $n(s_2) = Mary$ ,  $n(s_3) = Matt$ ,  $n(s_4) = Kevin$ , and  $n(s_5) = Mike$ .

### Images, Pre-Images, and Compositions of Mappings

The *image* of  $S$  in  $N$  is the set  $n(S) = \{n(s_1), n(s_2), \dots, n(s_5)\} \subseteq N$ . More generally, given a mapping from one abstract set into another,  $m : S_1 \rightarrow S_2$ ,

$$m(S_1) = \{m(\sigma) \in S_2 \mid \forall \sigma \in S_1\} \subseteq S_2. \quad (1.20)$$

This is read as “the image of the set  $S_1$  in  $S_2$  is the subset of  $S_2$  obtained by applying the mapping  $m$  to every element of  $S_1$ .” Here  $\sigma$  is a “dummy variable.” Its name is irrelevant. The use of the symbol  $m(S_1) \subseteq S_2$  here means that every element in  $m(S_1)$  is also in  $S_2$ , and the possibility exists that  $m(S_1) = S_2$ . The symbol  $\forall$  means “for all.”

Now consider the mapping from names to the Roman alphabet,  $\mathcal{A}$ , defined by extracting the first letter of each name as  $l : N \rightarrow \mathcal{A}$ . When this mapping is applied only to the names of students in the class,  $l : n(S) \rightarrow \{M, K, A\}$ . Explicitly,  $l(Matt) = l(Mike) = l(Mary) = M$ ,  $l(Kevin) = K$ , and  $l(Abigail) = A$ . The original set of students can then be broken into equivalence classes in which students who have the same first letter in their name are deemed equivalent. In this case the equivalence relation is defined by the *composed mapping*  $l(n(s_i))$ . First  $n$  is applied to  $s_i$ , and then  $l$  is applied to extract the first letter. This is denoted as  $(l \circ n)(s_i)$ .

It is also possible to look at any mapping of the form  $m : S_1 \rightarrow S_2$  from the opposite perspective, and ask which subset of elements of  $S_1$  map to the same element  $\tau \in S_2$ . Such elements are called the *pre-image* of  $\tau$  under the map. For example, the pre-image of the letter  $K$  under the mapping  $(l \circ n)(s_i)$  is  $\{s_4\}$  and the pre-image of  $M$  under the same composite map is  $\{s_2, s_3, s_5\}$ . This can be written as  $(l \circ n)^{-1}(M) = \{s_2, s_3, s_5\}$ . Note that  $(l \circ n)^{-1}$  is *not* an inverse mapping because  $(l \circ n)$  maps multiple elements of  $S$  to the same element of  $\{M, K, A\}$ , and therefore is not invertible. However, since applying  $(l \circ n)$  to  $S$  does “hit” all of the elements of  $\{M, K, A\}$ , the set of pre-images is sometimes denoted as

$$(l \circ n)^{-1}\{M, K, A\} \doteq \{\{s_1\}, \{s_4\}, \{s_2, s_3, s_5\}\}.$$

That is, the set of pre-images can be associated with the set of equivalence classes  $S/(l \circ n)$ .

### The Size of a Set and the Indicator Function

The number of elements in a finite set  $S$  is denoted as  $|S|$ . Given two finite sets,  $S_1$  and  $S_2$ , this has the properties that

$$|S_1 \cap S_2| \leq \min(|S_1|, |S_2|) \quad \text{and} \quad |S_1 \cup S_2| \geq \max(|S_1|, |S_2|)$$

with equality if and only if  $S_1 \subseteq S_2$ , or  $S_2 \subseteq S_1$ , or both (in which case  $S_1 = S_2$ ). These follow from the important equality

$$|S_1 \cup S_2| = |S_1| + |S_2| - |S_1 \cap S_2|. \quad (1.21)$$

From the definition of the empty set, it follows that  $|\emptyset| = 0$ . For continuous sets, such as the interior of a cube or sphere, expressions analogous to (1.21) hold where  $|\cdot|$  is replaced by  $Vol(\cdot)$ , the volume of the continuous set. The function  $|\cdot|$  takes sets as its arguments and returns non-negative real numbers,  $\mathbb{R}_{\geq 0}$ . This is not the only such function. For example, given a subset  $A \subset S$ , the *indicator function*,  $I_A : S \rightarrow \mathbb{R}_{\geq 0}$ , is defined for any  $x \in S$  as<sup>9</sup>

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases} \quad (1.22)$$

The indicator function has the properties

$$I_{A \cup B}(x) = I_A(x) + I_B(x) - I_{A \cap B}(x) \quad \text{and} \quad I_{A \cap B}(x) = I_A(x)I_B(x). \quad (1.23)$$

The concept of an indicator function is not limited to finite sets.

### Surjective, Injective, and Bijective Mappings

The mappings  $n : S \rightarrow N$ ,  $g : S \rightarrow G$ , and  $(l \circ n) \rightarrow \{M, K, A\}$  were from one set into another set of different size (having either more or fewer elements). In the case of  $(l \circ n)$  and  $g$ , every element of the set to which the arrow pointed was “used up.” In other words,  $g(S) = G$  and  $(l \circ n)(S) = \{M, K, A\}$ . More generally, if the image of a mapping  $m : S_1 \rightarrow S_2$  has the property that  $m(S_1) = S_2$ , then the mapping is called *onto*, or *surjective*. Therefore  $(l \circ n)$  and  $g$  are surjective. If a mapping  $m : S_1 \rightarrow S_2$  has the property that each element of the image  $m(S_1)$  corresponds to only one element in  $S_1$ , then  $m$  is called *one-to-one*, or *injective*. Stated another way, for an injective mapping  $m(\sigma_1) = m(\sigma_2) \implies \sigma_1 = \sigma_2$  for all  $\sigma_1, \sigma_2 \in S_1$ . Of the mappings examined above,  $n : S \rightarrow N$  is injective. A mapping that is both injective and surjective is called *bijective*, or *invertible*. None of the mappings  $n : S \rightarrow N$ ,  $g : S \rightarrow G$ , and  $(l \circ n) \rightarrow \{M, K, A\}$  can be bijective, because in each case the numbers of elements in the sets on both sides of the arrows are different. Therefore these mappings could not be bijective, and could not be inverted. As an example, if each student is assigned a number by the function  $\#(s_i) = i$ , then  $\# : S \rightarrow \{1, 2, 3, 4, 5\}$  is bijective. Also, if instead of the mapping  $n : S \rightarrow N$  (where  $N$  is the set of all names of people in the graduate program), a restricted mapping  $n : S \rightarrow n(S)$  is defined, then this will be bijective because  $n$  is injective, and the set to which this injective function maps has been pruned down to be the same size as the set from which it draws its arguments.

---

<sup>9</sup>Here  $\mathbb{R}_{>0}$  denotes the positive real numbers, and  $\mathbb{R}_{\geq 0}$  is therefore the non-negative real numbers. In some other texts these are referred to as  $\mathbb{R}^+$  and  $\mathbb{R}^+ \cup \{0\}$ , respectively.

When a mapping is of the form  $m : S_1 \rightarrow \mathbb{R}$  (the real numbers) or  $m : S_1 \rightarrow \mathbb{C}$  (the complex numbers), then the mapping is called a *function*. Sometimes the words *mapping* and *function* are used interchangeably, but when there is a difference, it is the one just mentioned. The concept of a set is not restricted to the case where there are a finite number of members. Indeed, most of the sets in this book are continuous sets. Continuous sets are usually called *spaces*.

## Products, Metrics, and Groups

It was already shown how a set can be “divided” into disjoint subsets by a mapping. It is also possible to form the product of two sets. Given two sets  $S_1$  and  $S_2$ , the *Cartesian product* is the set defined as

$$S_1 \times S_2 = \{(\sigma, \tau) \mid \sigma \in S_1, \tau \in S_2\}. \quad (1.24)$$

This is read as “ $S_1 \times S_2$  is the set consisting of all ordered pairs, the first entry of which runs over all elements of  $S_1$  and the second runs over all entries of  $S_2$ .” From this definition, in general  $S_1 \times S_2 \neq S_2 \times S_1$ , but  $|S_1 \times S_2| = |S_2 \times S_1| = |S_1| \cdot |S_2|$ .

The Cartesian product construction has several important simplifying effects when stating definitions. For example, suppose that some sense of distance exists between elements in a set  $S$ . Then a distance function is *not* defined on  $S$ , but rather on the Cartesian product of  $S$  with itself as  $d : S \times S \rightarrow \mathbb{R}_{\geq 0}$ . Recall that  $\mathbb{R}_{>0}$  denotes the positive real numbers, and  $\mathbb{R}_{\geq 0}$  denotes the non-negative real numbers. That is,  $d$  takes pairs of elements, each of which is drawn from  $S$  and returns a non-negative number. In addition, a valid *distance function* or *metric* must satisfy the following properties for any  $s_1, s_2, s_3 \in S$ :

$$d(s_1, s_2) \geq 0 \quad \text{with} \quad d(s_1, s_2) = 0 \quad \iff \quad s_1 = s_2 \quad (1.25)$$

$$d(s_1, s_2) = d(s_2, s_1) \quad (1.26)$$

and

$$d(s_1, s_2) + d(s_2, s_3) \geq d(s_1, s_3). \quad (1.27)$$

These properties are called positive definiteness, symmetry, and the triangle inequality.

As a concrete example of the concepts of Cartesian product and metric, consider

$$\mathbb{R}^n \doteq \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}}.$$

This is the usual space of  $n$ -dimensional vectors, and the usual metric,  $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ , is defined as

$$d(\mathbf{x}, \mathbf{y}) \doteq \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (1.28)$$

The Cartesian product construction also makes it possible to define *binary operations* that take in two elements of a set and return another element of that set:  $b : S \times S \rightarrow S$ . Such operations form the core of many definitions of mathematical objects. For example, a *group* is a nonempty set  $G$  together with a binary operation  $b : G \times G \rightarrow G$  such that there exists a special element  $e \in G$  with the property  $b(e, g) = b(g, e) = g$ ; for each

$g \in G$ , there is an element  $g^{-1} \in G$  such that  $b(g, g^{-1}) = b(g^{-1}, g) = e$ ; and for any three elements  $g_1, g_2, g_3 \in G$  the associative law holds:  $b(g_1, b(g_2, g_3)) = b(b(g_1, g_2), g_3)$ . As shorthand for this, the concept of a *group operation*,  $\circ$ , can be used to write  $b(g_1, g_2) = g_1 \circ g_2$ . Then, for example, the associative law can be written with fewer symbols as  $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ . A group will usually be denoted as  $(G, \circ)$ . In cases where the operation is obvious, the group can be referred to simply as  $G$ .

## Families of Sets and Valuations

A *family* is a set of sets,  $\mathcal{F} = \{S_1, S_2, S_3, \dots\}$ , that have common attributes. The family may be finite or infinite, and may even be uncountably infinite.<sup>10</sup> In the latter case, the constituent sets cannot be enumerated as was done in the preceding sentence. Unions and intersections can be taken over families as

$$\bigcup_{\alpha \in \mathcal{I}} S_\alpha \quad \text{and} \quad \bigcap_{\alpha \in \mathcal{I}} S_\alpha$$

where  $\mathcal{I}$  is the *indexing set*. For example, if  $\mathcal{F}$  is defined as in the first sentence of this paragraph, then

$$\bigcup_{\alpha \in \mathcal{I}} S_\alpha = S_1 \cup S_2 \cup \dots \quad \text{and} \quad \bigcap_{\alpha \in \mathcal{I}} S_\alpha = S_1 \cap S_2 \cap \dots$$

where the indexing set runs over all the subscripts,  $i$ .<sup>11</sup>

If  $\mathcal{L} = \{S_1, S_2, \dots\}$  is a special kind of family of subsets of a set  $S$  such that every element of  $\mathcal{L}$  can be constructed from finite unions or finite intersections of other elements of  $\mathcal{L}$ , and  $\forall i, j$   $S_i \cap S_j \in \mathcal{L}$  and  $S_i \cup S_j \in \mathcal{L}$ , then sometimes  $\mathcal{L}$  is called a *partially ordered lattice*. This means that subsets can be ordered according to which ones contain others. For example, if

$$S = \{s_1, s_2, s_3\} \quad \text{and} \quad \mathcal{L} = \{\{s_1\}, \{s_2\}, \{s_3\}, \{s_1, s_2\}, \{s_2, s_3\}, \{s_1, s_3\}, \{s_1, s_2, s_3\}\},$$

then

$$\{s_1\} \subset \{s_1, s_2\} \subset \{s_1, s_2, s_3\}, \quad \{s_2\} \subset \{s_1, s_2\} \subset \{s_1, s_2, s_3\}, \quad \text{etc.}$$

This is only a partial ordering because it is not possible to lay out all of the elements of  $\mathcal{L}$  in one sequential expression in which every element is contained in another. Any function  $\mu : \mathcal{L} \rightarrow \mathbb{R}_{\geq 0}$  that has the properties

$$\mu(S_i \cup S_j) = \mu(S_i) + \mu(S_j) - \mu(S_i \cap S_j) \quad \text{and} \quad \mu(\emptyset) = 0 \quad (1.29)$$

is called a *valuation* or additive measure on  $\mathcal{L}$ . Concrete examples of valuations include the volume of a finite body and the area of the surface that bounds a finite body. These can obviously be computed for the intersection of two bodies. For finite sets (1.21) is also an example of a valuation. While the set-indicator function has the similar looking property (1.23), its arguments are not sets in a partially ordered lattice, but rather are elements of arbitrary set.

<sup>10</sup>Uncountably infinite sets include, but are not limited to, continuous sets.

<sup>11</sup>A family of sets can also be defined by a non-countable index.

### 1.4.4 Transformation Groups

A special kind of bijective mapping from one set back into itself is a *transformation group*. This concept is defined formally in Volume 2. For now, a simple example suffices. Consider  $\mathbb{R}^2$ , the set of all two-dimensional real vectors. If  $\mathbf{x} \in \mathbb{R}^2$  and

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

denotes a  $2 \times 2$  rotation matrix, then  $R(\theta)\mathbf{x} \in \mathbb{R}^2$ . If  $\mathbb{R}^2$  is viewed as the union of an infinite number of concentric circles centered at the origin,  $\mathbf{0} = [0, 0]^T$ , then a point on the circle stays on the circle after the application of  $R(\theta)$ . Each of these circles is called an *orbit*. If  $G = \{R(\theta) \mid \forall \theta \in [0, 2\pi)\}$ , then the notation  $\mathbb{R}^2/G$  is used to denote the set of all such orbits. Membership in an orbit is an equivalence relation, and  $\mathbb{R}^2/G$  is the set of equivalence classes. The rotation matrices have the property

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2), \quad (1.30)$$

which means that they are closed under matrix multiplication, and furthermore,

$$R(\theta_1)[R(\theta_2)\mathbf{x}] = [R(\theta_1)R(\theta_2)]\mathbf{x}. \quad (1.31)$$

The property (1.30) implies closure (and even more than that, *commutativity*,  $R(\theta_1)R(\theta_2) = R(\theta_2)R(\theta_1)$ ). The property (1.31) is an example of a *group action*. Generally speaking, a group acting on a set either will divide that set into equivalence classes, or else it *acts transitively* (i.e., it can transform any element of the set into any other, in which case the whole set is one equivalence class).

The set  $G$  is an example of a *Lie group*. It is a continuous set of transformations that is closed under an associative operation (which in this case is matrix multiplication). It has an identity element ( $\theta = 0$ ) and every element  $R(\theta)$  has an inverse  $R^{-1}(\theta) = R(-\theta)$ . This group happens to be one-dimensional, commutative, and compact.

### 1.4.5 Understanding Commutative Diagrams

*Commutative diagrams* are used in many parts of modern mathematics to illustrate the relationship between different mappings. For example, consider the following example from linear algebra.<sup>12</sup> If  $\mathbf{x} \in \mathbb{R}^m$ , then pre-multiplication by a matrix  $B \in \mathbb{R}^{n \times m}$  produces a vector  $\mathbf{y} = B\mathbf{x} \in \mathbb{R}^n$ . If this vector in turn is pre-multiplied by  $A \in \mathbb{R}^{p \times n}$ , the result will be  $\mathbf{z} = A\mathbf{y} \in \mathbb{R}^p$ . The composite mapping has the properties

$$\mathbf{z} = A(B\mathbf{x}) = (AB)\mathbf{x}$$

that result from the associative property of matrix multiplication.

As a second example, consider two mappings  $f : U \rightarrow V$  and  $g : V \rightarrow W$  where  $U, V, W \subset \mathbb{R}$ . If  $y = f(x)$  and  $z = g(y)$ , they can be composed as  $z = (g \circ f)(x) = g(f(y))$ . In general, the mapping  $(g \circ f)$  will *not* be the same as  $(f \circ g)$ .

Both of the above examples can be illustrated with commutative diagrams:

---

<sup>12</sup>See the appendix for definitions.

$$\begin{array}{ccc}
 \mathbb{R}^m & \xrightarrow{B} & \mathbb{R}^n \\
 & \searrow AB & \downarrow A \\
 & & \mathbb{R}^p
 \end{array}
 \qquad
 \begin{array}{ccc}
 U & \xrightarrow{f} & V \\
 & \searrow g \circ f & \downarrow g \\
 & & W
 \end{array}
 \tag{1.32}$$

These diagrams have been given an equation number (as opposed to a figure caption) because they convey exactly the same information as equations.

Now, in light of the concept of a commutative diagram, consider the *chain rule*, which should be familiar to those who have gone through an undergraduate engineering or science program. Given a differentiable mapping  $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , classical multivariable calculus defines the differential

$$d\mathbf{f} = \mathbf{f}(\mathbf{x} + d\mathbf{x}) - \mathbf{f}(\mathbf{x})$$

(where  $d\mathbf{x}$  is infinitesimally small in the sense that  $\|d\mathbf{x}\|$  is almost zero), and provides the means to evaluate  $d\mathbf{f}$  as

$$d\mathbf{f} = D\mathbf{f}d\mathbf{x} \quad \text{where} \quad D\mathbf{f} = \left[ \frac{\partial f_i}{\partial x_j} \right] \in \mathbb{R}^{n \times m}. \tag{1.33}$$

The matrix  $D\mathbf{f}$  is called the *Jacobian*, and is sometimes denoted as  $J = \partial\mathbf{f}/\partial\mathbf{x}$ . For reasons that are explained in the appendix,  $\partial\mathbf{f}/\partial\mathbf{x}$  is not good notation. In contrast, it is better to write

$$D\mathbf{f} = \frac{\partial\mathbf{f}}{\partial\mathbf{x}^T},$$

where a raised  $T$  denotes the transpose of a vector. If  $m = n$ , the determinant of this matrix is denoted as

$$|D\mathbf{f}| = \det \frac{\partial\mathbf{f}}{\partial\mathbf{x}^T} = \frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)} = J(\mathbf{x}). \tag{1.34}$$

These are simply four different notations for the same thing. In different contexts each will have its advantages.

If  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ , then

$$\nabla_{\mathbf{x}}\phi \doteq \frac{\partial\phi}{\partial\mathbf{x}} \in \mathbb{R}^n. \tag{1.35}$$

This is the *gradient* of  $\phi(\mathbf{x})$ , and is a column vector (which is why the  $\mathbf{x}$  is not transposed in the denominator).

Given a second mapping,  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ , the composite function  $(\mathbf{g} \circ \mathbf{f}) : \mathbb{R}^m \rightarrow \mathbb{R}^p$  is evaluated by back substitution as  $(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = \mathbf{g}(\mathbf{y})$  where  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ . This can be written without introducing the variable  $\mathbf{y}$  as

$$(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x})).$$

The classical chain rule then specifies the following product of Jacobian matrices<sup>13</sup>

$$\boxed{d(\mathbf{g} \circ \mathbf{f}) = D\mathbf{g} D\mathbf{f} d\mathbf{x}} \quad \text{where} \quad \boxed{D\mathbf{g} = \left. \frac{\partial\mathbf{g}}{\partial\mathbf{y}^T} \right|_{\mathbf{y}=\mathbf{f}(\mathbf{x})} \in \mathbb{R}^{p \times n}} \tag{1.36}$$

<sup>13</sup>The notation  $\partial\mathbf{g}/\partial\mathbf{y}^T$  is explained in the appendix.

and  $D\mathbf{f}$  is defined as before. This can be visualized using a *commutative diagram* such as the one that follows.

$$\begin{array}{ccccc}
 \mathbf{x} \in \mathbb{R}^m & \xrightarrow{\mathbf{f}} & \mathbf{f}(\mathbf{x}) \in \mathbb{R}^n & \xrightarrow{\mathbf{g}} & (\mathbf{g} \circ \mathbf{f})(\mathbf{x}) \in \mathbb{R}^p \\
 \downarrow d & & \downarrow d & & \downarrow d \\
 d\mathbf{x} \in \mathbb{R}^m & \xrightarrow{D\mathbf{f}} & d\mathbf{f} \in \mathbb{R}^n & \xrightarrow{D\mathbf{g}} & d(\mathbf{g} \circ \mathbf{f}) \in \mathbb{R}^p
 \end{array} \tag{1.37}$$

The horizontal arrows at the top represent the application of the functions, and those at the bottom represent matrix multiplication. Both of these operations concatenate on the left side. The downward-pointing arrows denote the operation of taking a differential. The directions of the arrows indicate valid orders of operations. In this diagram three paths from start to finish are valid. These are “down-right-right,” “right-down-right,” and “right-right-down.” Each of these corresponds to the different quantities that can be equated where the tips of the arrows meet. In particular, these three paths give  $D\mathbf{g} D\mathbf{f} d\mathbf{x} = D\mathbf{g} d\mathbf{f} = d(\mathbf{g} \circ \mathbf{f})$ .

Of course, the chain rule can be understood perfectly well by (1.33) alone, but other more difficult concepts that will come later will be easier to understand with the assistance of commutative diagrams.

Equipped with these basic concepts, the remaining chapters are designed to contain sufficient descriptions to be understandable without prior knowledge (other than the concepts reviewed in the appendix) if read sequentially. In some instances it is possible to skip chapters, and then glean relevant definitions and results by occasionally skipping backwards using the pointers that are provided. More in-depth treatments of the topics reviewed in this introduction can be found in the classic books on modern algebra [3] and topology [9].

## 1.5 Transport Phenomena and Probability Flow

It will often be the case in problems discussed in later chapters that a time-evolving probability density function changes shape, and it is desirable to know how much probability is contained in a particular domain as a function of time.<sup>14</sup> A domain in  $\mathbb{R}^3$  with finite non-zero volume is called a finite body and is denoted as  $B$ . Its boundary is  $\partial B$ , and will be assumed here to be *smooth* (i.e., infinitely differentiable). The probability associated with the body  $B$  is

$$p(t) \doteq \int_{\mathbf{x} \in B \subset \mathbb{R}^n} f(\mathbf{x}, t) d(\mathbf{x})$$

where  $f(\mathbf{x}, t)$  is a probability density function. It could be that the body itself changes its size, shape, and/or location as a function of time in which case  $B = B(t)$ . When it is clear that it is static, then  $B = B_0 \doteq B(0)$ .

---

<sup>14</sup>The concept of a probability density function is defined and used in the next two chapters. The current section can be skipped if this concept is unfamiliar.

Here the notation  $d(\mathbf{x})$  is used to denote  $dx_1 dx_2 \cdots dx_n$ , the  $n$ -dimensional volume element. In  $\mathbb{R}^2$  it will sometimes be useful to denote this as  $dA$  (a differential area element) and in  $\mathbb{R}^3$  to call it  $dV$  (the usual three-dimensional differential volume element). The notation  $d(\mathbf{x})$  should not be confused with either the infinitesimal vector  $d\mathbf{x} = \mathbf{x}(t + dt) - \mathbf{x}(t)$ , or the metric  $d(\mathbf{x}, \mathbf{y})$  defined in (1.28). These are three very different things.

Often in applications it will be the case that  $f(\mathbf{x}, t)$  is the solution to a partial differential equation (i.e., a diffusion equation). And it would be convenient to use this fact to update  $p(t) \rightarrow p(t + dt)$ , without having to actually recompute the above  $n$ -dimensional integral.

The flow of probability density can be thought of as the flow of a fluid, or of heat. In this analogy, probability flows in and out of the body  $B$  by crossing its boundary. And in some applications, the body  $B$  itself may change with time. This makes the problem of rapidly updating the value of  $p(t)$  akin to problems in mechanics in which material and/or heat enter and leave a “control volume.” Such problems in  $\mathbb{R}^3$  involve the use of the divergence theorem and Stokes’ theorem (see the appendix for definitions). It makes sense, then, that the extension of these ideas to higher dimensions and in non-Euclidean spaces should be of interest in studying probability flow problems. And for this reason, the concept of differential forms and Stokes’ theorem as stated in (1.16) will be useful tools. But for now, some review of the mechanics of transport phenomena and heat flow will be instructive.

### 1.5.1 Continuum Mechanics

The field of mechanics is concerned with the interplay between forces, deformation, and motion of objects that have a smooth<sup>15</sup> distribution of mass over a finite volume. Such objects can be solid or fluid. And applied forces can be decomposed into normal and shear components. The essential difference between a solid and a fluid is that a solid resists all forces that attempt to deform it, while a fluid only resists normal forces and the rate with which shear forces are applied. A fluid will continue to deform (or flow) as long as a shear stress is applied, whereas a solid will not. However, both solids and fluids are continuous media that can be described using *continuum mechanics*. The review of continuum mechanics provided here follows the more detailed presentations in [6, 7, 8].

#### The Volume of a Deformable Body

Let  $B \subset \mathbb{R}^3$  denote a finite body. The points contained in  $B$ , denoted as  $\mathbf{x} \in B$ , can describe a solid object. Or, a fluid can flow in and out of  $B$  by crossing the boundary surface  $\partial B$ . The body  $B$  can change with time, and so  $B = B(t)$ . Denoting  $B_0 = B(0)$  as the initial body, and  $\mathbf{X} \in B_0$  as an arbitrary initial point,<sup>16</sup> then a deformation is a mapping  $\mathbf{x} : B_0 \rightarrow B$  that takes each point  $\mathbf{X} \in B_0$  and returns its new location  $\mathbf{x} \in B(t)$  at time  $t$ . In other words,  $\mathbf{x} = \mathbf{x}(\mathbf{X}, t)$ . The inverse of this mapping is  $\mathbf{X} = \mathbf{X}(\mathbf{x}, t)$ , and so

$$\mathbf{x} = \mathbf{x}(\mathbf{X}, t) \quad \text{and} \quad \mathbf{X} = \mathbf{X}(\mathbf{x}, t).$$

Computing the Jacobian matrices of these composed mappings gives

<sup>15</sup>Here all functions are  $C^\infty(\mathbb{R}^3)$ .

<sup>16</sup>Usually in this book upper-case letters are reserved for matrices or sets, but after 200 years of refinement,  $\mathbf{X}$  has become the standard notation for the initial/referential coordinates in mechanics problems. Since  $\mathbf{X}$  is bold, there should be no confusing  $\mathbf{X}$  with a matrix.



$$D\mathbf{x}D\mathbf{X} = D\mathbf{X}D\mathbf{x} = \mathbb{I}$$

where  $\mathbb{I}$  is the  $3 \times 3$  identity matrix.

The element of volume in the initial (or referential) state is related to the element of volume in the current state as

$$d(\mathbf{x}) = |D\mathbf{x}|d(\mathbf{X}) \quad \text{where} \quad |D\mathbf{x}| = \det[\partial x_i / \partial X_j] > 0.$$

The strict inequality  $|D\mathbf{x}| > 0$  holds for any physically feasible (or *admissible*) deformation. A second condition that every admissible deformation must satisfy is that it is injective. That is, two particles initially at two different locations in a body cannot be mapped under a deformation to the same point:

$$\mathbf{X}_1 \neq \mathbf{X}_2 \implies \mathbf{x}(\mathbf{X}_1, t) \neq \mathbf{x}(\mathbf{X}_2, t) \quad \forall \mathbf{X}_1, \mathbf{X}_2 \in B_0.$$

The volume of the body at the current time is

$$V(t) = \int_{B(t)} d(\mathbf{x}) = \int_{B_0} |D\mathbf{x}|d(\mathbf{X}).$$

And more generally, given a function of the current position of particles,  $f : B(t) \rightarrow \mathbb{R}$ , it can be represented in the referential coordinates as

$$\int_{\mathbf{x}(B_0, t)} f(\mathbf{x})d(\mathbf{x}) = \int_{B_0} f(\mathbf{x}(\mathbf{X}, t))|D\mathbf{x}|d(\mathbf{X}). \tag{1.38}$$

Or by defining  $\phi(\mathbf{X}) \doteq |D\mathbf{x}|(f \circ \mathbf{x})(\mathbf{X})$  and observing that  $B(t) = \mathbf{x}(B_0, t)$ , then the same thing can be written as

$$\int_{B(t)} \phi(\mathbf{X}(\mathbf{x}))|D\mathbf{X}|d(\mathbf{x}) = \int_{\mathbf{X}(B(t), t)} \phi(\mathbf{X})d(\mathbf{X}), \tag{1.39}$$

where  $\mathbf{X}(B(t), t) = B_0$ . These are both special cases of the *inverse function theorem* reviewed in the appendix.

In modern geometric terminology, the composed function  $f \circ \mathbf{x}$  (which for each fixed value of  $t$  is a function of  $\mathbf{X} \in B_0$ ), is called the *pull-back* of the function  $f$  (which is a function of  $\mathbf{x} \in B(t)$ ) via the mapping  $\mathbf{x} : B_0 \rightarrow B(t)$ . The word “pull-back” is used because it is describing the situation at the base end of the arrow in the mapping  $\mathbf{x} : B_0 \rightarrow B(t)$  starting with a function defined on the domain at the distal end of the arrow. In contrast, if  $\mathbf{c} : [0, 1] \rightarrow B_0$  is a curve segment in referential coordinates, then the curve segment that results after a deformation is  $\mathbf{x} \circ \mathbf{c} : [0, 1] \rightarrow B(t)$ . This is called the *push-forward* of the curve  $\mathbf{c}(t)$  via the mapping  $\mathbf{x} : B_0 \rightarrow B(t)$  because the new curve segment that results is obtained by following the direction of the arrow in the mapping.

Later it will be important to remember that the push-forward relates to curves and pull-backs relate to functions. And because tangent vectors can be associated with curves, and normal vectors can be associated with functions, push-forwards and pull-backs of vectors (and vector fields) can be defined. But it will not be until Chapters 6 and 7 that these concepts play an important role.

## Lagrangian vs. Eulerian Descriptions

In solid mechanics problems it is often convenient to use a *material* (or *Lagrangian*) *description* in which each material point is tracked from its initial position to its current position. For example, the velocity and acceleration of a material particle are calculated respectively as

$$\mathbf{v} = \frac{\partial \mathbf{x}}{\partial t} \quad \text{and} \quad \mathbf{a} = \frac{\partial^2 \mathbf{x}}{\partial t^2} \quad \text{where} \quad \mathbf{x} = \mathbf{x}(\mathbf{X}, t) \quad (1.40)$$

for a fixed value of  $\mathbf{X}$  corresponding to a particular material particle.

In contrast, in fluid mechanics problems it is often more convenient to examine the flow of material through a fixed point in space, rather than keeping track of individual material particles. This is known as the *spatial* (or *Eulerian*) *description*.

Associated with each body is a *mass density* function. This real-valued function is written in referential coordinates as  $\rho(\mathbf{X}, t)$ . In the spatial description, this same quantity would be written as

$$\rho^*(\mathbf{x}, t) \doteq \rho(\mathbf{X}(\mathbf{x}, t), t). \quad (1.41)$$

Other functions, which can be vector-valued or matrix/tensor-valued, can also be defined on a body. For any such function, which can be expressed in terms of its components as  $F(\mathbf{X}, t) = [F_{ij\dots}(\mathbf{X}, t)]$ , the full time derivative is

$$\frac{dF_{ij\dots}}{dt} = \frac{\partial F_{ij\dots}(\mathbf{X}, t)}{\partial t}.$$

In contrast, the same quantity viewed from the perspective of spatial coordinates will be

$$F^*(\mathbf{x}, t) \doteq F(\mathbf{X}(\mathbf{x}, t), t),$$

and the time derivative will be computed in this case according to the chain rule:

$$\frac{dF^*(\mathbf{x}, t)}{dt} = \frac{\partial F^*(\mathbf{x}, t)}{\partial t} + \sum_{i=1}^3 \frac{\partial F^*(\mathbf{x}, t)}{\partial x_i} \frac{\partial x_i}{\partial t}.$$

But from (1.40) this can be written as

$$\boxed{\frac{dF^*(\mathbf{x}, t)}{dt} = \frac{\partial F^*(\mathbf{x}, t)}{\partial t} + \mathbf{v}^* \cdot \nabla_{\mathbf{x}} F^*(\mathbf{x}, t)} \quad (1.42)$$

where  $\mathbf{v}^*(\mathbf{x}, t) \doteq \mathbf{v}(\mathbf{X}, t)$  and the  $\cdot$  is the scalar (dot) product.

If  $F^*$  is viewed as a  $k$ -dimensional array with three entries in each dimension, the  $\nabla_{\mathbf{x}}$  operation makes  $\nabla_{\mathbf{x}} F^*(\mathbf{x}, t)$  a  $(k+1)$ -dimensional array, and the  $\mathbf{v}^* \cdot$  operation reduces this back down to  $k$  dimensions.

For example, if  $F^*(\mathbf{x}, t)$  is replaced with  $\mathbf{v}^*(\mathbf{x}, t)$ , then the Eulerian description of the acceleration of fluid flowing through the position  $\mathbf{x}$  at time  $t$  becomes

$$\mathbf{a}^*(\mathbf{x}, t) = \frac{d\mathbf{v}^*(\mathbf{x}, t)}{dt} = \frac{\partial \mathbf{v}^*(\mathbf{x}, t)}{\partial t} + \mathbf{v}^* \cdot \nabla_{\mathbf{x}} \mathbf{v}^*(\mathbf{x}, t).$$

### Mass Balance

If particles of mass in a continuum are tracked and  $B(t)$  evolves so as to include all of the original particles in  $B_0$  and no others, then even if  $B$  changes in size and shape with time, it must be the case that

$$\int_{B(t)} \rho^*(\mathbf{x}, t) d(\mathbf{x}) = \int_{B_0} \rho(\mathbf{X}, 0) d(\mathbf{X}). \quad (1.43)$$

This is a *physical* statement rather than a mathematical one. Since the quantity on the right side is a constant, another way to write this same thing is

$$\frac{d}{dt} \int_{B(t)} \rho^*(\mathbf{x}, t) d(\mathbf{x}) = 0.$$

This is the conservation of mass integral.

The equality in (1.43) can be localized by observing (1.38) for  $f(\mathbf{x}) = \rho^*(\mathbf{x}, t)$  for each fixed  $t$  to write

$$\int_{B(t)} \rho^*(\mathbf{x}, t) d(\mathbf{x}) = \int_{B_0} \rho^*(\mathbf{x}(\mathbf{X}, t), t) |D\mathbf{x}| d(\mathbf{X}). \quad (1.44)$$

In other words, combining (1.43) and (1.44) yields

$$\int_{B_0} \{\rho^*(\mathbf{x}(\mathbf{X}, t), t) |D\mathbf{x}| - \rho(\mathbf{X}, 0)\} d(\mathbf{X}) = 0.$$

Observing that this must be true for any initial volume  $B_0$  and using the definition in (1.41) means that the integrand can be localized as

$$\boxed{\rho(\mathbf{X}, t) |D\mathbf{x}| = \rho(\mathbf{X}, 0)}. \quad (1.45)$$

In the spatial-coordinate version of the conservation of mass, an integral over a *fixed body*,  $B^* = B_0$ , is computed. The mass inside of this fixed body is

$$M = \int_{B^*} \rho^*(\mathbf{x}, t) d(\mathbf{x}).$$

Since the body is fixed, so too is its boundary. The time rate of change of mass inside of the fixed volume is

$$\frac{d}{dt} \int_{B^*} \rho^*(\mathbf{x}, t) d(\mathbf{x}) = \int_{B^*} \frac{\partial \rho^*(\mathbf{x}, t)}{\partial t} d(\mathbf{x}).$$

Since mass is neither created nor destroyed, and since the spatial volume  $B^*$  is fixed, the only way that  $dM/dt$  can be non-zero is if mass enters or leaves through the boundary of  $B$ . In other words,

$$\frac{dM}{dt} = - \int_{\partial B^*} \rho \mathbf{v} \cdot \mathbf{n} dS$$

where  $\mathbf{n}$  is the outward-pointing surface normal and  $\mathbf{v}(\mathbf{x}, t)$  is the velocity of the fluid. Combining the above results, and converting the surface integral to a volume integral using the divergence theorem in (1.18) gives

$$\int_{B^*} \left\{ \frac{\partial \rho^*}{\partial t} + \nabla_{\mathbf{x}} \cdot (\rho^* \mathbf{v}^*) \right\} d(\mathbf{x}) = 0.$$

Since this must hold for any fixed volume  $B^*$ , it can be localized as

$$\boxed{\frac{\partial \rho^*}{\partial t} + \nabla_{\mathbf{x}} \cdot (\rho^* \mathbf{v}^*) = 0.} \quad (1.46)$$

This is the continuity (or conservation of mass) equation in spatial coordinates.

### The Reynolds Transport Theorem

Given any scalar, vector, or tensor quantity that is not spontaneously created or destroyed, the same argument that was used above for mass density can be used to write a balance equation. Let  $F^*(\mathbf{x}, t)$  be the quantity of interest. Then [7, 8]

$$\frac{d}{dt} \int_{B^*} F^* d(\mathbf{x}) = \int_{B^*} \frac{\partial F^*}{\partial t} d(\mathbf{x}) + \int_{\partial B^*} (\mathbf{v}^* \cdot \mathbf{n}) F^* dS. \quad (1.47)$$

The right-hand side of the above equation can be written as a single volume integral by using the divergence theorem. The result is

$$\frac{d}{dt} \int_{B^*} F^* d(\mathbf{x}) = \int_{B^*} \left\{ \frac{\partial F^*}{\partial t} + \mathbf{v}^* \cdot (\nabla_{\mathbf{x}} F^*) \right\} d(\mathbf{x}). \quad (1.48)$$

Now suppose that  $F^* = \rho^* \Theta^*$  where  $\rho^* = \rho^*(\mathbf{x}, t)$  and  $\Theta^* = \Theta^*(\mathbf{x}, t)$ . Using the continuity equation (1.46), it can be shown (see Exercise 1.10) that (1.48) simplifies to

$$\boxed{\frac{d}{dt} \int_{B^*} \rho^* \Theta^* d(\mathbf{x}) = \int_{B^*} \rho^* \frac{d\Theta^*}{dt} d(\mathbf{x})} \quad (1.49)$$

where both of the time derivatives in this expression are full derivatives. This is the *Reynolds transport theorem* in its simplest form. It is used extensively in mechanics to “bring  $d/dt$  inside the integral.” Some specific cases are illustrated in the following subsections.

### Momentum Balance

Newton’s second law states that for a single particle,  $\mathbf{f} = d(m\mathbf{v})/dt$  where  $\mathbf{x} = d\mathbf{x}/dt$  is the position of the particle as measured in an inertial reference frame.<sup>17</sup> Based on this, it has been postulated that for a continuum, the time rate of change of the total momentum flowing into and out of a fixed control volume must be equal to the applied forces. These forces are broken into two categories: (a) those that act on the surface of the control volume (e.g., the forces imposed by restraints to keep the control volume fixed in space) and (b) those that act directly on the interior (e.g., gravity or electromagnetic forces). Given a fixed control volume, the *momentum balance* equation is written as

$$\int_{\partial B^*} \mathbf{t}^* dS + \int_{B^*} \rho^* \mathbf{b}^* d(\mathbf{x}) = \frac{d}{dt} \int_{B^*} \rho^* \mathbf{v}^* d(\mathbf{x}) \quad (1.50)$$

<sup>17</sup>That is, a reference frame that is not accelerating, which means that it is either not moving, or moving in pure translation with a constant velocity relative to a frame fixed in space.

where  $\mathbf{t}^*(\mathbf{x}, t)$  for  $\mathbf{x} \in \partial B^*$  is the so-called surface traction (force per unit area) acting on the boundary surface, and  $\mathbf{b}^*(\mathbf{x}, t)$  is the force per unit mass acting on each point in the volume,  $\mathbf{x} \in B^*$ .

The Reynolds transport theorem (1.49) can be used to transform the right side of (1.50) with  $\Theta^* = \mathbf{v}^*$  by bringing the time derivative under the integral, and the divergence theorem can be used to convert the surface integral to an integral over  $B^*$ . From there, the equations can be localized.

### Angular Momentum Balance

For a collection of particles, each obeying Newton's second law,  $\mathbf{f}_i = d(m_i \mathbf{v}_i)/dt$ , the angular momentum is defined as  $\mathbf{L} = \sum_i \mathbf{x}_i \times m_i \mathbf{v}_i$ . The time rate of change of this angular momentum is equal to the moment due to the applied forces:  $d\mathbf{L}/dt = \sum_i \mathbf{x}_i \times \mathbf{f}_i$ .

Angular momentum (or moment of momentum) is postulated to balance in a similar way for a continuum:

$$\int_{\partial B^*} \mathbf{x} \times \mathbf{t}^* dS + \int_{B^*} \mathbf{x} \times \rho^* \mathbf{b}^* d(\mathbf{x}) = \frac{d}{dt} \int_{B^*} \mathbf{x} \times \rho^* \mathbf{v}^* d(\mathbf{x}). \quad (1.51)$$

Again, the divergence theorem and the Reynolds transport theorem (now with  $\Theta^* = \mathbf{x} \times \mathbf{v}^*$ ) can be used to convert this to a localized form.

Continuum mechanics also takes into account the balance of energy that enters and exits a control volume. But this important topic will not be addressed here, since it does not relate to the remainder of the book.

### 1.5.2 Heat Flow and Entropy

In later chapters, concepts of probability flow and information-theoretic entropy are defined and used heavily. These concepts can be related to analogies in the physical world. While analogies are often useful in gaining understanding, limits exist where they break down. For example, the statement “doing the exercises in Chapter 1 of Chirikjian's book is a piece of cake” is an analogy (or metaphor) indicating how easy the problems are. However, the analogy breaks down at several levels, including the fact that eating cake carries calories, but solving exercises burns them. For this reason, physical quantities analogous to probability flow and information-theoretic entropy are reviewed in this section, but limits of these analogies should be kept in mind.

### Heat Conduction

In *heat conduction* problems, a function called the (absolute) *temperature*, which is denoted as  $\vartheta(\mathbf{x}, t)$ , is defined on a solid body  $B$  at each value of time. That is,  $\vartheta : B \times (\mathbb{R}_{\geq 0}) \rightarrow \mathbb{R}_{\geq 0}$ . If the body is surrounded by other solid material, then heat can pass through the boundary  $\partial B$  by a physical phenomenon known as *thermal conduction* [11, 12]. In other words, when a solid object that is cold is placed in contact with a solid object that is hot, over a period of time the temperature of the two bodies will tend to average out, where the exact value of the average will depend on the materials that constitute the bodies, their relative sizes, and their initial temperatures. This averaging process is due to the exchange of a physical quantity known as *heat*. The flow of heat is governed by *Fourier's law of heat conduction*:<sup>18</sup>

<sup>18</sup>The same equation governs molecular diffusion processes, but in that context it goes under the name of *Fick's law*.

$$\mathbf{q} = -K \operatorname{grad}(\vartheta) \quad (1.52)$$

where  $K : \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 3}$  is a symmetric *matrix-valued function* with positive eigenvalues (see appendix for definition) called the *thermal conductivity matrix*. The *gradient*  $\operatorname{grad}(\vartheta) = \nabla_{\mathbf{x}}\vartheta = \partial\vartheta/\partial\mathbf{x}$  points in the direction of maximal temperature. The negative sign then dictates that in the special case when  $K = k(\mathbf{x})\mathbb{I}$  (where  $\mathbb{I}$  is the  $3 \times 3$  identity matrix and  $k : \mathbb{R}^3 \rightarrow \mathbb{R}_{>0}$ ), the *heat current density*  $\mathbf{q}(\mathbf{x}, t)$  points in the direction of minimal temperature. But more generally, this direction can be altered by  $K(\mathbf{x})$  because the body may not be homogeneous, and different materials transmit heat more readily than others.

If there is no internal production of heat in the body, then there must be a balance of heat entering and leaving the body. In conduction problems, the only way for heat to enter or leave the body is by flow across its boundary. As heat enters the body, the average temperature rises, and as heat leaves, the average temperature falls. This is captured by the balance equation

$$\int_B c(\mathbf{x})\rho(\mathbf{x})\vartheta(\mathbf{x}, t) dV = - \int_0^t \int_{\partial B} \mathbf{q} \cdot \mathbf{n} dS dt$$

where  $c(\mathbf{x})$  is the heat capacity per unit mass and  $\rho(\mathbf{x})$  is the mass density per unit volume within the body.<sup>19</sup> The quantity  $\mathbf{q} \cdot \mathbf{n}$  is the heat flux crossing the boundary, where  $\mathbf{n}$  is the outward-pointing normal.

When the body itself does not change with time (i.e., it does not change in size, shape, or location), differentiating both sides of the above equation with respect to time gives

$$\int_B c\rho \frac{\partial\vartheta}{\partial t} dV = - \int_{\partial B} \mathbf{q} \cdot \mathbf{n} dS.$$

The term on the right can be converted to an integral over volume using the divergence theorem. Then, if everything is moved to the left side of the equation, this gives

$$\int_B \left\{ c\rho \frac{\partial\vartheta}{\partial t} + \operatorname{div}(\mathbf{q}) \right\} dV = 0$$

where  $\operatorname{div}(\mathbf{q})$  and  $\nabla_{\mathbf{x}} \cdot \mathbf{q}$  are simply different ways of writing the same thing. Since this equation holds for any body  $B$ , the integrand can be localized as

$$c\rho \frac{\partial\vartheta}{\partial t} = -\operatorname{div}(\mathbf{q}). \quad (1.53)$$

This equation is the *continuity equation for heat flow*.

Combining (1.53) and (1.52) results in the *heat equation*

$$\boxed{c\rho \frac{\partial\vartheta}{\partial t} = \operatorname{div}(K \operatorname{grad}(\vartheta))}. \quad (1.54)$$

When  $K = k\mathbb{I}$  and  $k, c, \rho$  are all constant, this reduces to

$$\frac{\partial\vartheta}{\partial t} = \kappa \nabla^2 \vartheta$$

where  $\kappa = k/c\rho$  and  $\nabla^2 \vartheta = \sum_{i=1}^3 \partial^2 \vartheta / \partial x_i^2$ . The solutions to this equation subject to particular initial conditions will be examined in detail in the next chapter.

<sup>19</sup>In the elementary problem considered here, the body is assumed to not change shape, and so  $\mathbf{x} = \mathbf{X}$ ,  $B = B^*$ , and  $\rho(\mathbf{x}, t) = \rho(\mathbf{x})$ .

## Thermodynamic Entropy

It was postulated in the mid-nineteenth century that a physical quantity called *entropy* exists. Given a body (or control volume),  $B^*$ , the total entropy inside of  $B^*$  is denoted as

$$S = \int_{B^*} s^*(\mathbf{x}, t) \rho^*(\mathbf{x}, t) d(\mathbf{x})$$

where the integrand on the right side is expressed using the Eulerian description, and  $s^*(\mathbf{x}, t)$  is the entropy per unit mass at the point  $\mathbf{x}$  and time  $t$ . The *Clausius–Duhem inequality* states [8]:

$$\frac{dS}{dt} \geq \int_{B^*} \frac{r}{\vartheta} \rho^* dV - \int_{\partial B^*} \frac{1}{\vartheta} \mathbf{q} \cdot \mathbf{n} dS \quad (1.55)$$

where  $r$  is the internal heat supply per unit mass and time, and all other quantities are the same as defined earlier.

If the system is closed, so that no heat is imported or exported across  $\partial B^*$ , and if a positive heat is produced internally (for example, due to a chemical reaction such as combustion), then the entropy of the system contained in  $B^*$  will increase. This is a version of the *Second Law of Thermodynamics*.

In future chapters, analogous expressions such as  $dS/dt \geq 0$  will be observed where  $S$  is information-theoretic entropy defined for a given probability density function that solves a diffusion equation analogous to the heat equation in (1.54).

## 1.6 Organization of this Book

Chapter 2 provides a detailed description of the Gaussian distribution on the real line, and in  $\mathbb{R}^n$ . The adaptation of Gaussian distributions to finite domains by folding or clipping and renormalizing is discussed. The relationship between the Gaussian distribution and the heat equation on the real line, on the circle, and on  $\mathbb{R}^n$  is examined. Symmetries of the heat equation are also discussed briefly.

With concrete examples of probability density functions in hand from Chapter 2, the general definitions and theorems of probability and information theory in Euclidean space are presented in Chapter 3. These include the concept of mean, variance, conditional expectation, Fisher information, Cramér–Rao bound, entropy power law, central limit theorem, etc.

Chapter 4 provides an introduction to stochastic differential equations (SDEs) from the perspective of mathematical modeling. Given an ordinary differential equation that is forced by noise, what will the ensemble behavior of the resulting sample paths be? The Wiener process is examined, the increments of which are used to define white-noise forcing. All other kinds of noise used throughout this book are built on this noise concept. The relationship between the Itô and Stratonovich forms of an SDE are explained. The corresponding Fokker–Planck equations are derived. It is shown how changes in coordinate systems affect these equations.

Ultimately, the main difference between this book and others on stochastic modeling is that the problems addressed here involve random processes that evolve on geometric objects. Therefore, Chapter 5 presents a self-contained review of geometry, starting with analytic geometry from a parametric and algebraic perspective. Then the local and global differential geometry of curves and surfaces in three-dimensional Euclidean space are developed.

A fundamental difference between high-dimensional spaces and  $\mathbb{R}^3$  is that the concept of the vector cross product needs to be modified. This is critical, for example, in the derivation of formulas for the high-dimensional analogues of surface area and the curl operator. Differential forms are a useful tool in this regard, and Chapter 6 serves as an introduction to this topic.

When describing the motion of complicated objects such as robots and biological organisms, keeping track of their constituent parts can be viewed as the motion of a point in a high-dimensional space. This necessarily transforms the original problem into one of geometry on manifolds. This is the subject of Chapter 7. Differential forms are shown to be the natural tool to use to integrate on manifolds, as well as to define intrinsic geometric features such as curvature.

Chapter 8 addresses stochastic differential equations on manifolds, and shows how Fokker–Planck equations are derived in this setting. Brownian motion on the sphere is used as an illustrative example.

The appendix reviews linear algebra, multivariate calculus, and systems theory.

Volume 2 in this collection will focus on the concept of Lie groups and will apply the methods developed in the current volume to that setting. It will also address how to solve problems in engineering and biology such as those described earlier in this chapter, as well as many more.

## 1.7 Chapter Summary

This chapter introduced a number of concepts from mathematics and mechanics. Terminology that will be used throughout the book has been established. Intuitive (mechanical) ideas related to fluid and heat flow were presented to serve as physical analogies that can be referred back to when examining more abstract problems. The exercises that follow will reinforce the ideas discussed here. Additional reading in any of the topics presented may be helpful. A substantial list of references is provided at the end of each chapter.

The next chapter will discuss the Gaussian distribution. This is an important topic in probability and statistics. The approach is concrete and explicit. Once familiarity with every aspect of the Gaussian distribution is mastered, then more general (and hence more abstract) presentations of probability, information theory, and stochastic processes will follow.

## 1.8 Exercises

1.1. Let  $f(\theta) = \theta$  for  $\theta \in [0, 2\pi)$ . Calculate the Fourier coefficients by hand and plot the Fourier series approximation  $f_N(\theta)$  in (1.12) for  $N = 2, 5, 100$ .

1.2. Indicate whether or not the following functions from  $\mathbb{R}$  to  $\mathbb{R}$  are injective, surjective, or bijective: (a)  $f(x) = x$ ; (b)  $f(x) = x^2$ ; (c)  $f(x) = x^3$ ; (d)  $f(x) = e^x$ .

1.3. Determine which of the following are valid metric functions on the plane  $\mathbb{R}^2$ :  
 (a)  $\phi_1(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + |x_2 - y_2|$ ; (b)  $\phi_2(\mathbf{x}, \mathbf{y}) = [(x_1 - y_1)^2 + (x_2 - y_2)^2]^{\frac{1}{2}}$ ;  
 (c)  $\phi_3(\mathbf{x}, \mathbf{y}) = [(x_1 - y_1)^3 + (x_2 - y_2)^3]^{\frac{1}{3}}$ ; (d)

$$\phi_4(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} \neq \mathbf{y} \\ 0 & \text{if } \mathbf{x} = \mathbf{y}. \end{cases}$$



1.4. Show that the matrices  $R(\theta)$  in (1.30) form a group under the operation of matrix multiplication.

1.5. Show that the set of matrices of the form  $g(x, y, \theta)$  in (1.1) form a group under the operation of matrix multiplication where  $(x, y, \theta) \in \mathbb{R}^2 \times [0, 2\pi)$ .

1.6. By recursively applying (1.29) prove that for any  $\mathcal{L} = \{S_1, S_2, \dots\}$ , a valuation satisfies the *inclusion-exclusion* relationship

$$\mu(S_1 \cup S_2 \cup \dots \cup S_n) = \sum_{k=1}^n \alpha_k \sum_{i_1 < i_2 < \dots < i_k} \mu(S_{i_1} \cap S_{i_2} \cap \dots \cap S_{i_k}) \quad (1.56)$$

where  $\alpha_k$  is a function of  $k$  that you will determine.

1.7. Let  $\mathbf{f} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  be defined by  $\mathbf{f}(\mathbf{x}) = [x_1^2 + 2x_2, x_1x_3]^T$  and  $\mathbf{g} : \mathbb{R}^2 \rightarrow \mathbb{R}$  be defined by  $g(\mathbf{y}) = y_1 \sin y_2$  for  $\mathbf{y} \in \mathbb{R}^2$  and  $\mathbf{x} \in \mathbb{R}^3$ . Calculate the Jacobians for these functions and demonstrate that the commutative diagram in (1.37) holds.

1.8. Show that when  $\alpha > 0$  and  $n \in \{1, 2, 3, \dots\}$ , each member of the family of functions  $\{\phi_{\alpha, n}(x)\}$  defined in (1.6) is smooth (i.e., all of its derivatives exist). What is the Taylor series of  $\phi_{\alpha, n}(x)$  about  $x = 0$ ?

1.9. Prove that for  $\mathbf{x} = \mathbf{x}(\mathbf{X}, t)$ , the Jacobian determinant  $j = |D\mathbf{x}|$  satisfies the differential equation

$$\frac{\partial j}{\partial t} = j \nabla_{\mathbf{x}} \cdot \mathbf{v}^*.$$

1.10. Prove the Reynolds transport theorem as stated in (1.49).

## References

1. Amari, S., Nagaoka, H., *Methods of Information Geometry*. Translations of Mathematical Monographs 191, American Mathematical Society, Providence, RI, 2000.
2. Arfken, G.B., Weber, H.J., *Mathematical Methods for Physicists*, 6th ed., Academic Press, San Diego, 2005.
3. Birkhoff, G., MacLane, S., *A Survey of Modern Algebra*, A.K. Peters, Wellesley, MA, 1997.
4. Joshi, A.W., *Matrices and Tensors in Physics*, 3rd ed., John Wiley and Sons, New York, 1995.
5. Kreyszig, E., *Advanced Engineering Mathematics*, 9th ed., John Wiley and Sons, New York, 2005.
6. Lai, W.M., Rubin, D., Krempl, E., *Introduction to Continuum Mechanics*, 3rd ed., Butterworth Heineman, New York, 1996.
7. Mace, G.E., *Continuum Mechanics*, Schaum's Outline Series, McGraw-Hill, New York, 1970.
8. Malvern, L.E., *Introduction to the Mechanics of a Continuous Medium*, Prentice-Hall, Englewood Cliffs, NJ, 1969.
9. Munkres, J.R., *Topology*, 2nd ed., Prentice-Hall, Upper Saddle River, NJ, 2000.
10. Neyfeh, A., *Problems in Perturbation*, John Wiley and Sons, New York, 1985.
11. Pinsky, M.A., *Introduction to Partial Differential Equations with Applications*, McGraw-Hill, New York, 1984.
12. Pitts, D.R., Sissom, L.E., *Heat Transfer*, Schaum's Outline Series, McGraw-Hill, New York, 1977.
13. Zhou, Y., Chirikjian, G.S., "Probabilistic models of dead-reckoning error in nonholonomic mobile robots," *Proceedings of the IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, Sept 14–19, 2003, pp. 1594–1599.

---

## Gaussian Distributions and the Heat Equation

In this chapter the Gaussian distribution is defined and its properties are explored. The chapter starts with the definition of a Gaussian distribution on the real line. In the process of exploring the properties of the Gaussian on the line, the Fourier transform and heat equation are introduced, and their relationship to the Gaussian is developed. The Gaussian distribution in multiple dimensions is defined, as are clipped and folded versions of this distribution. Some concepts from probability and statistics such as mean, variance, marginalization, and conditioning of probability densities are introduced in a concrete way using the Gaussian as the primary example. The properties of the Gaussian distribution are fundamental to understanding the concept of white noise, which is the driving process for all of the stochastic processes studied in this book.

The main things to take away from this chapter are:

- To become familiar with the Gaussian distribution and its properties, and to be comfortable in performing integrals involving multi-dimensional Gaussians;
- To become acquainted with the concepts of mean, covariance, and information-theoretic entropy;
- To understand how to marginalize and convolve probability densities, to compute conditional densities, and to fold and clip Gaussians;
- To observe that there is a relationship between Gaussian distributions and the heat equation;
- To know where to begin if presented with a diffusion equation, the symmetries of which are desired.

### 2.1 The Gaussian Distribution on the Real Line

#### 2.1.1 Defining Parameters

The Gaussian distribution  $\rho$  on the real line is any function of the form  $\rho(x - x_0)$  where

$$\rho(x) = ce^{-ax^2} \tag{2.1}$$

and  $c \in \mathbb{R}_{>0}$  is related to  $a \in \mathbb{R}_{>0}$  by the constraint that

$$I \doteq \int_{-\infty}^{\infty} \rho(x) dx = 1. \tag{2.2}$$

This constraint, together with the fact that  $\rho(x) \geq 0$  makes it a *probability density function* (or *pdf* for short). That is, any non-negative function satisfying (2.2) (not only those of the form in (2.1)) is a pdf.

The Gaussian distribution is the “bell curve” so often referred to when discussing statistical quantities. It is an infinitely differentiable function. Taking the first derivative gives

$$\frac{d\rho}{dx} = -2acxe^{-ax^2}.$$

From this it is clear that  $\rho(x)$  has a critical point at  $x = 0$ , and this is its only critical point. The second derivative of  $\rho(x)$  evaluated at  $x = 0$  is

$$\left. \frac{d^2\rho}{dx^2} \right|_{x=0} = -2ac,$$

which is always negative, indicating that  $x = 0$  is a maximum, and the maximal value that  $\rho(x)$  can attain is  $c$ . Furthermore, due to the negative sign in the exponential, the function  $\rho(x)$  decays to zero very rapidly as  $|x|$  increases. The Gaussian distribution is called *unimodal* because it has only one local maximum, or mode.

To determine the functional relationship between  $c$  and  $a$  that ensures that  $I = 1$ , the following trick can be used. First evaluate

$$I^2 = c^2 \left( \int_{-\infty}^{\infty} e^{-ax^2} dx \right)^2 = c^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-a(x^2+y^2)} dx dy.$$

Then, changing to the polar coordinates  $x = r \cos \theta$  and  $y = r \sin \theta$  it becomes clear that

$$I^2 = c^2 \int_0^{2\pi} \int_0^{\infty} e^{-ar^2} r dr d\theta.$$

The integral over  $\theta$  reduces to  $2\pi$  and the integral over  $r$  can be performed in closed form. The resulting relationship between  $c$  and  $a$  is then  $I^2 = c^2\pi/a = 1$ , or

$$c = \sqrt{\frac{a}{\pi}}. \quad (2.3)$$

The Gaussian distribution is an *even function*, and for any finite positive value of  $a$  it is also a “nice” function. An even function is one for which  $f_e(x) = f_e(-x)$  and an *odd function* is one for which  $f_o(x) = -f_o(-x)$ . Any function can be decomposed into a sum of even and odd functions as  $f(x) = f_e(x) + f_o(x)$  where

$$f_e(x) = \frac{1}{2}[f(x) + f(-x)] \quad \text{and} \quad f_o(x) = \frac{1}{2}[f(x) - f(-x)].$$

Furthermore, the product of two even functions is even, the product of two odd functions is even, and the product of one even and one odd function is odd. The integral of any well-behaved odd function over any finite interval that is symmetric around the origin is always zero. This can be seen as follows:

$$\int_{-b}^b f_o(x) dx = \int_{-b}^0 f_o(x) dx + \int_0^b f_o(x) dx,$$

but from the definition of an odd function,

$$\int_{-b}^0 f_o(x)dx = - \int_{-b}^0 f_o(-x)dx = - \int_0^b f_o(y)dy,$$

and so

$$\int_{-b}^b f_o(x)dx = 0.$$

For an even function

$$\int_{-b}^b f_e(x)dx = 2 \int_0^b f_e(x)dx.$$

For an even function, the product  $x \cdot f_e(x)$  must be an odd function, and since odd functions integrate to zero over any interval  $[-b, b]$ , it follows that

$$\int_{-\infty}^{\infty} x f_e(x)dx = \lim_{b \rightarrow \infty} \int_{-b}^b x f_e(x)dx = 0.$$

This limit would exist even if the upper and lower integrands go to  $\pm\infty$  at different rates because  $f_e(x)$ , like the other functions in this book, is restricted to be a “nice” function in the sense defined in (1.19), and hence it must decay to zero faster than  $1/x$  as  $x \rightarrow \pm\infty$ . More generally, the quantity  $\mu$  defined by the integral

$$\mu \doteq \int_{-\infty}^{\infty} x f(x)dx$$

for any probability density function,  $f(x)$ , is called the *mean*.

From the shift-invariance property of integration of an arbitrary integrable function on the real line,<sup>1</sup>

$$\int_{-\infty}^{\infty} f(x - x_0)dx = \int_{-\infty}^{\infty} f(x)dx,$$

it follows that for the special case of a Gaussian distribution shifted by  $\mu$ ,  $\rho(x - \mu)$ ,

$$\int_{-\infty}^{\infty} x \rho(x - \mu)dx = \int_{-\infty}^{\infty} (y + \mu) \rho(y)dy = 0 + \mu \cdot I = \mu.$$

The *median* of the Gaussian distribution is the point  $m$  for which

$$\int_{-\infty}^m \rho(x)dx = \int_m^{\infty} \rho(x)dx.$$

Due to the fact that the Gaussian distribution is an even function,  $m = 0$ .

In statistics it is useful to have indicators that describe how concentrated or how spread out a distribution is. One such indicator is the *variance*, defined as

$$\sigma^2 \doteq \int_{-\infty}^{\infty} x^2 f(x - \mu)dx. \quad (2.4)$$

---

<sup>1</sup>Another often-glossed-over property of integration of functions on the real line that will be useful later is invariance under inversion of the argument:

$$\int_{-\infty}^{\infty} f(-x)dx = \int_{-\infty}^{\infty} f(x)dx.$$

The square root of the variance is called the *standard deviation*. Note that this is different from

$$s \doteq \int_{-\infty}^{\infty} |x|f(x - \mu)dx, \quad (2.5)$$

which is called the *spread*. Of course, the concepts of mean, mode, variance, and spread are not limited to the study of Gaussian distributions. They can be calculated for any pdf.

For the Gaussian distribution in (2.1) with normalization (2.3), the mean, median, variance, and spread can be calculated in the following closed form:

$$\mu = m = 0, \quad \sigma^2 = \frac{1}{2a}, \quad \text{and} \quad s = \frac{1}{\sqrt{\pi a}}. \quad (2.6)$$

In general, non-Gaussian pdfs can have multiple modes, the mean and median need not be at the same point, and the relationship between spread and variance need not be so simple.

Since for a Gaussian these quantities are directly related to  $a$ , the Gaussian distribution can be redefined with  $\sigma^2$  or  $s$  incorporated into the definition. The most common choice is to use  $\sigma^2$ , in which case the Gaussian distribution with mean at  $\mu$  and standard deviation  $\sigma$  is denoted<sup>2</sup>

$$\rho(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}. \quad (2.7)$$

In some instances, such as in the following subsections, it will be more convenient to write this as  $\rho_{(\mu, \sigma^2)}(x)$ . Note: another common name for the Gaussian distribution is the *normal distribution*. Figure 2.1 shows a plot of the Gaussian distribution with  $\mu = 0$  and  $\sigma = 1$  plotted over the range  $[-3, 3]$ . Most (approximately 97 percent) of the probability density falls on this finite interval. Changing the value of  $\mu$  or  $\sigma$  would only shift or uniformly stretch this plot.

The integral

$$F(x; \mu, \sigma^2) = \int_{-\infty}^x \rho(\xi; \mu, \sigma^2)d\xi$$

is called the *cumulative distribution function*. This function is known to have a “closed-form” solution in terms of error integrals. In the limit as  $\sigma \rightarrow 0$ ,  $F(x; \mu, \sigma^2)$  exhibits a sharp transition from a value of 0 for  $x < \mu$  to a value of 1 for  $x > \mu$ . When  $\mu = 0$  this is idealized with the *Heaviside step function*

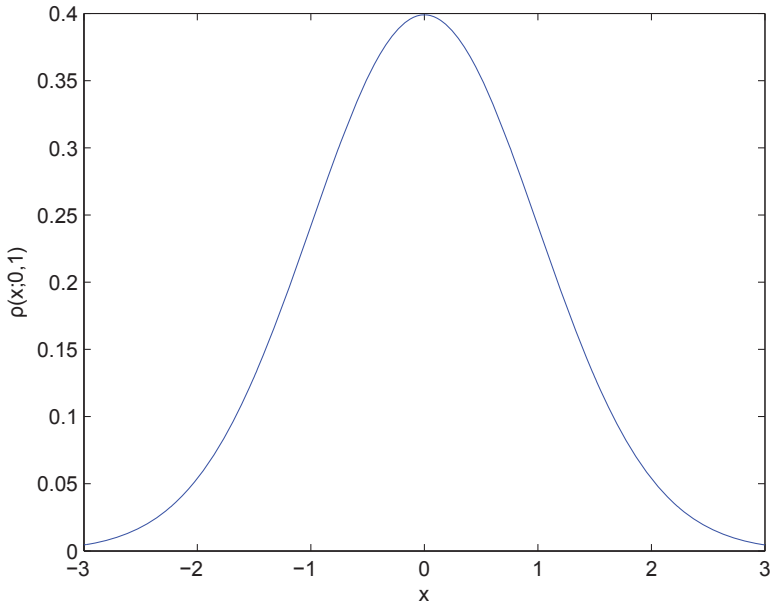
$$H(x) \doteq \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x \leq 0. \end{cases} \quad (2.8)$$

### 2.1.2 The Maximum Entropy Property

The entropy of a pdf  $f(x)$  is defined by the integral [23]

$$S(f) = - \int_{-\infty}^{\infty} f(x) \log f(x)dx \quad (2.9)$$

<sup>2</sup>The symbols  $f(x)$  and  $\rho(x)$  often will be used to denote generic pdfs, but when appended as  $\rho(x; \mu, \sigma^2)$ , this will always denote a Gaussian.



**Fig. 2.1.** The Gaussian Distribution  $\rho(x; 0, 1)$  Plotted over  $[-3, 3]$

where here  $\log = \log_e = \ln$ . This entropy is written as  $S(f)$  rather than  $S(f(x))$  because it is not a function of  $x$ , but rather it is a “functional” of  $f$ , since all dependence on  $x$  has been integrated out.

$S$  is computed in closed form for the Gaussian distribution as

$$S(\rho_{(\mu, \sigma^2)}) = \log(\sqrt{2\pi e} \sigma). \tag{2.10}$$

Interestingly, for any given value of variance, the Gaussian distribution is the pdf with maximal entropy. This can be shown by performing the following optimization:

$$\max_f S(f) \quad \text{subject to } f(x) \geq 0$$

and

$$\int_{-\infty}^{\infty} f(x) dx = 1, \quad \int_{-\infty}^{\infty} x f(x) dx = \mu, \quad \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \sigma^2. \tag{2.11}$$

To find the distribution that satisfies these conditions, Lagrange multipliers<sup>3</sup> are introduced to enforce constraints, and the following necessary conditions are calculated:

$$\frac{\partial C}{\partial f} = 0 \quad \text{where } C = -f \log f + \lambda_1 f + \lambda_2 x f + \lambda_3 (x - \mu)^2 f.$$

Performing the above calculation and solving for  $f$  and the  $\lambda_i$  that satisfy (2.11) gives  $f(x) = \rho_{(\mu, \sigma^2)}(x)$ . Note that the constraint  $f(x) \geq 0$  was not actively enforced in the above derivation, but the result satisfies this condition anyway.

What the above shows is that  $\rho_{(\mu, \sigma^2)}(x)$  extremizes the entropy subject to the given constraints. In other words,  $\rho_{(\mu, \sigma^2)}$  is a critical point of the functional  $S(f)$  subject

---

<sup>3</sup>See Section A.11.1 for a definition.

to the constraints (2.11). However, this could be a minimum, maximum, or point of inflection. To show that it actually maximizes the entropy (at least in a local sense), it is possible to define a perturbed version of this pdf as

$$f(x) = \rho_{(\mu, \sigma^2)}(x) \cdot [1 + \epsilon(x)] \quad (2.12)$$

where  $\epsilon(x)$  is arbitrary except for the fact that<sup>4</sup>  $|\epsilon(x)| \ll 1$  and it is defined such that  $f(x)$  satisfies (2.11). In other words,

$$\int_{-\infty}^{\infty} \rho_{(\mu, \sigma^2)}(x) \epsilon(x) dx = \int_{-\infty}^{\infty} x \rho_{(\mu, \sigma^2)}(x) \epsilon(x) dx = \int_{-\infty}^{\infty} (x - \mu)^2 \rho_{(\mu, \sigma^2)}(x) \epsilon(x) dx = 0.$$

Substituting (2.12) into (2.9) and using the Taylor series approximation  $\log(1 + \epsilon) \approx \epsilon - \epsilon^2/2$ ,

$$\begin{aligned} S(f) &= - \int_{-\infty}^{\infty} \rho_{(\mu, \sigma^2)}(x) \cdot [1 + \epsilon(x)] \log(\rho_{(\mu, \sigma^2)}(x) \cdot [1 + \epsilon(x)]) dx \\ &= - \int_{-\infty}^{\infty} \rho_{(\mu, \sigma^2)}(x) \cdot [1 + \epsilon(x)] \cdot [\log(\rho_{(\mu, \sigma^2)}(x)) + \log(1 + \epsilon(x))] dx \\ &= S(\rho_{(\mu, \sigma^2)}) - F(\epsilon^2) + O(\epsilon^3) \end{aligned}$$

where the functional  $F$  is always positive and the cross terms that are linear in  $\epsilon$  all vanish due to the integral constraints on  $\epsilon$ . This means that at least locally a Gaussian maximizes entropy. Determining the exact form of the functional  $F$  is left as an exercise.

### 2.1.3 The Convolution of Gaussians

The *convolution* of two pdfs on the real line is defined as

$$(f_1 * f_2)(x) \doteq \int_{-\infty}^{\infty} f_1(\xi) f_2(x - \xi) d\xi. \quad (2.13)$$

Sometimes this is written as  $f_1(x) * f_2(x)$ . Note that convolution on the real line is commutative:  $(f_1 * f_2)(x) = (f_2 * f_1)(x)$ . This is a direct consequence of the commutativity of addition:  $x + y = y + x$ .

In order for the convolution integral to exist,  $f_1(x)$  and  $f_2(x)$  must both decay to zero sufficiently fast as  $x \rightarrow \pm\infty$ . In addition, the scope here is restricted to “nice” functions in the sense of (1.19) with  $D = \mathbb{R}$ . Therefore these functions are infinitely differentiable and have integrals of their square and absolute values that are finite. It can be shown that the convolution integral will always exist for such “nice” functions, and furthermore

$$f_i \in \mathcal{N}(\mathbb{R}) \implies f_1 * f_2 \in \mathcal{N}(\mathbb{R}).$$

In (2.13)  $\xi$  is a dummy variable of integration, the name of which is unimportant. A geometric interpretation of (2.13) is as follows. First, the function  $f_2(x)$  is shifted along the real line in the positive direction by an amount  $\xi$ , resulting in  $f_2(x - \xi)$ . Then, the function  $f_1$  evaluated at the amount of shift,  $f_1(\xi)$ , is used to weight  $f_2(x - \xi)$ . Finally, all copies of the product  $f_1(\xi) f_2(x - \xi)$  are “added up” by integrating over all values of the shift. This has the effect of “smearing”  $f_2$  over  $f_1$ .

<sup>4</sup>To be concrete,  $\epsilon = 0.01 \ll 1$ . Then  $\epsilon^3 = 10^{-6}$  is certainly negligible in comparison to quantities that are on the order of 1.

In the case when  $f_1(x) = \delta(x)$ , i.e., the Dirac delta function, which is the probability density function with all of its mass concentrated at  $x = 0$ ,  $(\delta * f)(x) = f(x)$ . This is because the only shift that the delta function allows is  $\xi = 0$ . All other shifts are weighted by a value of zero, and therefore do not contribute. While  $\delta(x)$  is not a “nice” function, it is possible to approximate it with a Gaussian distribution with very small variance,  $\epsilon$ , which is a “nice” function. The approximation of the Dirac delta function as  $\delta(x) \approx \rho(x; 0, \epsilon)$  is deemed to be “good enough” if the integral of  $|\rho(x; 0, \epsilon) * f(x) - f(x)|$  and the integral of the square of this are both “small enough” when  $f(x)$  is a nice function.

The Gaussian distribution has the property that the convolution of two Gaussians is a Gaussian:

$$\rho(x; \mu_1, \sigma_1^2) * \rho(x; \mu_2, \sigma_2^2) = \rho(x; \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2). \quad (2.14)$$

The Dirac  $\delta$ -function can be viewed as the limit

$$\delta(x) = \lim_{\sigma \rightarrow 0} \rho(x; 0, \sigma^2). \quad (2.15)$$

It then follows from (2.14) that

$$\rho(x; \mu_1, \sigma_1^2) * \delta(x) = \rho(x; \mu_1, \sigma_1^2).$$

#### 2.1.4 The Fourier Transform of the Gaussian Distribution

The Fourier transform of a “nice” function  $f \in \mathcal{N}(\mathbb{R})$  is defined as

$$[\mathcal{F}(f)](\omega) \doteq \int_{-\infty}^{\infty} f(x)e^{-i\omega x} dx. \quad (2.16)$$

The shorthand  $\hat{f}(\omega) \doteq [\mathcal{F}(f)](\omega)$  will be used frequently.

The conditions for existence and properties of the Fourier transform of functions on the real line are described in detail in [6, 11, 15]. Tools for the computation of fast sampled versions of the Fourier transform of periodic functions can be found in many books such as [7, 10, 24]. From the definition of the Fourier transform, it can be shown that

$$\widehat{(f_1 * f_2)}(\omega) = \hat{f}_1(\omega)\hat{f}_2(\omega) \quad (2.17)$$

(i.e., the Fourier transform of the convolution is the product of Fourier transforms) and

$$f(x) = [\mathcal{F}^{-1}(\hat{f})](x) \doteq \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega x} d\omega. \quad (2.18)$$

This is called the *inverse Fourier transform* or *Fourier reconstruction formula*.

The proof of the property (2.17) is left as an exercise, whereas (2.18) is proven below. For more details about classical Fourier analysis and its extensions, see [8] and references therein.

The fact that a function is recovered from its Fourier transform is found by first observing that it is true for the special case of  $g(x) = e^{-ax^2}$  for  $a > 0$ . One way to calculate

$$\hat{g}(\omega) = \int_{-\infty}^{\infty} e^{-ax^2} e^{-i\omega x} dx$$

is to differentiate both sides with respect to  $\omega$ , which yields



$$\frac{d\hat{g}}{d\omega} = -i \int_{-\infty}^{\infty} x e^{-ax^2} e^{-i\omega x} dx = \frac{i}{2a} \int_{-\infty}^{\infty} \frac{dg}{dx} e^{-i\omega x} dx.$$

Integrating by parts, and observing that  $e^{-i\omega x}g(x)$  vanishes at the limits of integration yields

$$\frac{d\hat{g}}{d\omega} = -\frac{\omega}{2a}\hat{g}.$$

The solution of this first-order ordinary differential equation is of the form

$$\hat{g}(\omega) = \hat{g}(0)e^{-\frac{\omega^2}{4a}}$$

where

$$\hat{g}(0) = \int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}.$$

Having found the form of  $\hat{g}(\omega)$ , it is easy to see that  $g(x)$  is reconstructed from  $\hat{g}(\omega)$  using the inversion formula (2.18) (the calculation is essentially the same as for the forward Fourier transform). Likewise, the Gaussian function

$$\rho_{(0,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$$

has Fourier transform

$$\hat{\rho}_{(0,\sigma^2)}(\omega) = e^{-\frac{\sigma^2}{2}\omega^2}$$

and the reconstruction formula holds. As  $\sigma$  becomes small,  $\rho_{(0,\sigma^2)}(x)$  becomes like  $\delta(x)$ . From the property that  $(\delta * f)(x) = f(x)$ , the convolution theorem, and the above properties of Gaussian approximations to the Dirac  $\delta$ -function, (2.18) immediately follows.

### 2.1.5 Diffusion Equations

A one-dimensional linear diffusion equation with constant coefficients has the form

$$\frac{\partial u}{\partial t} = a \frac{\partial u}{\partial x} + b \frac{\partial^2 u}{\partial x^2} \quad (2.19)$$

where  $a \in \mathbb{R}$  is called the *drift coefficient* and  $b \in \mathbb{R}_{>0}$  is called the *diffusion coefficient*. When modeling diffusion phenomena in an infinite medium, the above diffusion equation for  $u(x, t)$  has initial conditions of the form  $u(x, 0) = f(x)$ . The boundary conditions

$$u(\pm\infty, 0) = \frac{\partial u}{\partial x}(\pm\infty, 0) = 0$$

are implicit in this problem, because otherwise the solutions will not be pdfs, or in the class  $\mathcal{N}(\mathbb{R})$ .

Note that (2.19) is a special case of the *Fokker-Planck equation*<sup>5</sup> which will be examined in great detail in Chapter 4. When the drift coefficient is zero, the diffusion equation is called the *heat equation*.

Taking the Fourier transform of  $u(x, t)$  for each value of  $t$  (i.e., treating time as a constant for the moment and  $x$  as the independent variable) produces  $\hat{u}(\omega, t)$ . Then applying the Fourier transform to both sides of (2.19) and the initial conditions results

<sup>5</sup>Also known as *Kolmogorov's forward equation*.

in a linear first-order ordinary differential equation with  $t$  as the independent variable, together with initial conditions, for each fixed frequency  $\omega$ :

$$\frac{d\hat{u}}{dt} = (ia\omega - b\omega^2)\hat{u} \quad \text{with} \quad \hat{u}(\omega, 0) = \hat{f}(\omega).$$

The solution to this initial value problem is of the form

$$\hat{u}(\omega, t) = \hat{f}(\omega)e^{(ia\omega - b\omega^2)t}.$$

Application of the inverse Fourier transform yields a solution. The above expression for  $\hat{u}(\omega, t)$  is a Gaussian with phase factor, and on inversion this becomes a shifted Gaussian:

$$[\mathcal{F}^{-1}(e^{iat\omega}e^{-b\omega^2t})](x) = \frac{1}{\sqrt{4\pi bt}} \exp\left(-\frac{(x+at)^2}{4bt}\right).$$

Using the convolution theorem in reverse then gives

$$u(x, t) = \frac{1}{\sqrt{4\pi bt}} \int_{-\infty}^{\infty} f(\xi) \exp\left(-\frac{(x+at-\xi)^2}{4bt}\right) d\xi. \quad (2.20)$$

### 2.1.6 Stirling's Formula

In probability theory for discrete variables, the *binomial distribution* is defined as

$$f(k; n, p) \doteq \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where} \quad \binom{n}{k} \doteq \frac{n!}{k!(n-k)!} \quad 0 \leq p \leq 1 \quad (2.21)$$

and  $k = 0, 1, 2, \dots, n$ , and the values  $\binom{n}{k}$  are called *binomial coefficients*. From the *binomial theorem*,

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k},$$

it follows that

$$\sum_{k=0}^n f(k; n, p) = (1-p+p)^n = 1,$$

and from the definition in (2.21)

$$\sum_{k=0}^n k \cdot f(k; n, p) = np \cdot \sum_{k'=0}^{n-1} f(k'; n-1, p) = np \quad \text{where} \quad k' = k-1.$$

The factorial  $n!$  can be approximated using the *Stirling series*:

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + \dots\right).$$

If the first term is kept, the result is *Stirling's formula*:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \quad (2.22)$$

Stirling's formula is used extensively in probability theory to establish limiting behaviors. In the current context, it can be used to show that the Gaussian distribution is the limiting distribution of the binomial distribution in the sense that [22]

$$\lim_{n \rightarrow \infty} \frac{f(k; n, p)}{\rho(k; np, np(1-p))} = 1 \quad \text{for finite } |k - np|/\sqrt{np(1-p)}. \quad (2.23)$$

## 2.2 The Multivariate Gaussian Distribution

The multivariate Gaussian distribution on  $\mathbb{R}^n$  is defined as<sup>6</sup>

$$\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma) \doteq \frac{1}{(2\pi)^{n/2} |\det \Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \quad (2.24)$$

This is the maximum entropy distribution subject to the constraints<sup>7</sup>

$$\int_{\mathbb{R}^n} \rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma) d\mathbf{x} = 1; \int_{\mathbb{R}^n} \mathbf{x} \rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma) d\mathbf{x} = \boldsymbol{\mu}; \int_{\mathbb{R}^n} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma) d\mathbf{x} = \Sigma. \quad (2.25)$$

The integral is calculated with respect to the differential volume element for  $\mathbb{R}^n$ , denoted above as  $d\mathbf{x} = dx_1 dx_2 \cdots dx_n$ . The above properties can be proved by changing coordinates as  $\mathbf{y} = \Sigma^{-\frac{1}{2}} (\mathbf{x} - \boldsymbol{\mu})$ , which reduces the problem to many one-dimensional integrals. The meaning of a fractional power of a matrix is reviewed in the appendix. Given a multi-dimensional coordinate transformation  $\mathbf{y} = \mathbf{y}(\mathbf{x})$  (which is written in components as  $y_i = y_i(x_1, \dots, x_n)$  for  $i = 1, \dots, n$ ), the following well-known integration rule (which is a restatement of (1.38) in different notation) holds:

$$\int_{\mathbf{y}(D)} F(\mathbf{y}) d\mathbf{y} = \int_D F(\mathbf{y}(\mathbf{x})) |\det J| d\mathbf{x} \quad (2.26)$$

where  $d\mathbf{x} = dx_1 dx_2 \cdots dx_n$ ,  $d\mathbf{y} = dy_1 dy_2 \cdots dy_n$ , and

$$J = \left[ \frac{\partial \mathbf{y}}{\partial x_1}, \dots, \frac{\partial \mathbf{y}}{\partial x_n} \right]$$

is the Jacobian matrix of the transformation and  $|\det J|$  gives a measure of *local volume change*.  $D$  is the domain of integration in terms of the coordinates  $\mathbf{x}$ , and  $\mathbf{y}(D)$  is the new domain to which each point in  $D$  is mapped under the transformation  $\mathbf{y}(\mathbf{x})$ . In the current context, the range of integrals over  $\mathbf{x}$  and  $\mathbf{y}$  are both copies of  $\mathbb{R}^n$ , i.e.,  $D = \mathbf{y}(D) = \mathbb{R}^n$ .

### 2.2.1 Conditional and Marginal Densities

A vector  $\mathbf{x} \in \mathbb{R}^n$  can be partitioned as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T \in \mathbb{R}^{n_1+n_2}$$

where  $\mathbf{x}_1 \in \mathbb{R}^{n_1}$  and  $\mathbf{x}_2 \in \mathbb{R}^{n_2}$ . The notation  $[\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ , which takes advantage of the fact that the “transpose of a transpose is the original,” has the benefit that it can be

<sup>6</sup>It is unfortunate that the notation for the one-dimensional case,  $\rho(x; \mu, \sigma^2)$ , is inconsistent with the multivariate case since  $\sigma^2$  becomes  $\Sigma$  (rather than  $\Sigma^2$ ), but this is the notation that is standard in the field.

<sup>7</sup>In Chapter 1 the notation  $d(\mathbf{x})$  was used to denote the volume element  $dx_1 dx_2 \cdots dx_n$ . In the expressions in this chapter, the parentheses will be dropped to reduce the amount of clutter, and  $d\mathbf{x}$  will be used as shorthand for  $d(\mathbf{x})$ . This will not cause trouble because  $\mathbf{x}(t + dt) - \mathbf{x}(t)$  does not appear in any of these calculations.

written on one line and included in a sentence, whereas it is difficult to do so for a column vector.

If  $f(\mathbf{x}) = f([\mathbf{x}_1^T, \mathbf{x}_2^T]^T)$  (which also will be referred to as  $f(\mathbf{x}_1, \mathbf{x}_2)$ ) is any pdf on  $\mathbb{R}^{n_1+n_2}$ , then the marginal density  $f_1(\mathbf{x}_1)$  is defined by integrating over all values of  $\mathbf{x}_2$ :

$$f_1(\mathbf{x}_1) = \int_{\mathbb{R}^{n_2}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2.$$

$f_2(\mathbf{x}_2)$  is obtained from  $f(\mathbf{x}_1, \mathbf{x}_2)$  in a similar way by integrating over all values of  $\mathbf{x}_1$ .

The mean and variance of  $f_1(\mathbf{x}_1)$  are obtained from the mean and variance of  $f(\mathbf{x})$  by observing that

$$\begin{aligned} \boldsymbol{\mu}_1 &= \int_{\mathbb{R}^{n_1}} \mathbf{x}_1 f_1(\mathbf{x}_1) d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} \mathbf{x}_1 \left( \int_{\mathbb{R}^{n_2}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_2}} \mathbf{x}_1 f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 d\mathbf{x}_1 \end{aligned}$$

and

$$\begin{aligned} \Sigma_{11} &= \int_{\mathbb{R}^{n_1}} (\mathbf{x}_1 - \boldsymbol{\mu}_1)(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T f_1(\mathbf{x}_1) d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} (\mathbf{x}_1 - \boldsymbol{\mu}_1)(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T \left( \int_{\mathbb{R}^{n_2}} f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \right) d\mathbf{x}_1 \\ &= \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_2}} (\mathbf{x}_1 - \boldsymbol{\mu}_1)(\mathbf{x}_1 - \boldsymbol{\mu}_1)^T f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 d\mathbf{x}_1. \end{aligned}$$

In other words, the mean vector and covariance matrix for the marginal density are obtained directly from those of the full density. For example,  $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T]^T$ .

Given a (multivariate) Gaussian distribution  $\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ , the associated covariance matrix can be written in terms of blocks as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where  $\Sigma_{11} = \Sigma_{11}^T$ ,  $\Sigma_{22} = \Sigma_{22}^T$ , and  $\Sigma_{21} = \Sigma_{12}^T$ . The block  $\Sigma_{ij}$  has dimensions  $n_i \times n_j$ . In other words,  $\Sigma_{ij} \in \mathbb{R}^{n_i \times n_j}$  where  $i$  and  $j$  can either be 1 or 2.

The *marginal density* that results from integrating the Gaussian distribution  $\rho(\mathbf{x}, \boldsymbol{\mu}, \Sigma)$  over all values of  $\mathbf{x}_2$  is

$$\int_{\mathbb{R}^{n_2}} \rho([\mathbf{x}_1^T, \mathbf{x}_2^T]^T; \boldsymbol{\mu}, \Sigma) d\mathbf{x}_2 = \rho(\mathbf{x}_1; \boldsymbol{\mu}_1, \Sigma_{11}). \quad (2.27)$$

This should not come as a surprise, since a Gaussian is defined completely by the values of its mean and covariance.

Another operation that is important in probability and statistics is that of conditioning. Given  $f(\mathbf{x}_1, \mathbf{x}_2)$ , the conditional density of  $\mathbf{x}_1$  given  $\mathbf{x}_2$  is

$$f(\mathbf{x}_1|\mathbf{x}_2) \doteq f(\mathbf{x}_1, \mathbf{x}_2)/f_2(\mathbf{x}_2). \quad (2.28)$$

Evaluating this expression using a Gaussian gives

$$\rho([\mathbf{x}_1^T, \mathbf{x}_2^T]^T; \boldsymbol{\mu}, \Sigma) / \rho(\mathbf{x}_2; \boldsymbol{\mu}_2, \Sigma_2) = \rho(\mathbf{x}_1; \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \quad (2.29)$$

The above formulas follow from decomposing  $\Sigma$  into a product of block lower triangular, block diagonal, and block upper triangular matrices as in Appendix A.4.3. Each of these can then be inverted in closed form resulting in explicit expressions for  $\Sigma^{-1}$  in terms of the blocks of  $\Sigma$ .

In summary, the set of Gaussian distributions has the remarkable property that it is closed under marginalization and conditioning, and as was demonstrated previously in the 1D case, it is also closed under convolution.

## 2.2.2 Multi-Dimensional Integrals Involving Gaussians

Several integral identities involving Gaussian distributions are used throughout this book. These are stated here and proved in the following subsections.

First, it is well known that

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi} \quad \implies \quad \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{x}\right) d\mathbf{x} = (2\pi)^{\frac{n}{2}}. \quad (2.30)$$

Here  $\mathbf{x} \in \mathbb{R}^n$  and  $d\mathbf{x} = dx_1 dx_2 \cdots dx_n$ . Note also that

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi}. \quad (2.31)$$

These identities are used below to prove

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{x}^T M \mathbf{x} - \mathbf{m}^T \mathbf{x}\right) d\mathbf{x} = (2\pi)^{n/2} |\det M|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\mathbf{m}^T M^{-1} \mathbf{m}\right) \quad (2.32)$$

and

$$\int_{\mathbb{R}^n} \mathbf{x}^T G \mathbf{x} \exp\left(-\frac{1}{2}\mathbf{x}^T A \mathbf{x}\right) d\mathbf{x} = (2\pi)^{n/2} \frac{\text{tr}(GA^{-1})}{|\det A|^{\frac{1}{2}}}. \quad (2.33)$$

These integrals have applications in the analysis of elastic network models of proteins [9].

### Proof of Equation (2.32)

Consider the integral

$$I = \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{x}^T M \mathbf{x} - \mathbf{m}^T \mathbf{x}\right) d\mathbf{x}.$$

Using the change of variables  $\mathbf{z} = M^{\frac{1}{2}}\mathbf{x} - M^{-\frac{1}{2}}\mathbf{m}$  implies that  $d\mathbf{z} = |\det M|^{\frac{1}{2}}d\mathbf{x}$  and  $\mathbf{x} = M^{-\frac{1}{2}}(\mathbf{z} + M^{-\frac{1}{2}}\mathbf{m})$ . Therefore

$$\begin{aligned} I &= \frac{1}{|\det M|^{\frac{1}{2}}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{z}^T \mathbf{z} + \frac{1}{2}\mathbf{m}^T M^{-1} \mathbf{m}\right) d\mathbf{z} \\ &= \frac{\exp\left(\frac{1}{2}\mathbf{m}^T M^{-1} \mathbf{m}\right)}{|\det M|^{\frac{1}{2}}} \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{z}^T \mathbf{z}\right) d\mathbf{z}. \end{aligned}$$

And so, (2.32) follows from (2.30).

**Proof of Equation (2.33)**

It is also convenient to have closed-form solutions for integrals of the form

$$J = \int_{\mathbb{R}^n} \mathbf{x}^T G \mathbf{x} \exp\left(-\frac{1}{2} \mathbf{x}^T A \mathbf{x}\right) d\mathbf{x}.$$

Let  $\mathbf{z} = A^{\frac{1}{2}} \mathbf{x}$ . Then

$$J = \frac{1}{|\det A|^{\frac{1}{2}}} \int_{\mathbb{R}^n} \mathbf{z}^T A^{-\frac{1}{2}} G A^{-\frac{1}{2}} \mathbf{z} \exp\left(-\frac{1}{2} \mathbf{z}^T \mathbf{z}\right) d\mathbf{z}.$$

Now let  $G' = A^{-\frac{1}{2}} G A^{-\frac{1}{2}}$ . Then it is clear that off-diagonal terms of  $G'$  do not contribute to this integral since odd moments of Gaussians are zero. Therefore,

$$\begin{aligned} J &= \frac{1}{|\det A|^{\frac{1}{2}}} \int_{\mathbb{R}^n} \sum_{i=1}^n g'_{ii} z_i^2 \exp\left(-\frac{1}{2} \mathbf{z}^T \mathbf{z}\right) d\mathbf{z} \\ &= \frac{1}{|\det A|^{\frac{1}{2}}} \sum_{i=1}^n g'_{ii} \int_{-\infty}^{\infty} z_i^2 e^{-\frac{1}{2} z_i^2} dz_i \int_{\mathbb{R}^{n-1}} \exp\left(-\frac{1}{2} \mathbf{y}_i^T \mathbf{y}_i\right) d\mathbf{y}_i \end{aligned}$$

where  $\mathbf{y}_i \in \mathbb{R}^{n-1}$  is the part of  $\mathbf{z} \in \mathbb{R}^n$  with the  $z_i$  component removed. The value of the integrals are independent of  $i$ , and

$$\sum_{i=1}^n g'_{ii} = \text{tr}(G') = \text{tr}(A^{-\frac{1}{2}} G A^{-\frac{1}{2}}) = \text{tr}(G A^{-1}),$$

and so, (2.33) follows.

### 2.3 The Volume of Spheres and Balls in $\mathbb{R}^n$

The volume of the  $(n - 1)$ -dimensional hyper-sphere with unit radius,  $S^{n-1} \subset \mathbb{R}^n$ , and of the open ball  $B^n \subset \mathbb{R}^n$  enclosed by  $S^{n-1}$  appear in a number of geometric and statistical applications. The argument used here for computing these volumes follows that given in [12]. Before proceeding, a note is in order regarding the use of the word “volume.” In the case of  $n = 3$ , the “volume” of the sphere  $S^2$  is its surface area, and in the case of  $n = 2$ , the “volume” of the circle  $S^1$  is its perimeter. In contrast, the “volume” of the ball  $B^2$  is the area on the interior of a circle, and the “volume” of  $B^3$  is the classical volume in  $\mathbb{R}^3$  bounded by the sphere  $S^2$ . In general, the volume of an  $n$ -dimensional manifold will be an  $n$ -dimensional measurement.

Consider the isotropic Gaussian distribution on  $\mathbb{R}^n$  with zero mean written as

$$\rho(\mathbf{x}; \boldsymbol{\mu} = \mathbf{0}, \Sigma = \sigma^2 I) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp(-\frac{1}{2} \|\mathbf{x}\|^2 / \sigma^2).$$

If  $\mathbf{x} = r \mathbf{u}$  where  $r$  and  $\mathbf{u} = \mathbf{u}(\phi_1, \phi_2, \dots, \phi_{n-1})$  represent “hyper-spherical” coordinates, then the Jacobian determinant relates the change from Cartesian coordinates as

$$d\mathbf{x} = \left| \det \left[ \frac{\partial \mathbf{x}}{\partial r}, \frac{\partial \mathbf{x}}{\partial \phi_1}, \dots, \frac{\partial \mathbf{x}}{\partial \phi_{n-1}} \right] \right| dr d\phi_1 \cdots d\phi_{n-1} = dV(\phi) r^{n-1} dr$$

where  $dV(\phi)$  is the volume element for the sphere  $S^{n-1}$ . The volume of  $S^{n-1}$  is then

$$\text{Vol}(S^{n-1}) = \int_{S^{n-1}} dV(\phi).$$

This can be computed directly by extending the usual spherical coordinates to higher dimensions in the natural way as

$$\mathbf{u}^{(2)} = \begin{pmatrix} \cos \phi_1 \\ \sin \phi_1 \end{pmatrix}; \quad \mathbf{u}^{(3)} = \begin{pmatrix} \cos \phi_1 \sin \phi_2 \\ \sin \phi_1 \sin \phi_2 \\ \cos \phi_2 \end{pmatrix}; \quad \mathbf{u}^{(3)} = \begin{pmatrix} \cos \phi_1 \sin \phi_2 \sin \phi_3 \\ \sin \phi_1 \sin \phi_2 \sin \phi_3 \\ \cos \phi_2 \sin \phi_3 \\ \cos \phi_3 \end{pmatrix}; \quad \text{etc.},$$

computing Jacobian determinants for each case, and then integrating over the appropriate range of angles,  $0 \leq \phi_1 < 2\pi$  and  $0 \leq \phi_i < \pi$  for  $1 < i \leq n-1$ . Or, the volume of the unit sphere can be calculated indirectly, as it is done below.

From the fact that  $\rho$  is a pdf,

$$\begin{aligned} 1 &= \int_{\mathbb{R}^n} \rho(\mathbf{x}; \mathbf{0}, \sigma^2 I) d\mathbf{x} \\ &= \int_0^\infty \int_{S^{n-1}} \rho(r\mathbf{u}; \mathbf{0}, \sigma^2 I) dV(u) r^{n-1} dr \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} \left( \int_0^\infty \exp(-r^2/(2\sigma^2)) r^{n-1} dr \right) \text{Vol}(S^{n-1}). \end{aligned}$$

Therefore, it must be that

$$\frac{1}{(2\pi)^{n/2} \sigma^n} \int_0^\infty \exp(-r^2/(2\sigma^2)) r^{n-1} dr = 1/\text{Vol}(S^{n-1})$$

for any value of  $\sigma$ . Letting  $s = r/(\sqrt{2}\sigma)$ , the integral on the left becomes

$$\int_0^\infty \exp(-r^2/(2\sigma^2)) r^{n-1} dr = 2^{n/2} \sigma^n \int_0^\infty \exp(-s^2) s^{n-1} ds = \frac{1}{2} 2^{n/2} \sigma^n \Gamma(n/2).$$

This can be taken as the definition of the *Gamma function*, or it can be viewed as the result of the change of coordinates  $t = s^2$  from the more standard definition

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt \tag{2.34}$$

with  $\alpha = n/2$ .

In any case, since the Gaussian pdf integrates to unity, the factors of  $2^{n/2} \sigma^n$  cancel, and it must be that  $\frac{1}{2} \Gamma(n/2) \text{Vol}(S^{n-1}) = (\pi)^{n/2}$ , or

$$\text{Vol}(S^{n-1}) = \frac{2(\pi)^{n/2}}{\Gamma\left(\frac{n}{2}\right)}. \tag{2.35}$$

This is the volume of a unit hyper-sphere  $S^{n-1} \subset \mathbb{R}^n$ . The volume of a hyper-sphere of radius  $r$  would be  $r^{n-1}$  times this quantity. The volume of the unit ball  $B^n \subset \mathbb{R}^n$  is then obtained by integrating over all of these spherical shells as

$$\text{Vol}(B^n) = \int_0^1 \text{Vol}(S^{n-1}) r^{n-1} dr = \frac{2(\pi)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} \int_0^1 r^{n-1} dr.$$

In other words,

$$\text{Vol}(B^n) = \frac{2(\pi)^{n/2}}{n \cdot \Gamma\left(\frac{n}{2}\right)} = \frac{(\pi)^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)}. \quad (2.36)$$

The first few values of  $\Gamma(n/2)$  are given in the following table:

**Table 2.1.** The First Few Half-Integer Values of the  $\Gamma$ -Function

$n$	$\Gamma(n/2)$
1	$\sqrt{\pi}$
2	1
3	$\sqrt{\pi}/2$
4	1
5	$3\sqrt{\pi}/4$
6	2

Note that for integer arguments,  $\Gamma(m) = (m - 1)!$ .

The shorthand notation

$$\text{Vol}(S^{n-1}) = \mathcal{O}_n \quad \text{and} \quad \text{Vol}(B^n) = \frac{\mathcal{O}_n}{n} \quad (2.37)$$

will be useful.

## 2.4 Clipped Gaussian Distributions

The Gaussian distribution has many interesting and useful properties. For example, it is the maximum entropy distribution of given mean and covariance, it satisfies a diffusion equation, as a family of parametric distributions it is closed under the operations of convolution and conditioning. In addition, its higher moments can be computed as closed-form integrals. It would be useful to take advantage of these properties when fitting a density to measured data on other domains such as spheres. However, a problem that immediately arises is that for compact domains, something must be done with the infinite tails of the Gaussian distribution. Two options are to wrap the tails around (resulting in a “folded” Gaussian), or to clip the tails. The folded Gaussian for the circle is discussed in Section 2.5. While this is a viable option in some cases, a more general procedure that can be used for other finite domains is clipping. In the subsections that follow, the properties of the univariate clipped Gaussian are obtained, and extended to the multi-dimensional case.

### 2.4.1 One-Dimensional Clipped Gaussian Distributions

Suppose that we want to clip the Gaussian distribution with mean at  $x = 0$  defined by

$$\rho(x; 0, \sigma_0) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-x^2/2\sigma_0^2}.$$

This is defined on the real line. By restricting it to the unit circle, which we identify with the interval  $[-\pi, \pi]$ , the mass is reduced from unity to



$$r(\sigma_0) \doteq \int_{-\pi}^{\pi} \rho(x; 0, \sigma_0) dx < 1. \quad (2.38)$$

An exact expression for  $r(\sigma_0)$  can be found in terms of the *error function*

$$\operatorname{erf}(x) \doteq \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (2.39)$$

However, if  $k\sigma_0 < \pi$  for  $k \geq 3$ , then  $r(\sigma_0) \approx 1$  is a good approximation.

The variance of a clipped Gaussian is then

$$\sigma^2 = \frac{1}{\sqrt{2\pi}\sigma_0 r(\sigma_0)} \int_{-\pi}^{\pi} x^2 e^{-x^2/2\sigma_0^2} dx = \frac{\sigma_0^2}{\sqrt{2\pi}r(\sigma_0)} \int_{-\pi/\sigma_0}^{\pi/\sigma_0} y^2 e^{-y^2/2} dy.$$

This can be written as

$$\sigma^2 = \frac{\sigma_0^2}{\sqrt{2\pi}r(\sigma_0)} \left[ \sqrt{2\pi} - \frac{2\pi}{\sigma_0} e^{-\pi^2/(2\sigma_0^2)} \right]$$

by using integration by parts. As  $\sigma_0 \rightarrow 0$ , then  $\sigma \rightarrow \sigma_0$ .

## 2.4.2 Multi-Dimensional Clipped Gaussian Distributions

The integral of a multi-dimensional Gaussian distribution over the interior of an ellipsoid defined by

$$\mathbf{x}^T \Sigma_0^{-1} \mathbf{x} = a^2$$

can be computed in closed form (using error integrals). We can therefore clip a multi-dimensional Gaussian distribution along the boundary of such an ellipsoid and renormalize the resulting distribution so as to be a pdf. In other words, a clipped Gaussian is defined relative to a Gaussian as

$$\rho_c(\mathbf{x}, \Sigma_0, a) \doteq \begin{cases} \rho(\mathbf{x}, \Sigma_0)/r(\Sigma_0, a) & \text{for } \mathbf{x}^T \Sigma_0^{-1} \mathbf{x} < a^2 \\ 0 & \text{otherwise} \end{cases} \quad (2.40)$$

where

$$r(\Sigma_0, a) \doteq \int_{\mathbf{x}^T \Sigma_0^{-1} \mathbf{x} < a^2} \rho(\mathbf{x}, \Sigma_0) d\mathbf{x}.$$

The covariance of a clipped Gaussian is then

$$\Sigma = \int_{\mathbf{x}^T \Sigma_0^{-1} \mathbf{x} < a^2} \mathbf{x}\mathbf{x}^T \rho_c(\mathbf{x}, \Sigma_0, a) d\mathbf{x}. \quad (2.41)$$

By making the change of variables  $\mathbf{y} = \Sigma_0^{-\frac{1}{2}} \mathbf{x}$ , it follows that

$$\Sigma = \Sigma_0^{\frac{1}{2}} \left[ \int_{\mathbf{y}^T \mathbf{y} < a^2} \mathbf{y}\mathbf{y}^T \rho_c(\mathbf{y}, I, a) d\mathbf{y} \right] \Sigma_0^{\frac{1}{2}}. \quad (2.42)$$

The above integral can be computed in closed form. This is done below for the three-dimensional case. The two-dimensional case is left as an exercise.

It will be convenient to define

$$f_0(a) \doteq \int_0^a e^{-r^2/2} dr = \sqrt{\frac{\pi}{2}} \operatorname{erf}(a/\sqrt{2})$$

$$f_1(a) \doteq \int_0^a r^2 e^{-r^2/2} dr = -ae^{-a^2/2} + f_0(a)$$

and

$$f_2(a) \doteq \int_0^a r^4 e^{-r^2/2} dr = 3f_1(a) - a^3 e^{-a^2/2}.$$

Then

$$m(\Sigma_0, a) \doteq \int_{\mathbf{x}^T \Sigma_0^{-1} \mathbf{x} < a^2} \exp\{-\mathbf{x}^T \Sigma_0^{-1} \mathbf{x}\} d\mathbf{x} = 4\pi f_1(a) \cdot |\Sigma_0|^{\frac{1}{2}}$$

and

$$r(\Sigma_0, a) = m(\Sigma_0, a)/(2\pi)^{\frac{3}{2}} |\Sigma_0|^{\frac{1}{2}} = \sqrt{\frac{2}{\pi}} f_1(a).$$

Using spherical coordinates,

$$\mathbf{y} = \begin{pmatrix} r \sin \theta \cos \phi \\ r \sin \theta \sin \phi \\ r \cos \theta \end{pmatrix},$$

$$\int_{\mathbf{y}^T \mathbf{y} < a^2} \mathbf{y} \mathbf{y}^T \rho_c(\mathbf{y}, I, a) d\mathbf{y} = \int_{r=0}^a \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \mathbf{y} \mathbf{y}^T \rho_c(\mathbf{y}, I, a) r^2 dr d\phi d\theta = \sqrt{\frac{2}{\pi}} \frac{f_2(a)}{3} I$$

where

$$f_2(a) = \int_0^a r^4 e^{-r^2/2} dr.$$

This can be computed in closed form using integration by parts. Therefore (2.42) reduces to

$$\Sigma = \frac{f_2(a)}{3 \cdot f_1(a)} \Sigma_0. \quad (2.43)$$

As  $a \rightarrow \infty$ ,  $\Sigma \rightarrow \Sigma_0$ .

## 2.5 Folded, or Wrapped, Gaussians

In some applications, data on the circle is given, and a corresponding concept of Gaussian distribution is needed. One approach that was discussed in the previous section that could be applied to this end is to “clip the tails” of a Gaussian outside of the range of values  $\theta \in [-\pi, \pi]$  and renormalize the result in order to make it a valid pdf. In contrast, the tails can be “wrapped around” the circle as

$$\rho_W(\theta; \mu, \sigma) \doteq \sum_{k=-\infty}^{\infty} \rho(\theta - 2\pi k; \mu, \sigma), \quad (2.44)$$

where if  $\mu$  is outside of the range  $[-\pi, \pi]$ , it can be “put back in the range” by subtracting  $2\pi N$  from it for some  $N \in \mathbb{Z}$  until it is in range.

If  $\sigma$  is very small and  $\mu = 0$ , only the  $k = 0$  term in the above sum needs to be retained, and there is no distinction between the original Gaussian restricted to the range  $\theta \in [-\pi, \pi]$ , the Gaussian clipped to this range, and the folded Gaussian. But as

$\sigma$  increases, so too do the values of  $|k|$  that need to be retained. As  $\sigma$  becomes very large, it becomes impractical to compute (2.44).

However, there is an alternative representation of the folded Gaussian that uses the fact that it is a periodic function. Recall that any  $2\pi$ -periodic function, i.e., a “function on the unit circle,” can be expanded in a Fourier series:

$$f(\theta) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{in\theta} \quad \text{where} \quad \hat{f}(n) = \int_0^{2\pi} f(\theta)e^{-in\theta} d\theta, \quad (2.45)$$

where  $e^{in\theta} = \cos n\theta + i \sin n\theta$  and  $i = \sqrt{-1}$ . Here  $\hat{f}(n)$  are called the *Fourier coefficients*, or circular Fourier transform. These coefficients can be computed in closed form for (2.44). This leads to the Fourier series representation of the folded Gaussian distribution:

$$\rho_W(\theta; \mu, \sigma) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-\frac{\sigma^2}{2}n^2} \cos(n(\theta - \mu)). \quad (2.46)$$

As  $\sigma$  becomes large, very close approximations can be achieved with the first couple of terms in the summation in (2.46). In contrast, as  $\sigma$  becomes very small, using very few of the terms in the series (2.44) will produce a very good approximation when  $\mu = 0$ .

The general theme that a Gaussian on a space other than the real line can be approximated well as a Gaussian restricted to a smaller domain when  $\sigma$  is small, or as a generalized Fourier series expansion when  $\sigma$  is large, will recur many times throughout this book.

Note that the above “folding” process is not restricted to Gaussian distributions; any well-behaved function,  $f(x)$ , defined on the line can be wrapped around the circle. The resulting folded function, which is  $2\pi$ -periodic, is related to the Fourier transform of the original non-periodic function on the real line through the *Poisson summation formula* [1]:

$$\sum_{n=-\infty}^{\infty} f(\theta + 2\pi n) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} [\mathcal{F}(f)](k)e^{ik\theta}. \quad (2.47)$$

In other words, the Fourier coefficients of the folded function are related to the Fourier transform of the original function as

$$\hat{f}(k) = [\mathcal{F}(f)](k).$$

## 2.6 The Heat Equation

In this section, the relationship between the Gaussian distribution and the heat equation (also called the diffusion equation) is developed.

Sometimes the exact solution of an equation is not as critical as knowing how its mean and covariance behave as a function of time. This is illustrated both in the one-dimensional and multi-dimensional settings in the following subsections.

### 2.6.1 The One-Dimensional Case

Consider the diffusion equation on the real line with time-varying diffusion and drift coefficients,  $k(t)$  and  $a(t)$ :

$$\frac{\partial f}{\partial t} = \frac{1}{2}k(t)\frac{\partial^2 f}{\partial x^2} - a(t)\frac{\partial f}{\partial x}. \quad (2.48)$$

The initial condition is  $f(x, 0) = \delta(x)$ . The solution  $f(x, t)$  can be obtained in closed form, following essentially the same procedure as in Section 2.1.5, and then the mean and variance can be computed from this solution as

$$\mu(t) = \int_{-\infty}^{\infty} xf(x, t)dx \quad \text{and} \quad \sigma^2(t) = \int_{-\infty}^{\infty} [x - \mu(t)]^2 f(x, t)dx. \quad (2.49)$$

Alternatively, the mean and variance of  $f(x, t)$  can be computed directly from (2.48) without actually knowing the solution  $f(x, t)$ . In fact, many properties of  $f(x, t)$  can be determined from (2.48) and the corresponding initial conditions without knowing  $f(x, t)$ . For example, integrating both sides of (2.48) with respect to  $x$  yields

$$\frac{d}{dt} \int_{-\infty}^{\infty} f(x, t)dx = 0.$$

This follows because

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial x} dx = f(x, t)|_{x=-\infty}^{\infty} \quad \text{and} \quad \int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial x^2} dx = \frac{\partial f}{\partial x} \Big|_{x=-\infty}^{\infty}$$

and under the boundary conditions that  $f(x, t)$  and  $\partial f/\partial x$  decay rapidly to zero as  $x \rightarrow \pm\infty$ , these terms become zero. Since the initial conditions are a delta function in  $x$ , it follows that

$$\int_{-\infty}^{\infty} f(x, t)dx = 1.$$

In other words, (2.48) preserves the initial mass of the distribution over all values of time after  $t = 0$ .

To compute  $\mu(t)$ , multiply both sides of (2.48) by  $x$  and integrate. On the one hand,

$$\int_{-\infty}^{\infty} x \frac{\partial f}{\partial t} dx = \frac{d}{dt} \int_{-\infty}^{\infty} xf(x, t)dx = \frac{d\mu}{dt}.$$

On the other hand,

$$\int_{-\infty}^{\infty} x \frac{\partial f}{\partial t} dx = \frac{1}{2}k(t) \int_{-\infty}^{\infty} x \frac{\partial^2 f}{\partial x^2} dx - a(t) \int_{-\infty}^{\infty} x \frac{\partial f}{\partial x} dx.$$

Evaluating both integrals on the right side by integrating by parts and using the conditions that both  $f(x, t)$  and  $\partial f/\partial x$  decay rapidly to zero as  $x \rightarrow \pm\infty$ , it becomes clear that

$$\frac{d\mu}{dt} = a(t) \quad \text{or} \quad \mu(t) = \int_0^t a(s)ds. \quad (2.50)$$

A similar argument shows that

$$\frac{d}{dt}(\sigma^2) = k(t) \quad \text{or} \quad \sigma^2(t) = \int_0^t k(s)ds. \quad (2.51)$$

### 2.6.2 The Multi-Dimensional Case

Consider the following time-varying diffusion equation without drift:

$$\frac{\partial f}{\partial t} = \frac{1}{2} \sum_{i,j=1}^n D_{ij}(t) \frac{\partial^2 f}{\partial x_i \partial x_j}, \quad (2.52)$$

where  $D_{ij}(t) = D_{ji}(t)$  are the *time-varying diffusion constants*. If  $f(\mathbf{x}, 0) = \delta(\mathbf{x})$ , then integrating (2.52) both sides over  $\mathbb{R}^n$  and using integration by parts in  $\mathbf{x}$  shows that the unit volume under the curve is preserved.

Multiplying both sides by  $x_k x_l$  and integrating over  $\mathbf{x} \in \mathbb{R}^n$  gives

$$\frac{d}{dt}(\sigma_{kl}) = \frac{1}{2} \sum_{i,j=1}^n D_{ij}(t) \int_{\mathbb{R}^n} x_k x_l \frac{\partial^2 f}{\partial x_i \partial x_j} d\mathbf{x}. \quad (2.53)$$

Let the integral over  $\mathbb{R}^{n-1}$  resulting from the exclusion of the integral over  $x_i$  be denoted as

$$\int_{\mathbf{x}-x_i} f(\mathbf{x}) d\mathbf{x} / dx_i = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_{i-1}=-\infty}^{\infty} \int_{x_{i+1}=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

so that

$$\int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} \left( \int_{\mathbf{x}-x_i} f(\mathbf{x}) d\mathbf{x} / dx_i \right) dx_i = \int_{\mathbf{x}-x_i} \left( \int_{-\infty}^{\infty} f(\mathbf{x}) dx_i \right) d\mathbf{x} / dx_i.$$

An integral over  $n-2$  degrees of freedom denoted by the integral with subscript  $\mathbf{x}-x_i-x_j$  follows in a similar way.

From integration by parts

$$\int_{\mathbb{R}^n} x_k x_l \frac{\partial^2 f}{\partial x_i \partial x_j} d\mathbf{x} = \int_{\mathbf{x}-x_i} \left[ x_k x_l \frac{\partial f}{\partial x_j} \Big|_{x_i=-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\partial}{\partial x_i} (x_k x_l) \frac{\partial f}{\partial x_j} dx_i \right] d\mathbf{x} / dx_i.$$

The assumption that  $f(\mathbf{x}, t)$  decays rapidly as  $\|\mathbf{x}\| \rightarrow \infty$  for all values of  $t$  makes the first term in the brackets disappear. Using the fact that  $\partial x_i / \partial x_j = \delta_{ij}$ , and integrating by parts again (over  $x_j$ ) reduces the above integral to

$$\int_{\mathbb{R}^n} x_k x_l \frac{\partial^2 f}{\partial x_i \partial x_j} d\mathbf{x} = \delta_{kj} \delta_{il} + \delta_{ik} \delta_{lj}.$$

Substituting this into (2.53) results in

$$\frac{d}{dt}(\sigma_{kl}) = D_{kl}(t) \quad \text{or} \quad \sigma_{kl}(t) = \int_0^t D_{kl}(s) ds. \quad (2.54)$$

Therefore, even without knowing the form of the time-varying pdf that solves (2.52) it is possible to obtain an exact expression for the covariance of the solution.

### 2.6.3 The Heat Equation on the Unit Circle

The heat equation on the circle is exactly the same as the heat equation on the real line (with  $\theta$  replacing  $x$  as the spatial variable). However, the topological constraint that  $\theta = \pm\pi$  represents the same point means that the long-time solution will be completely different than in the unconstrained case on the real line. Whereas the Fourier transform can be used to solve the heat equation on the line, the Fourier series expansion is used on the circle.

The result is that the solution on the line can be folded around the circle. In other words, the solution to the *heat equation on the circle* for constant diffusion coefficient  $k$ ,

$$\frac{\partial f}{\partial t} = \frac{1}{2}k \frac{\partial^2 f}{\partial \theta^2} \quad \text{subject to} \quad f(\theta, 0) = \delta(\theta),$$

is

$$f(\theta, t) = \sum_{k=-\infty}^{\infty} \rho(\theta - 2\pi k; 0, (kt)^{\frac{1}{2}}) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-ktn^2/2} \cos n\theta. \quad (2.55)$$

This is the folded Gaussian in (2.46) with  $\sigma^2 = kt$  and  $\mu = 0$ .

## 2.7 Gaussians and Multi-Dimensional Diffusions

In the previous section, the evolution of the mean and covariance of a diffusion equation was obtained without knowing the time-varying pdf. Here, the pdf is sought.

### 2.7.1 The Constant Diffusion Case

Consider the diffusion equation

$$\frac{\partial f}{\partial t} = \frac{1}{2} \sum_{i,j=1}^n D_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (2.56)$$

subject to the initial conditions  $f(\mathbf{x}, t) = \delta(\mathbf{x})$ , where  $D = [D_{ij}] = D^T$  is a constant matrix of diffusion constants.

Since diffusion equations preserve mass (see Section 2.6.2), it follows that

$$\int_{\mathbb{R}^n} f(\mathbf{x}, t) d\mathbf{x} = 1 \quad (2.57)$$

for all values of time,  $t \in \mathbb{R}_{>0}$ .

Try a solution of the form

$$f(\mathbf{x}, t) = c(t) \exp\left(-\frac{1}{2} \mathbf{x}^T A(t) \mathbf{x}\right) \quad (2.58)$$

where  $A(t) = \phi(t)A_0$  and  $A_0 = [\alpha_{ij}] = A_0^T$ . Then, from (2.57) and the formula (2.83) derived in the exercises, it follows that

$$c(t) = \left(\frac{\phi(t)}{2\pi}\right)^{n/2} |\det A_0|^{\frac{1}{2}}.$$

With this constraint in mind, substituting  $f(\mathbf{x}, t)$  into (2.56) produces the following conditions on  $\phi(t)$  and  $A_0$ :

$$n\phi' = -\phi^2 \sum_{i,j=1}^n D_{ij}\alpha_{ij}$$

$$\phi' \mathbf{x}^T A_0 \mathbf{x} = -\phi^2 \sum_{i,j=1}^n D_{ij} \left( \sum_{k=1}^n \alpha_{ik} x_k \right) \left( \sum_{l=1}^n \alpha_{jl} x_l \right)$$

where  $\phi' = d\phi/dt$ .

Both of the conditions (2.59) are satisfied if  $A_0 = \alpha_0 D^{-1}$  and  $\phi(t) = (\alpha_0 t)^{-1}$  for some arbitrary constant  $\alpha_0 \in \mathbb{R}_{>0}$ . But since  $A(t) = \phi(t)A_0 = t^{-1}D^{-1}$ , this constant does not matter.

Putting all of this together,

$$\boxed{f(\mathbf{x}, t) = \frac{1}{(2\pi t)^{n/2} |\det D|^{\frac{1}{2}}} \exp\left(-\frac{1}{2t} \mathbf{x}^T D^{-1} \mathbf{x}\right).} \quad (2.59)$$

Stated in another way, the solution to (2.56) is a time-varying Gaussian distribution with  $\Sigma(t) = tD$  when  $D$  is symmetric.

### 2.7.2 The Time-Varying Case

Consider again (2.56), but now let  $D = D(t)$ . Try a solution of the form

$$f(\mathbf{x}, t) = (2\pi)^{-n/2} |\det \Sigma(t)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1}(t) \mathbf{x}\right) \quad (2.60)$$

where  $\Sigma(t)$  is a time-varying covariance matrix, the form of which is as yet undetermined. This guess is simply  $f(\mathbf{x}, t) = \rho(\mathbf{x}; \mathbf{0}, \Sigma(t))$ .

The derivatives with respect to  $x_i$  are evaluated as before, using the chain rule. The time derivative is evaluated as follows:

$$\begin{aligned} \frac{\partial f}{\partial t} &= (2\pi)^{-n/2} \frac{d(|\det \Sigma|^{-\frac{1}{2}})}{dt} \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right) \\ &\quad + (2\pi)^{-n/2} |\det \Sigma|^{-\frac{1}{2}} \frac{d}{dt} \left[ \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right) \right] \\ &= -\frac{1}{2} (2\pi)^{-n/2} |\det \Sigma|^{-\frac{3}{2}} \frac{d(\det \Sigma)}{dt} \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right) \\ &\quad - \frac{1}{2} (2\pi)^{-n/2} |\det \Sigma|^{-\frac{1}{2}} \left( \mathbf{x}^T \frac{d}{dt} [\Sigma^{-1}] \mathbf{x} \right) \exp\left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right). \end{aligned}$$

On the other hand,

$$\frac{1}{2} \sum_{i,j=1}^n D_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{1}{2} \left\{ -\text{tr}(D\Sigma^{-1}) + \mathbf{x}^T (\Sigma^{-1} D \Sigma^{-1}) \mathbf{x} \right\} f(\mathbf{x}, t).$$

Therefore, if

$$|\det \Sigma|^{-1} \frac{d(\det \Sigma)}{dt} = \text{tr}(D\Sigma^{-1}) \quad \text{and} \quad \frac{d}{dt} [\Sigma^{-1}] = -\Sigma^{-1} D \Sigma^{-1}, \quad (2.61)$$

then (2.56) with variable diffusion coefficients will be satisfied. Since<sup>8</sup>

$$\frac{d}{dt}(\Sigma\Sigma^{-1}) = \mathbb{O} \quad \implies \quad \frac{d}{dt}[\Sigma^{-1}] = -\Sigma^{-1}\dot{\Sigma}\Sigma^{-1},$$

the second equality in (2.61) will be satisfied if  $D = \dot{\Sigma}$ . In this case the first equality in (2.61) becomes

$$\frac{d}{dt} \log(\det \Sigma) = \text{tr}(\dot{\Sigma}\Sigma^{-1}). \quad (2.62)$$

Under what conditions will this be true?

*Case 1:*

From Systems Theory (as reviewed in the appendix), if  $\Sigma = \exp(tS_0)$  where  $S_0 = S_0^T$  is constant, then

$$\det \Sigma = e^{\text{tr}(tS_0)} = e^{t(\text{tr}S_0)}.$$

Therefore, in this special case

$$\frac{d}{dt} \log(\det \Sigma) = \text{tr}(S_0).$$

Likewise, if  $\Sigma = \exp(tS_0)$ , then  $\text{tr}(\dot{\Sigma}\Sigma^{-1}) = \text{tr}(S_0)$ . Therefore, it can be concluded that a sufficient condition for the Gaussian in (2.60) to be a solution to (2.56) is if a constant symmetric matrix  $S_0$  can be found such that  $D(t) = S_0 \exp(tS_0)$ .

*Case 2:*

The condition in (2.62) will be satisfied if  $\Sigma = \sigma(t)\Sigma_0$  where  $\sigma(t)$  is a differentiable scalar function of time and  $\Sigma_0 = \Sigma_0^T$ . Substitution into (2.62) yields the condition

$$\frac{d}{dt} \log(\sigma^n \det \Sigma_0) = \dot{\sigma}\sigma^{-1} \text{tr}(\mathbb{I}).$$

Since  $\log(a \cdot b) = \log a + \log b$ , and  $\frac{d}{dt} \log a(t) = \dot{a}/a$ , the above condition becomes

$$\frac{1}{\sigma^n} n\sigma^{n-1}\dot{\sigma} = n\dot{\sigma}\sigma^{-1},$$

which is always true. Therefore any  $\sigma(t)$  will work.

A broader condition that encompasses both Case 1 and Case 2 is  $D(t) = \dot{S}(t) \exp S(t)$  where  $S = S^T$  and  $[\dot{S}, S] \doteq \dot{S}S - S\dot{S} = \mathbb{O}$ .

Under this condition,

$$\Sigma(t) = \int_0^t D(s) ds. \quad (2.63)$$

## 2.8 Symmetry Analysis of Evolution Equations

The concept of symmetry can have several meanings when applied to *evolution equations*.<sup>9</sup> For example, the diffusion matrix in the multi-dimensional heat equation might have symmetries in it other than the primary symmetry  $D = D^T$ . That kind of symmetry is reflected in the solution of the equation. Another kind of symmetry is that the equation itself can be solved when the independent variables undergo a non-linear change of coordinates. Both of these concepts of symmetry are addressed in this section.

<sup>8</sup>Here  $\mathbb{O} = \frac{d}{dt}(\mathbb{I})$  is the zero matrix.

<sup>9</sup>These are equations with a single partial derivative in time, and multiple partial derivatives in space. They include, but are not limited to, diffusion equations.



### 2.8.1 Symmetries in Parameters

Consider a drift-free diffusion in  $\mathbb{R}^n$  with constant diffusion matrix  $D = D^T$ , and let the solution be denoted as  $f(\mathbf{x}, t; D)$ . Since the dependence on  $D$  and  $t$  always appears as their product, the solution has a continuous scale symmetry of the form

$$f(\mathbf{x}, t; D) = f(\mathbf{x}, t/\alpha; \alpha D)$$

for any  $\alpha \in \mathbb{R}_{>0}$ .

In addition, since the solution is the Gaussian distribution in (2.59), it can be verified that

$$f(\mathbf{x}, t; D) = \beta^{n/2} f(\sqrt{\beta}\mathbf{x}, t; \beta D).$$

If  $D = \sigma^2 I$ , then any change of spatial coordinates of the form  $\mathbf{y} = Q\mathbf{x}$  where  $Q^T Q = I$  will preserve the solution:

$$f(Q\mathbf{x}, t; \sigma^2 I) = f(\mathbf{x}, t; \sigma^2 I).$$

In contrast, if  $n = 3$  and  $D = \text{diag}[\sigma_1^2, \sigma_1^2, \sigma_3^2]$  is the diagonal matrix with the indicated entries on the diagonal, then

$$R_3(\theta)^T D R_3(\theta) = D \quad \text{where} \quad R_3(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and so

$$f(R_3(\theta)\mathbf{x}, t; D) = f(\mathbf{x}, t; D).$$

These symmetries all involve simple transformations of the coordinates. Less obvious symmetries result by examining operators which, when applied to the equation of interest, leave it invariant in a sense that will be made precise.

### 2.8.2 Infinitesimal Symmetry Operators of the Heat Equation

Let  $Qf = 0$  denote any partial differential equation, where  $Q$  is a differential operator in temporal and spatial variables  $(t, \mathbf{x}) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^n$ . For example, for the heat equation on the real line where there is only one spatial variable  $(t, \mathbf{x})$  becomes  $(t, x)$  and

$$Q = \frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}$$

where the diffusion constant,  $k$ , is chosen to be  $k = 2$  here for convenience.

A body of literature exists that addresses the question of how to obtain new solutions of  $Qf = 0$  from old ones. In particular, if it is possible to find a first-order operator of the form

$$L = T(\mathbf{x}, t) \frac{\partial}{\partial t} + \sum_{i=1}^n X_i(\mathbf{x}, t) \frac{\partial}{\partial x_i} + Z(\mathbf{x}, t) \quad (2.64)$$

where  $T(\mathbf{x}, t)$ ,  $X_i(\mathbf{x}, t)$ , and  $Z(\mathbf{x}, t)$  are analytic functions such that

$$[L, Q]f(\mathbf{x}, t) = R(\mathbf{x}, t)Qf \quad \text{where} \quad [L, Q] = LQ - QL, \quad (2.65)$$

then  $f' \doteq Lf$  will solve  $Qf' = 0$ .

At first this might seem surprising, but since the condition in (2.64) reads  $LQf - QLf = RQf$ , and since  $Qf = 0$ , it must be that  $0 = QLf = Q(Lf) = Qf'$ .

Following [2, 3, 4, 17, 19, 21], the infinitesimal operators that transform solutions of the heat equation into new solutions are presented below. In this case there is one spatial dimension and so

$$L = T(x, t) \frac{\partial}{\partial t} + X(x, t) \frac{\partial}{\partial x} + Z(x, t).$$

Some mundane calculus yields

$$\begin{aligned} LQf &= \left( \frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2} \right) \left( T(x, t) \frac{\partial}{\partial t} + X(x, t) \frac{\partial}{\partial x} + Z(x, t) \right) \\ &= \left( \frac{\partial T}{\partial t} \right) \left( \frac{\partial f}{\partial t} \right) + T \left( \frac{\partial^2 f}{\partial t^2} \right) + \left( \frac{\partial X}{\partial t} \right) \left( \frac{\partial f}{\partial x} \right) + X \left( \frac{\partial^2 f}{\partial t \partial x} \right) + \left( \frac{\partial Z}{\partial t} \right) f \\ &\quad + Z \left( \frac{\partial f}{\partial t} \right) - \left( \frac{\partial^2 T}{\partial x^2} \right) \left( \frac{\partial f}{\partial t} \right) - 2 \left( \frac{\partial T}{\partial x} \right) \left( \frac{\partial^2 f}{\partial t \partial x} \right) - T \left( \frac{\partial^3 f}{\partial t \partial x^2} \right) \\ &\quad - \left( \frac{\partial^2 X}{\partial x^2} \right) \left( \frac{\partial f}{\partial x} \right) - 2 \left( \frac{\partial X}{\partial x} \right) \left( \frac{\partial^2 f}{\partial x^2} \right) - X \left( \frac{\partial^3 f}{\partial x^3} \right) \\ &\quad - \left( \frac{\partial^2 Z}{\partial x^2} \right) f - 2 \left( \frac{\partial Z}{\partial x} \right) \left( \frac{\partial f}{\partial x} \right) - Z \left( \frac{\partial^2 f}{\partial x^2} \right) \end{aligned}$$

and

$$\begin{aligned} LQf &= \left( T(x, t) \frac{\partial}{\partial t} + X(x, t) \frac{\partial}{\partial x} + Z(x, t) \right) \left( \frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2} \right) f \\ &= T \frac{\partial^2 f}{\partial t^2} - T \frac{\partial^3 f}{\partial x^2 \partial t} + X \frac{\partial^2 f}{\partial x \partial t} - X \frac{\partial^3 f}{\partial x^3} + Z \frac{\partial f}{\partial t} - Z \frac{\partial^2 f}{\partial x^2}. \end{aligned}$$

Note that every term in  $LQf$  can also be found in  $QLf$ . Subtracting, and reorganizing the terms that result, yields

$$\begin{aligned} [Q, L]f &= \left( \frac{\partial T}{\partial t} - \frac{\partial^2 T}{\partial x^2} \right) \frac{\partial f}{\partial t} + \left( \frac{\partial X}{\partial t} - \frac{\partial^2 X}{\partial x^2} - 2 \frac{\partial Z}{\partial x} \right) \frac{\partial f}{\partial x} \\ &\quad + \left( -2 \frac{\partial T}{\partial x} \right) \frac{\partial^2 f}{\partial x \partial t} + \left( -2 \frac{\partial X}{\partial x} \right) \frac{\partial^2 f}{\partial x^2} + \left( \frac{\partial Z}{\partial t} - \frac{\partial^2 Z}{\partial x^2} \right) f. \end{aligned}$$

Since  $[Q, L] = -[L, Q]$ , (2.64) is the same as computing  $[Q, L]f = -RQf$  where

$$RQf = R \frac{\partial f}{\partial t} - R \frac{\partial^2 f}{\partial x^2}.$$

Then equating the coefficients in front of each term involving  $f$ , the following five equations result:

$$\frac{\partial T}{\partial t} - \frac{\partial^2 T}{\partial x^2} = -R \tag{2.66}$$

$$2 \frac{\partial X}{\partial x} = -R \tag{2.67}$$

$$\frac{\partial X}{\partial t} - \frac{\partial^2 X}{\partial x^2} - 2 \frac{\partial Z}{\partial x} = 0 \tag{2.68}$$

$$\frac{\partial T}{\partial x} = 0 \quad (2.69)$$

$$\frac{\partial Z}{\partial t} - \frac{\partial^2 Z}{\partial x^2} = 0. \quad (2.70)$$

These equations completely determine the structure of the operator  $L$  that transforms solutions into solutions.

Starting with (2.69), the restriction  $T(x, t) = T(t)$  must be observed. Then, using this result in (2.66) means  $-R(x, t) = T'(t)$ . This in turn can be substituted into (2.67) to yield

$$X(x, t) = \frac{1}{2}T'(t)x + c_1(t)$$

where  $c_1(t)$  is a yet-to-be-determined function resulting from integration over  $x$ . Substituting this into (2.68) forces the form of  $Z(x, t)$  to be

$$Z(x, t) = \frac{1}{8}T''(t)x^2 + \frac{1}{2}c_1'(t)x + c_2(t).$$

Substituting this into (2.70) forces

$$T'''(t) = 0; \quad c_1''(t) = 0; \quad c_2'(t) = \frac{1}{4}T''(t).$$

It follows that

$$T(t) = a_0t^2 + b_0t + c_0; \quad c_1(t) = \alpha_0t + \beta_0; \quad c_2(t) = \frac{1}{2}a_0t + \gamma_0$$

where  $a_0, b_0, c_0, \alpha_0, \beta_0, \gamma_0$  are all free constants.

This means that any  $L$  with the following form will map solutions of the heat equation into solutions:

$$\begin{aligned} T(x, t) &= a_0t^2 + b_0t + c_0 \\ X(x, t) &= (a_0t + b_0/2)x + \alpha_0t + \beta_0 \\ Z(x, t) &= \frac{1}{4}a_0x^2 + \frac{1}{2}\alpha_0x + \frac{1}{2}a_0t + \gamma_0. \end{aligned}$$

In fact, the space of all allowable  $L$  operators is a vector space with elements of the form

$$L = a_0L_1 + b_0L_2 + c_0L_3 + \alpha_0L_4 + \beta_0L_5 + \gamma_0L_6$$

where the following serves as a basis:

$$L_1 = t^2 \frac{\partial}{\partial t} + xt \frac{\partial}{\partial x} + \frac{1}{4}x^2 + \frac{1}{2}t \quad (2.71)$$

$$L_2 = t \frac{\partial}{\partial t} + \frac{1}{2}x \frac{\partial}{\partial x} \quad (2.72)$$

$$L_3 = \frac{\partial}{\partial t} \quad (2.73)$$

$$L_4 = t \frac{\partial}{\partial x} + \frac{1}{2}x \quad (2.74)$$

$$L_5 = \frac{\partial}{\partial x} \quad (2.75)$$

$$L_6 = 1. \quad (2.76)$$

In addition to being a vector space, operators of the form of  $L$  given above are also closed under the Lie bracket,  $[\cdot, \cdot]$ . In other words,  $[L_i, L_j]$  for any  $i, j \in \{1, \dots, 6\}$  will result in a linear combination of these same basis elements. This makes the space of all  $L$  operators that map solutions of the heat equation into solutions a *Lie algebra* [16]. This concept will be defined more rigorously in the appendix and in Volume 2.

### 2.8.3 Non-Linear Transformations of Coordinates

Consider the heat equation

$$\frac{\partial f}{\partial t} = \frac{\partial f^2}{\partial x^2}$$

and assume that an  $f(x, t)$  has been obtained that satisfies this equation. For the moment, the initial conditions will be left unspecified.

The following matrices can be defined [17]:

$$B = B(u, v, w) = \begin{pmatrix} 1 & v & 2w + uv/2 \\ 0 & 1 & u \\ 0 & 0 & 1 \end{pmatrix} \quad \text{where } u, v, w \in \mathbb{R} \quad (2.77)$$

and

$$A = A(\alpha, \beta, \gamma, \delta) = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad \text{where } \alpha, \beta, \gamma, \delta \in \mathbb{R} \quad \text{and } \alpha\delta - \beta\gamma = 1. \quad (2.78)$$

It is clear that since  $\det A = 1$  by definition, then the product of two such matrices also satisfies this condition:  $\det(A_1 A_2) = \det A_1 \det A_2 = 1$ . Likewise, the form of the  $B$  matrices are preserved under matrix multiplication, and

$$B(u, v, w)B(u', v', w') = B(u + u', v + v', w + w' + (vu' - uv')/4).$$

These are examples of *matrix Lie groups* which, roughly speaking, are groups of continuous transformations, the elements of which are matrices. The group operation is matrix multiplication.

It can be shown (see Exercise 2.18) that transformations of the following form convert solutions into solutions [17]:

$$(T_1(B)f)(x, t) = \exp \frac{1}{2} \left[ b_{13} + b_{23}x + \frac{1}{2}b_{23}^2 t \right] f(x + b_{12} + b_{23}t, t) \quad (2.79)$$

and

$$(T_2(A)f)(x, t) = \exp \left( -\frac{x^2\beta/4}{\delta + t\beta} \right) (\delta + t\beta)^{-\frac{1}{2}} f \left( \frac{x}{\delta + t\beta}, \frac{\gamma + t\alpha}{\delta + t\beta} \right). \quad (2.80)$$

In other words, if  $f(x, t)$  is a solution to the heat equation, then so too are  $f_1(x, t) = (T_1(B)f)(x, t)$  and  $f_2(x, t) = (T_2(A)f)(x, t)$ . This means that applying these transformations twice with different permissible matrices  $A_i$  and  $B_i$  will also take solutions into solutions:

$$f_1(x, t) = (T_1(B_2)T_1(B_1)f)(x, t) = (T_1(B_2)(T_1(B_1)f))(x, t)$$

and

$$f_2(x, t) = (T_2(A_2)T_2(A_1)f)(x, t) = (T_2(A_2)(T_2(A_1)f))(x, t).$$

This gets really interesting when these definitions are combined with the closure property under multiplication of matrices of the same kind since

$$T_1(B_2)T_1(B_1) = T_1(B_2B_1) \quad \text{and} \quad T_2(A_2)T_2(A_1) = T_2(A_2A_1). \quad (2.81)$$

What this means is that there are two independent sets of three-parameter transformations that can map solutions into solutions. And furthermore, these can be combined since  $(T_2(A)(T_1(B)f))(x, t)$  and  $(T_1(B)(T_2(A)f))(x, t)$  must also be solutions.

In Volume 2 this example will be revisited as an example of a six-dimensional Lie group, where the  $A$  matrices and  $B$  matrices each independently form three-dimensional subgroups.

## 2.9 Chapter Summary

Many aspects of the Gaussian distribution were reviewed. These include the parametrization of multi-dimensional Gaussians by their mean and covariance, the form of marginals and conditionals of Gaussians, the properties of Gaussians under convolution, the maximum entropy property, and the relationship between Gaussians and diffusion/heat equations.<sup>10</sup> Finally, a brief review of the theory of symmetry analysis of partial differential equations, as applied to diffusion equations, was presented. This forms the first of many links between the topic of diffusion equations and Lie groups that will be forged throughout these books.

The connection between Lie group methods and partial differential equations has a long history dating back to the 1950s [25, 26, 27, 28]. In addition to those references cited earlier in this chapter, significant progress on this topic was made through the 1970s and 1980s including [5, 13, 14, 18, 20]. These approaches have been used for very complicated partial differential equations, such as in [21].

The next chapter will serve as a more formal introduction to probability and information theory. With the concrete example of the Gaussian distribution in mind, it should be easier to tackle these problems. Furthermore, the maximum entropy property of Gaussians, as well as their role in the central limit theorem will justify what might appear to be a preoccupation with Gaussians in the current chapter.

## 2.10 Exercises

2.1. Verify (2.6) by performing the integrals in the definitions of  $\sigma^2$  and  $s$ .

2.2. Verify (2.10).

2.3. Verify (2.14) by: (a) directly computing the convolution integral in (2.13); (b) using the convolution property of the Fourier transform (2.17).

---

<sup>10</sup>Note that although these equations were written in Cartesian coordinates in this chapter, it is possible to convert to polar, spherical, or other coordinates. For covariance matrices with symmetry, this can be more convenient. See [8] for a detailed discussion of different curvilinear coordinate systems.

2.4. Using the same reasoning as in Section 2.1.2, compute: (a) the maximum entropy distribution on the real line subject to the constraint that it has a specified value of the spread (rather than variance); (b) the maximum entropy distribution on the finite interval  $[a, b]$  subject to no constraints.

2.5. What is the exact expression for the functional  $F(\epsilon^2)$  in Section 2.1.2?

2.6. Prove that for any suitable  $f_1(x)$  and  $f_2(x)$ , the convolution theorem (2.17) holds. Hint: a change of variables and a change in the order in which integrals are performed will be required.

2.7. Verify (2.27). Hint: Use the block decomposition in (A.70) to obtain an explicit expression for  $\Sigma^{-1}$ .

2.8. Verify (2.29). Hint: Use the property of the exponential function  $e^{a+b} = e^a e^b$ .

2.9. Calculate the exact form of  $r(\sigma_0)$  in (2.38) in terms of the error function in (2.39).

2.10. Work out the covariance matrix for the 2D clipped Gaussian in analogy with the 3D case presented in (2.43).

2.11. Following the steps in Section 2.1.5, derive the closed-form solution  $f(x, t)$  that satisfies (2.48) subject to the initial conditions  $f(x, 0) = \delta(x)$ .

2.12. Using (2.49), show that the mean and variance of  $f(x, t)$  computed from (2.48) in (2.50) and (2.51) are the same as computed directly from the closed-form solution of  $f(x, t)$  obtained in the previous exercise.

2.13. Verify (2.46) analytically by computing the Fourier coefficients  $\hat{\rho}_W(n; \mu, \sigma)$  of (2.44).

2.14. Using integration by parts, prove (2.50) and (2.51).

2.15. Show that the matrices in (2.77) and (2.78) are invertible.

2.16. Find the nine basis operators  $\{L_i\}$  that take solutions of

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (2.82)$$

into other solutions.

2.17. Find the thirteen basis operators  $\{L_i\}$  that take solutions of

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

into other solutions.

2.18. Verify that transformations of the form in (2.79) and (2.80) will transform one solution into another. Hint: Use the chain rule.

2.19. Verify that the two equations in (2.81) hold. That is, first compute two concatenated transformations, and then compute the single transformation resulting from the matrix products, and compare.

2.20. Show that for  $A \in \mathbb{R}^{n \times n}$  with  $A = A^T > 0$ ,

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{x}^T A \mathbf{x}\right) d\mathbf{x} = (2\pi)^{n/2} |\det A|^{-\frac{1}{2}}. \quad (2.83)$$

Hint: Decompose  $A = QAQ^T$  where  $Q$  is orthogonal and  $A$  is the diagonal matrix consisting of eigenvalues of  $A$ , which are all positive.

2.21. Verify (2.59) by substituting (2.58) into (2.56) and using the chain rule.

2.22. Can a  $2 \times 2$  matrix  $S(t)$  be constructed such that  $\Sigma = \exp S$  which does not fall into Case 1 or Case 2? If so, provide an example. If not, explain why not.

2.23. Determine conditions under which the time-dependent diffusion with drift

$$\frac{\partial f}{\partial t} = \frac{1}{2} \sum_{k,l=1}^n D_{kl}(t) \frac{\partial^2 f}{\partial x_k \partial x_l} - \sum_{k=1}^n d_k(t) \frac{\partial f}{\partial x_k} \quad (2.84)$$

will have a solution of the form

$$f(\mathbf{x}, t) = c(t) \exp \left[ -\frac{1}{2} [\mathbf{x} - \mathbf{a}(t)]^T C(t) [\mathbf{x} - \mathbf{a}(t)] \right]. \quad (2.85)$$

2.24. Show that the following transformations take solutions of (2.82) into solutions [17]:

$$T_1(\mathbf{w}, \mathbf{z}, \omega) f(\mathbf{x}, t) = \exp \left[ \frac{1}{2} \mathbf{x} \cdot \mathbf{w} + \frac{1}{4} t \|\mathbf{w}\|^2 + \omega \right] f(\mathbf{x} + t\mathbf{w} + \mathbf{z}, t) \quad (2.86)$$

where  $\mathbf{w}, \mathbf{z} \in \mathbb{R}^2$  and  $\omega \in \mathbb{R}$ ;

$$T_2(A) f(\mathbf{x}, t) = \exp \left[ -\frac{1}{4} (\delta + t\beta)^{-1} \beta \|\mathbf{x}\|^2 \right] (\delta + t\beta)^{-1} f \left( (\delta + t\beta)^{-1} \mathbf{x}, \frac{\gamma + t\alpha}{\delta + t\beta} \right) \quad (2.87)$$

where  $A \in \mathbb{R}^{2 \times 2}$  with  $\det A = 1$ ;

$$T_3(\theta) f(\mathbf{x}, t) = f(R^T(\theta)\mathbf{x}, t) \quad \text{where} \quad R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (2.88)$$

## References

1. Benedetto, J.J., Zimmermann, G., "Sampling multipliers and the Poisson summation formula," *J. Fourier Anal. Appl.*, 3, pp. 505–523, 1997.
2. Bluman, G., *Construction of Solutions to Partial Differential Equations by the Use of Transformation Groups*, PhD Dissertation, Caltech, 1967.
3. Bluman, G., Cole, J., "The general similarity solution of the heat equation," *J. Math. Mech.*, 18, pp. 1025–1042, 1969.
4. Bluman, G., Cole, J., *Similarity Methods for Differential Equations*, Springer-Verlag, New York, 1974.
5. Boyer, C., "The maximal kinematical invariance group for an arbitrary potential," *Helv. Phys. Acta*, 47, pp. 589–605, 1974.
6. Bracewell, R.N., *The Fourier Transform and Its Applications*, 2nd ed., McGraw-Hill, New York, 1986.
7. Burrus, C.S., Parks, T.W., *DFT/FFT and Convolution Algorithms*, John Wiley and Sons, New York, 1985.

8. Chirikjian, G.S., Kyatkin, A.B., *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, Boca Raton, FL, 2001.
9. Chirikjian, G.S., "A methodology for determining mechanical properties of macromolecules from ensemble motion data," *Trends Anal. Chem.*, 22, pp. 549–553, 2003.
10. Cooley, J.W., Tukey, J., "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, 19, pp. 297–301, 1965.
11. Fourier, J.B.J., *Théorie Analytique de la Chaleur*, F. Didot, Paris, 1822.
12. Gray, A., *Tubes*, 2nd ed., Birkhäuser, Boston, 2004.
13. Kalnins, E., Miller, W., Jr., "Symmetry and separation of variables for the heat equation," *Proc. Conf. on Symmetry, Similarity and Group-Theoretic Methods in Mechanics*, pp. 246–261, University of Calgary, Calgary, Canada, 1974.
14. Kalnins, E., Miller, W., Jr., "Lie theory and separation of variables 4: The groups  $SO(2, 1)$  and  $SO(3)$ ," *J. Math. Phys.*, 15, pp. 1263–1274, 1974.
15. Körner, T.W., *Fourier Analysis*, Cambridge University Press, London, 1988 (reprinted 1993).
16. Lie, S., "Über die Integration durch bestimmte Integral von einer Klasse linearer partieller Differentialgleichungen," *Arch. Math.*, 4(3), Kristiana, p. 328, 1881.
17. Miller, W., Jr., *Symmetry and Separation of Variables*, Encyclopedia of Mathematics and Its Applications, Vol. 4, G.-C. Rota, ed., Addison-Wesley, Reading MA, 1974.
18. Olver, P.J., *Applications of Lie Groups to Differential Equations*, Springer-Verlag, New York, 1986.
19. Ovsiannikov, L.V., *Group Analysis of Differential Equations*, Academic Press, New York, 1982.
20. Patera, J., Winternitz, P., "A new basis for the representations of the rotation group: Lamé and Heun polynomials," *J. Math. Phys.*, 14, pp. 1130–1139, 1973.
21. Poluyanov, L.V., Aguilar, A., González, M., *Group Properties of the Acoustic Differential Equation*, Taylor & Francis, London, 1995.
22. Rényi, A., *Probability Theory*, North-Holland, Amsterdam, 1970.
23. Shannon, C.E., "A mathematical theory of communication," *Bell Syst. Tech. J.*, 27, pp. 379–423 and 623–656, July and October, 1948.
24. Van Loan, C., *Computational Frameworks for the Fast Fourier Transform*, SIAM, Philadelphia, 1992.
25. Weisner, L., "Group-theoretic origin of certain generating functions," *Pacific J. Math.*, 5, pp. 1033–1039, 1955.
26. Weisner, L., "Generating functions for Hermite functions," *Can. J. Math.*, 11, pp. 141–147, 1959.
27. Weisner, L., "Generating functions for Bessel functions," *Can. J. Math.*, 11, pp. 148–155, 1959.
28. Winternitz, P., Fris, I., "Invariant expansions of relativistic amplitudes and subgroups of the proper Lorentz group," *Soviet Physics JNP*, 1, pp. 636–643, 1965.



---

## Probability and Information Theory

This chapter serves as an introduction to concepts from elementary probability theory and information theory in the concrete context of the real line and multi-dimensional Euclidean space. The probabilistic concepts of mean, variance, expected value, marginalization, conditioning, and conditional expectation are reviewed. In this part of the presentation there is some overlap with the previous chapter, which has some pedagogical benefit. There will be no mention of Borel measurability,  $\sigma$ -algebras, filtrations, or martingales, as these are treated in numerous other books on probability theory and stochastic processes such as [1, 14, 15, 32, 27, 48]. The presentation here, while drawing from these excellent works, will be restricted only to those topics that are required either in the mathematical and computational modeling of stochastic physical systems, or the determination of properties of solutions to the equations in these models.

Basic concepts of information theory are addressed such as measures of distance, or “divergence,” between probability density functions, and the properties of “information” and entropy. All pdfs treated here will be differentiable functions on  $\mathbb{R}^n$ . Therefore the entropy and information measures addressed in this chapter are those that are referred to in the literature as the “differential” or “continuous” version.

It is shown that the amount of information contained in pdfs decreases with convolution, while the entropy contained in them increases. Information theory and Fourier analysis are both used in this chapter to derive the central limit theorem, which states that under suitable conditions iterated convolutions converge to Gaussian distributions. All of the concepts presented here will be extended to the context of Lie groups in Volume 2.

For the reader who is already familiar with probability and information theory, the main point that should be taken away from this chapter is the non-standard notation that is used. Rather than  $E[\mathbf{X}]$  denoting the expected value of the random variable<sup>1</sup>  $\mathbf{X}$ , the notation  $\langle \mathbf{x} \rangle = \int \mathbf{x}\rho(\mathbf{x})d\mathbf{x}$  is used here where  $\rho(\mathbf{x})$  is the probability density function for  $\mathbf{X}$ . More generally,  $\langle f(\mathbf{x}) \rangle = \int f(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}$  for any function,  $f(\mathbf{x})$  (which need not be scalar valued). Instead of denoting the Shannon entropy as  $H(\mathbf{X})$ , it is denoted here as  $S(\rho)$ . Fisher information is denoted as  $F(\rho)$ . Defining these quantities

---

<sup>1</sup>A random variable is a mathematical object such as a scalar, vector, or matrix that does not have a specific fixed value, but rather can take on any of a number of values. That is, each time the random variable is queried (or interrogated), it can return a different value. But the distribution of these values is fixed, meaning that as the number of queries goes to infinity, the underlying distribution that defines the random variable is observed. The space over which these values are defined can be discrete or continuous.

in terms of probability densities rather than random variables will lead to conveniences when concepts from Lie theory are added to the mix.

For the reader who is not familiar with probability and information theory the main things to take away from this chapter are:

- To know that the definitions of convolution, mean, covariance, and marginal and conditional densities, are fully general, and apply to a wide variety of probability density functions (not only Gaussians);
- To understand the definitions and properties of (continuous/differential) information-theoretic entropy, including how it scales and how it behaves under convolution;
- To become familiar with the concepts of conditional expectation and marginal entropy;
- To understand the fundamental inequalities of information theory such as the Cramér–Rao bound and the de Bruijn identity, and the entropy power inequality;
- To be able to follow the statement of the central limit theorem, the conditions under which it holds, and to have an idea of the various ways that its proof can be approached.

### 3.1 Probability Theory in Euclidean Space

This section reviews basic concepts from probability theory on the real line,  $\mathbb{R}$ , and in multi-dimensional Euclidean space,  $\mathbb{R}^n$ .

#### 3.1.1 Basic Definitions and Properties of Probability Density Functions

In classical probability theory a deterministic vector has a specific unambiguous value, and is denoted in lower case as  $\mathbf{x} \in \mathbb{R}^n$ . In contrast, a random vector, which can take on any of a variety of vector values, is denoted in upper case as  $\mathbf{X} \in \mathbb{R}^n$ . Some of these values will be more likely encountered than others. The relative likelihood that a specific deterministic value will be encountered is characterized by a *probability density function* (or pdf for short). A pdf on  $\mathbb{R}^n$  is a non-negative real-valued function that integrates to unity:<sup>2</sup>

$$\rho(\mathbf{x}) \geq 0 \quad \forall \quad \mathbf{x} \in \mathbb{R}^n \quad \text{and} \quad \int_{\mathbb{R}^n} \rho(\mathbf{x}) d\mathbf{x} = 1 \quad (3.1)$$

where  $d\mathbf{x} = dx_1 dx_2 \cdots dx_n$  is the usual integration measure on  $\mathbb{R}^n$ .

The probability that  $\mathbf{X} \in B \subset \mathbb{R}^n$  is then computed as

$$P[\mathbf{X} \in B] = \int_B \rho(\mathbf{x}) d\mathbf{x}.$$

Furthermore, the *expected value* (or *expectation*) of a function of  $X$  is computed as

$$E[\alpha(\mathbf{X})] = \int_{\mathbb{R}^n} \alpha(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x} \quad (3.2)$$

where  $\alpha(\cdot)$  can be a scalar, vector, or matrix/tensor-valued function of vector-valued argument. The mean and covariance are special cases of expected values:

$$\boldsymbol{\mu} = E[\mathbf{X}] \quad \text{and} \quad \boldsymbol{\Sigma} = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T].$$

---

<sup>2</sup>Pdfs will be denoted as  $\rho(\cdot)$  and  $f(\cdot)$ .

The expectation operator is linear due to the linearity of integration, i.e.,

$$E[b_1\alpha_1(\mathbf{X}) + b_2\alpha_2(\mathbf{X})] = b_1E[\alpha_1(\mathbf{X})] + b_2E[\alpha_2(\mathbf{X})]. \quad (3.3)$$

This standard notation, while perfectly fine for addressing problems in  $\mathbb{R}^n$ , will lead to some difficulties when considering stochastic modeling problems on Lie groups. This is because lower case and upper case letters have specific meanings in that context that are separate from anything having to do with probabilistic concepts. Therefore, an alternative (but equivalent) formalism to that used in standard probability theory will be used throughout this chapter. Namely, there will be no mention of random variables. Instead, only probability density functions and the domains on which these pdfs are evaluated will appear in equations. Since the very concept of “expected value” is defined in the context of random variables, there will not be any notation of the form  $E[\alpha(\mathbf{X})]$  throughout the remainder of the book. Instead, (3.2) will be written in the shorthand  $\langle\alpha(\mathbf{x})\rangle$ . In words, this is the average of  $\alpha(\mathbf{x})$  over the ensemble where the relative frequency of occurrence of each value of  $\mathbf{x}$  is specified by the probability density  $\rho(\mathbf{x})$ . In this notation, (3.3) is written as

$$\langle b_1\alpha_1(\mathbf{x}) + b_2\alpha_2(\mathbf{x}) \rangle = b_1\langle\alpha_1(\mathbf{x})\rangle + b_2\langle\alpha_2(\mathbf{x})\rangle.$$

While this is not a particularly difficult change of notation, it is important to keep in mind when translating between statements presented here and the corresponding statements in other sources.

### 3.1.2 Change of Variables

Suppose that a probability density function  $\rho_X(\mathbf{x})$  is given corresponding to the random vector  $\mathbf{X} \in \mathbb{R}^n$ , and it is known that another vector is related to  $\mathbf{X}$  by the invertible function  $\mathbf{Y} = \mathbf{f}(\mathbf{X}) \in \mathbb{R}^n$ . In order to obtain the probability density function  $\rho_Y(\mathbf{y})$ , equal amounts of probability under the two parameterizations need to be equated. If  $D$  is an arbitrary domain and  $\mathbf{f}(D)$  denotes the image of this domain under the bijective differentiable mapping  $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , then we know from the inverse function theorem as stated in (2.26) that

$$\rho_X(\mathbf{x}) = \rho_Y(\mathbf{f}(\mathbf{x}))|J(\mathbf{x})| \quad (3.4)$$

where  $J = \partial\mathbf{y}/\partial\mathbf{x}^T$ ,  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ , and  $F(\mathbf{y}) = \rho_Y(\mathbf{y})$ .

Writing (3.4) in another way,

$$\rho_Y(\mathbf{y}) = \rho_X(\mathbf{f}^{-1}(\mathbf{y}))|J(\mathbf{f}^{-1}(\mathbf{y}))|. \quad (3.5)$$

For example, for the affine transformation,  $\mathbf{y} = \mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{a}$ ,

$$\rho_X(\mathbf{x}) = \rho_Y(A\mathbf{x} + \mathbf{a})|A|,$$

or

$$\rho_Y(\mathbf{y}) = \rho_X(A^{-1}(\mathbf{y} - \mathbf{a}))|A|.$$

### 3.1.3 Marginalization, Conditioning, and Convolution

Another generic operation on pdfs is *marginalization*:

$$\rho(x_1, x_2, \dots, x_m) = \int_{x_{m+1}=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \rho(x_1, x_2, \dots, x_n) dx_{m+1} \cdots dx_n.$$

This is written here with the variables being integrated over as the last  $n - m$  vector entries, resulting in a pdf on a lower dimensional Euclidean space consisting of the remaining  $m$  components. But marginalization can be over *any* of the dimensions, not necessarily the last ones. Finally, the operation of *conditioning* is defined as

$$\rho(x_1, x_2, \dots, x_m | x_{m+1}, x_{m+2}, \dots, x_n) = \rho(x_1, x_2, \dots, x_n) / \rho(x_{m+1}, x_{m+2}, \dots, x_n)$$

where the denominator in this expression is the result of marginalizing  $\rho(x_1, x_2, \dots, x_n)$  over the first  $m$  dimensions (rather than the last  $n - m$ ).

Since addition in  $\mathbb{R}^n$  is well defined, random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  can be added. If their corresponding pdfs are  $\rho_X(\mathbf{x})$  and  $\rho_Y(\mathbf{x})$ , then the pdf of  $\mathbf{X} + \mathbf{Y}$  will be the *convolution*:

$$\rho_{X+Y}(\mathbf{x}) = (\rho_X * \rho_Y)(\mathbf{x}) = \int_{\mathbb{R}^n} \rho_X(\boldsymbol{\xi}) \rho_Y(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi}.$$

### 3.1.4 Mean and Covariance

Whereas the above definitions generalize easily to domains other than  $\mathbb{R}^n$  (including continuous domains such as Lie groups or manifolds, and discrete sets such as permutation groups or graphs), the structure of  $\mathbb{R}^n$  makes it convenient to define additional concepts that are useful in the analysis of data in this space. In particular, in  $\mathbb{R}^n$  the *mean* of a pdf is

$$\boldsymbol{\mu} = \langle \mathbf{x} \rangle = \int_{\mathbb{R}^n} \mathbf{x} \rho(\mathbf{x}) d\mathbf{x}, \quad \text{or} \quad \langle \mathbf{x} - \boldsymbol{\mu} \rangle = \int_{\mathbb{R}^n} (\mathbf{x} - \boldsymbol{\mu}) \rho(\mathbf{x}) d\mathbf{x} = \mathbf{0}. \quad (3.6)$$

Note that  $\boldsymbol{\mu}$  minimizes the cost function

$$c(\mathbf{x}) = \int_{\mathbb{R}^n} \|\mathbf{x} - \mathbf{y}\|^2 f(\mathbf{y}) d\mathbf{y} \quad (3.7)$$

where  $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$  is the 2-norm in  $\mathbb{R}^n$ .

The covariance about the mean is the  $n \times n$  matrix defined as

$$\Sigma = \langle (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \rangle = \int_{\mathbb{R}^n} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \rho(\mathbf{x}) d\mathbf{x}. \quad (3.8)$$

It follows from this definition that

$$\int_{\mathbb{R}^n} \mathbf{x} \mathbf{x}^T \rho(\mathbf{x}) d\mathbf{x} = \Sigma + \boldsymbol{\mu} \boldsymbol{\mu}^T. \quad (3.9)$$

If  $\mathbf{z} \in \mathbb{R}^m$  is defined relative to  $\mathbf{x} \in \mathbb{R}^n$  by the transformation  $\mathbf{z} = A\mathbf{x} + \mathbf{a}$ , where  $A \in \mathbb{R}^{m \times n}$  is a matrix and  $\mathbf{a} \in \mathbb{R}^m$  is a vector, then it is easy to see from the linearity of the operation of integration that<sup>3</sup>

$$\begin{aligned} \boldsymbol{\mu}_Z = \langle \mathbf{z} \rangle &= \int_{\mathbb{R}^n} (A\mathbf{x} + \mathbf{a}) \rho(\mathbf{x}) d\mathbf{x} \\ &= A \left( \int_{\mathbb{R}^n} \mathbf{x} \rho(\mathbf{x}) d\mathbf{x} \right) + \mathbf{a} \left( \int_{\mathbb{R}^n} \rho(\mathbf{x}) d\mathbf{x} \right) \\ &= A\boldsymbol{\mu}_X + \mathbf{a} \end{aligned}$$

<sup>3</sup>Here capital subscripts  $X$  and  $Z$  are used rather than lower case  $x$  and  $z$  so as not to confuse quantities such as  $\boldsymbol{\mu}_Z$  with a parameterized family of vectors.

and

$$\begin{aligned}
 \Sigma_Z &= \langle (\mathbf{z} - \boldsymbol{\mu}_Z)(\mathbf{z} - \boldsymbol{\mu}_Z)^T \rangle \\
 &= \int_{\mathbb{R}^n} (A\mathbf{x} + \mathbf{a} - \boldsymbol{\mu}_Z)(A\mathbf{x} + \mathbf{a} - \boldsymbol{\mu}_Z)^T \rho(\mathbf{x}) d\mathbf{x} \\
 &= \int_{\mathbb{R}^n} (A[\mathbf{x} - \boldsymbol{\mu}_X])(A[\mathbf{x} - \boldsymbol{\mu}_X])^T \rho(\mathbf{x}) d\mathbf{x} \\
 &= \int_{\mathbb{R}^n} A[\mathbf{x} - \boldsymbol{\mu}_X][\mathbf{x} - \boldsymbol{\mu}_X]^T A^T \rho(\mathbf{x}) d\mathbf{x} \\
 &= A \left( \int_{\mathbb{R}^n} [\mathbf{x} - \boldsymbol{\mu}_X][\mathbf{x} - \boldsymbol{\mu}_X]^T \rho(\mathbf{x}) d\mathbf{x} \right) A^T \\
 &= A\Sigma_X A^T.
 \end{aligned}$$

These calculations are true regardless of whether or not the transformation is invertible (i.e., we did not even have to limit the discussion to the case when  $m = n$  and  $|A| > 0$ ).

Pdfs are often used to describe distributions of errors. If these errors are concatenated, they “add” by convolution:

$$(\rho_1 * \rho_2)(\mathbf{x}) = \int_{\mathbb{R}^n} \rho_1(\boldsymbol{\xi}) \rho_2(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi}. \quad (3.10)$$

The mean and covariance of convolved distributions are found as

$$\boldsymbol{\mu}_{1*2} = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 \quad \text{and} \quad \Sigma_{1*2} = \Sigma_1 + \Sigma_2. \quad (3.11)$$

In other words, these quantities can be propagated without explicitly performing the convolution computation, or even knowing the full pdfs. This is independent of the parametric form of the pdf, i.e., it works for non-Gaussians just as well as Gaussians.

If the scalar random variables  $X_1, X_2, \dots, X_n$  are all *independent* of each other, then the corresponding probability density function is separable:

$$\rho(x_1, x_2, \dots, x_n) = \rho_1(x_1) \rho_2(x_2) \cdots \rho_n(x_n). \quad (3.12)$$

When this happens, the covariance matrix will be diagonal.

### 3.1.5 Parametric Distributions

Every “well-behaved” pdf that either decays to zero rapidly as a function of distance from the mean, or takes the value zero outside of a bounded domain has a well-defined mean and covariance. The most important example is the multi-dimensional Gaussian distribution on  $\mathbb{R}^n$  with mean  $\boldsymbol{\mu}$  and covariance  $\Sigma$ :

$$\rho_G(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-n/2} |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]. \quad (3.13)$$

While the Gaussian distribution in (2.7) is by far the most important and commonly used pdf, it is certainly not the only one. For example, the multi-dimensional *Cauchy distribution* in  $\mathbb{R}^n$  is defined as [20, 51]

$$\rho_C(\mathbf{x}; \mathbf{c}, S) = \Gamma\left(\frac{n+1}{2}\right) (\pi)^{-(n+1)/2} |S|^{-\frac{1}{2}} [1 + (\mathbf{x} - \mathbf{c})^T S^{-1} (\mathbf{x} - \mathbf{c})]^{-(n+1)/2}. \quad (3.14)$$

However, this distribution has the drawback that the tails are so “heavy” that the integral in the definition of covariance does not converge, and hence the covariance is meaningless. Certainly in the one-dimensional case the integral

$$I(B) = \int_{-B}^B \frac{x^2}{1 + bx^2} dx$$

diverges as  $B \rightarrow \infty$ .

Since the Gaussian is so special, its properties from Chapter 2 are summarized again here. In addition to being parameterized by the mean vector and covariance matrix (which guarantees that these quantities exist), the Gaussian distribution has the following nice properties [8]:

- Closure under convolution, i.e.,

$$\rho_G(\mathbf{x}; \boldsymbol{\mu}_1, \Sigma_1) * \rho_G(\mathbf{x}; \boldsymbol{\mu}_2, \Sigma_2) = \rho_G(\mathbf{x}; \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \Sigma_1 + \Sigma_2).$$

- Closure under marginalization. In particular, it is not difficult to see that if  $\mathbf{x}_m$  denotes the first  $m$  entries of the vector  $\mathbf{x} \in \mathbb{R}^n$ , then the  $m \times m$  covariance matrix

$$\Sigma_m = \int_{\mathbb{R}^n} \mathbf{x}_m \mathbf{x}_m^T \rho(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^m} \mathbf{x}_m \mathbf{x}_m^T \tilde{\rho}(\mathbf{x}_m) d\mathbf{x}_m$$

where

$$\tilde{\rho}(\mathbf{x}_m) = \int_{x_{m+1}=-\infty}^{\infty} \int_{x_{m+2}=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} \rho(\mathbf{x}) dx_{m+1} dx_{m+2} \cdots dx_n.$$

In other words, the covariance matrix of the marginal of a Gaussian distribution is the part of the covariance matrix of the original Gaussian distribution corresponding to the variables that remain after the marginalization.

- Closure under conditioning (i.e., the product of two Gaussians is a Gaussian (to within a scale factor), and when the quotient can be normalized to be a pdf, this pdf will be a Gaussian).
- The central limit theorem (i.e., the convolution of a large number of well-behaved pdfs tends to the Gaussian distribution (see, e.g., [15])). This will be proved later in this chapter.
- Gaussians are solutions to a heat/diffusion equation with  $\delta$ -function as initial conditions.
- The Fourier transform of a Gaussian is a Gaussian.
- The Gaussian is an even function of its argument.
- Gaussians are the maximum entropy distribution for given mean and covariance.

## 3.2 Conditional Expectation

Consider a bivariate probability density function  $f(x_1, x_2)$  where  $\mathbf{x} = [x_1, x_2]^T \in \mathbb{R}^2$  and denote the marginal densities as

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \quad \text{and} \quad f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1.$$

The expected value of any function  $\phi(x_1, x_2)$  is defined as

$$\langle \phi \rangle \doteq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x_1, x_2) f(x_1, x_2) dx_1 dx_2.$$

If  $\phi(x_1, x_2) = \phi_1(x_1)$  is independent of  $x_2$ , then the integral over  $x_2$  passes through and

$$\langle \phi \rangle = \int_{-\infty}^{\infty} \phi_1(x_1) f_1(x_1) dx_1.$$

Likewise, if  $\phi(x_1, x_2) = \phi_1(x_1)\phi_2(x_2)$ , then

$$\langle \phi \rangle = \langle \phi_1 \rangle \cdot \langle \phi_2 \rangle.$$

There is no harm in keeping the arguments of the function and writing  $\langle \phi(x_1, x_2) \rangle$  as long as it is understood that this is no longer a function of  $x_1$  or  $x_2$  since both of these variables have been integrated out.

The conditional density  $f(x_1|x_2)$ , which is read as “the probability density of  $x_1$  given that the value of  $x_2$  is known,” and  $f(x_2|x_1)$  satisfy Bayes’ rule:

$$f(x_1|x_2)f_2(x_2) = f(x_1, x_2) = f(x_2|x_1)f_1(x_1).$$

And integration over one variable gives

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1|x_2)f_2(x_2) dx_2 \quad \text{and} \quad f_2(x_2) = \int_{-\infty}^{\infty} f(x_2|x_1)f_1(x_1) dx_1.$$

Note that  $f(x_1|x_2)$  is a pdf in  $x_1$  for any choice of  $x_2$ , but it is not a pdf on the  $x_1$ - $x_2$  plane, nor is it a pdf in the variable  $x_2$  for fixed value of  $x_1$ . This can be stated as

$$\int_{-\infty}^{\infty} f(x_1|x_2) dx_1 = \int_{-\infty}^{\infty} f(x_1, x_2)/f_2(x_2) dx_1 = f_2(x_2)/f_2(x_2) = 1$$

but

$$\int_{-\infty}^{\infty} f(x_1|x_2) dx_2 \neq 1 \quad \text{and} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1|x_2) dx_2 dx_1 \neq 1.$$

The *conditional expectation* of any function  $\phi(x_1)$  given  $x_2$  is defined as [27, 32]

$$\boxed{\langle \phi(x_1)|x_2 \rangle \doteq \frac{1}{f_2(x_2)} \int_{-\infty}^{\infty} \phi(x_1) f(x_1, x_2) dx_1.} \quad (3.15)$$

For example, if  $\phi(x_1) = x_1$ , then the *conditional mean* results.

Note that all dependence on  $x_1$  is integrated out in the definition of  $\langle \phi(x_1)|x_2 \rangle$ , and so this is a function of  $x_2$  only. It is also easy to see that from the linearity of the operation of integration,

$$\langle a\phi(x_1) + b\psi(x_1)|x_2 \rangle = a\langle \phi(x_1)|x_2 \rangle + b\langle \psi(x_1)|x_2 \rangle$$

for arbitrary constants  $a$  and  $b$ . Furthermore, it follows directly from the definition (3.15) that

$$\langle \phi_1(x_1)\phi_2(x_2)|x_2 \rangle = \langle \phi_1(x_1)|x_2 \rangle \phi_2(x_2) \quad (3.16)$$

$$\langle \phi_1(x_1) + \phi_2(x_2)|x_2 \rangle = \langle \phi_1(x_1)|x_2 \rangle + \phi_2(x_2) \quad (3.17)$$

Now let  $\psi(x_2) = \langle \phi_1(x_1)|x_2 \rangle$ . Taking the (unconditional) expectation of  $\psi(x_2)$  yields

$$\begin{aligned} \langle \langle \phi_1(x_1)|x_2 \rangle \rangle &= \int_{-\infty}^{\infty} \langle \phi_1(x_1)|x_2 \rangle f_2(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} \left( \frac{1}{f_2(x_2)} \int_{-\infty}^{\infty} \phi_1(x_1) f(x_1, x_2) dx_1 \right) f_2(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} \phi_1(x_1) \left( \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \right) dx_1 \\ &= \int_{-\infty}^{\infty} \phi_1(x_1) f_1(x_1) dx_1 \\ &= \langle \phi_1(x_1) \rangle. \end{aligned} \quad (3.18)$$

Note, however, that in general  $\langle \langle \phi(x_1, x_2)|x_2 \rangle \rangle \neq \langle \phi(x_1, x_2) \rangle$ .

In the case of a tri-variate distribution<sup>4</sup> of the form  $f(x_1, x_2, x_3)$  the following definitions can be made:

$$\langle \phi_1(x_1)|x_2, x_3 \rangle = \frac{1}{f_{23}(x_2, x_3)} \int_{-\infty}^{\infty} \phi_1(x_1) f(x_1, x_2, x_3) dx_1$$

and

$$\langle \phi_{12}(x_1, x_2)|x_3 \rangle = \frac{1}{f_3(x_3)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_1(x_1, x_2) f(x_1, x_2, x_3) dx_1 dx_2.$$

Now let  $\psi(x_2, x_3) = \langle \phi_1(x_1)|x_2, x_3 \rangle$ . Taking the expectation of  $\psi(x_2, x_3)$  conditioned on  $x_2$  and using the above formulas with the appropriate renaming of variables gives  $I \doteq \langle \langle \phi_1(x_1)|x_2, x_3 \rangle |x_2 \rangle$ , which simplifies to

$$\begin{aligned} I &= \frac{1}{f_2(x_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \langle \phi_1(x_1)|x_2, x_3 \rangle f(x_1, x_2, x_3) dx_1 dx_3 \\ &= \frac{1}{f_2(x_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ \frac{1}{f_{23}(x_2, x_3)} \int_{-\infty}^{\infty} \phi_1(x'_1) f(x'_1, x_2, x_3) dx'_1 \right] f_2(x_1, x_2, x_3) dx_1 dx_3 \\ &= \frac{1}{f_2(x_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_1(x'_1) f(x'_1, x_2, x_3) dx'_1 \left( \frac{1}{f_{23}(x_2, x_3)} \int_{-\infty}^{\infty} f(x_1, x_2, x_3) dx_1 \right) dx_3 \\ &= \langle \phi_1(x_1)|x_2 \rangle. \end{aligned} \quad (3.19)$$

Extrapolating this to higher dimensions, the following can be written:

$$\boxed{\langle \langle \phi_1(x_1)|x_2, x_3, \dots, x_n \rangle |x_2, x_3, \dots, x_{n-1} \rangle = \langle \phi_1(x_1)|x_2, x_3, \dots, x_{n-1} \rangle.} \quad (3.20)$$

This is a pattern that can be recursively applied downward to obtain  $\langle \phi_1(x_1)|x_2 \rangle$  after  $n - 2$  conditional expectation operations.

<sup>4</sup>A distribution in one variable is called univariate. A distribution in two variables is called bi-variate. A distribution in three variables is called tri-variate. And a distribution in any number of variables more than one is called multi-variate.



Everything stated above in the context of bivariate and tri-variate distributions can be generalized to higher dimensions where  $\mathbf{x} \in \mathbb{R}^n$  can be partitioned as  $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$  or  $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{x}_3^T]^T$  with the vectors  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  with  $\sum_i n_i = n$ . Then the vectors  $\mathbf{x}_i$  can take the place of  $x_i$  in the above formulas.

In fact, there was nothing special about the structure of  $\mathbb{R}^n$  that was used. And so the above concepts generalize nicely to other spaces. For example,  $x_1$  and  $x_2$  could equally be the angles  $\phi$  and  $\theta$  that parameterize position on the unit sphere, as long as the correct integration measure is used. Such issues will be discussed in future chapters after sufficient geometric concepts are established.

### 3.2.1 Jensen's Inequality and Conditional Expectation

If  $\Phi(x)$  is a *convex function* [3, 39] on  $\mathbb{R}$ , i.e.,

$$\boxed{\Phi(tx + (1-t)y) \leq t\Phi(x) + (1-t)\Phi(y) \quad \forall t \in [0, 1]} \quad (3.21)$$

then Jensen's inequality [26] states

$$\Phi\left(\int_{-\infty}^{\infty} \phi(x)f(x)dx\right) \leq \int_{-\infty}^{\infty} \Phi(\phi(x))f(x)dx$$

for an arbitrary measurable function  $\phi(x)$  and pdf  $f(x)$ . Jensen's inequality can be stated for more general domains than the real line as

$$\boxed{\Phi(\langle\phi\rangle) \leq \langle\Phi \circ \phi\rangle} \quad (3.22)$$

where  $(\Phi \circ \phi)(x) = \Phi(\phi(x))$ . As a direct consequence, if  $\phi(x) = f_2(x)/f_1(x)$ ,  $f(x) = f_1(x)$ , and  $\Phi(y) = -\log y$ , the following property of the *Kullback–Leibler divergence* (which is defined in the first equality below and denoted as  $D_{KL}(f_1\|f_2)$ ) is observed:

$$\begin{aligned} D_{KL}(f_1\|f_2) &\doteq \int_{-\infty}^{\infty} f_1(x) \log \frac{f_1(x)}{f_2(x)} dx \\ &= - \int_{-\infty}^{\infty} f_1(x) \log \frac{f_2(x)}{f_1(x)} dx \\ &\geq - \log \int_{-\infty}^{\infty} f_1(x) \frac{f_2(x)}{f_1(x)} dx \\ &= - \log 1 = 0, \end{aligned}$$

and likewise for domains other than the real line.

Jensen's inequality also holds for condition expectation. Given a multivariate pdf  $f(x, y)$  with variables partitioned as  $x$  and  $y$ , this can be written as

$$\Phi(\langle\phi(x)|y\rangle) \leq \langle\Phi(\phi(x))|y\rangle. \quad (3.23)$$

In particular, if  $\Phi(x) = x^2$  and using the property of conditional expectation in (3.18) gives

$$\langle\langle\phi(x)|y\rangle^2\rangle \leq \langle\langle\phi^2(x)|y\rangle\rangle = \langle\phi^2(x)\rangle. \quad (3.24)$$

### 3.2.2 Convolution and Conditional Expectation

Consider the very special joint probability density function

$$f(x_1, x_2) = \rho_1(x_1)\rho_2(x_2 - x_1)$$

where  $\rho_i(\cdot)$  are themselves univariate pdfs. The marginal densities of  $f(x_1, x_2)$  are then

$$f_1(x_1) = \rho_1(x_1) \quad \text{and} \quad f_2(x_2) = (\rho_1 * \rho_2)(x_2).$$

Note that  $f(x_1, x_2)$  is not separable into a product of marginals in  $x_1$  and  $x_2$ , but if  $u = x_1$  and  $v = x_2 - u$ , then  $\tilde{f}(u, v) = \rho_1(u)\rho_2(v)$ . And the area element in the  $x_1$ - $x_2$  plane is equal to that in the  $u$ - $v$  plane:  $dx_1 dx_2 = dudv$ . These properties of this change of coordinates are used below.

Using the invariance of integration on the real line under shifts and inversions of the argument, together with the commutative nature of addition, the convolution can be written in the following equivalent forms:

$$\begin{aligned} (\rho_1 * \rho_2)(x_2) &= \int_{-\infty}^{\infty} \rho_1(x_1)\rho_2(x_2 - x_1)dx_1 \\ &= \int_{-\infty}^{\infty} \rho_1(x_2 - v)\rho_2(v)dv \\ &= \int_{-\infty}^{\infty} \rho_2(v)\rho_1(x_2 - v)dv \\ &= (\rho_2 * \rho_1)(x_2) \end{aligned}$$

where the substitution  $v = x_2 - x_1$  has been used.

Denoting the derivative of  $f_2(x_2)$  with respect to  $x_2$  as  $f_2'(x_2)$ , it follows that [2, 27, 28]

$$\begin{aligned} \frac{f_2'(x_2)}{f_2(x_2)} &= \int_{-\infty}^{\infty} \frac{\rho_1(x_1)\rho_2'(x_2 - x_1)}{f_2(x_2)} dx_1 \\ &= \int_{-\infty}^{\infty} \frac{\rho_1(x_2 - v)\rho_2'(v)}{f_2(x_2)} dv \\ &= \frac{1}{f_2(x_2)} \int_{-\infty}^{\infty} \frac{\rho_2'(v)}{\rho_2(v)} \rho_2(v)\rho_1(x_2 - v)dv \\ &= \frac{1}{f_2(x_2)} \int_{-\infty}^{\infty} \frac{\rho_2'(x_1)}{\rho_2(x_1)} \rho_2(x_1)\rho_1(x_2 - x_1)dx_1 \\ &= \left\langle \frac{\rho_2'(v)}{\rho_2(v)} \middle| x_2 \right\rangle. \end{aligned}$$

The variable name  $v$  used here is currently irrelevant because it has been integrated out.

Due to the commutativity of convolution on the real line, the roles of  $\rho_1$  and  $\rho_2$  can be interchanged, and

$$\frac{f_2'(x_2)}{f_2(x_2)} = \left\langle \frac{\rho_1'(u)}{\rho_1(u)} \middle| x_2 \right\rangle$$

where  $u$  is another dummy variable of integration, the name of which is irrelevant.

Multiplying the first of these expressions by  $1 - \beta$  and the second by  $\beta$  and adding together:

$$\begin{aligned} \frac{f'_2(x_2)}{f_2(x_2)} &= \beta \left\langle \frac{\rho'_1(u)}{\rho_1(u)} \middle| x_2 \right\rangle + (1 - \beta) \left\langle \frac{\rho'_2(v)}{\rho_2(v)} \middle| x_2 \right\rangle \\ &= \left\langle \beta \frac{\rho'_1(u)}{\rho_1(u)} + (1 - \beta) \frac{\rho'_2(v)}{\rho_2(v)} \middle| x_2 \right\rangle \end{aligned}$$

for arbitrary  $\beta \in [0, 1]$ . This, together with Jensen's inequality in the form of (3.24), can be used to show [2, 7, 27, 28]:

$$\begin{aligned} \left\langle \left( \frac{f'_2(x_2)}{f_2(x_2)} \right)^2 \right\rangle &= \left\langle \left\langle \beta \frac{\rho'_1(u)}{\rho_1(u)} + (1 - \beta) \frac{\rho'_2(v)}{\rho_2(v)} \middle| x_2 \right\rangle^2 \right\rangle \\ &\leq \left\langle \left( \beta \frac{\rho'_1(u)}{\rho_1(u)} + (1 - \beta) \frac{\rho'_2(v)}{\rho_2(v)} \right)^2 \right\rangle \\ &= \beta^2 \left\langle \frac{\rho'_1(u)}{\rho_1(u)} \right\rangle^2 + (1 - \beta)^2 \left\langle \frac{\rho'_2(v)}{\rho_2(v)} \right\rangle^2. \end{aligned} \quad (3.25)$$

The reason why the expectation of the cross term that results from completing the square is zero in the final step leading to (3.25) is left as an exercise.

### 3.3 Information Theory

Given a probability density function (pdf)  $f(\mathbf{x})$  describing the distribution of states of a random vector  $\mathbf{X} \in \mathbb{R}^n$ , the information-theoretic entropy is defined as<sup>5</sup>

$$S(f) \doteq - \int_{\mathbf{x}} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}. \quad (3.26)$$

This is a measure of dispersion of a pdf. However, it is very different from the variance and spread of a pdf as defined in Chapter 2. To demonstrate this difference, consider two  $n$ -dimensional boxes of equal shape and size that are “cut out” of  $\mathbb{R}^n$ . In general the value of  $f(\mathbf{x})$ , as well as the integrand in (3.26) will be different for the two boxes. A new pdf can be defined by swapping the values of the original pdf between the boxes. This new pdf will have the same entropy as the original, but in general it will have different covariance. One way to think of this is that entropy is a measure of variation in the “height” of a pdf viewed as a graph  $y = f(\mathbf{x})$  in  $\mathbb{R}^{n+1}$ , while covariance is a measure of spatial dispersion in the variables  $\mathbf{x} \in \mathbb{R}^n$ .

Note that the standard in the literature is to denote the entropy of the random variable  $\mathbf{X}$  as  $H(\mathbf{X})$ . However, the notation  $S(f)$  (which stands for the entropy of the pdf that fully describes the random variable  $\mathbf{X}$ ) generalizes more easily to the Lie group setting addressed in Volume 2.

<sup>5</sup>In information theory, this would be called *differential entropy*. It is referred to here as *continuous entropy* to denote the difference between this and the discrete case. In this chapter log denotes  $\log_e$ , though many of the properties of entropy hold for any base.

Many operations can be performed on probability density functions including marginalization, conditioning, and convolution. Entropies for each can be evaluated. The subsections that follow provide inequalities that can be used to illustrate the relationships between the entropies in these cases. These relationships can be used to compute bounds on entropy in cases where it is difficult to compute directly.

### 3.3.1 Entropy of Conditional and Marginal Density Functions

Generally speaking, the entropy of a pdf  $f = f(x_1, \dots, x_n)$  is bounded from above by the sum of entropies of corresponding marginal densities. For example,

$$-\int_{x_1} \cdots \int_{x_n} f(x_1, \dots, x_n) \log f(x_1, \dots, x_n) dx_1 \cdots dx_n \leq -\sum_{i=1}^n \int_{x_i} f_i(x_i) \log f_i(x_i) dx_i \quad (3.27)$$

where

$$f_i(x_i) = \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_n} f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Equality in (3.27) holds if and only if  $f(x_1, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdots f_n(x_n)$ .

The result in (3.27) can be written as

$$S(f) \leq \sum_{i=1}^n S(f_i).$$

Likewise, given the pdf  $f(\mathbf{x}, \mathbf{y})$  with marginals  $f_X(\mathbf{x})$  and  $f_Y(\mathbf{y})$ ,

$$S(f) \leq S(f_X) + S(f_Y). \quad (3.28)$$

In fact, (3.27) can be obtained by recursively applying (3.28).

Recall from probability theory that the conditional probability  $f_{X|Y}(\mathbf{x}|\mathbf{y})$  (which is a pdf in  $\mathbf{x}$  for each fixed value of  $\mathbf{y}$ ) is related to the joint and marginal probabilities as

$$f(\mathbf{x}, \mathbf{y}) = f_{X|Y}(\mathbf{x}|\mathbf{y})f_Y(\mathbf{y}).$$

A *conditional entropy* is defined as

$$S(f_{X|Y}; f) \doteq -\int_{\mathbf{x}} \int_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \log f_{X|Y}(\mathbf{x}|\mathbf{y}) d\mathbf{y} d\mathbf{x}, \quad (3.29)$$

and a *marginal entropy* is defined as

$$\begin{aligned} S(f_Y; f) &\doteq -\int_{\mathbf{x}} \int_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \log f_Y(\mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &= -\int_{\mathbf{x}} \int_{\mathbf{y}} f_{X|Y}(\mathbf{x}|\mathbf{y}) f_Y(\mathbf{y}) \log f_Y(\mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &= -\int_{\mathbf{y}} \left( \int_{\mathbf{x}} f_{X|Y}(\mathbf{x}|\mathbf{y}) d\mathbf{x} \right) f_Y(\mathbf{y}) \log f_Y(\mathbf{y}) d\mathbf{y} \\ &= -\int_{\mathbf{y}} f_Y(\mathbf{y}) \log f_Y(\mathbf{y}) d\mathbf{y} \\ &= S(f_Y). \end{aligned} \quad (3.30)$$

Note, however, that in general

$$S(f_{X|Y}; f) \neq S(f_{X|Y}). \quad (3.31)$$

Stated in words, this says that the conditional entropy is *not* the entropy of the conditional pdf.

Calculating the entropy of  $f$  using the definition in (3.26) with  $\mathbf{z} = [\mathbf{x}^T, \mathbf{y}^T]^T$ ,

$$\begin{aligned} S(f) &= - \int_{\mathbf{x}} \int_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \log f(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &= - \int_{\mathbf{x}} \int_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \log [f_{X|Y}(\mathbf{x}|\mathbf{y}) f_Y(\mathbf{y})] d\mathbf{y} d\mathbf{x} \\ &= - \int_{\mathbf{x}} \int_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \{ \log f_{X|Y}(\mathbf{x}|\mathbf{y}) + \log f_Y(\mathbf{y}) \} d\mathbf{y} d\mathbf{x} \\ &= - \int_{\mathbf{y}} f_Y(\mathbf{y}) \log f_Y(\mathbf{y}) d\mathbf{y} - \int_{\mathbf{x}} \int_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \log f_{X|Y}(\mathbf{x}|\mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &= S(f_Y) + S(f_{X|Y}; f). \end{aligned}$$

Of course, the conditional density of  $\mathbf{y}$  given  $\mathbf{x}$  could have been computed just as easily as  $\mathbf{x}$  given  $\mathbf{y}$ , and so it is also true that

$$\boxed{S(f) = S(f_X) + S(f_{Y|X}; f)}. \quad (3.32)$$

If two independent random variables are added (and so their pdfs convolve), the resulting entropy generally will be greater than that of either of the original functions:

$$S(f * \rho) \geq \max\{S(f), S(\rho)\}.$$

Another lower bound on the entropy of convolved distributions is the “entropy power inequality” for pdfs on  $\mathbb{R}^n$  (see Shannon [43][Theorem 15 and Appendix 6], and [5, 13, 45, 46]):

$$\boxed{N(f_1 * f_2) \geq N(f_1) + N(f_2) \quad \text{where} \quad N(f) = \exp \left[ \frac{2}{n} S(f) \right]}. \quad (3.33)$$

The above inequalities will be proved in subsequent sections in this chapter.

The general results from information theory presented in this section follow from the properties of the natural logarithm function. The logarithm function has the property that  $\log(f_1 \cdot f_2) = \log(f_1) + \log(f_2)$ . In addition, it is strictly increasing so that for all  $f, f_1, f_2 > 0$ :  $\log f_1 > \log f_2$  if and only if  $f_1 > f_2$ ;  $\log f_1 < \log(f_1 + f_2)$ ; its negative is convex so that  $a \log f_1 + (1 - a) \log f_2 \leq \log(a f_1 + (1 - a) f_2)$  for all  $0 \leq a \leq 1$ ; and it exhibits sublinear growth:  $\log f \leq f - 1$ .

### 3.3.2 Entropy and Gaussian Distributions

The information-theoretic entropy of one-dimensional and  $n$ -dimensional Gaussian distributions

$$\rho_{(0,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad \text{and} \quad \rho_{(\mathbf{0},\Sigma)}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{x}^T \Sigma^{-1}\mathbf{x}\right)$$

are respectively [43]

$$S(\rho_{(0,\sigma^2)}) = \log(\sqrt{2\pi}e\sigma)$$

and

$$S(\rho_{(\mathbf{0},\Sigma)}) = \log\{(2\pi e)^{n/2}|\Sigma|^{\frac{1}{2}}\} \quad (3.34)$$

where  $\log = \log_e$ .

The entropy of a Gaussian distribution is greater than the entropy of any other distribution over  $\mathbb{R}^n$  with the same mean and variance (thus it is known as the maximum-entropy distribution). This means that if the covariances of an arbitrary distribution are computed, the entropy can be immediately bounded from above using (3.34).

### 3.3.3 Mutual Information

Given a pdf  $f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2)$ , which may or may not be Gaussian, where  $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ ,  $\mathbf{x}_2 \in \mathbb{R}^{n_2}$ , and  $\mathbf{x} \in \mathbb{R}^n$  where  $n = n_1 + n_2$ , the mutual information is defined as the functional<sup>6</sup>

$$I(f_1, f_2; f) \doteq \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_2}} f(\mathbf{x}_1, \mathbf{x}_2) \log\left(\frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)}\right) d\mathbf{x}_1 d\mathbf{x}_2 = I(f_2, f_1; f). \quad (3.35)$$

This can be related to the joint and marginal entropies as

$$I(f_1, f_2; f) = S(f_1) + S(f_2) - S(f). \quad (3.36)$$

When  $f(\mathbf{x}_1, \mathbf{x}_2) = f(\mathbf{x}_1) \cdot f(\mathbf{x}_2)$ , it follows that  $I(f_1, f_2; f_1 \cdot f_2) = 0$ .

### 3.3.4 Information-Theoretic Measures of Divergence

Given two probability density functions  $f_1$  and  $f_2$  on  $\mathbb{R}^n$ , the *Kullback–Leibler divergence* between them is defined as

$$D_{KL}(f_1 \| f_2) \doteq \int_{\mathbb{R}^n} f_1(\mathbf{x}) \log\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right) d\mathbf{x}. \quad (3.37)$$

This has the properties that for non-pathological (i.e., smooth, absolutely integrable, and square integrable) pdfs  $D_{KL}(f_1 \| f_2) \geq 0$  with equality indicating that  $f_1 = f_2$  up to a set of measure zero. And it is bounded from below by the 1-norm in the following way [31]:

$$\frac{1}{4} \left( \int_{\mathbb{R}^n} |f_1(\mathbf{x}) - f_2(\mathbf{x})| d\mathbf{x} \right)^2 \leq D_{KL}(f_1 \| f_2).$$

Note that while this is a useful measure of how much two pdfs diverge, it is not a metric (i.e., a function for evaluating distances between pdfs) because it is not symmetric,  $D_{KL}(f_2 \| f_1) \neq D_{KL}(f_1 \| f_2)$ , and it does not satisfy the triangle inequality. The Fisher information divergence between two pdfs is defined as

<sup>6</sup>In the literature this would be denoted as  $I(X_1; X_2)$ .

$$D_{FI}(f_1 \| f_2) \doteq \int_{\mathbb{R}^n} \left\| \frac{1}{f_1} \nabla f_1 - \frac{1}{f_2} \nabla f_2 \right\|^2 f_1 d\mathbf{x}. \quad (3.38)$$

This is also not a “distance” function in the sense that it is not symmetric in the arguments and does not satisfy the triangle inequality. In the one-dimensional case, this can be written as

$$D_{FI}(f_1 \| f_2) = \int_{-\infty}^{\infty} \left( \frac{1}{f_1} \frac{df_1}{dx} - \frac{1}{f_2} \frac{df_2}{dx} \right)^2 f_1 dx = 4 \int_{-\infty}^{\infty} \left( \frac{d}{dx} \sqrt{\frac{f_1}{f_2}} \right)^2 f_2 dx.$$

Now consider how the divergence measures in (3.37) and (3.38) vary under changes of coordinates. For example, if  $\mathbf{x} = \mathbf{x}(\phi)$ , then the pdf in the coordinates  $\phi$  corresponding to  $f_i(\mathbf{x})$  is  $\tilde{f}_i(\phi) = f_i(\mathbf{x}(\phi)) |J(\phi)|$  where  $J(\phi) = d\mathbf{x}/d\phi^T$  is the  $n \times n$  Jacobian of this coordinate transformation. (For a detailed discussion of coordinate transformations see Chapter 5.) In this way,

$$\int_{\mathbb{R}^n} \tilde{f}_i(\phi) d\phi = \int_{\mathbb{R}^n} f_i(\mathbf{x}) d\mathbf{x}.$$

It is easy to see that

$$D_{KL}(\tilde{f}_1(\phi) \| \tilde{f}_2(\phi)) = D_{KL}(f_1(\mathbf{x}) \| f_2(\mathbf{x})).$$

Likewise, writing the chain rule as

$$\nabla_{\phi}^T \tilde{f}(\phi) = (\nabla_{\mathbf{x}}^T f)|_{\mathbf{x}(\phi)} J(\phi),$$

where  $\nabla_{\mathbf{x}}^T = [\partial/\partial x_1, \partial/\partial x_2, \dots, \partial/\partial x_n]$  is a row vector, it follows that

$$\begin{aligned} D_{FI}(\tilde{f}_1 \| \tilde{f}_2) &= \int_{\phi \in \mathbb{R}^n} \left\| \frac{1}{\tilde{f}_1} \nabla_{\phi} \tilde{f}_1 - \frac{1}{\tilde{f}_2} \nabla_{\phi} \tilde{f}_2 \right\|^2 \tilde{f}_1 d\phi \\ &= \int_{\phi \in \mathbb{R}^n} \left\| \left( \frac{1}{f_1} \nabla_{\mathbf{x}}^T f_1 - \frac{1}{f_2} \nabla_{\mathbf{x}}^T f_2 \right) \frac{J(\phi)}{|J(\phi)|} \right\|^2 f_1 \cdot |J(\phi)| d\phi. \end{aligned}$$

This will be equal to  $D_{FI}(f_1 \| f_2)$  if

$$J(\phi) J^T(\phi) = |J(\phi)|^2 \mathbb{I}.$$

Such is always the case for one dimension, and it holds in multiple dimensions if  $J$  is orthogonal.

### 3.3.5 Fisher Information

By definition, a parametric multivariate probability density function on  $\mathbb{R}^n$  that depends on parameters  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_m]^T \in \mathbb{R}^m$ , satisfies

$$\int_{\mathbb{R}^n} f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = 1 \quad \text{and} \quad f(\mathbf{x}; \boldsymbol{\theta}) \geq 0.$$

For example, the multivariate Gaussian distribution depends on the parameters  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \Sigma)$ , which represent  $m = n + n(n+1)/2$  independent numbers.

The *Fisher information matrix* for the pdf  $f(\mathbf{x}; \boldsymbol{\theta})$  is defined as the matrix with entries

$$F_{ij}(\boldsymbol{\theta}; f) \doteq \int_{\mathbb{R}^n} \frac{1}{f} \frac{\partial f}{\partial \theta_i} \frac{\partial f}{\partial \theta_j} d\mathbf{x}. \quad (3.39)$$

When it is clear which pdf is used to define  $F_{ij}(\boldsymbol{\theta}; f)$ , this can be abbreviated as  $F_{ij}(\boldsymbol{\theta})$ .

In the special case when

$$f(\mathbf{x}; \boldsymbol{\theta}) = \prod_{k=1}^n f_k(x_k; \boldsymbol{\theta}),$$

the Fisher information is additive:

$$F_{ij}(\boldsymbol{\theta}; f) = \sum_{k=1}^n F_{ij}(\boldsymbol{\theta}, f_k).$$

As a special case, when  $m = n$  and  $f(\mathbf{x}; \boldsymbol{\theta}) = f(\mathbf{x} - \boldsymbol{\theta})$ , the Fisher information matrix evaluated at  $\boldsymbol{\theta} = \mathbf{0}$  becomes

$$F(f) \doteq F(\mathbf{0}, f) = \int_{\mathbb{R}^n} \frac{1}{f} (\nabla_{\mathbf{x}} f)(\nabla_{\mathbf{x}} f)^T d\mathbf{x}. \quad (3.40)$$

In the one-dimensional case (3.40) reduces to

$$F(f) = \int_{-\infty}^{\infty} [f'(x)]^2 / f dx \quad (3.41)$$

where  $f'(x) = df/dx$ .

### 3.3.6 Information and Convolution

The ‘‘information’’  $F(f)$  is a measure of the sharpness of the pdf  $f(x)$  in the sense that rapid fluctuations in  $f(x)$  cause  $F(f)$  to increase, whereas a blurred version of  $f(x)$  will have smaller derivatives and a lower value of  $F(f)$ . It follows that  $F(f)$  is reduced under convolution in the same way that entropy is increased.

Recall the following bound on the information of two convolved pdfs from (3.25):

$$F(f_1 * f_2) \leq \beta^2 F(f_1) + (1 - \beta)^2 F(f_2) \quad \text{where } 0 \leq \beta \leq 1. \quad (3.42)$$

If  $\beta = 0$  the right side reduces to  $F(f_2)$  and if  $\beta = 1$  it reduces to  $F(f_1)$ . Therefore,

$$F(f_1 * f_2) \leq \min\{F(f_1), F(f_2)\}. \quad (3.43)$$

However, it can be the case that a value of  $\beta$  in (3.42) other than 0 or 1 yields the tightest bound. Note that the right-hand side in (3.42) is quadratic in  $\beta$ , which is minimized at a value of

$$\beta = \frac{F(f_2)}{F(f_1) + F(f_2)}.$$

This can be verified by either completing the square or setting the derivative with respect to  $\beta$  to zero. Substituting this optimal value of  $\beta$  into (3.42) gives

$$F(f_1 * f_2) \leq \frac{F(f_1)F(f_2)}{F(f_1) + F(f_2)} \quad \text{or} \quad \frac{1}{F(f_1 * f_2)} \geq \frac{1}{F(f_1)} + \frac{1}{F(f_2)}. \quad (3.44)$$



Alternative bounds on the information contained in the convolution of two pdfs on the real line can be obtained by using the Cauchy–Schwarz inequality, as was done by Brown [7]. The version of the Cauchy–Schwarz inequality that is applicable here is

$$\left(\int_{-\infty}^{\infty} a(t)b(t)dt\right)^2 \leq \left(\int_{-\infty}^{\infty} a^2(t)dt\right) \left(\int_{-\infty}^{\infty} b^2(t)dt\right) \tag{3.45}$$

where  $a(t)$  and  $b(t)$  are arbitrary functions whose absolute values and squares are integrable. Equation (3.45) is also called the *Cauchy–Bunyakovsky–Schwarz*, or CBS, inequality. It can be applied directly to the evaluation of  $F(f_1 * f_2)$ . By definition,

$$\begin{aligned} F(f_1 * f_2) &= \int_{-\infty}^{\infty} \frac{\left(\int_{-\infty}^{\infty} f_1'(z-t)f_2(t)dt\right)^2}{(f_1 * f_2)(z)} dz \\ &= \int_{-\infty}^{\infty} \frac{\left(\int_{-\infty}^{\infty} f_1'(z-t)/[f_1(z-t)]^{\frac{1}{2}} \cdot [f_1(z-t)]^{\frac{1}{2}} f_2(t)dt\right)^2}{(f_1 * f_2)(z)} dz. \end{aligned}$$

For each fixed  $z$  in the integral in the numerator, letting  $a(t) = f_1'(z-t)/[f_1(z-t)]^{\frac{1}{2}}$  and  $b(t) = [f_1(z-t)]^{\frac{1}{2}} f_2(t)$ , and using the CBS inequality results in

$$\begin{aligned} F(f_1 * f_2) &\leq \int_{-\infty}^{\infty} \frac{\left(\int_{-\infty}^{\infty} [f_1'(z-t)]^2/f_1(z-t)dt\right) \cdot \left(\int_{-\infty}^{\infty} f_1(z-t)[f_2(t)]^2dt\right)}{(f_1 * f_2)(z)} dz \\ &= \int_{-\infty}^{\infty} \frac{\left(\int_{-\infty}^{\infty} [f_1'(t')]^2/f_1(t')dt'\right) \cdot \left(\int_{-\infty}^{\infty} f_1(z-t)[f_2(t)]^2dt\right)}{(f_1 * f_2)(z)} dz \\ &= F(f_1) \cdot \int_{-\infty}^{\infty} \frac{(f_1 * f_2^2)(z)}{(f_1 * f_2)(z)} dz. \end{aligned}$$

The key point in the above proof is that integration over the whole real line is invariant under shifts and inversions of the argument of the function, which allows the change of variables  $t' = z - t$  and  $F(f_1)$  to be taken outside of the integral over  $z$ . Unfortunately, the above is not a tight bound.

Revisiting the CBS inequality (3.45), if  $a(t) \geq 0$  for all values of  $t$ , then it is possible to define  $j(t) = [a(t)]^{\frac{1}{2}}$  and  $k(t) = [a(t)]^{\frac{1}{2}} b(t)$ , and since  $j(t)k(t) = a(t)b(t)$  [22],

$$\begin{aligned} \left(\int_{-\infty}^{\infty} a(t)b(t)dt\right)^2 &\leq \left(\int_{-\infty}^{\infty} j^2(t)dt\right) \left(\int_{-\infty}^{\infty} k^2(t)dt\right) \\ &= \left(\int_{-\infty}^{\infty} a(t)dt\right) \left(\int_{-\infty}^{\infty} a(t)[b(t)]^2dt\right). \end{aligned} \tag{3.46}$$

Using this version of the CBS inequality, and letting  $b(t) = f_1'(z-t)/[f_1(z-t)]$  and  $a(t) = f_1(z-t)f_2(t)$ , Brown [7] showed

$$\begin{aligned}
F(f_1 * f_2) &= \int_{-\infty}^{\infty} \frac{\left( \int_{-\infty}^{\infty} [f_1'(z-t)/f_1(z-t)] \cdot [f_1(z-t)f_2(t)] dt \right)^2}{(f_1 * f_2)(z)} dz \\
&\leq \int_{-\infty}^{\infty} \frac{\left( \int_{-\infty}^{\infty} \left[ \frac{f_1'(z-t)}{f_1(z-t)} \right]^2 [f_1(z-t)f_2(t)] dt \right) \left( \int_{-\infty}^{\infty} f_1(z-\tau)f_2(\tau) d\tau \right)}{(f_1 * f_2)(z)} dz \\
&= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \left\{ \frac{[f_1'(z-t)]^2}{f_1(z-t)} \right\} f_2(t) dt \right) dz \\
&= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \{ [f_1'(z-t)]^2 / f_1(z-t) \} dz \right) f_2(t) dt \\
&= F(f_1) \int_{-\infty}^{\infty} f_2(t) dt \\
&= F(f_1).
\end{aligned}$$

Since convolution is commutative, this is equivalent to (3.43), which is not as tight as (3.44).

### 3.3.7 Shift and Scaling Properties

In this subsection the behavior of Shannon entropy under changes of coordinates is examined. Scaling properties of Fisher information can be computed in an analogous way.

#### The One-Dimensional Case

Consider the entropy of a pdf  $f(x)$ , and the entropy of the shifted version of this pdf:  $f_a(x) \doteq f(x-a)$ . Due to the invariance of integration of any integrable function on the line,

$$\begin{aligned}
S(f_a) &= - \int_{-\infty}^{\infty} f(x-a) \log f(x-a) dx \\
&= - \int_{-\infty}^{\infty} f(x) \log f(x) dx \\
&= S(f).
\end{aligned}$$

Now consider the scaled version of the pdf  $f(x)$  defined as

$$f_s(x) \doteq \frac{1}{s} f(x/s) \quad \text{where } s > 0.$$

If  $s > 1$ , this is a more “spread out” version of  $f$ , and if  $s < 1$ , then this is a more “concentrated” version of  $f$ . It can be verified easily that  $f_s(x)$  is indeed a pdf by making the change of coordinates  $y = x/s$  and replacing the integral over  $x$  with that over  $y$ .

Likewise, the entropy of  $f_s(x)$  is calculated as

$$\begin{aligned}
S(f_s) &= - \int_{-\infty}^{\infty} \frac{1}{s} f(x/s) \log \left[ \frac{1}{s} f(x/s) \right] dx \\
&= - \int_{-\infty}^{\infty} f(y) \log \left[ \frac{1}{s} f(y) \right] dy \\
&= S(f) + \log s.
\end{aligned}$$

### The Multi-Dimensional Case

The multi-dimensional case proceeds in a similar way as in the one-dimensional case. Given a pdf  $f(\mathbf{x})$ , a shifted version is  $f_{\mathbf{a}}(\mathbf{x}) = f(\mathbf{x} - \mathbf{a})$ . And

$$S(f_{\mathbf{a}}) = - \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{a}) \log f(\mathbf{x} - \mathbf{a}) d\mathbf{x} = - \int_{\mathbb{R}^n} f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x} = S(f).$$

Now consider the scaled version of the pdf  $f(\mathbf{x})$  defined as

$$f_A(\mathbf{x}) = \frac{1}{\det A} f(A^{-1}\mathbf{x}) \quad \text{where} \quad \det A > 0.$$

If  $\det A > 1$ , this is a more “spread out” version of  $f$ , and if  $\det A < 1$ , then this is a more “concentrated” version of  $f$ . It can be verified easily that  $f_A(\mathbf{x})$  is indeed a pdf by making the change of coordinates  $\mathbf{y} = A^{-1}\mathbf{x}$  and replacing the integral over  $\mathbf{x}$  with that over  $\mathbf{y}$ .

The entropy of  $f_A(\mathbf{x})$  is calculated as

$$\begin{aligned}
S(f_A) &= - \int_{\mathbb{R}^n} \frac{1}{\det A} f(A^{-1}\mathbf{x}) \log \left[ \frac{1}{\det A} f(A^{-1}\mathbf{x}) \right] d\mathbf{x} \\
&= - \int_{\mathbb{R}^n} f(\mathbf{y}) \log \left[ \frac{1}{\det A} f(\mathbf{y}) \right] d\mathbf{y} \\
&= S(f) + \log \det A.
\end{aligned}$$

## 3.4 Parameter Estimation

Let  $f(\mathbf{x}; \boldsymbol{\theta})$  be any member of a family of pdfs in the variable  $\mathbf{x} \in \mathbb{R}^n$  characterized by a vector value  $\boldsymbol{\theta} \in \mathbb{R}^m$ . That is,

$$\int_{\mathbb{R}^n} f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = 1. \quad (3.47)$$

The whole family of pdfs is parameterized by letting  $\boldsymbol{\theta}$  take a range of values in  $\mathbb{R}^m$ .

For example, the family of multivariate Gaussian distributions  $\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$  is parameterized by  $\boldsymbol{\mu}, \Sigma$ . If both  $\boldsymbol{\mu}$  and  $\Sigma$  are unknown, then  $\boldsymbol{\theta}$  would be  $n + n(n + 1)/2$ -dimensional (since  $\Sigma = \Sigma^T$ ); if the mean is known, then  $\boldsymbol{\theta}$  would be  $n(n + 1)/2$ -dimensional; if the mean is the only unknown, then  $\boldsymbol{\theta}$  can take any value in  $\mathbb{R}^n$ .

### 3.4.1 Unbiased Estimators

Let  $\mathbf{v} : \mathbb{R}^n \rightarrow \mathbb{R}^p$  be any well-behaved vector-valued function<sup>7</sup> of  $\mathbf{x}$ . Then

$$\langle \mathbf{v}(\mathbf{x}) \rangle = \int_{\mathbb{R}^n} \mathbf{v}(\mathbf{x}) f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} \doteq \boldsymbol{\psi}(\boldsymbol{\theta}), \quad (3.48)$$

where the equality on the right simply means that the dependence on  $\mathbf{x}$  has been integrated out, and the result is defined as  $\boldsymbol{\psi}(\boldsymbol{\theta}) \in \mathbb{R}^p$ . It is sometimes convenient to rewrite (3.48) as

$$\int_{\mathbb{R}^n} [\mathbf{v}(\mathbf{x}) - \boldsymbol{\psi}(\boldsymbol{\theta})] f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = \mathbf{0}. \quad (3.49)$$

Given a set of sampled data,  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ , a goal often encountered in practice is to find the particular member of the family of parametric distributions that best fits the data. For example, values of  $\hat{\boldsymbol{\theta}}$  could be obtained by solving the equation

$$\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{i=1}^N \mathbf{v}(\mathbf{x}_i) \quad (3.50)$$

for a large value of  $N$ . If  $m = p$  and  $\boldsymbol{\psi}(\boldsymbol{\theta}) \approx \boldsymbol{\theta}$ , then in this context  $\mathbf{v}(\mathbf{x})$  is called an *estimator* of  $\boldsymbol{\theta}$ , and  $\hat{\boldsymbol{\theta}}$  is called the *estimate* of  $\boldsymbol{\theta}$ . If  $\boldsymbol{\psi}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ , then  $\mathbf{v}(\mathbf{x})$  is called an *unbiased estimator* and  $\hat{\boldsymbol{\theta}}$  is called an *unbiased estimate*.

The samples  $\{\mathbf{x}_i\}$  in (3.50) are assumed to be drawn at random from the distribution  $f(\mathbf{x}; \boldsymbol{\theta})$  for some unknown, but fixed, value of  $\boldsymbol{\theta}$ . The law of large numbers states that the underlying pdf is observed as the number of samples goes to infinity, and so the estimate  $\hat{\boldsymbol{\theta}}$  obtained in this way should become better as  $N$  becomes larger. If the estimator  $\mathbf{v}(\mathbf{x})$  is unbiased, then  $\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\theta}}$  and obtaining  $\hat{\boldsymbol{\theta}}$  from (3.50) becomes trivial. When using other estimators the estimation problem becomes one of inverting the function  $\boldsymbol{\psi}$ .

### 3.4.2 The Cramér–Rao Bound

The Fisher information matrix in (3.39) can be written in the following alternative forms:

$$F = \int_{\mathbb{R}^m} \left[ \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{x}; \boldsymbol{\theta}) \right] \left[ \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{x}; \boldsymbol{\theta}) \right]^T f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} \quad (3.51)$$

$$= - \int_{\mathbb{R}^m} f(\mathbf{x}; \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}, \quad (3.52)$$

where  $\partial f / \partial \boldsymbol{\theta}$  is interpreted as a column vector, and  $\partial f / \partial \boldsymbol{\theta}^T = [\partial f / \partial \boldsymbol{\theta}]^T$ . Here  $\log(\cdot)$  is the scalar natural logarithm function.

Differentiation of both sides of (3.49) with respect to  $\boldsymbol{\theta}^T$  gives

$$\frac{\partial}{\partial \boldsymbol{\theta}^T} \int_{\mathbb{R}^n} [\mathbf{v}(\mathbf{x}) - \boldsymbol{\psi}(\boldsymbol{\theta})] f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} = \int_{\mathbb{R}^n} [\mathbf{v}(\mathbf{x}) - \boldsymbol{\psi}(\boldsymbol{\theta})] \frac{\partial f(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d\mathbf{x} - \frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}^T} = \mathbb{O},$$

where the derivative is taken under the integral and the product rule for differentiation and the fact that  $f$  is a pdf in  $\mathbf{x}$  is used. Here  $\mathbb{O}$  is the  $m \times m$  zero matrix resulting from the computation of  $\partial \mathbf{0} / \partial \boldsymbol{\theta}^T$ .

<sup>7</sup>The class of “nice” functions extends to those that are vector valued by simply restricting each component of the vector to be nice, i.e.,  $v_i \in \mathcal{N}(\mathbb{R}^n)$ .

The above equation can be written as

$$\frac{\partial \psi}{\partial \boldsymbol{\theta}^T} = \int_{\mathbb{R}^n} \mathbf{a}(\mathbf{x}, \boldsymbol{\theta}) \mathbf{b}^T(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} \in \mathbb{R}^{p \times m} \quad (3.53)$$

where

$$\mathbf{a}(\mathbf{x}, \boldsymbol{\theta}) = [f(\mathbf{x}; \boldsymbol{\theta})]^{1/2} [\mathbf{v}(\mathbf{x}) - \boldsymbol{\psi}(\boldsymbol{\theta})] \quad \text{and} \quad \mathbf{b}(\mathbf{x}, \boldsymbol{\theta}) = [f(\mathbf{x}; \boldsymbol{\theta})]^{1/2} \frac{\partial}{\partial \boldsymbol{\theta}} \log f(\mathbf{x}; \boldsymbol{\theta}).$$

Referring back to the first equality in (3.52), it is clear that

$$F = \int_{\mathbb{R}^n} \mathbf{b}(\mathbf{x}, \boldsymbol{\theta}) [\mathbf{b}(\mathbf{x}, \boldsymbol{\theta})]^T d\mathbf{x} \quad (3.54)$$

and the error covariance for  $\mathbf{v}(\mathbf{x})$ , denoted as

$$C = \langle [\mathbf{v}(\mathbf{x}) - \boldsymbol{\psi}(\boldsymbol{\theta})][\mathbf{v}(\mathbf{x}) - \boldsymbol{\psi}(\boldsymbol{\theta})]^T \rangle,$$

is computed explicitly as

$$C = \int_{\mathbb{R}^n} \mathbf{a}(\mathbf{x}, \boldsymbol{\theta}) [\mathbf{a}(\mathbf{x}, \boldsymbol{\theta})]^T d\mathbf{x}. \quad (3.55)$$

Following [12], the multiplication of (3.53) on the left by the transpose of an arbitrary constant vector  $\boldsymbol{\alpha} \in \mathbb{R}^p$  and on the right by an arbitrary constant column vector  $\boldsymbol{\beta} \in \mathbb{R}^m$  gives

$$\boldsymbol{\alpha}^T \frac{\partial \psi}{\partial \boldsymbol{\theta}^T} \boldsymbol{\beta} = \int_{\mathbb{R}^n} \boldsymbol{\alpha}^T \mathbf{a}(\mathbf{x}, \boldsymbol{\theta}) \mathbf{b}^T(\mathbf{x}, \boldsymbol{\theta}) \boldsymbol{\beta} d\mathbf{x}. \quad (3.56)$$

Then regrouping terms in the expression on the right and squaring, and using the Cauchy–Schwarz inequality gives

$$\begin{aligned} \left( \int_{\mathbb{R}^n} \boldsymbol{\alpha}^T (\mathbf{a} \mathbf{b}^T) \boldsymbol{\beta} d\mathbf{x} \right)^2 &= \left( \int_{\mathbb{R}^n} (\boldsymbol{\alpha}^T \mathbf{a}) (\mathbf{b}^T \boldsymbol{\beta}) d\mathbf{x} \right)^2 \\ &\leq \left( \int_{\mathbb{R}^n} (\boldsymbol{\alpha}^T \mathbf{a})^2 d\mathbf{x} \right) \left( \int_{\mathbb{R}^n} (\boldsymbol{\beta}^T \mathbf{b})^2 d\mathbf{x} \right) \\ &= \left( \int_{\mathbb{R}^n} \boldsymbol{\alpha}^T \mathbf{a} \mathbf{a}^T \boldsymbol{\alpha} d\mathbf{x} \right) \left( \int_{\mathbb{R}^n} \boldsymbol{\beta}^T \mathbf{b} \mathbf{b}^T \boldsymbol{\beta} d\mathbf{x} \right). \end{aligned}$$

But from (3.54), (3.55), and (3.56), this can be written as

$$\left( \boldsymbol{\alpha}^T \frac{\partial \psi}{\partial \boldsymbol{\theta}^T} \boldsymbol{\beta} \right)^2 \leq (\boldsymbol{\alpha}^T C \boldsymbol{\alpha}) (\boldsymbol{\beta}^T F \boldsymbol{\beta}).$$

Making the choice of  $\boldsymbol{\beta} = F^{-1} [\partial \boldsymbol{\psi}^T / \partial \boldsymbol{\theta}] \boldsymbol{\alpha}$  yields

$$\left( \boldsymbol{\alpha}^T \frac{\partial \psi}{\partial \boldsymbol{\theta}^T} F^{-1} \frac{\partial \boldsymbol{\psi}^T}{\partial \boldsymbol{\theta}} \boldsymbol{\alpha} \right)^2 \leq (\boldsymbol{\alpha}^T C \boldsymbol{\alpha}) \left( \boldsymbol{\alpha}^T \frac{\partial \psi}{\partial \boldsymbol{\theta}^T} F^{-1} \frac{\partial \boldsymbol{\psi}^T}{\partial \boldsymbol{\theta}} \boldsymbol{\alpha} \right).$$

This simplifies to

$$\boldsymbol{\alpha}^T \left( C - \frac{\partial \psi}{\partial \boldsymbol{\theta}^T} F^{-1} \frac{\partial \boldsymbol{\psi}^T}{\partial \boldsymbol{\theta}} \right) \boldsymbol{\alpha} \geq 0 \quad \text{for arbitrary } \boldsymbol{\alpha} \in \mathbb{R}^n. \quad (3.57)$$

This means that the term in parentheses is a positive definite matrix. This statement is often denoted as

$$C \geq \frac{\partial \psi}{\partial \boldsymbol{\theta}^T} F^{-1} \frac{\partial \psi^T}{\partial \boldsymbol{\theta}}, \quad (3.58)$$

which is *not* an inequality in the entries of the matrices, but rather simply shorthand for (3.57), or equivalently, the statement that all of the eigenvalues of  $C - (\partial \psi / \partial \boldsymbol{\theta}^T) F^{-1} (\partial \psi^T / \partial \boldsymbol{\theta})$  are greater than or equal to zero.

While the above holds true for any estimator, in the case of an unbiased estimator it simplifies because then  $\partial \psi / \partial \boldsymbol{\theta}^T = \partial \psi^T / \partial \boldsymbol{\theta} = \mathbb{I}$ .

### 3.4.3 Demonstration with Gaussian Distributions

Note that (3.58) is true for *any* estimator. In the special case when  $m = n$  and  $f(\mathbf{x}; \boldsymbol{\theta}) = f(\mathbf{x} - \boldsymbol{\theta})$ , and  $\hat{\boldsymbol{\theta}} = \boldsymbol{\mu}$ , the Cramér–Rao bound becomes

$$\Sigma \geq F^{-1}. \quad (3.59)$$

When  $f(\mathbf{x} - \boldsymbol{\mu}) = \rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$  is a Gaussian distribution with known covariance  $\Sigma$ ,

$$\frac{\partial f}{\partial \mu_i} = \mathbf{e}_i^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$$

and the identity in (2.33) can be used to show that the Fisher information matrix becomes  $F = \Sigma^{-1}$ , and therefore the inequality in (3.59) becomes an equality.

### 3.4.4 The de Bruijn Identity

When written in terms of probability densities, the *de Bruijn identity* states [10]

$$\frac{d}{dt} S(\alpha * f_{0,t}) = \frac{1}{2} F(\alpha * f_{0,t}). \quad (3.60)$$

Here  $f_{0,t}(x) = \rho(x; 0, t)$  is the Gaussian distribution with zero mean and variance  $t > 0$  that solves the heat equation in Section 2.7.1 (in the 1D case with unit diffusion constant),  $\alpha(x)$  is an arbitrary differentiable pdf, and  $F(\cdot)$  denotes the Fisher information as defined in (3.41). It follows from (3.60) that

$$\lim_{t \rightarrow 0} f_{0,t}(x) = \delta(x) \quad \implies \quad \left. \frac{d}{dt} S(\alpha * f_{0,t}) \right|_{t=0} = \frac{1}{2} F(\alpha). \quad (3.61)$$

The derivation of (3.60) itself is relatively straightforward. Following the presentation in Cover and Thomas [10, pp. 672–673], but using different notation and a different order of operations,

$$\begin{aligned} \frac{d}{dt} S(\alpha * f_{0,t}) &= - \frac{d}{dt} \int_{-\infty}^{\infty} (\alpha * f_{0,t})(x) \log[(\alpha * f_{0,t})(x)] dx \\ &= - \int_{-\infty}^{\infty} \left\{ \left[ \frac{\partial}{\partial t} (\alpha * f_{0,t}) \right] \cdot \log(\alpha * f_{0,t}) + (\alpha * f_{0,t}) \cdot \left[ \frac{\partial}{\partial t} \log(\alpha * f_{0,t}) \right] \right\} dx \\ &= - \int_{-\infty}^{\infty} \left\{ \left( \alpha * \frac{\partial f_{0,t}}{\partial t} \right) \cdot \log(\alpha * f_{0,t}) + \alpha * \frac{\partial f_{0,t}}{\partial t} \right\} dx. \end{aligned}$$

The Gaussian distribution  $f_{0,t}$  is precisely the one corresponding to one-dimensional Brownian motion:

$$f_{0,t}(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} \implies \frac{\partial f_{0,t}}{\partial t} = \frac{1}{2} \frac{\partial^2 f_{0,t}}{\partial x^2}. \quad (3.62)$$

Now in general for a convolution product

$$\frac{\partial}{\partial x}[(\phi_1 * \phi_2)(x)] = (\phi_1 * \phi_2')(x)$$

as long as  $\phi_2'(x) = \partial\phi_2/\partial x$  is well behaved. This is certainly true in the present case, and means that

$$\begin{aligned} \frac{d}{dt}S(\alpha * f_{0,t}) &= -\frac{1}{2} \int_{-\infty}^{\infty} \left\{ \left( \alpha * \frac{\partial^2 f_{0,t}}{\partial x^2} \right) \cdot \log(\alpha * f_{0,t}) + \alpha * \frac{\partial^2 f_{0,t}}{\partial x^2} \right\} dx \\ &= -\frac{1}{2} \int_{-\infty}^{\infty} \left\{ \frac{\partial^2}{\partial x^2} (\alpha * f_{0,t}) \cdot \log(\alpha * f_{0,t}) + \frac{\partial^2}{\partial x^2} (\alpha * f_{0,t}) \right\} dx. \end{aligned}$$

The second term disappears because, from the fundamental theorem of calculus,

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2} (\alpha * f_{0,t}) dx = \frac{\partial}{\partial x} (\alpha * f_{0,t})(x) \Big|_{x=-\infty}^{\infty} = 0$$

since  $\alpha(x)$  and its derivatives decay to zero as  $x \rightarrow \pm\infty$ .

Using integration by parts on the term that remains,

$$\begin{aligned} \frac{d}{dt}S(\alpha * f_{0,t}) &= -\frac{1}{2} \frac{\partial}{\partial x} (\alpha * f_{0,t}) \cdot \log(\alpha * f_{0,t}) \Big|_{-\infty}^{\infty} \\ &\quad + \frac{1}{2} \int_{-\infty}^{\infty} \frac{1}{(\alpha * f_{0,t})} \left[ \frac{\partial}{\partial x} (\alpha * f_{0,t}) \right]^2 dx. \end{aligned}$$

Again, as long as  $\alpha(x)$  decays rapidly enough as  $x \rightarrow \pm\infty$ , the term on the left will evaluate to zero. And the integral on the right is  $F(\alpha * f_{0,t})$ , and so (3.60) results.

### 3.4.5 The Entropy Power Inequality

The statement of the entropy power inequality dates back to Shannon's original paper, though complete and rigorous proofs came later [45, 5]. Shannon defined the entropy power of a pdf  $p(x)$  on  $\mathbb{R}^n$  as  $N(p) = \exp(2S(p)/n)/2\pi e$  where  $S(p)$  is the entropy of  $p$ . The entropy power inequality then states

$$\boxed{N(p * q) \geq N(p) + N(q)} \quad (3.63)$$

with equality if and only if  $p$  and  $q$  are both Gaussian distributions with covariance matrices that are a scalar multiple of each other.

Variations on the theme as well as different methods of proof have appeared in the literature since that time [9, 13, 34, 46]. Here the proofs of [45, 5] are reviewed for the 1D case. In the literature usually the one-dimensional case is proven, and then mathematical induction is used to extend to higher dimensions.

Let  $f_{\sigma^2(t)}(x)$  denote a Gaussian distribution with zero mean with variance  $\sigma^2(t)$ , and let  $\sigma^2(0) = 0$ . Given differentiable pdfs  $p(x)$  and  $q(x)$ , define  $p_t \doteq p * f_{\sigma_1^2(t)}$  and  $q_t \doteq q * f_{\sigma_2^2(t)}$ . Following Stam and Blachman [5, 45], let

$$V(t) = \frac{\exp[2 \cdot S(p_t)] + \exp[2 \cdot S(q_t)]}{\exp[2 \cdot S(p_t * q_t)]} = (\exp[2 \cdot S(p_t)] + \exp[2 \cdot S(q_t)]) \exp[-2 \cdot S(p_t * q_t)]. \quad (3.64)$$

As  $t \rightarrow 0$ ,  $V(0) \rightarrow [N(p) + N(q)]/N(p * q)$ . Therefore, if it can be proven that  $V(0) \leq 1$ , then (3.63) will hold in the one-dimensional case.

Taking the time derivative of (3.64), and using the chain rule and product rule,

$$\begin{aligned} \frac{dV}{dt} &= \left( 2 \exp[2 \cdot S(p_t)] \frac{d}{dt} S(p_t) + 2 \exp[2 \cdot S(q_t)] \frac{d}{dt} S(q_t) \right) \exp[-2 \cdot S(p_t * q_t)] \\ &\quad - 2 (\exp[2 \cdot S(p_t)] + \exp[2 \cdot S(q_t)]) \exp[-2 \cdot S(p_t * q_t)] \frac{d}{dt} S(p_t * q_t). \end{aligned}$$

Using the de Bruijn identity (3.60) and the chain rule,

$$\frac{d}{dt} S(q_t) = \frac{dS(q_t)}{d(\sigma_2^2)} \frac{d(\sigma_2^2)}{dt} = \frac{1}{2} F(q_t) \frac{d(\sigma_2^2)}{dt}$$

and likewise for  $p_t$ . Furthermore, since convolution on the real line is commutative,

$$p_t * q_t = p * f_{\sigma_1^2} * q * f_{\sigma_2^2} = p * q * f_{\sigma_1^2} * f_{\sigma_2^2} = p * q * f_{\sigma_1^2 + \sigma_2^2}.$$

Therefore,

$$\begin{aligned} \frac{dV}{dt} &= \left( \exp[2 \cdot S(p_t)] F(p_t) \frac{d(\sigma_1^2)}{dt} + \exp[2 \cdot S(q_t)] F(q_t) \frac{d(\sigma_2^2)}{dt} \right) \exp[-2 \cdot S(p_t * q_t)] \\ &\quad - (\exp[2 \cdot S(p_t)] + \exp[2 \cdot S(q_t)]) \exp[-2 \cdot S(p_t * q_t)] F(p_t * q_t) \frac{d(\sigma_1^2 + \sigma_2^2)}{dt}. \end{aligned}$$

Multiplying both sides by  $\exp[2 \cdot S(p_t * q_t)]$  and choosing  $\sigma_1^2(t)$  and  $\sigma_2^2(t)$  such that

$$\frac{d(\sigma_1^2)}{dt} = \exp[2 \cdot S(p_t)] \quad \text{and} \quad \frac{d(\sigma_2^2)}{dt} = \exp[2 \cdot S(q_t)], \quad (3.65)$$

$$\begin{aligned} \exp[2 \cdot S(p_t * q_t)] \frac{dV}{dt} &= (\exp[2 \cdot S(p_t)])^2 F(p_t) + (\exp[2 \cdot S(q_t)])^2 F(q_t) \\ &\quad - (\exp[2 \cdot S(p_t)] + \exp[2 \cdot S(q_t)])^2 F(p_t * q_t). \end{aligned} \quad (3.66)$$

But from the general inequality

$$(\alpha_1 + \alpha_2)^2 F(f_1 * f_2) \leq \alpha_1^2 F(f_1) + \alpha_2^2 F(f_2) \quad (3.67)$$

(which is equivalent to (3.42)) it follows from (3.66) with  $\alpha_1 = \exp[2 \cdot S(p_t)]$ ,  $\alpha_2 = \exp[2 \cdot S(q_t)]$ ,  $f_1 = p_t$ , and  $f_2 = q_t$  that

$$\frac{dV}{dt} \geq 0.$$



Equality holds in this expression if and only if  $p$  and  $q$  are Gaussians. In that case  $V$  is a constant. Otherwise,  $V$  is a strictly increasing function. Therefore,

$$V(\infty) \geq V(0)$$

with equality holding only for Gaussians.

Since the entropy of the convolution of two functions is no less than the entropy of either of the original functions, and since the exponential function is always positive, the choice in (3.65) implies that  $\sigma_i^2(\infty) = \infty$ . Furthermore, the scaled pdf  $\sigma_1 p_t(\sigma_1 x)$  will have entropy  $S(p_t) - \log \sigma_1$ , as discussed in Section 3.3.7. But since by definition

$$p_t(z) = (p * f_{\sigma_1^2})(z) = \frac{1}{\sqrt{2\pi} \sigma_1} \int_{-\infty}^{\infty} p(y) \exp \left[ -\frac{1}{2}(z - y)^2 / \sigma_1^2 \right] dy,$$

making the substitutions  $z = \sigma_1 x$  and  $y = \sigma_1 \xi$  yields

$$\sigma_1 p_t(\sigma_1 x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma_1 p(\sigma_1 \xi) \exp \left[ -\frac{1}{2}(x - \xi)^2 \right] d\xi.$$

And since  $\sigma_1 p(\sigma_1 x)$  becomes more and more like a delta function as  $\sigma_1 \rightarrow \infty$ , it follows that

$$\lim_{\sigma_1 \rightarrow \infty} \sigma_1 p_t(\sigma_1 x) = f_{0,1}(x)$$

where  $f_{0,t}(x)$  is defined in (3.62). Therefore,

$$\lim_{t \rightarrow \infty} S(p_t) = \frac{1}{2} \log 2\pi e \sigma_1^2,$$

and similarly,

$$\lim_{t \rightarrow \infty} S(q_t) = \frac{1}{2} \log 2\pi e \sigma_2^2$$

and

$$\lim_{t \rightarrow \infty} S(p_t * q_t) = \frac{1}{2} \log 2\pi e (\sigma_1^2 + \sigma_2^2).$$

Substituting these into (3.64) gives

$$\lim_{t \rightarrow \infty} V(t) = 1,$$

and since  $V(0) \leq V(\infty)$ , this proves (3.63) for the case of  $n = 1$ .

### 3.4.6 Entropy of a Weighted Sum of Disjoint PDFs

Let  $\rho(x)$  be a probability density of the form

$$\rho(x) = \sum_{i=1}^n w_i \rho_i(x)$$

where<sup>8</sup>

$$w_i \geq 0 \quad \text{and} \quad \sum_{i=1}^n w_i = 1,$$

---

<sup>8</sup>The set of values  $\{w_i\}$  satisfying these conditions is called a *partition of unity*.

and each  $\rho_i(x)$  is a pdf that is disjoint from the others in the sense that

$$\int_x \rho_i^{\frac{1}{2}}(x)\rho_j^{\frac{1}{2}}(x)dx = \delta_{ij}.$$

In other words, each  $\rho_j(x)$  has an associated region where it is positive and a region where it is zero, and no two of these functions are positive on the same region.

The entropy of  $\rho(x)$  can be computed using the fact that if  $x$  is in the region where  $\rho_j(x)$  is not zero, then

$$w_j\rho_j(x) = \sum_{i=1}^n w_i\rho_i(x) \implies \log(w_j\rho_j(x)) = \log\left(\sum_{i=1}^n w_i\rho_i(x)\right).$$

Then

$$\begin{aligned} -S(x) &= \int_x \rho(x) \log \rho(x) dx = \int_x \left(\sum_{i=1}^n w_i\rho_i(x)\right) \log\left(\sum_{i=1}^n w_i\rho_i(x)\right) dx \\ &= \sum_{j=1}^n w_j \int_x \rho_j(x) \log w_j\rho_j(x) dx \\ &= \sum_{j=1}^n w_j \int_x \rho_j(x) (\log w_j + \log \rho_j(x)) dx \\ &= \sum_{j=1}^n w_j \log w_j \int_x \rho_j(x) dx + \sum_{j=1}^n w_j \int_x \rho_j(x) \log \rho_j(x) dx \\ &= \sum_{j=1}^n w_j \log w_j - \sum_{j=1}^n w_j S_j. \end{aligned}$$

Multiplying by  $-1$ , this result can be written as

$$\boxed{S(\rho) = -\sum_{j=1}^n w_j \log w_j + \sum_{j=1}^n w_j S(\rho_j)}. \quad (3.68)$$

The weights  $\{w_i\}$  can be viewed as a probability distribution function on a finite set, and so (3.68) can be viewed as a statement relating the entropy of this distribution, the (weighted) average of the continuous entropies of the family of probability density functions  $\{\rho_i(x)\}$ , and the entropy of  $\rho(x)$ . However, it should be noted that this equality *does not* hold if the pdfs overlap.

### 3.4.7 Change of Coordinates

Given a pdf,  $\rho_Y(\mathbf{y})$ , and a change of variables  $\mathbf{y} = \mathbf{y}(\mathbf{x})$  with Jacobian matrix  $J(\mathbf{x}) = \partial\mathbf{y}/\partial\mathbf{x}^T$ , then from (3.4) we have  $\rho_X(\mathbf{x}) = \rho_Y(\mathbf{y}(\mathbf{x}))|J(\mathbf{x})|$ . However, when computing the entropy in new coordinates it generally will not retain its value,

$$S(\rho) \neq S(\rho'),$$

although a sufficient condition for equality is the case when  $|J(\mathbf{x})| = 1$  for all values of  $\mathbf{x}$ .

In contrast, to compute the same value of entropy in the new coordinate system, choosing  $f(\mathbf{y}) = -\rho(\mathbf{y}) \log \rho(\mathbf{y})$  gives  $f'(\mathbf{x}) = -\rho(\mathbf{y}(\mathbf{x})) \log \rho(\mathbf{y}(\mathbf{x})) |J(\mathbf{x})|$ , the integral of which produces the same value of entropy. However, this is a somewhat unnatural thing to do since  $\rho(\mathbf{y}(\mathbf{x}))$  is not a pdf without the Jacobian factor.

### 3.4.8 Computation of Entropy via Discretization

Often a probability density function is represented as a histogram, which is effectively an average of the pdf over small intervals that are joined together. Similarly, discrete probabilities that result by integrating probability densities over regularly spaced bins can be stored at lattice points, from which approximations of entropy can be computed. In this subsection the issue of how the computed value of entropy varies based on discretization parameters is addressed.

Given a probability density function,  $\rho(x)$ , a corresponding histogram with *compact support*<sup>9</sup>  $[x_{min}, x_{max}]$  and  $N$  bins of size  $\nu = (x_{max} - x_{min})/N$  is written as

$$\rho_H(x) = \sum_{i=0}^{N-1} \bar{\rho}_i \cdot W(x, x_{min} + i\nu, x_{min} + (i+1)\nu)$$

where  $W(x, a, b)$  is a window function equal to 1 on  $a \leq x < b$  and zero otherwise. Here

$$\bar{\rho}_i = \frac{1}{\nu} \int_{x_{min} + i\nu}^{x_{min} + (i+1)\nu} \rho(x) dx$$

is the average value of  $\rho(x)$  over the  $i$ th bin. From the definition of a pdf,

$$\nu \sum_{i=0}^{N-1} \bar{\rho}_i = 1.$$

Note that the original pdf can be written as

$$\rho(x) = \sum_{i=0}^{N-1} (1 + \epsilon_i(x)) \bar{\rho}_i \cdot W(x, x_{min} + i\nu, x_{min} + (i+1)\nu) \quad (3.69)$$

where  $\epsilon_i(x)$  is a function describing the deviation of  $\rho(x)$  from  $\rho_H(x)$  at each value of  $x$  in the  $i$ th bin. Since by definition  $\bar{\rho}_i$  is an average over the bin, it must be the case that

$$\int_{x_{min} + i\nu}^{x_{min} + (i+1)\nu} \epsilon_i(x) dx = 0.$$

As the number of bins becomes large, the magnitude of  $|\epsilon_i(x)|$  must become smaller if  $\rho(x)$  is continuous.

Using the form (3.69) and properties of the window function and log function, the continuous entropy of  $\rho$  can be written as

---

<sup>9</sup>A function is said to have compact support if it takes a value of zero outside of a compact domain.

$$\begin{aligned}
S(\rho) &= - \int_x \rho(x) \log \rho(x) dx \\
&= - \int_x \sum_{i=0}^{N-1} (1 + \epsilon_i(x)) \bar{\rho}_i \log[(1 + \epsilon_i(x)) \bar{\rho}_i] \cdot W(x, x_{min} + i\nu, x_{min} + (i+1)\nu) dx \\
&= -\nu \sum_{i=0}^{N-1} \bar{\rho}_i \log \bar{\rho}_i - \sum_{i=0}^{N-1} \int_{bin_i} (1 + \epsilon_i(x)) \bar{\rho}_i \log(1 + \epsilon_i(x)) dx.
\end{aligned}$$

For  $|\epsilon_i(x)| \ll 1$ , the approximation  $\log(1 + \epsilon_i(x)) \approx \epsilon_i(x)$  is good. This means that for relatively small bins,

$$S(\rho) = -\nu \sum_{i=0}^{N-1} \bar{\rho}_i \log \bar{\rho}_i - \sum_{i=0}^{N-1} \bar{\rho}_i \int_{bin_i} |\epsilon_i(x)|^2 dx + O(\max_x \|\epsilon(x)\|^3).$$

The first term on the right is the entropy of the histogram,  $S(\rho_H)$ . The second term is a negative quantity that dominates the third-order terms for sufficiently small bin sizes. Therefore, for sufficiently small bin sizes

$$\boxed{S(\rho) \leq S(\rho_H)}. \quad (3.70)$$

Now the question of how discrete and continuous entropies relate can be addressed. The probability contained in bin  $i$  can be written as  $p_i = \nu \bar{\rho}_i$ . This means that for rather small bins

$$S(\rho) \approx S(\rho_H) \approx \log \nu - \sum_{i=0}^{N-1} p_i \log p_i.$$

The last term can be called *discrete entropy*, and can be denoted as

$$\boxed{S(\{p_i\}) \doteq - \sum_{i=0}^{N-1} p_i \log p_i.} \quad (3.71)$$

Hence, the absolute value of discrete entropy depends on the bin size. Whereas  $S(\rho)$  can take negative values (and approaches a value of negative infinity as  $\rho$  becomes a Dirac delta function), the discrete entropy is always bounded from below by zero:

$$S(\{p_i\}) \geq 0. \quad (3.72)$$

Given two pdfs,  $\rho^{(1)}(x)$  and  $\rho^{(2)}(x)$ , on the same domain and applying the same histogram rules to both results in discrete probabilities  $\{p_i^{(1)}\}$  and  $\{p_i^{(2)}\}$ . The *entropy difference* between the two continuous and two discrete entropies approaches zero as the number of discretizations becomes large:

$$S(\rho^{(2)}) - S(\rho^{(1)}) = S(\{p_i^{(2)}\}) - S(\{p_i^{(1)}\}) \text{ as } N \rightarrow \infty. \quad (3.73)$$

This can be viewed as one of the justifications for using lattice models for computing the (statistical mechanical) entropy differences for physical systems. However, what should also be clear from this discussion is that only entropies of the same kind (i.e., continuous or discrete) should be compared with each other. Otherwise, physically meaningless numbers such as  $\log \nu$  will enter and render the result meaningless.

### 3.5 The Classical Central Limit Theorem

The classical central limit theorem for the real line, when stated in terms of probability densities, is as follows. Let  $\rho_j(x) = \rho(x)$  for  $j = 1, \dots, n$  for some positive integer  $n$ . If

$$\int_{-\infty}^{\infty} x\rho(x)dx = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x^2\rho(x)dx = 1/n$$

and all moments

$$\langle x^k \rangle = \int_{-\infty}^{\infty} x^k \rho(x) dx$$

are bounded, then

$$\boxed{\lim_{n \rightarrow \infty} (\rho_1 * \rho_2 * \dots * \rho_n)(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.} \tag{3.74}$$

In the subsections that follow, several very different ways to approach the proof of the central limit theorem are provided.

#### 3.5.1 The Central Limit Theorem (Fourier Version)

The proof of this statement follows by taking the Fourier transform of the  $n$ -fold convolution, which results in the  $n$ th power of  $\hat{\rho}(\omega)$ , and recognizing that

$$\begin{aligned} \hat{\rho}(\omega) &= \int_{-\infty}^{\infty} \rho(x) e^{-i\omega x} dx \\ &= \int_{-\infty}^{\infty} \rho(x) \left( \lim_{m \rightarrow \infty} \sum_{k=0}^m (-i\omega)^k x^k / k! \right) dx \\ &= \lim_{m \rightarrow \infty} \sum_{k=0}^m (-i\omega)^k / k! \int_{-\infty}^{\infty} x^k \rho(x) dx \\ &= \lim_{m \rightarrow \infty} \sum_{k=0}^m \langle x^k \rangle \omega^k (-i)^k / k!. \end{aligned}$$

Then

$$\begin{aligned} \mathcal{F} \left( \lim_{n \rightarrow \infty} (\rho_1 * \rho_2 * \dots * \rho_n)(x) \right) &= \lim_{n \rightarrow \infty} \left( \prod_{j=1}^n \hat{\rho}_j(\omega) \right) \\ &= \lim_{n \rightarrow \infty} [\hat{\rho}(\omega)]^n \\ &= \lim_{n \rightarrow \infty} \left[ \lim_{m \rightarrow \infty} \sum_{k=0}^m \langle x^k \rangle \omega^k (-i)^k / k! \right]^n \\ &= \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \left[ \sum_{k=0}^m \langle x^k \rangle \omega^k (-i)^k / k! \right]^n. \end{aligned}$$

This last step requires that the sequence defined by the terms in brackets converges, which is guaranteed by the assumption that the moments are all bounded.

Now, if the above is approximated at finite  $m$  and  $n$ , the *multinomial expansion*

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1, k_2, \dots, k_m} \frac{n!}{k_1! k_2! \cdots k_m!} x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m} \quad \text{where} \quad \sum_{l=1}^m k_l = n$$

can be used. The sums in this expansion are over all sequences of non-negative integers constrained as indicated by the equality on the right side above.

For any finite  $m$ , the limit as  $n \rightarrow \infty$  in the multinomial expansion is dominated by the first two terms:

$$\lim_{n \rightarrow \infty} \left[ \sum_{k=0}^m \langle x^k \rangle \omega^k (-i)^k / k! \right]^n = \lim_{n \rightarrow \infty} [1 - \langle x \rangle \omega i - \langle x^2 \rangle \omega^2 / 2 + \cdots]^n$$

$$= \lim_{n \rightarrow \infty} [1 - \omega^2 / 2n + \cdots]^n \tag{3.75}$$

$$= e^{-\omega^2 / 2}. \tag{3.76}$$

The limit over  $m$  then becomes irrelevant. Taking the inverse Fourier transform then gives (3.74). Clearly if the terms  $+\cdots$  in (3.75) are small enough, then the limit will be the same regardless of whether or not  $\rho_i(x) = \rho(x)$  for all  $i$ .

### 3.5.2 The Central Limit Theorem (RMSD Error Version)

Consider the class of probability density functions for which each member can be described as a weighted sum of Gaussian distributions. In other words, each member of this class will be of the form

$$f(x) = \sum_{i=1}^N a_i \rho(x; \mu_i, \sigma_i^2) \tag{3.77}$$

where  $\rho(x; \mu_i, \sigma_i^2)$  is a Gaussian with mean  $\mu_i$  and variance  $\sigma_i^2$ . In order for  $f(x)$  to be a pdf,

$$\int_{-\infty}^{\infty} f(x) dx = \sum_{i=1}^N a_i = 1$$

since each  $\rho(x; \mu_i, \sigma_i^2)$  is a pdf. Sometimes pdfs of the form (3.77) are called *multi-Gaussian distributions* (which should not be confused with the multivariate Gaussian distributions in (2.24)).

The mean and variance of  $f(x)$  are calculated as

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \sum_{i=1}^N a_i \int_{-\infty}^{\infty} x \rho(x; \mu_i, \sigma_i^2) dx \\ &= \sum_{i=1}^N a_i \mu_i \end{aligned}$$

and

$$\begin{aligned}
 \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\
 &= \sum_{i=1}^N a_i \int_{-\infty}^{\infty} (x - \mu)^2 \rho(x; \mu_i, \sigma_i^2) dx \\
 &= \sum_{i=1}^N a_i \int_{-\infty}^{\infty} (y + \mu_i - \mu)^2 \rho(y; 0, \sigma_i^2) dy \\
 &= \sum_{i=1}^N a_i \int_{-\infty}^{\infty} [y^2 + 2(\mu_i - \mu)y + (\mu_i - \mu)^2] \rho(y; 0, \sigma_i^2) dy \\
 &= \sum_{i=1}^N a_i [\sigma_i^2 + (\mu_i - \mu)^2].
 \end{aligned}$$

In summary,

$$\mu = \sum_{i=1}^N a_i \mu_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^N a_i [\sigma_i^2 + (\mu_i - \mu)^2] \quad \text{where} \quad \sum_{i=1}^N a_i = 1. \quad (3.78)$$

Now consider the case when  $\mu = 0$  and  $\sigma^2 = 1$  and define the *root-mean-square “distance”* (or *deviation*) (abbreviated as *RMSD*) between  $f(x)$  and  $\rho(x; 0, 1)$  as

$$d[f, \rho] \doteq \left( \int_{-\infty}^{\infty} |\rho(x; 0, 1) - f(x)|^2 dx \right)^{\frac{1}{2}}. \quad (3.79)$$

Unlike  $D_{KL}(f \parallel \rho)$  and  $D_{FI}(f \parallel \rho)$ , the RMSD  $d[f, \rho]$  actually is a valid distance/metric function on the set of univariate pdfs in the sense that it satisfies all of the properties (1.25)–(1.27). The closer this number is to zero, the closer  $f(x)$  will be to a Gaussian distribution with zero mean and unit variance.

Note that for a scaled version of  $f(x)$ ,  $f_s(x) \doteq f(x/s)/s$ , the mean will be  $\mu_{f_s} = s\mu_f$  and the variance will be  $\sigma_{f_s}^2 = s^2\sigma_f^2$ . (In the present case  $\mu_f = 0$  and  $\sigma_f^2 = 1$ , but the above observation is true more generally.) If  $s > 1$ , this scaling has the effect of widening and shortening the pdf, whereas if  $s < 1$ , it makes  $f_s(x)$  more “concentrated” or “tightly focused.” This is true for any pdf.

In light of this scaling operation, the central limit theorem can be viewed in the following way. Choosing the scale  $s = 1/\sqrt{2}$ , and using the fact that variances add as a result of convolution, the central limit theorem says

$$d[(f_{1/\sqrt{2}} * f_{1/\sqrt{2}}), \rho] \leq d[f, \rho]. \quad (3.80)$$

In other words, self-convolution causes a multi-Gaussian to look more like a Gaussian than it did before the convolution.

Since  $f_{1/\sqrt{2}}(x)$  is a weighted sum of Gaussians, the convolution  $(f_{1/\sqrt{2}} * f_{1/\sqrt{2}})(x)$  can be computed in closed form, as can the RMSD expressions on both sides in (3.80). Comparison of the resulting expressions can be used to verify that (3.80) holds, thereby proving the central limit theorem.

Explicitly,

$$(f_{1/\sqrt{2}} * f_{1/\sqrt{2}})(x) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j \rho(x; (\mu_i + \mu_j)/\sqrt{2}, (\sigma_i^2 + \sigma_j^2)/2).$$

Then, using the result (3.86) from the exercises,  $d^2[(f_{1/\sqrt{2}} * f_{1/\sqrt{2}}), \rho]$  and  $d^2[f, \rho]$  can be calculated in closed form. In principle, inequalities provided in [22] can then be manipulated to verify that (3.80) holds.

### 3.5.3 The Central Limit Theorem (Information-Theoretic Version)

The definitions and properties of information-theoretic entropy can be used to try to prove the central limit theorem in ways that are independent of Fourier analysis without assuming that the pdfs have a particular form. For example, if

$$f_{1,n}(x) \doteq (f_1 * f_2 * \cdots * f_n)(x)$$

where each  $f_i(x)$  has mean  $\mu/n$  and variance  $\sigma^2/n$ , then one information-theoretic argument would be to try to show that

$$\lim_{n \rightarrow \infty} D_{KL}(f_{1,n} \| \rho_{\mu, \sigma^2}) \rightarrow 0. \quad (3.81)$$

Another information-theoretic argument would be to try to show that

$$\lim_{n \rightarrow \infty} |S(f_{1,n}) - S(\rho_{\mu, \sigma^2})|^2 / |S(\rho_{\mu, \sigma^2})|^2 \rightarrow 0 \quad (3.82)$$

or more generally to show that  $S(f_{1,n})$  approaches  $S(\rho_{\mu, \sigma^2})$ . Stated in words, (3.82) says that convolutions increase entropy until it asymptotes at its maximal possible value for given mean and variance.

Another information-theoretic argument considers Fisher information. Whereas entropy increases under convolutions, information is lost under convolutions. Convolution “smooths” pdfs and destroys details. The Gaussian distribution contains the least Fisher information of any distribution for given values of mean and variance. Therefore, another information-theoretic strategy to prove the central limit theorem would be to show that

$$\lim_{n \rightarrow \infty} F(f_{1,n}) \rightarrow F(\rho_{\mu, \sigma^2}). \quad (3.83)$$

The use of entropy-theoretic approaches to the central limit theorem originated with Linnik [33] and has been refined and described more fully in [1, 2, 7, 27, 28, 29, 30, 45, 47].

### 3.5.4 Limitations of the Central Limit Theorem

The condition that the moments  $\langle x^k \rangle$  must be bounded is a rather severe condition. It is not so bad when  $\rho(x)$  has finite support and variance  $1/n$  because then these moments will all be decaying functions. For example, even in the extreme case when

$$\rho(x) = \frac{1}{2} \delta(x + 1/n) + \frac{1}{2} \delta(x - 1/n)$$

the higher moments are bounded.

However, there are cases in which the premise of the central limit theorem is violated, but the result holds nonetheless. For example, the normal distribution  $N(0, 1/n)$  for any fixed and finite value of  $n$  has moments of the form



$$\langle x^{2k} \rangle = \frac{(2k)!}{2^k k!} (1/n)^{2k} \quad \text{and} \quad \langle x^{2k+1} \rangle = 0 \quad (3.84)$$

for all positive integers  $k$ . Therefore as  $k \rightarrow \infty$ , it follows from Stirling's formula (2.22) that the even moments grow rapidly and are not bounded, and the conditions assumed in the Fourier proof will not be satisfied. Nonetheless, Gaussian distributions are closed under convolution and since variances add under convolution, the  $n$ -fold convolution of Gaussian distributions with variance  $1/n$  will result in a Gaussian distribution with variance of unity for any positive value of  $n$ .

On the other hand, there are distributions that do not have bounded variance, and repeated convolution of these distributions with themselves will not converge to a Gaussian distribution. One example is the Cauchy distribution in (3.14). See [40, 51] for other examples.

### 3.6 An Alternative Measure of Dispersion

Entropy is a measure of dispersion or disorder. However, this is not the only such measure. For example, it is possible to define a measure of dispersion in Fourier space as [21]

$$D(f) = - \int_{-\infty}^{\infty} \log |\hat{f}(\omega)| d\omega. \quad (3.85)$$

Since  $f(x) \geq 0$ ,

$$|\hat{f}(\omega)| = \left| \int_{-\infty}^{\infty} f(x) e^{i\omega x} dx \right| \leq \int_{-\infty}^{\infty} f(x) |e^{i\omega x}| dx = \int_{-\infty}^{\infty} f(x) dx = 1,$$

and so the integrand in (3.85) is always negative and hence  $D(f)$  is always positive.

It becomes immediately obvious that

$$\begin{aligned} D(f_1 * f_2) &= - \int_{-\infty}^{\infty} \log |\hat{f}_1(\omega) \hat{f}_2(\omega)| d\omega \\ &= - \int_{-\infty}^{\infty} \log (|\hat{f}_1(\omega)| \cdot |\hat{f}_2(\omega)|) d\omega \\ &= - \int_{-\infty}^{\infty} \{ \log |\hat{f}_1(\omega)| + \log |\hat{f}_2(\omega)| \} d\omega \\ &= D(f_1) + D(f_2). \end{aligned}$$

As was discussed earlier, the Gaussian distribution is the pdf that maximizes entropy subject to constraints on the value of the mean and variance. A natural question then becomes, "What distribution maximizes  $D(f)$  in (3.85) subject to these same constraints?"

### 3.7 Chapter Summary

This chapter presented a broad summary of probability and information theory. Information-theoretic measures of the "divergence" between two probability density functions were reviewed together with classical inequalities such as the Cramér–Rao bound, entropy power inequality, and de Bruijn identity. In the proof of the de Bruijn identity,

properties of the heat equation reminiscent of Chapter 2 were employed. The prominent role of the Gaussian distribution as the special pdf to which others converge under iterated convolution was established in the central limit theorem. This, together with its convenient parametrization and relationship to the heat equation make the Gaussian ideal in the context of the problems that follow. In particular, random sampling from a Gaussian distribution (which is an operation built into software packages such as MATLAB<sup>TM</sup>) is a convenient (and physically motivated) way to generate random noise. It is this noise that is superimposed onto an ordinary differential equation to result in a stochastic differential equation, which is one of the subjects of the next chapter.

This chapter serves as an elementary introduction to probability and information theory. More in-depth references are provided at the end of this chapter. Classical references in information theory include [11, 16, 17, 25, 38]. The study and application of information-theoretic inequalities remains an area of investigation. See, for example, the recent references [19, 24, 35, 44, 50]. The topics in geometry presented in later chapters of this book make it possible to understand the connections between probability, information theory, and geometry that have begun to emerge in recent years that are described in [4, 6, 23, 36, 37, 41, 49].

### 3.8 Exercises

3.1. Show that for  $a > 0$ ,

$$\int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} [ax^2 - 2bx + c] \right\} dx = \left( \frac{2\pi}{a} \right)^{\frac{1}{2}} e^{-\frac{1}{2}(c-b^2/a)}.$$

3.2. Show that for Gaussian distributions  $\rho(x; \mu_i, \sigma_i^2)$  with  $\mu_i \in \mathbb{R}$  and  $\sigma_i \in \mathbb{R}_{>0}$ ,

$$\int_{-\infty}^{\infty} \rho(x; \mu_1, \sigma_1^2) \rho(x; \mu_2, \sigma_2^2) dx = \frac{1}{(2\pi)^{\frac{1}{2}} (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}} \exp \left( -\frac{1}{2} \left[ \frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2} (\mu_1 - \mu_2)^2 \right]^2 \right). \quad (3.86)$$

3.3. If the mean and covariance of  $\rho(\mathbf{x})$  are known, what will the mean and covariance of

$$\rho_{(A, \mathbf{b})}(\mathbf{x}) = |\det A|^{-1} \rho(A^{-1}(\mathbf{x} - \mathbf{b}))$$

be?

3.4. Suppose that the convolution product  $(\rho_1 * \rho_2)(\mathbf{x})$  has already been computed for two pdfs,  $\rho_1$  and  $\rho_2$  on  $\mathbb{R}^n$ . Now suppose that new pdfs  $\rho_{(A_i, \mathbf{b}_i)}(\mathbf{x}) = |\det A_i|^{-1} \rho_i(A_i^{-1}(\mathbf{x} - \mathbf{b}_i))$  are defined for  $i = 1, 2$ . What are the most general conditions under which the original convolution can be evaluated by a change of variables to produce  $(\rho_{(A_1, \mathbf{b}_1)} * \rho_{(A_2, \mathbf{b}_2)})(\mathbf{x})$ , thereby circumventing the direct calculation of the convolution from scratch?

3.5. Prove both equalities in (3.16) and (3.17) for arbitrary integrable scalar-valued function  $\phi_i(x_i)$ . If instead of being scalar-valued functions, if  $\phi_i(x_i)$  are matrix valued, will these equalities still hold?

3.6. Verify that: (a)  $\langle 1|x \rangle = 1$ ; (b)  $\langle \phi(x)|x \rangle = \phi(x)$ ; and (c)

$$\langle\langle\phi_1(x_1)\phi_2(x_2)|x_2\rangle x_1\rangle = \langle\phi_1(x_1)|x_2\rangle \cdot \langle\phi_2(x_2)|x_1\rangle. \quad (3.87)$$

3.7. Given the pdf  $f(x, y, z)$ , prove that

$$\langle\Phi(\langle\langle\phi(x)|y, z\rangle z\rangle)\rangle \leq \langle\Phi(\phi(x))\rangle. \quad (3.88)$$

3.8. The explicit meaning of the penultimate step in the derivation of (3.25) is

$$\left\langle \left( \beta \frac{\rho'_1(u)}{\rho_1(u)} + (1 - \beta) \frac{\rho'_2(v)}{\rho_2(v)} \right)^2 \right\rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \beta \frac{\rho'_1(u)}{\rho_1(u)} + (1 - \beta) \frac{\rho'_2(v)}{\rho_2(v)} \right)^2 \rho_1(u) \rho_2(v) du dv.$$

When completing the square, why does the cross term integrate to zero?

3.9. For the multivariate Gaussian distribution  $\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ , if  $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$  with  $\mathbf{x}_i \in \mathbb{R}^{n_i}$ , calculate the following: (a) entropy of the full density  $S(\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma))$ ; (b) entropy of the marginal density  $S(\rho_i(\mathbf{x}_i; \boldsymbol{\mu}_i, \Sigma_i))$ ; (c) the marginal entropy  $S(\rho_i; \rho)$ .

3.10. For two multivariate Gaussian distributions  $\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$  and  $\rho'(\mathbf{x}; \boldsymbol{\mu}', \Sigma')$  compute: (a) the Kullback–Leibler divergence,  $D_{KL}(\rho\|\rho')$ ; and (b) the Fisher information divergence,  $D_{FI}(\rho\|\rho')$ .

3.11. Let  $\rho(\mathbf{x}, \boldsymbol{\theta})$  be a pdf in the variable  $\mathbf{x} \in \mathbb{R}^n$  for each different value of  $\boldsymbol{\theta}$ . Show that the Fisher information matrix and Kullback–Leibler distance are related as follows:

$$F_{ij}(\boldsymbol{\theta}_0) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} D_{KL}(\rho(\mathbf{x}, \boldsymbol{\theta})\|\rho(\mathbf{x}, \boldsymbol{\theta}_0)) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \quad (3.89)$$

3.12. Show that the Fisher information divergence is invariant under Euclidean (rigid-body) transformations of the form  $(Ef_i)(\mathbf{x}) = f_i(R^T(\mathbf{x} - \mathbf{t}))$  where  $R$  is an arbitrary rotation matrix satisfying  $RR^T = I$  and  $\mathbf{t}$  is an arbitrary translation vector. In other words, show that

$$D_{FI}(Ef_1\|Ef_2) = D_{FI}(f_1\|f_2).$$

3.13. Prove that the equalities in (3.52) are the same as the definition in (3.39) in the special case when the dimensions of  $\mathbf{x}$  and  $\boldsymbol{\theta}$  are the same and  $f(\mathbf{x}; \boldsymbol{\theta}) = f(\mathbf{x} - \boldsymbol{\theta})$ .

3.14. Prove that if  $\boldsymbol{\theta}$  is the vector made up of the  $n$  entries in  $\boldsymbol{\mu}$  and the  $n(n+1)/2$  independent entries in  $\Sigma$ , then the Fisher information matrix for a Gaussian distribution on  $\mathbb{R}^n$  with unknown mean and variance is the  $[n + n(n+1)/2] \times [n + n(n+1)/2]$  matrix with entries [18, 42]

$$F_{jk} = \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_j} \Sigma^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_k} + \frac{1}{2} \text{tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k} \right). \quad (3.90)$$

## References

1. Applebaum, D., *Probability and Information: An Integrated Approach*, 2nd ed., Cambridge University Press, London, 2008.
2. Barron, A.R., “Entropy and the central-limit-theorem,” *Ann. Prob.*, 14, pp. 336–342, 1986.
3. Bertsekas, D., *Convex Analysis and Optimization*, Athena Scientific, 2003.

4. Bhattacharya, R., Patrangenaru, V., "Nonparametric estimation of location and dispersion on Riemannian manifolds," *J. Stat. Plann. Inference*, 108, pp. 23–36, 2002.
5. Blachman, N.M., "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, 11, pp. 267–271, 1965.
6. Braunstein, S.L., Caves, C.M., "Statistical distance and the geometry of quantum states," *Phys. Rev. Lett.*, 72, pp. 3439–3443, 1994.
7. Brown, L.D., "A proof of the Central Limit Theorem motivated by the Cramér–Rao inequality," in G. Kallianpur, P.R. Krishnaiah, and J.K. Ghosh, eds., *Statistics and Probability: Essays in Honour of C.R. Rao*, pp. 141–148, North-Holland, New York, 1982.
8. Chirikjian, G.S., Kyatkin, A.B., *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, Boca Raton, FL, 2001.
9. Costa, M.H., "A new power inequality," *IEEE Trans. Inf. Theory*, 31, pp. 751–760, 1985.
10. Cover, T.M., Thomas, J.A., *Elements of Information Theory*, Wiley-Interscience, 2nd ed., Hoboken, NJ, 2006.
11. Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ, 1946.
12. Crassidis, J.L., Junkins, J.L., *Optimal Estimation of Dynamic Systems*, Chapman & Hall/CRC, London, 2004.
13. Dembo, A., Cover, T.M., Thomas, J.A., "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, 37, pp. 1501–1518, 1991.
14. Doob, J.L., *Stochastic Processes*, Wiley, New York, 1953.
15. Feller, W., *Introduction to Probability Theory and its Applications*, John Wiley & Sons, New York, 1971.
16. Fisher, R.A., "On the mathematical foundations of theoretical statistics," *Philos. Trans. R. Soc. London Ser. A*, 222, pp. 309–368, 1922.
17. Fisher, R.A., "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, 22, pp. 700–725, 1925.
18. Frieden, B.R., *Science from Fisher Information*, Cambridge University Press, New York, 2004.
19. Gardner, R.J., "The Brunn–Minkowski inequality," *Bull. Amer. Math. Soc.*, 39, pp. 355–405, 2002.
20. Gnedenko, B.V., Kolmogorov, A.N., *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading, MA, 1954 (and 1968).
21. Grenander, U., *Probabilities on Algebraic Structures*, John Wiley & Sons, New York, 1963 (reprinted by Dover, 2008).
22. Hardy, G.I., Littlewood, J.E., Pólya, G., *Inequalities*, 2nd ed., Cambridge University Press, London, 1952.
23. Hendricks, H., "A Cramér–Rao type lower bound for estimators with values in a manifold," *J. Multivariate Anal.*, 38, pp. 245–261, 1991.
24. Itoh, Y., "An application of the convolution inequality for the Fisher information," *Ann. Inst. Stat. Math.*, 41, pp. 9–12, 1989.
25. Jaynes, E.T., *Probability Theory: The Logic of Science*, Cambridge University Press, London, 2003.
26. Jensen, J.L.W.V., "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Math.*, 30, pp. 175–193, 1906.
27. Johnson, O., *Information Theory and the Central Limit Theorem*, Imperial College Press, London, 2004.
28. Johnson, O.T., "Entropy inequalities and the Central Limit Theorem," *Stochastic Process. Appl.*, 88, pp. 291–304, 2000.
29. Johnson, O., "A conditional entropy power inequality for dependent variables," *IEEE Trans. Inf. Theory*, 50, pp. 1581–1583, 2004.
30. Johnson, O., Barron, A., "Fisher information inequalities and the central limit theorem," *Probability Theory Related Fields*, 129, pp. 391–409, 2004.
31. Kullback, S., *Information Theory and Statistics*, Dover, New York, 1997 (originally published in 1958).

32. Lawler, G.F., *Introduction to Stochastic Processes*, 2nd ed., CRC Press, Boca Raton, FL, 2006.
33. Linnik, Y.V., “An information-theoretic proof of the Central Limit Theorem with the Lindeberg condition,” *Theory Probab. Its Appl.*, 4, pp. 288–299, 1959.
34. Madiman, M., Barron, A., “Generalized entropy power inequalities and monotonicity properties of information,” *IEEE Trans. Inf. Theory*, 53, pp. 2317–2329, 2007.
35. Nikolov, B., Frieden, B.R., “Limitation on entropy increase imposed by Fisher information,” *Phys. Rev. E*, 49, pp. 4815–4820 Part A, 1994.
36. Pennec, X., “Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements,” *J. Math. Imaging Vision*, 25, pp. 127–154, 2006.
37. Pennec, X., “Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements,” *IEEE Workshop on Nonlinear Signal and Image Processing*, 1999.
38. Rao, C.R., “Information and the accuracy attainable in the estimation of statistical parameters,” *Bull. Calcutta Math. Soc.*, 37, pp. 81–89, 1945.
39. Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
40. Samorodnitsky, G., Taqqu, M.S., *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, 1994.
41. Scharf, L.L., McWhorter, L., “Geometry of the Cramér–Rao bound,” *Signal Process.*, 31, pp. 301–311, 1993.
42. Schervish, M.J., *Theory of Statistics*, Springer, New York, 1995.
43. Shannon, C.E., Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.
44. Smith, S.T., “Covariance, subspace, and intrinsic Cramér–Rao bounds in signal processing,” *IEEE Trans. Acoustics Speech Signal Process.*, 53, pp. 1610–1630, 2005.
45. Stam, A.J., “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” *Inf. Control*, 2, pp. 101–112, 1959.
46. Verdú, S., Guo, D., “A simple proof of the entropy-power inequality,” *IEEE Trans. Inf. Theory*, 52, pp. 2165–2166, 2006.
47. Villani, C., “Entropy production and convergence to equilibrium,” in *Entropy Methods For The Boltzmann Equation*, Lecture Notes in Mathematics, Volume: 1916, pp. 1–70, Springer, Berlin, 2008.
48. Williams, D., *Probability with Martingales*, Cambridge University Press, London, 1991.
49. Xavier, J., Barroso, V., “Intrinsic variance lower bound (IVLB): an extension of the Cramér–Rao bound to Riemannian manifolds,” *IEEE International Conference on Acoustics, Speech, and Signal Processing 2005 Proceedings (ICASSP '05)*, Vol. 5, pp. 1033–1036, March 18–23, 2005.
50. Zamir, R., Feder, M., “A generalization of the entropy power inequality with applications,” *IEEE Trans. Inf. Theory*, 39, pp. 1723–1728, 1993.
51. Zolotarev, V.M., *One-Dimensional Stable Distributions*, Translations of Mathematical Monographs, Vol. 65, American Mathematical Society, Providence, RI, 1986.

---

## Stochastic Differential Equations

The chapter begins with Section 4.1 in which motivational examples of random walks and stochastic phenomena in nature are presented. In Section 4.2 the concept of random processes is introduced in a more precise way. In Section 4.3 the concept of a Gaussian and Markov random process is developed. In Section 4.4 the important special case of white noise is defined. White noise is the driving force for all of the stochastic processes studied in this book. Other sections in this chapter define Itô and Stratonovich stochastic differential equations (SDEs), their properties and corresponding Fokker–Planck equations, which describe how probability densities associated with SDEs evolve over time. In particular, Section 4.7 examines the Fokker–Planck equation for a particular kind of SDE called an Ornstein–Uhlenbeck process. And Section 4.8 examines how SDEs and Fokker–Planck equations change their appearance when different coordinate systems are used.

The main points that the reader should take away from this chapter are:

- Whereas a deterministic system of ordinary differential equations that satisfies certain conditions (i.e., the Lipschitz conditions) are guaranteed to have a unique solution for any given initial conditions, when random noise is introduced the resulting “stochastic differential equation” will not produce repeatable solutions.
- It is the ensemble behavior of the sample paths obtained from numerically solving a stochastic differential equation many times that is important.
- This ensemble behavior can be described either as a stochastic integral (of which there are two main types, called Itô and Stratonovich), or by using a partial differential equation akin to the diffusion equations studied in Chapter 2, which is called the Fokker–Planck (or forward Kolmogorov) equation.
- Two different forms of the Fokker–Planck equation exist, corresponding to the interpretation of the solution of a given SDE as being either an Itô or Stratonovich integral, and an analytical apparatus exists for converting between these forms.
- Multi-dimensional SDEs in  $\mathbb{R}^n$  can be written in Cartesian or curvilinear coordinates, but care must be taken when converting between coordinate systems because the usual rules of multivariable calculus do not apply in some situations.

### 4.1 Motivating Examples

Motivational examples are provided in this section to introduce the usefulness of stochastic differential equations as a modeling tool. First the discrete-time discrete-space case

of a random walker on the integers is discussed. Then, continuous-time continuous-space Brownian motion is discussed as a limiting case of the discrete theory.

#### 4.1.1 The Discrete Random Walker

A *random (or stochastic) process* is a random variable that varies with time. One of the simplest examples of a random process is a random walk on the integers. Imagine a random walker who moves from one integer value to either of the two adjacent ones with equal probability. Motion occurs only at integer units of time,<sup>1</sup>  $n \in \mathbb{Z}$ . At  $n = 0$ , the walker starts at the position  $k = 0$ . Then, at  $n = 1$ , the walker will be at either  $k = +1$  or  $k = -1$ . Starting at this new location, the walker will change to a new location defined by  $\Delta k = \pm 1$  at  $n = 2$ . The process repeats. The fundamental question becomes, “What is the probability that the walker will be at any specific integer if this process is repeated  $n$  times?”

This discrete-time discrete-space random process can be addressed without using the methods developed later in this section (which are for continuous-time continuous-space random processes).<sup>2</sup> The continuous case can be viewed as a limiting process of the discrete one, and so it is worth working out the solution to the integer random walker up front.

Since the space on which the random walker is moving is discrete, it follows that a discrete probability distribution will describe the state, rather than using a probability density function. At each discrete value of time,  $n$ , if the walker starts at  $k \in \mathbb{Z}$ , the probability of movement to the adjacent integer location will be

$$\Delta p(\Delta k, n) \doteq \frac{1}{2}(\delta_{\Delta k, 1} + \delta_{\Delta k, -1}), \quad (4.1)$$

where  $\delta_{ij}$  is the Kronecker delta function that takes a value of 1 when  $i = j$  and zero otherwise and  $\Delta k$  denotes the change in value of  $k$  from time  $n$  to time  $n + 1$ . Although  $\Delta p(\Delta k, n)$  is written as depending on the discrete time variables  $n \in \{0, 1, 2, \dots\}$ , it is actually constant with respect to this variable in the current problem.

In the same way that convolution on the real line was used in the previous chapters to determine the distribution corresponding to the sum of continuous random variables, convolution on the integers is used to determine the distribution of integer positions that the discrete random walker attains at time  $n + 1$  relative to the distribution at time  $n$ :

$$p(k, n + 1) = (p * \Delta p)(k, n) \doteq \sum_{j \in \mathbb{Z}} p(j, n) \Delta p(k - j, n). \quad (4.2)$$

Starting with the known initial location  $k = 0$ , it follows that  $p(k, 0) = \delta_{k, 0}$ . Using (4.2) recursively,

$$p(k, 1) = \frac{1}{2}(\delta_{k, 1} + \delta_{k, -1})$$

$$p(k, 2) = \frac{1}{4}(\delta_{k, 2} + 2\delta_{k, 0} + \delta_{k, -2})$$

<sup>1</sup>In some books a random discrete-time “process” is called a “sequence.” In this book the word “process” refers to both the discrete and continuous time.

<sup>2</sup>It is also possible to define continuous-time discrete-space processes (e.g., random walks on graphs), and discrete-time continuous-space processes, but neither of these will be addressed in this book.

$$p(k, 3) = \frac{1}{8}(\delta_{k,3} + 3\delta_{k,1} + 3\delta_{k,-1} + \delta_{k,-3})$$

and so on.

It should not take too much convincing to believe that the pattern that emerges for the coefficients in front of each Kronecker delta is the same as Pascal's triangle, which describes the binomial coefficients. The only small twist is that the binomial coefficients are spread out over a range of values of  $k$  from  $-n$  to  $n$ . In other words, the pattern above generalizes to [31]

$$p(k, n) = \frac{1}{2^n} \sum_{j=-n}^n \binom{n}{\frac{n-j}{2}} \frac{(-1)^n + (-1)^j}{2} \delta_{k,j}. \quad (4.3)$$

The reason for the multiplicative term  $((-1)^n + (-1)^j)/2$  is that adjacent to every value of  $k$  for which  $p(k, n)$  takes the value  $\binom{n}{\frac{n-k}{2}}$  are the zero values  $p(k \pm 1, n) = 0$ . This can be observed above for  $p(k, 1)$ ,  $p(k, 2)$ , and  $p(k, 3)$ . When  $((-1)^n + (-1)^j)/2 = 1$ ,  $n$  and  $j$  are said to have the same *parity*.

The distribution  $p(k, n)$  can be thought of as a special case of the binomial distribution from Section 2.1.6 with  $p(k, n) = f(n - k/2; n, 1/2)$  when  $k$  and  $n$  have the same parity, and  $p(k, n) = 0$  when  $k$  and  $n$  have different parity. It follows from (2.23) that as  $n \rightarrow \infty$ , this is approximated well as a Gaussian distribution when  $k$  and  $n$  have the same parity.

Equation (4.2) can be viewed as describing the evolution of the probability distribution of the discrete random variable  $k(n)$  that evolves in discrete time  $n \in \mathbb{Z}$ . As an alternative to posing the problem in terms of probability distributions, it is possible to write an equation describing the discrete random process  $k(n)$  directly. Such an equation takes the form

$$k(n+1) = k(n) + \nu(n) \quad \text{where} \quad k(0) = 0 \quad (4.4)$$

and  $\nu(n)$  is the “random noise” that has the distribution  $\Delta p(\Delta k, n)$ . Equation (4.4) can be considered to be a stochastic difference equation. This random noise model is, in a sense, equivalent to flipping a coin at each discrete value of time, with heads corresponding to a +1 motion and tails corresponding to a -1 motion. If this is done  $n$  times, a single random path defined by discrete values of  $k(\tau)$  for  $\tau \in \{0, 1, \dots, n\}$  will be generated. If the same experiment is performed many times, an ensemble of random paths will be generated. According to the law of large numbers, the statistical properties of  $k(n)$  in a very large ensemble of random paths should be captured well by (4.3). Therefore, each random path in the ensemble can be thought of as “sampling” the distribution  $p(k, n)$  according to the rules set forth by  $\Delta p(\Delta k, n)$  (or equivalently,  $\nu(n)$ ). In other words,  $p(k, n)$  is not sampled completely at random, but rather according to the adjacency constraint that  $\Delta k = \pm 1$ . Each random path  $k(\tau)$  for  $\tau \in \{0, 1, \dots, n\}$  can be called a *sample path* corresponding to  $\{p(k, \tau) | \tau \in \{0, 1, \dots, n\}\}$ , since many such  $k(\tau)$  paths reproduce the statistics of this time-evolving set of probability distributions.

#### 4.1.2 Continuous-Time Brownian Motion in Continuous Space

The discrete-time discrete-space model of random motion described in the previous subsection can be used to motivate mathematical models of continuous-time random motion in continuous space. If each integer value  $k$  is divided by a fixed number  $N$ , then



a probability density function on the real line with zero mean and variance  $n/N$  can be defined from  $p(k, n)$  when  $n$  is large. This is accomplished by dividing the real line into bins, each of which is centered on  $k/N$ , and ranging from  $(k-1)/N$  to  $(k+1)/N$ . Since the discrete probability  $p(k, n)$  falls inside of this bin, and the size of each bin is  $2/N$ , as  $n \rightarrow \infty$  the resulting histogram will converge to a Gaussian according to (2.23). This histogram is stretched and squashed by a factor of two relative to the distributions in (2.23) because the range of values is  $-n \leq k \leq n$ , and the  $n+1$  non-zero values are distributed evenly over this range.

In the same way that a pdf on the real line is generated from this histogram process, the discrete time parameter,  $n$ , can be replaced with  $t = n/N$ , which can be viewed as a continuous parameter sampled at closely spaced discrete values.

Note that the same Gaussian distribution with variance  $t = n/N$  that was obtained as the limiting case of a binomial distribution could have been obtained by the  $n$ -fold convolution of Gaussian distributions, each with variance  $1/N$ . In other words, a noise model that will generate the same long-time statistical behavior as coin flipping is one in which values are sampled from a continuous Gaussian distribution, provided the correct value of variance is used. With this in mind, the continuous version of (4.4) is

$$x(t + \Delta t) - x(t) = n(t)\Delta t \quad \text{where} \quad x(0) = 0. \quad (4.5)$$

Here  $x(t)$  can take on continuous values, and at each instant in time  $n(t)$  is defined by randomly sampling values from a Gaussian distribution. Each of these samples is drawn without knowledge of the sample values that were taken at previous times.

In the limit as  $\Delta t \rightarrow 0$ , (4.5) becomes a *stochastic differential equation* (or SDE for short). The noise  $n(t)\Delta t$  is denoted as  $dw$ , which is called white noise. The random process  $x(t)$ , which can be written as

$$x(t) = \int_0^t n(\tau)d\tau,$$

is continuous due to the smoothing effect of the integral, but is not differentiable because  $n(\tau)$  is producing values at each value of  $\tau$  that are unrelated to those that come before and after. For reasons that will be explained later, the distribution  $\rho(x, t)$  will be Gaussian. This makes sense intuitively since noise that is driving the motion of  $x$  is Gaussian. And in fact, (4.5) is not far from the models used by Einstein [6] and Langevin [20] in establishing the physical theory of Brownian motion at the beginning of the twentieth century.

A one-dimensional SDE will more generally be thought of as the limiting case of an equation of the form

$$x(t + \Delta t) - x(t) = a(x, t)\Delta t + b(x, t)n(t)\Delta t \quad \text{where} \quad x(0) = x_0 \quad (4.6)$$

as  $\Delta t \rightarrow 0$ . The methods in this chapter establish the tools for obtaining the corresponding probability density function,  $\rho(x, t)$ , for such an equation. Furthermore, SDEs can evolve in multi-dimensional Euclidean space, or on surfaces such as spheres. In order to handle the subtle issues that arise in these generalized settings, terminology and results from the theory of stochastic processes are required. This is the subject of the following sections.

## 4.2 Stationary and Non-Stationary Random Processes

A *random process* is a time-varying random variable. A random process can be a scalar,  $x(t) \in \mathbb{R}$ , vector,  $\mathbf{x}(t) \in \mathbb{R}^d$ , or matrix,  $X(t) \in \mathbb{R}^{d_1 \times d_2}$ . Since a scalar can be thought of as a one-dimensional vector, and a matrix (by the  $\vee$  operation in Appendix A.7) can be identified with a vector, the discussion will be limited to the vector case for convenience. The value of the random process *without history*,  $\mathbf{x}(t)$ , can be thought of as being drawn from a time-varying pdf:  $p(\mathbf{x}, t)$ .<sup>3</sup> In other words, for each fixed value  $t = t_0$ , the random vector  $\mathbf{x}(t_0)$  is sampled from the distribution  $p(\mathbf{x}, t_0)$ . On the other hand, in more general situations the value of  $\mathbf{x}(t)$  could be influenced by both its own value at prior times, as well as the explicit values of those prior times. If the value of  $\mathbf{x}(t_i)$  is recorded at a cascade of prior times,  $t = t_1 > t_2 > \dots > t_n$ , then the pdf describing  $\mathbf{x}(t)$  actually would be one on the product space  $[\mathbf{x}_1, t_1] \times [\mathbf{x}_2, t_2] \times \dots \times [\mathbf{x}_n, t_n]$ . Let  $p(\mathbf{x}, t; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  denote the joint pdf on this product space. This is the pdf for a random process  $\mathbf{x}(t)$ .

There should be no confusion between this and  $p(\mathbf{x}, t)$ . Even though they are both “ $p$ ,” they are different functions and that difference is clear from their arguments. To denote their difference by giving them subscripts is possible, but then  $p(\mathbf{x}, t; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  would be written as  $p_{1,2,\dots,n}(\mathbf{x}, t; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$ , which contains exactly the same information as the arguments themselves, but makes equations involving these pdfs substantially longer. For this reason, these subscripts will not be used when the meaning is clear.

Another alternative way to write  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  would be as  $p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n; t_1, t_2, t_3, \dots, t_n)$ . While perhaps this would be more consistent with the notation used in Chapter 3 in the sense that the variables defining the domain of the pdf appear before the semicolon and the parameters defining the shape of the pdf appear after it, separating the time variables from the corresponding spatial variable would lead to other problems. Therefore, the notation  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  will be used, which is consistent with the literature.

Explicitly, if all of the  $\mathbf{x}_i$ 's have been measured for  $t_i = t_2, \dots, t_n$ , then the pdf describing the statistical behavior of  $\mathbf{x}(t)$  at  $t = t_1$  would be described by the conditional density

$$p(\mathbf{x}, t | \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n) \doteq \frac{p(\mathbf{x}, t; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)}{p(\mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)}. \quad (4.7)$$

That is, the values of  $\mathbf{x}(t)$  would be drawn from the density  $p(\mathbf{x}, t | \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  where all  $\mathbf{x}_i$  and  $t_i$  for  $i \geq 2$  are fixed, because these times have already passed and the values of  $\mathbf{x}_i$  and  $t_i$  have been recorded. Usually for physical systems, the memory that the system has is limited to the prior instant in time, leading to the *Markov property* that will be discussed later.

### 4.2.1 Weak and Strong Stationarity

For fixed times  $t_1 > t_2 > \dots > t_n$ , the function  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  denotes the probability density function of the composite vector  $[\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T$ . If each  $\mathbf{x}_i \in \mathbb{R}^d$ , then this composite vector is  $n \cdot d$ -dimensional. If this pdf is invariant under uniform shifts in time, so that for any  $t_0$ :

<sup>3</sup>The symbol  $p(\cdot)$  is used here rather than  $\rho(\cdot)$  or  $f(\cdot)$  to avoid confusion with Gaussians or solutions to heat equations.

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n) = p(\mathbf{x}_1, t_1 - t_0; \mathbf{x}_2, t_2 - t_0; \mathbf{x}_3, t_3 - t_0; \dots; \mathbf{x}_n, t_n - t_0) \quad (4.8)$$

then the process  $\mathbf{x}(t)$  is called *strongly stationary* or *strictly stationary* [4]. On the other hand, if (4.8) is not satisfied but the mean and covariance of the pdf  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  are the same as the mean and covariance of  $p(\mathbf{x}_1, t_1 - t_0; \mathbf{x}_2, t_2 - t_0; \mathbf{x}_3, t_3 - t_0; \dots; \mathbf{x}_n, t_n - t_0)$ , then  $\mathbf{x}(t)$  is called *weakly stationary*, or *wide-sense stationary*. (In some treatments the condition that the mean remains fixed is relaxed [4], but when referring to a weakly stationary process in the current work, constancy of the mean will be taken as part of the definition.)

For example, if  $\mathbf{x}(t)$  is a memoryless process drawn at random from a probability density function  $p(\mathbf{x}, t) = p_0(\mathbf{x})$  that is completely independent of time, then  $\mathbf{x}(t)$  will be a strongly stationary random process. On the other hand, if  $\mathbf{x}(t)$  is drawn from a time-varying density  $p(\mathbf{x}, t)$ , but the mean and covariance of  $p$  are constant, then  $\mathbf{x}(t)$  will be a weakly stationary process without memory. For example, if  $p_1(\mathbf{x})$  and  $p_2(\mathbf{x})$  are constant pdfs, and  $p_1(\mathbf{x}) \neq p_2(\mathbf{x})$  but  $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}$  and  $\Sigma_1 = \Sigma_2 = \Sigma$ , then for  $0 \leq \alpha(t) \leq 1$ ,

$$p(\mathbf{x}, t) = [1 - \alpha(t)]p_1(\mathbf{x}) + \alpha(t)p_2(\mathbf{x}) \quad (4.9)$$

will also be a pdf with mean and variance  $\boldsymbol{\mu}, \Sigma$ . And a process defined by drawing values from such a pdf will be a weakly stationary process without memory.

The property of strong stationarity<sup>4</sup> of a process implies weak stationarity, but not the other way around.

### 4.2.2 Non-Stationary Processes

If the pdf describing a random process without memory has mean and covariance that change with time, then that process is not stationary. For example, the solutions to the heat/diffusion equations discussed in Chapter 2,  $f(\mathbf{x}, t)$ , are not the pdfs of stationary processes. This is because these pdfs “spread out” as a function of time.

Many of the stochastic differential equations and corresponding Fokker–Planck equations that will be derived later describe processes that are not stationary. However, the input noise that is used to define these processes will not only be stationary, but strongly so.

## 4.3 Gaussian and Markov Processes

In this section basic stochastic processes are reviewed. See [13, 18, 23] for in-depth treatments. The treatment here follows [11, 40].

Let  $p(\mathbf{x}, t)d\mathbf{x}$  be the probability that the random process  $\mathbf{x}(t)$  is contained in the  $d$ -dimensional voxel with volume  $d\mathbf{x} = dx_1 \cdots dx_d$  centered at  $\mathbf{x} \in \mathbb{R}^d$ . The distinction between the stochastic process,  $\mathbf{x}(t)$ , and the domain in which it moves (also denoted as  $\mathbf{x}$ ) will be clear because in the former, the dependence on time will be denoted, whereas the latter does not depend on time.

Let  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_n, t_n)d\mathbf{x}_1 \cdots d\mathbf{x}_n$  be the probability that for each time  $t_i$ , each  $\mathbf{x}(t_i)$  is in the voxel centered at  $\mathbf{x}_i$  for each  $i = 1, \dots, n$ . Hence,  $p(\mathbf{x}, t)$  is a probability density function on  $\mathbb{R}^d$  for each fixed  $t$ , while  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_n, t_n)$  is a pdf on  $\mathbb{R}^{d \cdot n} = \mathbb{R}^d \times \mathbb{R}^d \times \dots \times \mathbb{R}^d$  for each fixed choice of  $(t_1, \dots, t_n)^T \in \mathbb{R}^n$ . Let the times be ordered such that

<sup>4</sup>The word “stationary” is an adjective, whereas “stationarity” is the corresponding noun.

$$t_1 > t_2 > \dots > t_n.$$

It is important to be clear about how these times are ordered. In some books they are ordered in the opposite way. In the treatment here the times will be ordered from most recent (largest times) to those that are furthest in the past (smallest times).

By integrating the pdf  $p(\mathbf{x}_1, t_1; \dots; \mathbf{x}_n, t_n)$  over the last  $n - k$  copies of  $\mathbb{R}^d$  (which erases all historical information), the following general relationship is obtained:

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_k, t_k) = \int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_n, t_n) d\mathbf{x}_{k+1} \dots d\mathbf{x}_n. \quad (4.10)$$

For the case when  $d = 1$ , a closed-form example of (4.10) is easily verified for the *Gaussian process*:

$$p(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = \frac{\exp\left[-\frac{1}{2} \sum_{i,j=1}^n \Sigma_{ij}^{-1} (x_i - \mu_i)(x_j - \mu_j)\right]}{[(2\pi)^n \det \Sigma]^{\frac{1}{2}}} \quad (4.11)$$

where  $\Sigma$  is an  $n \times n$  covariance matrix with elements  $\Sigma_{ij}$  and  $\mu_i = \langle x_i(t) \rangle$  are the components of the mean of  $p$ . For a general Gaussian process with no additional restrictions,  $\Sigma = \Sigma(t_1, \dots, t_n)$  and  $\boldsymbol{\mu} = \boldsymbol{\mu}(t_1, \dots, t_n)$ .

More generally, the two pdfs  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_k, t_k)$  and  $p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_n, t_n)$  are related by the definition of the conditional probability density function  $p(\cdot|\cdot)$

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_n, t_n) =$$

$$p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}; \dots; \mathbf{x}_n, t_n) p(\mathbf{x}_{k+1}, t_{k+1}; \mathbf{x}_{k+2}, t_{k+2}; \dots; \mathbf{x}_n, t_n). \quad (4.12)$$

Direct consequences of the definition in (4.12) and the observation in (4.10) are that

$$\int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}; \dots; \mathbf{x}_n, t_n) \times \\ p(\mathbf{x}_{k+1}, t_{k+1}; \mathbf{x}_{k+2}, t_{k+2}; \dots; \mathbf{x}_n, t_n) d\mathbf{x}_1 \dots d\mathbf{x}_n = 1$$

and

$$\int_{\mathbb{R}^d} \dots \int_{\mathbb{R}^d} p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_k, t_k | \mathbf{x}_{k+1}, t_{k+1}; \dots; \mathbf{x}_n, t_n) \times \\ p(\mathbf{x}_{k+1}, t_{k+1}; \mathbf{x}_{k+2}, t_{k+2}; \dots; \mathbf{x}_n, t_n) d\mathbf{x}_{k+1} \dots d\mathbf{x}_n = p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots; \mathbf{x}_k, t_k).$$

A *Markov process* is one with conditional probability density functions which satisfies the condition

$$p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2; \dots; \mathbf{x}_n, t_n) = p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2). \quad (4.13)$$

That is, it is a process with memory limited to only the preceding step. For a Markov process, the *Chapman-Kolmogorov equation*

$$p(\mathbf{x}_1, t_1; \mathbf{x}_3, t_3) = \int_{\mathbb{R}^d} p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) p(\mathbf{x}_2, t_2 | \mathbf{x}_3, t_3) d\mathbf{x}_2 \quad (4.14)$$

is satisfied. This results directly from making the substitution of (4.13) into (4.12) and integrating.

A *strongly stationary Markov process* is one which is both strongly stationary and Markov:

$$p(\mathbf{x}_{i-1}, t_{i-1} | \mathbf{x}_i, t_i) = p(\mathbf{x}_{i-1}, 0 | \mathbf{x}_i, t_i - t_{i-1}). \quad (4.15)$$

For such a process, there is no reason to “carry around” the zero and the following shorthand notation is sometimes used:

$$p(\mathbf{x}_{i-1} | \mathbf{x}_i, t) \doteq p(\mathbf{x}_{i-1}, 0; \mathbf{x}_i, t).$$

Using this notation for a strongly stationary Markov process, the Chapman–Kolmogorov equation is written as [11, 18]

$$p(\mathbf{x}_1 | \mathbf{x}_3, t) = \int_{\mathbb{R}^d} p(\mathbf{x}_1 | \mathbf{x}_2, s) p(\mathbf{x}_2 | \mathbf{x}_3, t - s) d\mathbf{x}_2 \quad (4.16)$$

for any times  $s < t$ .

Just because a stochastic process is strongly stationary and Markovian does not make it Gaussian, and vice versa. Consider the Gaussian random process defined by the joint pdf in (4.11). It is clear that, since a Gaussian is defined by its mean and covariance, a weakly stationary Gaussian process will also be strongly stationary. And so it suffices to simply refer to them as stationary Gaussian processes (dropping the adjectives “weak” or “strong”). Stationarity of a Gaussian process simply means that  $\Sigma(t_1, \dots, t_n) = \Sigma(t_1 - t_0, \dots, t_n - t_0)$  and  $\boldsymbol{\mu}(t_1, \dots, t_n) = \boldsymbol{\mu}(t_1 - t_0, \dots, t_n - t_0)$  for any  $t_0 < t_n$ .

Under what conditions will a Gaussian process be a Markov process? Recall that marginal densities of Gaussian distributions were discussed in Section 2.2. In particular, the mean and covariance of a conditional Gaussian distribution were given in (2.29). The constraint that the Gaussian process in (4.11) be Markov is (4.13), which can be calculated explicitly using (2.29). Stochastic processes that are simultaneously stationary, Markov, and Gaussian form the foundation for the most common kinds of stochastic differential equations. <sup>i</sup>

## 4.4 Wiener Processes and Stochastic Differential Equations

### 4.4.1 An Informal Introduction

Let  $w_1(t), \dots, w_m(t)$  denote  $m$  independent stochastic processes with the property that for any non-negative real numbers  $s_i$  and  $t_i$  with  $t_i > s_i$  the *increment*  $w_i(t_i) - w_i(s_i)$  for each  $i = 1, \dots, m$  has a zero-mean Gaussian probability density function

$$\rho_i(x_i; s_i, t_i) = \frac{1}{\sqrt{2\pi(t_i - s_i)}} e^{-x_i^2/2(t_i - s_i)}. \quad (4.17)$$

This pdf has variance  $\sigma_i^2 = |t_i - s_i|$ , and  $w_i(t)$  is called a *Wiener process* of strength  $\sigma_i^2$ . The semicolon in the definition (4.17) separates the variable  $x_i$  (that describes the domain on which the pdf is defined) from the variables  $s_i, t_i$  (that describe properties of the pdf). Independence means that  $[w_i(t_i) - w_i(s_i)][w_j(t_j) - w_j(s_j)]$  has the joint distribution

$$\rho_{ij}(x_i, x_j; s_i, t_i, s_j, t_j) = \rho_i(x_i; s_i, t_i) \rho_j(x_j; s_j, t_j), \quad (4.18)$$

and likewise the joint distribution of three or more variables would be a product of three or more univariate distributions.

Let  $A(\mathbf{x})$  be any smooth function of  $\mathbf{x} \in \mathbb{R}^d$ . For example, it can be a scalar function such as  $A(\mathbf{x}) = \mathbf{x} \cdot \mathbf{x}$  or  $A(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ ; it can be a column or row vector function such as

$A(\mathbf{x}) = A_0\mathbf{x}$  or  $A(\mathbf{x}) = \mathbf{x}^T A_0$ ; or it can be a tensor function such as  $A(\mathbf{x}) = \mathbf{x}\mathbf{x}^T$ , just to name a few of the possibilities. Recall that the ensemble average of  $A(\mathbf{x}(t))$ , where  $\mathbf{x}(t) \in \mathbb{R}^d$  is a stochastic process with a corresponding probability density function  $\rho(\mathbf{x}, t)$  (which need not be Gaussian), is defined by the equality

$$\langle A(\mathbf{x}(t)) \rangle = \int_{\mathbb{R}^d} A(\mathbf{x})\rho(\mathbf{x}, t)d\mathbf{x}. \quad (4.19)$$

This statement is true at each value of time. That is,  $\langle A(\mathbf{x}(t)) \rangle$  is an average over many trials, each of which is evaluated at the same point in time in the trial. Clearly, because each ensemble averaging procedure is at a fixed time,

$$\begin{aligned} \int_0^T \langle A(\mathbf{x}(t)) \rangle dt &= \int_0^T \left[ \int_{\mathbb{R}^d} A(\mathbf{x})\rho(\mathbf{x}, t)d\mathbf{x} \right] dt \\ &= \int_{\mathbb{R}^d} \left[ \int_0^T A(\mathbf{x})\rho(\mathbf{x}, t)dt \right] d\mathbf{x} \\ &= \left\langle \int_0^T A(\mathbf{x}(t))dt \right\rangle. \end{aligned} \quad (4.20)$$

Both (4.19) and (4.20) are different from the *ergodic property* (also called *ergodicity*), which is a hypothesis stating that for some physical systems, averages over a spatial ensemble of identical copies of a system and time averages of a single copy over a sufficiently long period,  $[0, T]$ , will yield the same statistical behavior. In other words,  $\mathbf{x}(t)$  is ergodic if

$$\langle A(\mathbf{x}(t)) \rangle = \frac{1}{T} \int_0^T A(\mathbf{x}(t))dt. \quad (4.21)$$

Equations (4.19) and (4.21) should not be confused: (4.19) is a definition that is always valid, and (4.21) is an assumption that needs to be stated.

Now, with (4.17)–(4.19) in mind, it follows that

$$\langle w_i(t_i) - w_i(s_i) \rangle = \int_{-\infty}^{\infty} x_i \rho_i(x_i; s_i, t_i) dx_i = 0.$$

In the special case when  $t_i = s_i + dt_i$ , this can be written as

$$\langle dw_i(t_i) \rangle = 0 \quad (4.22)$$

where  $dw_i(t_i) \doteq w_i(t_i + dt_i) - w_i(t_i)$  and  $dt_i$  is an infinitesimal amount of time.

The *Wiener stochastic integral* is defined as [42]

$$\int_0^t F(\tau)dw(\tau) \doteq \lim_{\Delta t \rightarrow 0} \sum_{n=0}^{1/\Delta t} F(n\Delta t)[w((n+1)\Delta t) - w(n\Delta t)].$$

An immediate consequence of this definition and (4.22) is that the mean value of the integral of a deterministic function  $F(t)$  against  $dw(t)$  is zero:

$$\left\langle \int_0^t F(\tau)dw(\tau) \right\rangle = \int_0^t \langle F(\tau)dw(\tau) \rangle = \int_0^t F(\tau)\langle dw(\tau) \rangle = 0. \quad (4.23)$$

Furthermore, if  $i \neq j$ ,

$$\begin{aligned} \langle [w_i(t_i) - w_i(s_i)][w_j(t_j) - w_j(s_j)] \rangle &= \int_{\mathbb{R}^2} x_i x_j \rho_{ij}(x_i, x_j; s_i, t_i, s_j, t_j) dx_i dx_j \\ &= \left( \int_{-\infty}^{\infty} x_i \rho_i(x_i; s_i, t_i) dx_i \right) \left( \int_{-\infty}^{\infty} x_j \rho_j(x_j; s_j, t_j) dx_j \right) \\ &= 0 \cdot 0 = 0. \end{aligned}$$

In contrast, if  $i = j$ ,

$$\begin{aligned} \langle (w_i(t_i) - w_i(s_i))^2 \rangle &= \int_{\mathbb{R}^2} x_i^2 \rho_{ij}(x_i, x_j; s_i, t_i, s_j, t_j) dx_i dx_j \\ &= \left( \int_{-\infty}^{\infty} x_i^2 \rho_i(x_i; s_i, t_i) dx_i \right) \left( \int_{-\infty}^{\infty} \rho_j(x_j; s_j, t_j) dx_j \right) \\ &= |t_i - s_i| \cdot 1 = |t_i - s_i|. \end{aligned}$$

These are summarized as

$$\boxed{\langle [w_i(t_i) - w_i(s_i)][w_j(t_j) - w_j(s_j)] \rangle = |t_i - s_i| \delta_{ij}.} \quad (4.24)$$

If the definition  $dw(t) = w(t + dt) - w(t)$  is made, then from setting  $t_i = s_i + dt_i$  and  $t_j = s_j + dt_j$  it follows that

$$\boxed{\langle dw_i(t_i) dw_j(t_j) \rangle = dt_i \delta_{ij} = \delta(t_i - t_j) dt_i dt_j \delta_{ij}} \quad (4.25)$$

where the Dirac delta function can be viewed as having a value of  $1/dt$  over the interval  $[0, dt]$ . The property (4.25) indicates that the *increments of a Wiener process* are uncorrelated. The symbol  $dw_i(t)$  is often referred to as (unit strength) *white noise*.<sup>5</sup>

Equation (4.25) can be quite useful. For example, given the deterministic function  $F(t)$ , it can be used to write

$$\left\langle \int_0^t F(\tau) [dw(\tau)]^2 \right\rangle = \int_0^t F(\tau) \langle [dw(\tau)]^2 \rangle = \int_0^t F(\tau) d\tau. \quad (4.26)$$

Equation (4.25) also can be used to simplify the following integral involving the deterministic functions  $F_1(t)$  and  $F_2(t)$ :

$$\begin{aligned} \left\langle \int_0^t F_1(\tau_1) dw(\tau_1) \int_0^t F_2(\tau_2) dw(\tau_2) \right\rangle &= \int_0^t \int_0^t F_1(\tau_1) F_2(\tau_2) \langle dw(\tau_1) dw(\tau_2) \rangle \\ &= \int_0^t \int_0^t F_1(\tau_1) F_2(\tau_2) \delta(\tau_1 - \tau_2) d\tau_1 d\tau_2 \\ &= \int_0^t F_1(\tau_1) F_2(\tau_1) d\tau_1. \end{aligned} \quad (4.27)$$

Sometimes ensemble averages of the form  $\langle dw(t) dt \rangle$  are encountered. But this vanishes because  $\langle dw(t) dt \rangle = \langle dw(t) \rangle dt = 0$ . It is also possible to compute higher-order

<sup>5</sup>In most books  $dw$  is written as  $dW$  (for Wiener) or  $dB$  (for Brownian). Lowercase  $dw$  is used here so as not to confuse it with a matrix quantity.

correlations such as  $\langle dw_i(t_i)dw_j(t_j)dw_k(t_k) \rangle$ . Odd powers in any index will integrate to zero, and those that are even powers in all indices will result in higher powers of  $dt$  that are also effectively equal to zero. Therefore, (4.24) and (4.25) will be most useful in the study of stochastic differential equations, with ensemble averages of all higher-power products in  $dw$  vanishing. In fact, these and other important properties were formalized and abstracted by Norbert Wiener in the early twentieth century (see [42]). This is summarized below.

#### 4.4.2 Abstracted Definitions

The vector  $\mathbf{w}(t) = [w_1, \dots, w_m]^T$  denotes an  $m$ -dimensional *Wiener process* (also called a Brownian motion process) with the following properties. All of the components  $w_j(t)$  have zero ensemble (time) average, are taken to be zero at time zero, and are stationary and independent processes. Denoting an ensemble average as  $\langle \cdot \rangle$ , these properties are written as

$$\begin{aligned} \langle w_j(t) \rangle &= 0 \quad \forall \quad t \geq 0 \\ w_j(0) &= 0 \\ \langle [w_j(t_1 + t) - w_j(t_2 + t)]^2 \rangle &= \langle [w_j(t_1) - w_j(t_2)]^2 \rangle \quad \forall \quad t_1, t_2, t_1 + t, t_2 + t \geq 0 \\ \langle [w(t_i) - w(t_j)][w(t_k) - w(t_l)] \rangle &= 0 \quad \forall \quad t_i > t_j \geq t_k > t_l \geq 0. \end{aligned}$$

From these defining properties, it is clear that for the Wiener process,  $w_j(t)$ ,

$$\begin{aligned} \langle [w_j(t_1 + t_2)]^2 \rangle &= \langle [w_j(t_1 + t_2) - w_j(t_1) + w_j(t_1) - w_j(0)]^2 \rangle \\ &= \langle [w_j(t_1 + t_2) - w_j(t_1)]^2 \rangle + \langle [w_j(t_1) - w_j(0)]^2 \rangle = \langle [w_j(t_1)]^2 \rangle + \langle [w_j(t_2)]^2 \rangle. \end{aligned}$$

For the equality

$$\langle [w_j(t_1 + t_2)]^2 \rangle = \langle [w_j(t_1)]^2 \rangle + \langle [w_j(t_2)]^2 \rangle \quad (4.28)$$

to hold for all values of time  $t_1, t_2$ , it must be the case that [19]

$$\langle [w_j(t - s)]^2 \rangle = \sigma_j^2 |t - s| \quad (4.29)$$

for some positive real number  $\sigma_j^2$ . This together with the absolute value signs ensures that  $\langle [w_j(t - s)]^2 \rangle > 0$ .

The correlation of a scalar-valued Wiener process with itself at two different times  $t$  and  $s$  with  $0 \leq s \leq t$  is calculated as

$$\begin{aligned} \langle w_j(s)w_j(t) \rangle &= \langle w_j(s)(w_j(s) + w_j(t) - w_j(s)) \rangle = \\ &= \langle [w_j(s)]^2 \rangle + \langle (w_j(s) - w_j(0))(w_j(t) - w_j(s)) \rangle = \sigma_j^2 s. \end{aligned}$$

As before, the notation  $dw_j$  is defined by

$$dw_j(t) \doteq w_j(t + dt) - w_j(t). \quad (4.30)$$

Hence, from the definitions and discussion above,

$$\langle dw_j(t) \rangle = \langle w_j(t + dt) \rangle - \langle w_j(t) \rangle = 0$$

and

$$\langle [dw_j(t)]^2 \rangle = \langle (w_j(t + dt) - w_j(t))(w_j(t + dt) - w_j(t)) \rangle$$



$$\begin{aligned}
&= \langle [w_j(t+dt)]^2 \rangle - 2\langle w_j(t)w_j(t+dt) \rangle + \langle [w_j(t)]^2 \rangle \\
&= \sigma_j^2(t+dt-2t+t) = \sigma_j^2 dt.
\end{aligned}$$

For the  $m$ -dimensional Wiener process  $\mathbf{w}(t)$ , each component is uncorrelated with the others for all values of time. This is written together with what we already know from above as

$$\boxed{\langle w_i(s)w_j(t) \rangle = \sigma_j^2 \delta_{ij} \min(s,t) \quad \text{and} \quad \langle dw_i(t_i)dw_j(t_j) \rangle = \sigma_j^2 \delta_{ij} dt_j.} \quad (4.31)$$

The *unit strength* Wiener process has  $\sigma_j^2 = 1$ , corresponding to (4.24) and (4.25). Throughout the presentations in this text, all Wiener processes are taken to have unit strength. This does not cause any loss of generality, because if a non-unit strength is required, it can be achieved simply by multiplying by a scalar strength factor.

## 4.5 The Itô Stochastic Calculus

This section reviews the Itô stochastic calculus, and closely follows the presentations in [11] and [19], which should be consulted in the case that the brief introduction presented here is insufficient.

In the usual calculus, the *Riemann integral* of a continuous function  $f : [a, b] \rightarrow \mathbb{R}$  is obtained as a limit of the form

$$\int_a^b f(x)dx \doteq \lim_{n \rightarrow \infty} \sum_{i=1}^n f(y_i(x_i, x_{i-1}))(x_i - x_{i-1}) \quad (4.32)$$

where

$$a = x_0 < x_1 < x_2 < \dots < x_n = b.$$

Here  $y_i(x_i, x_{i-1})$  is any function such that as  $n$  increases

$$\max_i |x_i - x_{i-1}| \rightarrow 0 \quad \text{and} \quad x_{i-1} \leq y_i(x_i, x_{i-1}) \leq x_i. \quad (4.33)$$

Then the limit in (4.32) will exist. Note that there is some flexibility in how to choose  $y_i(x_i, x_{i-1})$ , and as long as  $n$  is large enough, and the conditions (4.33) are observed, the limit in (4.32) will converge to the one and only answer.<sup>6</sup>

Similarly, given two continuous functions,  $f$  and  $g$ , with  $g$  being monotonically increasing, the *Riemann–Stieltjes integral* can be defined as

$$\int_a^b f(x)dg(x) \doteq \lim_{n \rightarrow \infty} \sum_{i=1}^n f(y_i(x_i, x_{i-1}))(g(x_i) - g(x_{i-1})) \quad (4.34)$$

under the same conditions as before. And this too will converge. Moreover, if  $g(x)$  is continuously differentiable, this can be evaluated as

$$\int_a^b f(x)dg(x) = \int_a^b f(x)g'(x)dx.$$

---

<sup>6</sup>Of course this is a different issue than that of quadrature rules that attempt to rapidly and accurately evaluate functions in a certain class by sampling as few points as possible. References to the literature on numerical quadrature schemes are discussed in [3], and are not the subject here.

However, when it comes to time-dependent stochastic problems, it is desirable to calculate integrals of the form  $\int_a^b f(t)dw(t)$  where the Wiener increments  $dw(t)$  are the discontinuous functions defined in Section 4.4.2. Furthermore,  $f(t)$  also might be a discontinuous function. Nevertheless, for the sorts of discontinuities encountered in stochastic modeling problems, it is still possible to obtain a meaningful answer for  $\int_a^b f(t)dw(t)$ , provided particular rules are followed.

One such rule is that of the Itô stochastic calculus, which defines the following *Itô integral* [11]:<sup>7</sup>

$$\int_{t_0}^t f(\tau)dw(\tau) \doteq \lim_{n \rightarrow \infty} \sum_{i=1}^n f(t_{i-1})[w(t_i) - w(t_{i-1})] \tag{4.35}$$

(where equality is interpreted in the mean-squared sense discussed in Section 1.2.2). Note that this is akin to choosing  $y_i(x_i, x_{i-1}) = x_{i-1}$  in (4.34). However, unlike the case when  $f$  and  $g$  are continuous where any  $y_i(x_i, x_{i-1})$  satisfying  $x_{i-1} \leq y_i(x_i, x_{i-1}) \leq x_i$  will result in the same value of the integral, *a different choice for the rule  $y_i(x_i, x_{i-1})$  will result in a different answer for the value of the integral.* In particular, the seemingly innocuous replacement of  $f(t_{i-1})$  with  $f((t_{i-1} + t_i)/2)$  in (4.35) converts it from being an Itô integral to being a *Stratonovich integral*, the value of which can be different. (The Stratonovich integral and its properties will be discussed in detail in Section 4.6.)

In some stochastic modeling problems, the Itô integral is the most natural. For example, in simulating financial markets, orders are placed based on the information available at the current time, and so evaluating  $f(t_{i-1})$  at the beginning of each time interval makes sense. And the structure of the Itô integral makes it easy to calculate expectations. However, it does have some counterintuitive features. For example, in usual calculus

$$\int_a^b xdx = \frac{1}{2}(b^2 - a^2) \quad \text{and} \quad \int_a^b f(x)df(x) = \frac{1}{2}(f(b)^2 - f(a)^2),$$

and more generally

$$\int_a^b [f(x)]^n df(x) = \frac{1}{n+1}([f(b)]^{n+1} - [f(a)]^{n+1}). \tag{4.36}$$

However, in the Itô stochastic calculus, evaluating (4.35) with  $f(t) = w(t)$ , and using the properties of the Wiener process stated in Section 4.4.2, it can be shown that (see [11] for the detailed derivation)

$$\int_{t_0}^t w(\tau)dw(\tau) = \frac{1}{2}[w(t)^2 - w(t_0)^2 - (t - t_0)].$$

And more generally [11]

$$\int_{t_0}^t [w(\tau)]^n dw(\tau) = \frac{1}{n+1}([w(t)]^{n+1} - [w(t_0)]^{n+1}) - \frac{n}{2} \int_{t_0}^t [w(t)]^{n-1} dt. \tag{4.37}$$

These and other counterintuitive aspects of the Itô stochastic calculus must be kept in mind when using it as a tool in modeling problems.

---

<sup>7</sup>Note that unlike in the discussion in Section 4.3, the times are now ordered as  $t_0 < t_1 < \dots < t_{n-1}$ .

### 4.5.1 Itô Stochastic Differential Equations in $\mathbb{R}^d$

Consider the system of  $d$  SDEs:

$$dx_i(t) = h_i(x_1(t), \dots, x_d(t), t)dt + \sum_{j=1}^m H_{ij}(x_1(t), \dots, x_d(t), t)dw_j(t) \quad \text{for } i = 1, \dots, d. \tag{4.38}$$

If  $H_{ij}(\mathbf{x}, t) \equiv 0$ , and if the vector function with components  $h_i(\mathbf{x})$  satisfies the *Lipschitz condition*,<sup>8</sup> then (4.38) simply becomes the system of (possibly non-linear) ordinary differential equations  $\dot{\mathbf{x}} = \mathbf{h}(\mathbf{x}, t)$ . A unique solution can always be obtained for such a system given any initial conditions. The reason for writing the  $dx_i$  and  $dt$  terms separately in (4.38) is because, strictly speaking, a Wiener process is not differentiable, and so  $dw_j/dt$  is not defined. But, as was discussed in Section 4.4.2, the increments  $dw_j(t)$  have some very well defined properties.

In the more general case when  $H_{ij}(\mathbf{x}, t) \neq 0$  (for at least some values of  $(\mathbf{x}, t)$ ), then (4.38) is called an *Itô SDE* if its solution,

$$x_i(t) - x_i(0) = \int_0^t h_i(x_1(\tau), \dots, x_d(\tau), \tau)d\tau + \sum_{j=1}^m \int_0^t H_{ij}(x_1(\tau), \dots, x_d(\tau), \tau)dw_j(\tau), \tag{4.39}$$

is interpreted as in (4.35). Or, equivalently, the second integral in the above expression is defined to satisfy the condition

$$\lim_{n \rightarrow \infty} \left\langle \left[ \int_0^t H_{ij}(\mathbf{x}(\tau), \tau)dw_j(\tau) - \sum_{k=1}^n H_{ij}(\mathbf{x}(t_{k-1}), t_{k-1})[w_j(t_k) - w_j(t_{k-1})] \right]^2 \right\rangle = 0 \tag{4.40}$$

where  $t_0 = 0 < t_1 < t_2 < \dots < t_n = t$ .

The first integral automatically satisfies

$$\lim_{n \rightarrow \infty} \left\langle \left[ \int_0^t h_i(\mathbf{x}(\tau), \tau)d\tau - \frac{1}{n} \sum_{k=1}^n h_i(\mathbf{x}(t_{k-1}), t_{k-1}) \right]^2 \right\rangle = 0 \tag{4.41}$$

because as long as  $h_i(\mathbf{x}, t)$  is not pathological, the limit can pass through the expectation and the term inside of the brackets becomes zero because

$$\int_0^t h_i(\mathbf{x}(\tau), \tau)d\tau = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n h_i(\mathbf{x}(t_{k-1}), t_{k-1})$$

is the classical *Riemann integral* that is known to hold for continuous integrands. In contrast, this is not true for the term inside of the brackets in (4.40) because the Wiener increments are discontinuous, and the integrand only has meaning when interpreted in the sense of an ensemble average.

---

<sup>8</sup>A vector-valued function  $\mathbf{h}(\mathbf{x})$  is said to be Lipschitz if there exists a finite constant  $c \in \mathbb{R}_{>0}$  such that  $\|\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)\| \leq c \cdot \|\mathbf{x}_1 - \mathbf{x}_2\|$  for all  $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ . Functions that are everywhere differentiable are necessarily Lipschitz.

### 4.5.2 Numerical Approximations

Whereas the above are exact mathematical statements, the goal of simulation is to assign numerically computed values to these integrals. This necessarily involves some level of approximation, because the exact limits as  $n \rightarrow \infty$  cannot be achieved when simulating the behavior of SDEs on a computer.

In numerical practice *sample paths* are generated from  $t = 0$  to a particular end time  $t = T$ , and the values  $t_k$  are taken to be  $t_k = Tk/n$  for a finite value of  $n$ . The resulting numerical approximation to (4.39) evaluated at discrete points in time is

$$\begin{aligned} \hat{x}_i(T) - x_i(0) &= \frac{1}{n} \sum_{k=1}^n h_i(\hat{x}_1(t_{k-1}), \dots, \hat{x}_d(t_{k-1}), t_{k-1}) \\ &+ \sum_{j=1}^m \sum_{k=1}^n H_{ij}(\hat{x}_1(t_{k-1}), \dots, \hat{x}_d(t_{k-1}), t_{k-1}) [w_j(t_k) - w_j(t_{k-1})]. \end{aligned} \quad (4.42)$$

This is the baseline method for numerical stochastic integration. It is called *Euler–Maruyama* integration [12, 22, 26]. The increment  $[w_j(t_k) - w_j(t_{k-1})]$  for each value of  $k$  and  $j$  is drawn randomly from a Gaussian distribution with variance of  $t_k - t_{k-1} = 1/n$ . Or, what is equivalent to this, is to draw samples from a Gaussian distribution with unit variance and then scale the samples by  $1/\sqrt{n}$ .

As  $n$  becomes very large, (4.40) becomes more true in the sense that

$$\lim_{n \rightarrow \infty} \mathcal{E}(n) = 0 \quad \text{where} \quad \mathcal{E}(n) \doteq \frac{1}{n} \sum_{k=1}^n \left\langle [\hat{x}_i(t_k) - x_i(t_k)]^2 \right\rangle \quad (4.43)$$

for each  $i \in [1, \dots, d]$ . Other more sophisticated methods converge faster than Euler–Maruyama integration in the sense that  $\mathcal{E}(n) \rightarrow 0$  more rapidly as  $n$  becomes large. Nevertheless, Euler–Maruyama integration will be the method used throughout this text because of its simplicity. For more sophisticated treatments of numerical methods for stochastic differential equations see [1, 16, 22, 25, 26, 30, 32]. A particularly easy-to-follow presentation is that due to D. Higham [12], which has links to computer code that can be freely downloaded.

An important point to keep in mind when performing stochastic simulations using any integration scheme is that an individual sample path is meaningless. It is only the behavior of an ensemble that has meaning. Therefore, when evaluating the accuracy of a numerical approximation method for computing stochastic integrals, it is only the convergence of the ensemble properties, such as (4.43), that are important.

In practice, not only the end value  $\hat{x}_i(T)$  is of interest, but rather all values  $\hat{x}_i(t_k)$  are, and so (4.42) is calculated along a whole *sample path* using the Euler–Maruyama approach by observing that the increments follow the rule

$$\begin{aligned} \hat{x}_i(t_k) - \hat{x}_i(t_{k-1}) &= \frac{1}{n} h_i(\hat{x}_1(t_{k-1}), \dots, \hat{x}_d(t_{k-1}), t_{k-1}) \\ &+ \sum_{j=1}^m H_{ij}(\hat{x}_1(t_{k-1}), \dots, \hat{x}_d(t_{k-1}), t_{k-1}) [w_j(t_k) - w_j(t_{k-1})], \end{aligned} \quad (4.44)$$

which is basically a localized version of Itô’s rule, and provides a numerical way to evaluate (4.38) at discrete values of time.

This amounts to dividing up the interval  $[0, t]$  into  $n$  subintervals  $[t_{k-1}, t_k]$  for  $k = 1, \dots, n$ , and evaluating  $H_{ij}(\mathbf{x}(t), t)$  at the first point of each interval. This is an important thing to observe.

Figure 4.1 shows six sample paths of a Wiener process over the period of time  $0 \leq t \leq 1$  generated using the Matlab<sup>TM</sup> code provided in [12]. Note that  $w(0) = 0$ , as must be the case by definition. White noise forcing for SDEs is obtained as the difference  $dw(t) = w(t+dt) - w(t)$  at each value of time. This difference, or increment, is simulated numerically by sampling from a Gaussian distribution with unit variance, and then multiplying by  $\sqrt{dt}$ . In the terminology of random variables, a normally distributed (i.e., Gaussian) random variable,  $X$ , with mean  $\mu$  and variance  $\sigma^2$  is denoted as  $X \sim N(\mu, \sigma^2)$ , which can be interpreted as “ $X$  is drawn from  $N(\mu, \sigma^2)$ .” Within this terminology,  $dw(t) \sim \sqrt{dt} \cdot N(0, 1)$ , which means that  $dw(t)$  is computed by first sampling and then scaling by  $\sqrt{dt}$ . Numerical software such as Matlab<sup>TM</sup> have built-in pseudo-random-number generators that perform the sampling step.

When numerically simulating sample paths of SDEs, it is this step of sampling the white noise that is central. After that point, the SDE can be evaluated using the rules of stochastic calculus (Itô or Stratonovich), with the integrals approximated as a finite sum. As with usual numerical integration, various levels of accuracy can be achieved at the expense of greater computational effort. The Euler–Maruyama method is a baseline method that can be quickly implemented and is computationally inexpensive. Other more sophisticated methods can be obtained in the references provided earlier in this subsection. But as will be demonstrated in Section 4.5.6, it is possible to derive deterministic equations for the evolution of probability density that do not require simulation of the SDE.

### 4.5.3 Mathematical Properties of the Itô Integral

Returning now to the “exact” mathematical treatment of SDEs interpreted by Itô’s rule, recall that all equalities are interpreted in the sense of (4.40) being true. In other words, the statement

$$\int_0^t F(\tau) dw_j(\tau) = \lim_{n \rightarrow \infty} \sum_{k=1}^n F(t_{k-1}) [w_j(t_k) - w_j(t_{k-1})] \quad (4.45)$$

is not strictly true. But if we understand this to be shorthand for

$$\lim_{n \rightarrow \infty} \left\langle \left[ \int_0^t F(\tau) dw_j(\tau) - \sum_{k=1}^n F(t_{k-1}) [w_j(t_k) - w_j(t_{k-1})] \right]^2 \right\rangle = 0, \quad (4.46)$$

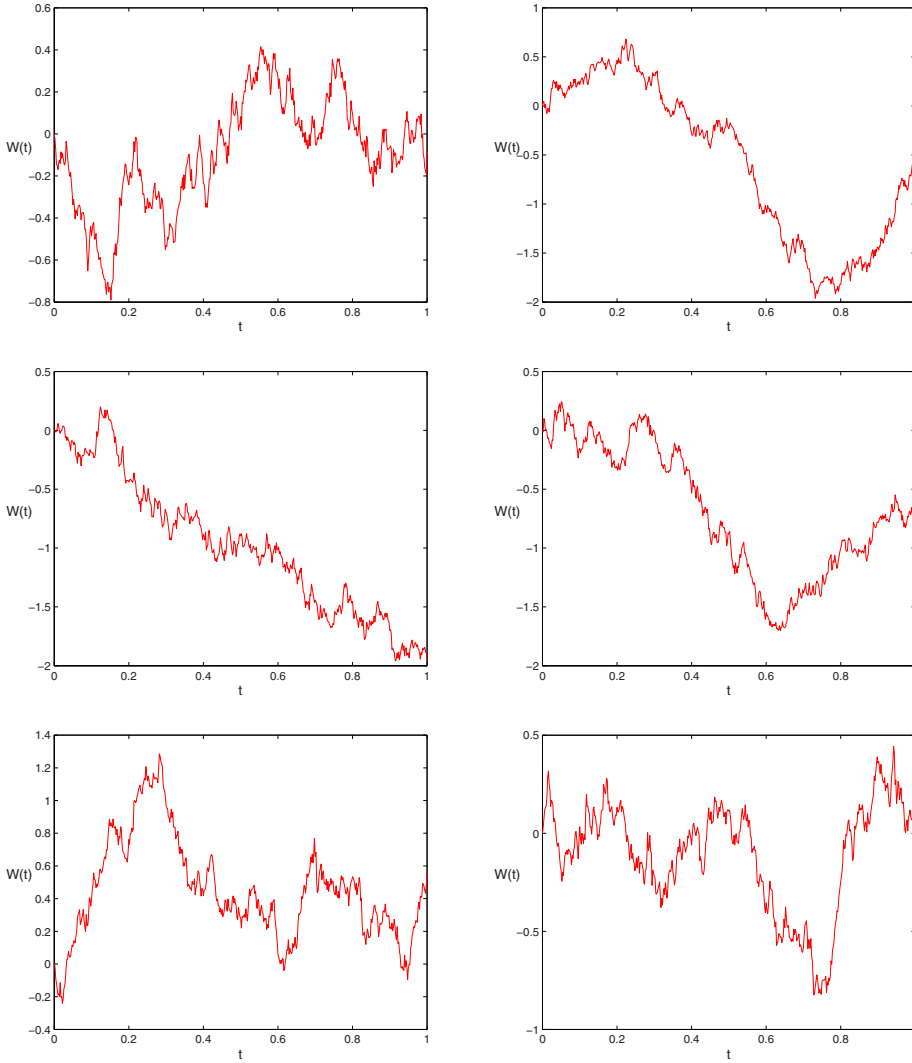
then a number of “equalities” will follow (in the same sense that (4.45) itself is an “equality”).

For example, the following is often stated in books on the Itô calculus:

$$\int_0^t w(\tau) dw(\tau) = \frac{1}{2} \{ [w(\tau)]^2 - t \}. \quad (4.47)$$

Where does this come from? Working backwards, if (4.47) is true, it means nothing more than the statement that

$$\left\langle \left[ \int_0^t w(\tau) dw(\tau) - \frac{1}{2} \{ [w(t)]^2 - t \} \right]^2 \right\rangle = 0,$$



**Fig. 4.1.** Sample Paths of a Wiener Process

or equivalently,

$$\left\langle \left[ \lim_{n \rightarrow \infty} \sum_{k=1}^n w(t_{k-1}) [w_j(t_k) - w_j(t_{k-1})] - \frac{1}{2} \{ [w(t)]^2 - t \} \right]^2 \right\rangle = 0.$$

Expanding the square, and changing the order of summation and expectations, which is acceptable due to (4.20), the result is

$$\left\langle \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \sum_{k=1}^n \sum_{l=1}^m w(t_{k-1})w(t_{l-1})[w_j(t_k) - w_j(t_{k-1})][w_j(t_l) - w_j(t_{l-1})] - \frac{1}{2} \{[w(t)]^2 - t\} \lim_{n \rightarrow \infty} \sum_{k=1}^n w(t_{k-1})[w_j(t_k) - w_j(t_{k-1})] + \frac{1}{4} \{[w(t)]^2 - t\}^2 \right\rangle = 0. \quad (4.48)$$

For an alternative treatment of this calculation, see Gardiner [11, p. 84].

#### 4.5.4 Evaluating Expectations is Convenient for Itô Equations

Using the same sorts of manipulations, it can be shown that for any *non-anticipating functions*<sup>9</sup> [11], integrals analogous to those in (4.23), (4.26), (4.27) (with a stochastic function,  $F(T)$ , replacing a deterministic one) can be written as

$$\left\langle \int_0^t F(\tau) dw(\tau) \right\rangle = \int_0^t \langle F(\tau) dw(\tau) \rangle = \int_0^t \langle F(\tau) \rangle \langle dw(\tau) \rangle = 0 \quad (4.49)$$

$$\left\langle \int_0^t F(\tau) [dw(\tau)]^2 \right\rangle = \int_0^t \langle F(\tau) [dw(\tau)]^2 \rangle = \int_0^t \langle F(\tau) \rangle \langle [dw(\tau)]^2 \rangle = \int_0^t \langle F(\tau) \rangle d\tau \quad (4.50)$$

and

$$\begin{aligned} \left\langle \int_0^t F_1(\tau_1) dw(\tau_1) \int_0^t F_2(\tau_2) dw(\tau_2) \right\rangle &= \int_0^t \int_0^t \langle F_1(\tau_1) F_2(\tau_2) dw(\tau_1) dw(\tau_2) \rangle \\ &= \int_0^t \int_0^t \langle F_1(\tau_1) F_2(\tau_2) \rangle \langle dw(\tau_1) dw(\tau_2) \rangle \\ &= \int_0^t \langle F_1(\tau) F_2(\tau) \rangle d\tau. \end{aligned} \quad (4.51)$$

Again, these “equalities” are only true in the sense of (4.46).

Now consider an Itô stochastic differential equation that generates a random variable  $\mathbf{x}(t)$ . Since this is generated by an Itô integral, the value of  $\mathbf{x}(t)$  depends only on  $\mathbf{x}(t-dt)$  and  $d\mathbf{x}(t)$  since  $\mathbf{x}(t) = \mathbf{x}(t-dt) + [\mathbf{x}(t) - \mathbf{x}(t-dt)]$ . It follows that  $\mathbf{x}(t)$  is a Markov process and so the Chapman–Kolmogorov equation applies. Furthermore, since in the infinitesimally short period of time,  $dt$ , the conditional probability density  $p(\mathbf{x}|\mathbf{y}, dt)$  will be very much like a delta function when  $x = x(t)$  and  $y = x(t-dt)$ . This means that, for example,

$$\langle \mathbf{h}(\mathbf{x}(t)) \rangle = \int_{\mathbb{R}^d} \mathbf{h}(\mathbf{y}) p(\mathbf{x}|\mathbf{y}, dt) d\mathbf{y} = \mathbf{h}(\mathbf{x}).$$

Also,

$$\langle H(\mathbf{x}(t)) d\mathbf{w}(t) \rangle = \left( \int_{\mathbb{R}^d} H(\mathbf{y}) p(\mathbf{x}|\mathbf{y}, dt) d\mathbf{y} \right) \langle d\mathbf{w}(t) \rangle = \mathbf{0}$$

and

<sup>9</sup>A function  $F(t)$  is called non-anticipating if it is statistically independent of  $w(s) - w(t)$  for all  $s > t$ . An immediate consequence is that  $\langle F(t)[w(s) - w(t)] \rangle = \langle F(t) \rangle \langle w(s) - w(t) \rangle$ .

$$\begin{aligned}\langle H_{ij}(\mathbf{x}(t))H_{kl}(\mathbf{x}(t))dw_j(t)dw_l(t) \rangle &= \left\langle \int_{\mathbb{R}^d} H_{ij}(\mathbf{y})H_{kl}(\mathbf{y})p(\mathbf{x}|\mathbf{y}, dt)d\mathbf{y} \right\rangle \langle dw_j(t)dw_l(t) \rangle \\ &= H_{ij}(x_1, \dots, x_d, t)H_{kl}(x_1, \dots, x_d, t)dt.\end{aligned}$$

From these properties and (4.38) the following shorthand can be used:

$$\begin{aligned}\langle dx_i(t) \rangle &= \langle h_i(x_1(t), \dots, x_d(t), t) \rangle dt + \sum_{j=1}^m \langle H_{ij}(x_1(t), \dots, x_d(t), t) \rangle \langle dw_j(t) \rangle \\ &= h_i(x_1, \dots, x_d, t)dt\end{aligned}\tag{4.52}$$

and

$$\begin{aligned}\langle dx_i(t)dx_k(t) \rangle &= \left\langle \left( h_i(x_1(t), \dots, x_d(t), t)dt + \sum_{j=1}^m H_{ij}(x_1(t), \dots, x_d(t), t)dw_j(t) \right) \times \right. \\ &\quad \left. \left( h_k(x_1(t), \dots, x_d(t), t)dt + \sum_{l=1}^m H_{kl}(x_1(t), \dots, x_d(t), t)dw_l(t) \right) \right\rangle \\ &= \sum_{j=1}^m \sum_{l=1}^m H_{ij}(x_1, \dots, x_d, t)H_{kl}(x_1, \dots, x_d, t) \langle dw_j(t)dw_l(t) \rangle \\ &= \sum_{j=1}^m H_{ij}(x_1, \dots, x_d, t)H_{kj}(x_1, \dots, x_d, t)dt.\end{aligned}\tag{4.53}$$

Equations (4.52) and (4.53) are essential in the derivation of the Fokker–Planck equation that will follow shortly.

#### 4.5.5 Itô's Rule

In the usual multivariate calculus, the differential of a vector-valued function of vector argument,  $\mathbf{y} = \mathbf{f}(\mathbf{x})$  is given by  $d\mathbf{y} = D\mathbf{f}d\mathbf{x}$  where the entries of the Jacobian matrix  $D\mathbf{f}$  are  $D\mathbf{f} = \partial f_i / \partial x_j$ . This Jacobian matrix (which is often denoted as  $J$  for convenience) is reviewed in Section 1.4.5. In contrast, when transforming between coordinate systems using the *Itô stochastic calculus*, this no longer applies.

The sample paths,  $\mathbf{x}(t)$ , generated by an SDE are not differentiable, though they are continuous. Given a smooth function  $\mathbf{f}(\mathbf{x})$ , and an increment  $d\mathbf{x}$ , the behavior of which is defined by an SDE, then the quantity  $d\mathbf{y} = \mathbf{f}(\mathbf{x} + d\mathbf{x}) - \mathbf{f}(\mathbf{x})$  can be calculated by expanding  $\mathbf{f}(\mathbf{x} + d\mathbf{x})$  in a Taylor series around  $\mathbf{x}$ . Explicitly in component form this gives

$$dy_i = \sum_j \frac{\partial f_i}{\partial x_j} dx_j + \frac{1}{2} \sum_{k,l} \frac{\partial^2 f_i}{\partial x_k \partial x_l} dx_k dx_l + \text{h.o.t.'s}.\tag{4.54}$$

The higher order terms (h.o.t.'s) are third order and higher in the increments  $dx_i$ . Substituting an SDE of the form (4.38) into (4.54) gives *Itô's rule*:

$$dy_i = \left( \sum_j \frac{\partial f_i}{\partial x_j} h_j(\mathbf{x}, t) + \frac{1}{2} \sum_{k,l} \frac{\partial^2 f_i}{\partial x_k \partial x_l} [H(\mathbf{x}, t)H^T(\mathbf{x}, t)]_{kl} \right) dt + \sum_{k,l} \frac{\partial f_i}{\partial x_k} H_{kl}(\mathbf{x}, t)dw_l.\tag{4.55}$$



The reason why the higher order terms disappear is that the sense of equality used here is that of *equality under expectation*. In other words,  $a = b$  is shorthand for  $\langle ac \rangle = \langle bc \rangle$  for any deterministic  $c$ . And taking expectations using the results of the previous subsection means that all terms that involve third-order and higher powers of  $dw_i$  as well as products such as  $dt dw_i$  will vanish.

#### 4.5.6 The Fokker–Planck Equation (Itô Version)

The goal of this section is to review the derivation of the Fokker–Planck equation, which governs the evolution of the pdf  $f(\mathbf{x}, t)$  for a system of the form in (4.38) which is forced by a Wiener process. The derivation reviewed here has a similar flavor to the arguments used in classical variational calculus (see, for example, [3] or Volume 2) in the sense that functionals of  $f(\mathbf{x})$  and its derivatives,  $m(f(\mathbf{x}), f'(\mathbf{x}), \dots, \mathbf{x})$ , are projected against an “arbitrary” function  $\epsilon(\mathbf{x})$ , and hence integrals of the form

$$\int_{\mathbb{R}^d} m_i(f(\mathbf{x}), f'(\mathbf{x}), \dots, \mathbf{x}) \epsilon(\mathbf{x}) d\mathbf{x} = 0 \quad (4.56)$$

are localized to

$$m(f(\mathbf{x}), f'(\mathbf{x}), \dots, \mathbf{x}) = 0 \quad (4.57)$$

using the “arbitrariness” of the function  $\epsilon(\mathbf{x})$ . The details of this procedure are now examined.

To begin, let  $\mathbf{x} = \mathbf{x}(t)$  and  $\mathbf{y} = \mathbf{x}(t - dt)$  where  $dt$  is an infinitesimal time increment. Using the properties of  $p(\mathbf{x}|\mathbf{y}, dt)$  in (4.52) and (4.53), it follows that

$$\int_{\mathbb{R}^d} (x_i - y_i) p(\mathbf{x}|\mathbf{y}, dt) d\mathbf{y} = \langle x_i - y_i \rangle = h_i(\mathbf{x}, t) dt \quad (4.58)$$

and

$$\int_{\mathbb{R}^d} (x_i - y_i)(x_j - y_j) p(\mathbf{x}|\mathbf{y}, dt) d\mathbf{y} = \langle (x_i - y_i)(x_j - y_j) \rangle = \sum_{k=1}^m H_{ik}(\mathbf{x}, t) H_{kj}^T(\mathbf{x}, t) dt. \quad (4.59)$$

Using the Chapman–Kolmogorov equation, (4.16), together with the definition of partial derivative gives

$$\begin{aligned} \frac{\partial p(\mathbf{x}|\mathbf{y}, t)}{\partial t} &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} [p(\mathbf{x}|\mathbf{y}, t + \Delta t) - p(\mathbf{x}|\mathbf{y}, t)] \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[ \int_{\mathbb{R}^n} p(\mathbf{x}|\boldsymbol{\xi}, t) p(\boldsymbol{\xi}|\mathbf{y}, \Delta t) d\boldsymbol{\xi} - p(\mathbf{x}|\mathbf{y}, t) \right]. \end{aligned}$$

Let  $\epsilon(\mathbf{x})$  be an arbitrary compactly supported function for which  $\partial\epsilon/\partial x_i$  and  $\partial^2\epsilon/\partial x_j\partial x_k$  are continuous for all  $i, j, k = 1, \dots, n$ . Then the projection of  $\partial p/\partial t$  against  $\epsilon(\mathbf{y})$  can be expanded as

$$\begin{aligned} \int_{\mathbb{R}^d} \frac{\partial p(\mathbf{x}|\mathbf{y}, t)}{\partial t} \epsilon(\mathbf{y}) d\mathbf{y} &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left[ \int_{\mathbb{R}^d} \epsilon(\mathbf{y}) d\mathbf{y} \int_{\mathbb{R}^d} p(\mathbf{x}|\boldsymbol{\xi}, t) p(\boldsymbol{\xi}|\mathbf{y}, \Delta t) d\boldsymbol{\xi} \right. \\ &\quad \left. - \int_{\mathbb{R}^n} p(\mathbf{x}|\boldsymbol{\xi}, t) \epsilon(\boldsymbol{\xi}) d\boldsymbol{\xi} \right]. \end{aligned}$$

Inverting the order of integration on the left-hand side results in

$$\int_{\mathbb{R}^d} \frac{\partial p(\mathbf{x}|\mathbf{y}, t)}{\partial t} \epsilon(\mathbf{y}) d\mathbf{y} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{\mathbb{R}^d} p(\mathbf{x}|\boldsymbol{\xi}, t) \left[ \int_{\mathbb{R}^d} p(\boldsymbol{\xi}|\mathbf{y}, \Delta t) \epsilon(\mathbf{y}) d\mathbf{y} - \epsilon(\boldsymbol{\xi}) \right] d\boldsymbol{\xi}.$$

Expanding the function  $\epsilon(\mathbf{y})$  in its Taylor series about  $\boldsymbol{\xi}$ :

$$\epsilon(\mathbf{y}) = \epsilon(\boldsymbol{\xi}) + \sum_{i=1}^d (y_i - \xi_i) \frac{\partial \epsilon}{\partial \xi_i} + \frac{1}{2} \sum_{i,j=1}^d (y_i - \xi_i)(y_j - \xi_j) \frac{\partial^2 \epsilon}{\partial \xi_j \partial \xi_k} + \dots$$

and substituting this series into the previous equation results in

$$\int_{\mathbb{R}^d} \frac{\partial p(\mathbf{x}|\mathbf{y}, t)}{\partial t} \epsilon(\mathbf{y}) d\mathbf{y} = \int_{\mathbb{R}^d} \left[ \sum_{i=1}^n \frac{\partial \epsilon}{\partial y_i} h_i(\mathbf{y}, t) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 \epsilon}{\partial y_i \partial y_j} \sum_{k=1}^m H_{ik} H_{kj}^T \right] p(\mathbf{x}|\mathbf{y}, t) d\mathbf{y}$$

when (4.58) and (4.59) are observed.

The final step is to integrate the two terms on the right-hand side of the above equation by parts to generate

$$\int_{\mathbb{R}^d} \left\{ \frac{\partial p(\mathbf{x}|\mathbf{y}, t)}{\partial t} + \sum_{i=1}^d \frac{\partial}{\partial y_i} (h_i(\mathbf{y}, t) p(\mathbf{x}|\mathbf{y}, t)) - \frac{1}{2} \sum_{k=1}^m \sum_{i,j=1}^d \frac{\partial^2}{\partial y_i \partial y_j} (H_{ik} H_{kj}^T p(\mathbf{x}|\mathbf{y}, t)) \right\} \epsilon(\mathbf{y}) d\mathbf{y} = 0. \quad (4.60)$$

Using the standard localization argument (4.56)  $\implies$  (4.57), and using  $f(\mathbf{x}, t)$  as shorthand for the transition probability  $p(\mathbf{x}|\mathbf{y}, t)$ , the term in braces becomes

$$\boxed{\frac{\partial f(\mathbf{x}, t)}{\partial t} + \sum_{i=1}^d \frac{\partial}{\partial x_i} (h_i(\mathbf{x}, t) f(\mathbf{x}, t)) - \frac{1}{2} \sum_{k=1}^m \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} (H_{ik}(\mathbf{x}, t) H_{kj}^T(\mathbf{x}, t) f(\mathbf{x}, t)) = 0.} \quad (4.61)$$

This can also be written as

$$\frac{\partial f(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^d \frac{\partial}{\partial x_i} (h_i(\mathbf{x}, t) f(\mathbf{x}, t)) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} \left( \sum_{k=1}^m H_{ik}(\mathbf{x}, t) H_{kj}^T(\mathbf{x}, t) f(\mathbf{x}, t) \right), \quad (4.62)$$

or symbolically in vector form (with the dependence of functions on  $\mathbf{x}$  and  $t$  suppressed) as

$$\frac{\partial f}{\partial t} = -\nabla_{\mathbf{x}} \cdot (\mathbf{h}f) + \frac{1}{2} \text{tr} [(\nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^T)(HH^T f)] \quad (4.63)$$

where  $(\nabla_{\mathbf{x}} \nabla_{\mathbf{x}}^T)_{ij} = \partial^2 / \partial x_i \partial x_j$ .

## 4.6 The Stratonovich Stochastic Calculus

The *Stratonovich stochastic integral* is defined as [11, 35]

$$\boxed{\int_{t_0}^t f(\tau) \circ dw(\tau) \doteq \lim_{n \rightarrow \infty} \sum_{i=1}^n f((t_{i-1} + t_i)/2) [w(t_i) - w(t_{i-1})].} \quad (4.64)$$

Here the function  $f(t)$  can be of the form  $f(t) = F(\mathbf{x}(t), t)$  where  $\mathbf{x}(t)$  is governed by a stochastic differential equation which itself is defined by an integral like the one in (4.64).

The inclusion of the symbol  $\circledast$  inside the integral is to distinguish it from the Itô integral, because in general

$$\int_{t_0}^t f(\tau) \circledast dw(\tau) \neq \int_{t_0}^t f(\tau) dw(\tau).$$

Though these two integrals are generally not equal, it is always possible to convert one into the other.

One of the benefits of the Stratonovich calculus is that [11]

$$\int_{t_0}^t [w(\tau)]^n \circledast dw(\tau) = \frac{1}{n+1} ([w(t)]^{n+1} - [w(t_0)]^{n+1}),$$

which, unlike (4.37), is akin to the answer in usual calculus in (4.36). In fact the Stratonovich calculus generally behaves like the usual calculus, which makes it easy to use. Furthermore, due to the inherent continuity of random motions associated with physical problems, the “midpoint” approach in the evaluation of  $f(t)$  in (4.64) is natural.

However, *unlike the Itô integral, the Stratonovich approach has the drawback that it is extremely difficult to evaluate expected values such as was done in the Itô case in (4.52) and (4.53)*. In order to “get the benefit of both worlds” it is important to know how to convert an Itô equation into a Stratonovich equation, and vice versa. When calculus operations are required, conversion from Itô to the Stratonovich form can be performed, and then regular calculus can be used. Or, if expectation operations are required, a Stratonovich equation can be converted to Itô form, and then the expectation can be taken. Being able to weave back and forth between these two forms makes it much easier to address stochastic modeling problems.

Consider the system of  $d$  SDEs:

$$\boxed{dx_i(t) = h_i^s(x_1(t), \dots, x_d(t), t)dt + \sum_{j=1}^m H_{ij}^s(x_1(t), \dots, x_d(t), t) \circledast dw_j(t) \text{ for } i = 1, \dots, d.} \quad (4.65)$$

This is called a *Stratonovich SDE* if its solution is interpreted as the integral

$$x_i(t) - x_i(0) = \int_0^t h_i^s(x_1(\tau), \dots, x_d(\tau), \tau)d\tau + \sum_{j=1}^m \int_0^t H_{ij}^s(x_1(\tau), \dots, x_d(\tau), \tau) \circledast dw_j(\tau). \quad (4.66)$$

In vector form this is written as

$$\mathbf{x}(t) - \mathbf{x}(0) = \int_0^t \mathbf{h}^s(\mathbf{x}(\tau), \tau)d\tau + \int_0^t H^s(\mathbf{x}(\tau), \tau) \circledast d\mathbf{w}(\tau).$$

Note that the coefficient functions  $h_i^s(\mathbf{x}, t)$  and  $H_{ij}^s(\mathbf{x}, t)$  have a superscript “s” in order to distinguish them from the coefficient functions  $h_i(\mathbf{x}, t)$  and  $H_{ij}(\mathbf{x}, t)$  in an Itô SDE.

Now the interconversion between the two forms will be summarized following the arguments in Gardiner [11]. Suppose that corresponding to the Stratonovich SDE (4.65) there is an Itô SDE for  $\mathbf{x}(t)$  defined by drift and diffusion coefficients  $h_i(\mathbf{x}, t)$  and  $H_{ij}(\mathbf{x}, t)$ . With this,  $\mathbf{x}(t)$  can be viewed as the solution to an Itô SDE, and so Itô’s rule

can be used to expand out  $H_{ij}^s(\mathbf{x}((t_{i-1}+t_i)/2), (t_{i-1}+t_i)/2)$  in (4.66) to evaluate the integral according to the rule (4.64). This is because  $\mathbf{x}((t_{i-1}+t_i)/2) \approx \mathbf{x}(t_{i-1}) + \frac{1}{2}d\mathbf{x}(t_{i-1})$  and the  $H_{ij}^s$  is defined to be differentiable in all arguments. Expanding everything out in a multi-dimensional Taylor series and using Itô's rule then establishes the following equivalence between Itô and Stratonovich integrals:

$$\int_0^t H^s(\mathbf{x}(\tau), \tau) \circ d\mathbf{w}(\tau) = \int_0^t H^s(\mathbf{x}(\tau), \tau) d\mathbf{w}(\tau) + \frac{1}{2} \sum_{i=1}^d \mathbf{e}_i \sum_{j=1}^m \sum_{k=1}^d \int_0^t \frac{\partial H_{ij}^s}{\partial x_k} H_{kj} d\tau \quad (4.67)$$

where  $\{\mathbf{e}_i\}$  is the natural basis for  $\mathbb{R}^d$ . This means that if we *choose* to set  $H_{ij} = H_{ij}^s$ , then  $\mathbf{x}(t)$  as defined in the Itô and Stratonovich forms will be equal if the drift terms are chosen appropriately.

In general if  $\{x_1, \dots, x_d\}$  is a set of Cartesian coordinates, given the Stratonovich equation (4.65), the corresponding Itô equation will be (4.38) where

$$h_i(\mathbf{x}, t) = h_i^s(\mathbf{x}, t) + \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^d \frac{\partial H_{ij}^s}{\partial x_k} H_{kj}^s \quad \text{and} \quad H_{ij} = H_{ij}^s. \quad (4.68)$$

This important relationship allows for the conversion between Itô and Stratonovich forms of an SDE. Using it in the reverse direction is trivial once (4.68) is known:

$$h_i^s(\mathbf{x}, t) = h_i(\mathbf{x}, t) - \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^d \frac{\partial H_{ij}}{\partial x_k} H_{kj} \quad \text{and} \quad H_{ij}^s = H_{ij}. \quad (4.69)$$

Starting with the Stratonovich SDE (4.65), and using (4.68) to obtain the equivalent Itô SDE, the Fokker–Planck equation resulting from the derivation of the Itô version can be used as an indirect way of obtaining the Stratonovich version of the Fokker–Planck equation:

$$\frac{\partial f}{\partial t} = - \sum_{i=1}^d \frac{\partial}{\partial x_i} (h_i^s f) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left[ \sum_{k=1}^m H_{ik}^s \frac{\partial}{\partial x_j} (H_{jk}^s f) \right]. \quad (4.70)$$

In the next section, a special kind of SDE is reviewed, which happens to be the same in both the Itô and Stratonovich forms.

## 4.7 Multi-Dimensional Ornstein–Uhlenbeck Processes

Consider a forced mechanical system consisting of a spring, mass, and damper that is governed by the second-order linear differential equation

$$m\ddot{x} + c\dot{x} + kx = f(t). \quad (4.71)$$

Here  $m$  is the mass,  $c$  is the damping constant,  $k$  is the stiffness of the linear spring, and  $f(t)$  is an external forcing function applied to the system. This is a model that is widely used to describe systems such as an automobile with shock absorbers as it passes over a bump in the road (which supplies the forcing), or a civil structure such as bridge or building subjected to forcing supplied by wind or an earthquake.

When  $f(t)$  is a random forcing, this model is also used to describe Brownian motion at the molecular level. In that context,  $f(t)dt = \sigma dw$  is a white noise forcing with strength  $\sigma$  and  $c$  is the damping, both of which are supplied by the surrounding liquid. For extremely small particles, the inertial term become negligible, and (4.71) reduces to

$$dx = -\frac{k}{c}x + \frac{\sigma}{c}dw. \quad (4.72)$$

In the case when the mass is not negligible, the second-order scalar equation in (4.71) can be converted to two first-order *state space* equations by defining  $x_1 \doteq x$  and  $x_2 \doteq \dot{x}$ . Then (4.71) becomes

$$\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = -\begin{pmatrix} 0 & -1 \\ k/m & c/m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dt + \begin{pmatrix} 0 \\ 1/m \end{pmatrix} dw. \quad (4.73)$$

Both (4.72) and (4.73), as well as more complicated models involving multiple springs, masses, and dampers, generalize to the following stochastic differential equation [37, 40]:

$$\boxed{d\mathbf{x} = -\Gamma\mathbf{x}dt + Cdw.} \quad (4.74)$$

Here  $\mathbf{x} \in \mathbb{R}^d$ ,  $\Gamma = [\gamma_{ij}] \in \mathbb{R}^{d \times d}$ ,  $C = [c_{ij}] \in \mathbb{R}^{d \times m}$ , and  $d\mathbf{w} \in \mathbb{R}^m$  is a vector of uncorrelated unit-strength white noises. That is,

$$\langle dw_i(t)dw_j(t) \rangle = \delta_{ij}dt \quad \text{and} \quad \langle dw_i(t_j)dw_i(t_k) \rangle = \delta_{jk}dt.$$

The SDE in (4.74) is called an *Ornstein–Uhlenbeck process*, or *O-U process*. Note that the coefficient matrix function  $H(\mathbf{x}, t) = C$  in this case is constant. *Whenever  $H(\mathbf{x}, t)$  is constant, it turns out that the Itô and Stratonovich SDEs are equivalent.* Therefore, for the O-U process, there is no need to call it an “Itô O-U process” or a “Stratonovich O-U process.” Furthermore, there is only one Fokker–Planck equation.

The Fokker–Planck equation corresponding to (4.74) that describes the evolution of the probability density  $f(\mathbf{x}, t)$  for this process is obtained by substituting (4.74) in (4.61). The result is

$$\frac{\partial f}{\partial t} = \sum_{i,j} \frac{\partial}{\partial x_i} (\gamma_{ij}x_j f) + \frac{1}{2} \sum_{i,j,k} \frac{\partial^2}{\partial x_i \partial x_j} (c_{ik}c_{kj}^T f). \quad (4.75)$$

This equation was originally derived in a special case by Fokker [7] and Planck [29], and was generalized by Kolmogorov [18].

#### 4.7.1 Steady-State Conditions

By defining the matrix  $B = [b_{ij}] = CC^T$ , we can write the Fokker–Planck equation (4.75) in the case when  $\partial f/\partial t \rightarrow 0$  as

$$0 = \sum_{i,j} \gamma_{ij} \frac{\partial}{\partial x_i} (x_j f) + \frac{1}{2} \sum_{i,j} b_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}. \quad (4.76)$$

The first term on the right-hand side of the above equation that multiplies  $\gamma_{ij}$  can be expanded as

$$\frac{\partial}{\partial x_i} (x_j f) = \frac{\partial x_j}{\partial x_i} f + x_j \frac{\partial f}{\partial x_i} = \delta_{ij} f + x_j \frac{\partial f}{\partial x_i}.$$

Here  $\delta_{ij}$  is the Kronecker delta defined to be equal to one when  $i = j$  and zero otherwise.

This means that (4.76) can be rewritten as

$$0 = \sum_{ij} \left[ \gamma_{ij} (\delta_{ij} f + x_j \frac{\partial f}{\partial x_i}) + \frac{1}{2} b_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right].$$

Observing that  $\sum_{ij} \gamma_{ij} \delta_{ij} = \sum_i \gamma_{ii}$  allows this to be written as

$$0 = \text{tr}(\Gamma) f + \sum_{ij} \left( \gamma_{ij} x_j \frac{\partial f}{\partial x_i} + \frac{1}{2} b_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right). \tag{4.77}$$

Following Risken [33], let us assume a steady-state solution of the form of a Gaussian:

$$f(\mathbf{x}) = c_0 \exp\left(-\frac{1}{2} \mathbf{x}^T A \mathbf{x}\right) = c_0 \exp\left(-\frac{1}{2} \sum_{kl} a_{kl} x_k x_l\right) \tag{4.78}$$

where  $A = A^T > 0$  is the inverse of the covariance matrix and  $c_0$  is the normalizing constant such that  $f(\mathbf{x})$  is a probability density function.

In order to check if such a solution is valid, simply substitute it into (4.77) and determine if equality can be made to hold.

First, observe that for this assumed solution,

$$\begin{aligned} \frac{\partial f}{\partial x_i} &= -\frac{1}{2} f(\mathbf{x}) \sum_{kl} a_{kl} \left( \frac{\partial x_k}{\partial x_i} x_l + x_k \frac{\partial x_l}{\partial x_i} \right) \\ &= -\frac{1}{2} f(\mathbf{x}) \sum_{kl} a_{kl} (\delta_{ki} x_l + x_k \delta_{li}) \\ &= -f(\mathbf{x}) \sum_l a_{il} x_l. \end{aligned}$$

Next observe that

$$\begin{aligned} \frac{\partial^2 f}{\partial x_i \partial x_j} &= \frac{\partial}{\partial x_j} \left[ \frac{\partial f}{\partial x_i} \right] \\ &= -\frac{\partial}{\partial x_j} \left[ f(\mathbf{x}) \sum_l a_{il} x_l \right] \\ &= f(\mathbf{x}) \sum_{kl} (a_{jk} x_k) (a_{il} x_l) - f(\mathbf{x}) \sum_l \left( a_{il} \frac{\partial x_l}{\partial x_j} \right). \end{aligned}$$

The last term in the above expression is simplified by observing that

$$\sum_l a_{il} \frac{\partial x_l}{\partial x_j} = \sum_l a_{il} \delta_{lj} = a_{ij}.$$

This means that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = f(\mathbf{x}) \sum_{kl} (a_{jk} a_{il} x_l x_k - a_{ij}).$$

Using these facts, (4.77) can be used to write the condition

$$0 = \sum_i \gamma_{ii} - \sum_{ijl} \gamma_{ij} a_{il} x_j x_l + \frac{1}{2} \sum_{ijkl} b_{ij} (a_{jk} a_{il} x_l x_k - a_{ij}). \quad (4.79)$$

Equating the coefficients at each power of  $x$  to zero results in a sufficient condition for the assumed solution to work. For the zeroth power of  $x$ :

$$\sum_i \gamma_{ii} - \sum_{ij} \frac{1}{2} b_{ij} a_{ij} = 0. \quad (4.80)$$

Matching the quadratic powers in  $x$  gives

$$0 = - \sum_{ijl} \gamma_{ij} a_{il} x_j x_l + \frac{1}{2} \sum_{ijkl} b_{ij} (a_{jk} a_{il} x_l x_k).$$

Note that  $\sum_{jk} b_{ij} a_{jk} a_{il}$  can be written as  $\sum_{jk} a_{il} b_{ij} a_{jk}$  or as  $\sum_{jk} a_{li} b_{ij} a_{jk}$  since  $a_{il} = a_{li}$ . This is the  $lk$ th element of the product  $ABA$ . Recall that  $A = A^T$  and  $B = B^T$ , which means that  $(ABA)^T = ABA$ . In contrast,  $\sum_i \gamma_{ij} a_{il}$  are not the elements of a symmetric matrix. However, by observing that

$$\sum_{ijl} \gamma_{ij} a_{il} x_j x_l = \frac{1}{2} \sum_{ijl} (\gamma_{ij} a_{il} + (\gamma_{ij} a_{il})^T) x_j x_l,$$

the original quantity can be replaced with a symmetric one. Then equating all the coefficients in front of the  $x_j x_l$  terms results in

$$\sum_i [\gamma_{ij} a_{il} + (\gamma_{ij} a_{il})^T] = \sum_{ik} a_{ji} b_{ik} a_{kl}.$$

Written in matrix form this is

$$\Gamma^T A + (\Gamma^T A)^T = ABA.$$

Recognizing that  $(\Gamma^T A)^T = A^T \Gamma = A\Gamma$ , and multiplying on the left and right of both sides of the equation by  $A^{-1}$  gives

$$\boxed{\Gamma A^{-1} + A^{-1} \Gamma^T = B.} \quad (4.81)$$

Hence, the condition that the assumed steady-state solution is valid boils down to solving a linear-algebraic matrix equation. The explicit solution method is discussed in the following subsection.

#### 4.7.2 Steady-State Solution

The solution method presented here follows [33]. The way to solve (4.81) when  $\Gamma$  has distinct eigenvalues is to expand this known matrix in the *spectral decomposition*

$$\Gamma = U \Lambda V^T = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{v}_i^T.$$

Here  $U = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  and  $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$  are matrices such that the columns satisfy the following equations:

$$\Gamma \mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \Gamma^T \mathbf{v}_i = \lambda_i \mathbf{v}_i.$$

In other words, we can write  $\Gamma = U\Lambda U^{-1}$  and  $\Gamma^T = V\Lambda V^{-1}$ . Since this kind of decomposition is unique up to ordering of the eigenvalues and the normalization of the eigenvectors, the equality

$$(U\Lambda U^{-1})^T = V\Lambda V^{-1}$$

can be made to hold when  $U^{-T} = V$ . This means that

$$VU^T = U^T V = \mathbb{I} = UV^T = V^T U. \quad (4.82)$$

$\Lambda$  is a diagonal matrix with entries  $\lambda_i$ , which are the *eigenvalues* of  $\Gamma$  (which are the same as the eigenvalues of  $\Gamma^T$ ). If  $\Gamma = \Gamma^T$  then  $U = V$ , and they are orthogonal, whereas in the general case  $U$  and  $V$  are not orthogonal matrices.

Substituting the spectral decomposition of  $\Gamma$  into (4.81) gives

$$U\Lambda V^T A^{-1} + A^{-1}V\Lambda U^T = B.$$

Multiplying on the left by  $V^T$  and on the right by  $V$  gives

$$V^T U\Lambda V^T A^{-1} V + V^T A^{-1} V\Lambda U^T V = V^T B V.$$

Using (4.82), this reduces to

$$\Lambda V^T A^{-1} V + V^T A^{-1} V \Lambda = V^T B V.$$

If  $C' \doteq V^T A^{-1} V$  and  $B' \doteq V^T B V$ , then the original problem is transformed to one of finding  $C'$  such that

$$\Lambda C' + C' \Lambda = B'$$

where  $C'$  and  $B'$  are symmetric matrices. This problem can be written in component form as

$$\sum_{j=1}^n (\lambda_i \delta_{ij} c'_{jk} + c'_{ij} \lambda_j \delta_{jk}) = b'_{ik}.$$

Using the properties of the Kronecker delta, this reduces to

$$\lambda_i c'_{ik} + c'_{ik} \lambda_k = b'_{ik}.$$

Hence,

$$c'_{ik} = \frac{b'_{ik}}{\lambda_i + \lambda_k}.$$

$A^{-1}$  is then recovered from  $C'$  by observing from (4.82) that

$$UC'U^T = U(V^T A^{-1} V)U^T = A^{-1}.$$

Therefore, as was done in [33], we can write  $A^{-1}$  as

$$A^{-1} = \sum_{i,j} \frac{1}{\lambda_i + \lambda_j} (\mathbf{v}_i^T B \mathbf{v}_j) \mathbf{u}_i \mathbf{u}_j^T. \quad (4.83)$$

This is the covariance for the assumed Gaussian. Note that this steady-state solution washes out any initial conditions. Regardless of whether  $f(\mathbf{x}, 0)$  was initially either more tightly focused or more spread out than this steady-state solution, the O-U process will drive it to become the Gaussian with this covariance. Therefore, the O-U process is not a diffusion process, but rather a *return-to-equilibrium* process.



### 4.7.3 Detailed Balance and the Onsager Relations

The concept of *detailed balance* is a physical argument that reflects in the transition probability for a Fokker–Planck (or Chapman–Kolmogorov) equation a condition that systems of pairwise colliding particles must satisfy. Namely, if elastic particles collide and bounce off of each other in such a way that preserves linear and angular momentum, and if their velocities are tracked before and after the collision, it should be the case that if time is reversed, the collision viewed as time goes backwards must also obey the laws of Newtonian mechanics. This imposes the following condition on the transition probability [2, 8, 9, 11, 28, 39]:

$$p(\mathbf{r}', \mathbf{v}', dt | \mathbf{r}, \mathbf{v}, 0) p_s(\mathbf{r}, \mathbf{v}) = p(\mathbf{r}, -\mathbf{v}, dt | \mathbf{r}', -\mathbf{v}', 0) p_s(\mathbf{r}', -\mathbf{v}'). \quad (4.84)$$

Here  $\mathbf{v} = \dot{\mathbf{r}}$  is velocity and  $\mathbf{r}$  is position.  $p_s(\mathbf{r}, \mathbf{v})$  is a stationary solution to a Fokker–Planck equation (assuming that one exists as  $t \rightarrow \infty$ ) in which the spatial variable is  $\mathbf{x} = (\mathbf{r}^T, \mathbf{v}^T)^T$ , and  $p(\mathbf{r}, \mathbf{v}, t + dt | \mathbf{r}_0, \mathbf{v}_0, t) = p(\mathbf{r}, \mathbf{v}, dt | \mathbf{r}_0, \mathbf{v}_0, 0)$  is the solution to the same Fokker–Planck equation at time  $dt$  with initial conditions  $p(\mathbf{r}, \mathbf{v}, 0 | \mathbf{r}_0, \mathbf{v}_0, 0) = \delta(\mathbf{r} - \mathbf{r}_0) \delta(\mathbf{v} - \mathbf{v}_0)$ .

It is possible to write (4.84) in the equivalent form

$$\boxed{p(\mathbf{x}', dt | \mathbf{x}, 0) p_s(\mathbf{x}) = p(\varepsilon \mathbf{x}, dt | \varepsilon \mathbf{x}', 0) p_s(\varepsilon \mathbf{x}')} \quad (4.85)$$

where the matrix  $\varepsilon = \text{diag}[\varepsilon_1, \dots, \varepsilon_n]$ ,  $n = \dim(\mathbf{x})$ , and  $\varepsilon_i \in \{-1, +1\}$ . A value of  $+1$  corresponds to positional (or “even”) variables, and a value of  $-1$  corresponds to velocity (or “odd”) variables. Note: the terms even/odd need not have anything to do with the evenness/oddness of the subscripts with which the scalar components of the variables are labeled.

If  $p(\mathbf{x}, dt | \mathbf{x}_0, 0)$  satisfies the Fokker–Planck equation (4.61) with drift and diffusion coefficients that do not depend explicitly on time, and if  $p(\mathbf{x}, 0 | \mathbf{x}_0, 0) = \delta(\mathbf{x} - \mathbf{x}_0)$ , then it can be shown that the conditions of detailed balance in (4.85) are equivalent to [11, 33]

$$p_s(\mathbf{x}) = p_s(\varepsilon \mathbf{x}) \quad (4.86)$$

$$[\varepsilon_i h_i(\varepsilon \mathbf{x}) + h_i(\mathbf{x})] p_s(\mathbf{x}) = \sum_j \frac{\partial}{\partial x_j} [H_{ij}(\mathbf{x}) p_s(\mathbf{x})] \quad (4.87)$$

$$\varepsilon_i \varepsilon_j H_{ij}(\varepsilon \mathbf{x}) = H_{ij}(\mathbf{x}). \quad (4.88)$$

These are somewhat more convenient than (4.85) because in many situations  $p(\mathbf{x}, dt | \mathbf{x}_0, 0)$  is not known in closed form but  $p_s(\mathbf{x})$ ,  $h_i(\mathbf{x})$ , and  $H_{ij}(\mathbf{x})$  are.

Condition (4.86) follows from (4.85) because for  $dt = 0$ ,

$$p(\mathbf{x}', 0 | \mathbf{x}, 0) = \delta(\mathbf{x}' - \mathbf{x}) = \delta(\varepsilon \mathbf{x}' - \varepsilon \mathbf{x}).$$

Conditions (4.87) and (4.88) can be obtained by expanding  $p(\mathbf{x}, dt | \mathbf{x}_0, 0)$  in a Taylor series in  $dt$  and using the Chapman–Kolmogorov equation, in analogy with what was done in the derivation of the Fokker–Planck equation. See [2, 11, 28, 39] for details.

The condition of detailed balance in (4.85) has been generalized to other, more abstract, Markov processes, but the discussion here is restricted to physical arguments. As a concrete example, consider the Fokker–Planck equation corresponding to the Ornstein–Uhlenbeck process in (4.74) is (4.75). It is clear that if it originates from

a mechanical system such as (4.73), there will be an equal number of even and odd variables. Furthermore, the steady-state Gaussian solution (4.78) with  $A$  calculated in (4.83) should satisfy (4.86)–(4.88). For the Ornstein–Uhlenbeck process, these conditions respectively correspond to

$$\varepsilon A \varepsilon = A; \quad \varepsilon \Gamma \varepsilon + \Gamma = B A^{-1}; \quad \varepsilon B \varepsilon = B.$$

Combining these with (4.81) and rearranging terms gives the *Onsager relations* [2, 28, 39]:

$$\boxed{\varepsilon(\Gamma A^{-1}) = (\Gamma A^{-1})^T \varepsilon.} \quad (4.89)$$

As an example that demonstrates the usefulness of (4.89), consider the multi-dimensional version of (4.71),

$$M \ddot{\mathbf{x}} + C \dot{\mathbf{x}} + K \mathbf{x} = \mathbf{f}(t), \quad (4.90)$$

where  $M$ ,  $C$ , and  $K$  are all symmetric positive definite  $n \times n$  matrices and  $\mathbf{f} dt = S d\mathbf{w}$  is a stochastic forcing vector where  $S \in \mathbb{R}^{n \times n}$  is arbitrary and  $d\mathbf{w}$  is a vector, each element of which is a unit-strength white noise that is uncorrelated with the others.

It is possible to write (4.90) in the form (4.74) by introducing a new variable  $\mathbf{y} = [\mathbf{x}^T, \dot{\mathbf{x}}^T]^T$ . Alternatively, in physical applications it is more common to use coordinates  $\mathbf{z} = [\mathbf{x}^T, \mathbf{p}^T]^T$  where  $\mathbf{p} = M \dot{\mathbf{x}}$ . Whereas  $\mathbf{y}$  is called a *state-space* variable,  $\mathbf{z}$  is called a *phase-space* variable. It is also possible to use other variables such as  $\mathbf{p}' = M^{\frac{1}{2}} \dot{\mathbf{x}}$ . Here the phase-space formulation will be used, and the result will be of the form in (4.74) with  $\mathbf{z}$  taking the place of  $\mathbf{x}$ . In this case, the matrices  $\Gamma$  and  $B$  are

$$\Gamma = \begin{bmatrix} \mathbb{O} & -M^{-1} \\ K & CM^{-1} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \mathbb{O} & \mathbb{O} \\ \mathbb{O} & SS^T \end{bmatrix}.$$

Writing the candidate matrix  $A$  and its inverse, which can be taken as being symmetric without loss of generality, in terms of blocks as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \quad \text{and} \quad A^{-1} = \begin{bmatrix} A'_{11} & A'_{12} \\ A'_{12}^T & A'_{22} \end{bmatrix},$$

the conditions (4.81) are expressed block-by-block as

$$\begin{aligned} M^{-1} A'_{12}{}^T &= -A'_{12} M^{-1} \\ A'_{11} K + A'_{12} M^{-1} C &= M^{-1} A'_{22} \\ K A'_{12} + A'_{12}{}^T K + A'_{22} M^{-1} C + C M^{-1} A'_{22} &= SS^T. \end{aligned}$$

The condition  $\varepsilon A \varepsilon = A$  (or equivalently  $\varepsilon A^{-1} \varepsilon = A^{-1}$ ) gives  $A_{12} = A'_{12} = 0$ . This simplifies the last two of the above equations to

$$A'_{11} K = M^{-1} A'_{22} \quad \text{and} \quad A'_{22} M^{-1} C + C M^{-1} A'_{22} = SS^T.$$

The Onsager relations (4.89) written out in block-matrix form also give  $A'_{11} K = M^{-1} A'_{22}$ , but in addition give

$$C M^{-1} A'_{22} = A'_{22} M^{-1} C.$$

Combining these equations gives

$$A'_{11}CK^{-1} + K^{-1}CA'_{11} = K^{-1}SS^TK^{-1} \quad \text{and} \quad A'_{22}M^{-1}C + CM^{-1}A'_{22} = SS^T. \quad (4.91)$$

And so in this particular problem the Onsager relations provide a tool for converting a system of matrix equations (4.81) of dimension  $2n \times 2n$  into two matrix equations of the same kind, each of dimension  $n \times n$ . Since full-rank linear systems of equations have unique solutions, any solution to these equations will be “the” solution. By inspection, it is clear that given some scalar constant  $\beta$ ,

$$2C = \beta SS^T \quad \iff \quad A = A_{11} \oplus A_{22} = \beta K \oplus M^{-1}. \quad (4.92)$$

The condition on the left indicates that viscous/dissipative forces and stochastic fluctuations forcing the system are balanced in a particular way. This is a statement of the *fluctuation–dissipation theorem* which will be revisited in the context of statistical mechanics in Volume 2.

## 4.8 SDEs and Fokker–Planck Equations Under Coordinate Changes

The purpose of this section is to address problems associated with changing coordinate systems in stochastic modeling problems. This sort of geometric problem will be unavoidable when considering SDEs that describe processes that evolve on (possibly high-dimensional) surfaces rather than unconstrained translational motion in  $\mathbb{R}^n$ . But even when modeling problems in  $\mathbb{R}^n$ , geometric issues will arise. The general problems associated with coordinate changes will be formalized later in this section, but they are first illustrated here with the concrete example of Brownian motion in the plane.

### 4.8.1 Brownian Motion in the Plane

From the presentation earlier in this chapter, it should be clear that the following two-dimensional SDE and Fokker–Planck equation describe the same process:

$$dx = dw \quad \iff \quad \frac{\partial f}{\partial t} = \frac{1}{2} \left( \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} \right).$$

In this case, it does not matter if the SDE is interpreted as an Itô or Stratonovich equation. The above equations describe isotropic translational diffusion in the plane.

As a physical problem, the behavior should be independent of the coordinate system used. Therefore, if instead of Cartesian coordinates, a change of variables  $x_1 = r \cos \phi$  and  $x_2 = r \sin \phi$  is made, it should be possible to describe the same process in terms of SDEs and Fokker–Planck equations in the polar coordinates  $(r, \phi)$ . Since there is no ambiguity in how to do this change of coordinates for the Fokker–Planck equation (since the usual Newton–Leibniz calculus is well understood by all), this is a good place to begin.

### Coordinate Changes and the Fokker–Planck Equation

Let  $\tilde{f}(r, \phi; t) = f(r \cos \phi, r \sin \phi; t)$ . Then it is clear from the classical chain rule that

$$\frac{\partial \tilde{f}}{\partial r} = \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial r} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial r}$$

and

$$\frac{\partial \tilde{f}}{\partial \phi} = \frac{\partial f}{\partial x_1} \frac{\partial x_1}{\partial \phi} + \frac{\partial f}{\partial x_2} \frac{\partial x_2}{\partial \phi}.$$

If the Jacobian of the coordinate change is calculated as

$$J(r, \phi) = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \phi} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix},$$

then the Jacobian determinant is  $|J| = r$ .

It is clear from the above equations that

$$\begin{pmatrix} \frac{\partial \tilde{f}}{\partial r} \\ \frac{\partial \tilde{f}}{\partial \phi} \end{pmatrix} = J^T(r, \phi) \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = J^{-T}(r, \phi) \begin{pmatrix} \frac{\partial \tilde{f}}{\partial r} \\ \frac{\partial \tilde{f}}{\partial \phi} \end{pmatrix}$$

where  $J^{-T} = (J^{-1})^T = (J^T)^{-1}$ .

In component form this means that

$$\frac{\partial f}{\partial x_1} = \cos \phi \frac{\partial \tilde{f}}{\partial r} - \frac{\sin \phi}{r} \frac{\partial \tilde{f}}{\partial \phi}$$

and

$$\frac{\partial f}{\partial x_2} = \sin \phi \frac{\partial \tilde{f}}{\partial r} + \frac{\cos \phi}{r} \frac{\partial \tilde{f}}{\partial \phi}.$$

Applying this rule twice,

$$\begin{aligned} \frac{\partial^2 f}{\partial x_1^2} &= \cos \phi \frac{\partial}{\partial r} \left( \cos \phi \frac{\partial \tilde{f}}{\partial r} - \frac{\sin \phi}{r} \frac{\partial \tilde{f}}{\partial \phi} \right) - \frac{\sin \phi}{r} \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\partial \tilde{f}}{\partial r} - \frac{\sin \phi}{r} \frac{\partial \tilde{f}}{\partial \phi} \right) \\ &= \cos^2 \phi \frac{\partial^2 \tilde{f}}{\partial r^2} - \sin \phi \cos \phi \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial \tilde{f}}{\partial \phi} \right) + \frac{\sin^2 \phi}{r} \frac{\partial \tilde{f}}{\partial r} - \\ &\quad \frac{\sin \phi \cos \phi}{r} \frac{\partial^2 \tilde{f}}{\partial \phi \partial r} + \frac{\sin \phi \cos \phi}{r^2} \frac{\partial \tilde{f}}{\partial \phi} + \frac{\sin^2 \phi}{r^2} \frac{\partial^2 \tilde{f}}{\partial \phi^2} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 f}{\partial x_2^2} &= \sin \phi \frac{\partial}{\partial r} \left( \sin \phi \frac{\partial \tilde{f}}{\partial r} + \frac{\cos \phi}{r} \frac{\partial \tilde{f}}{\partial \phi} \right) + \frac{\cos \phi}{r} \frac{\partial}{\partial \phi} \left( \sin \phi \frac{\partial \tilde{f}}{\partial r} + \frac{\cos \phi}{r} \frac{\partial \tilde{f}}{\partial \phi} \right) \\ &= \sin^2 \phi \frac{\partial^2 \tilde{f}}{\partial r^2} + \sin \phi \cos \phi \frac{\partial}{\partial r} \left( \frac{1}{r} \frac{\partial \tilde{f}}{\partial \phi} \right) + \frac{\cos^2 \phi}{r} \frac{\partial \tilde{f}}{\partial r} + \\ &\quad \frac{\sin \phi \cos \phi}{r} \frac{\partial^2 \tilde{f}}{\partial \phi \partial r} - \frac{\sin \phi \cos \phi}{r^2} \frac{\partial \tilde{f}}{\partial \phi} + \frac{\cos^2 \phi}{r^2} \frac{\partial^2 \tilde{f}}{\partial \phi^2}. \end{aligned}$$

Therefore,

$$\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} = \frac{\partial^2 \tilde{f}}{\partial r^2} + \frac{1}{r} \frac{\partial \tilde{f}}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \tilde{f}}{\partial \phi^2},$$

and so

$$\frac{\partial f}{\partial t} = \frac{1}{2} \left( \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} \right) \iff \frac{\partial \tilde{f}}{\partial t} = \frac{1}{2} \left( \frac{\partial^2 \tilde{f}}{\partial r^2} + \frac{1}{r} \frac{\partial \tilde{f}}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \tilde{f}}{\partial \phi^2} \right). \quad (4.93)$$

The next question is, if  $d\mathbf{x} = d\mathbf{w}$  is interpreted as a Stratonovich or Itô SDE, what will the corresponding SDEs in polar coordinates look like?

### Coordinate Conversion and the Stratonovich SDE

The Stratonovich case is straightforward, since it obeys the usual Newton–Leibniz calculus, and so  $d\mathbf{x} = J(r, \phi)[dr, d\phi]^T$ . This then means that  $[dr, d\phi]^T = J^{-1}(r, \phi)d\mathbf{w}$ , which is written in component form as

$$dr = \cos \phi \circledast dw_1 + \sin \phi \circledast dw_2 \quad (4.94)$$

$$d\phi = -\frac{\sin \phi}{r} \circledast dw_1 + \frac{1}{r} \cos \phi \circledast dw_2.$$

### Coordinate Conversion and the Itô SDE (Approach 1)

How can the corresponding Itô equation in polar coordinates be found? First, from *Itô's rule* in (4.55) and the functional relationship between Cartesian and polar coordinates, it follows that

$$dx_1 = \cos \phi dr - r \sin \phi d\phi - \frac{1}{2} r \cos \phi (d\phi)^2 = dw_1 \quad (4.95)$$

$$dx_2 = \sin \phi dr + r \cos \phi d\phi - \frac{1}{2} r \sin \phi (d\phi)^2 = dw_2$$

where the rightmost equalities in each of the above come from the original SDE  $d\mathbf{x} = d\mathbf{w}$ .

Second, the form of the Itô SDE that is sought is, by definition, of the form

$$\begin{aligned} dr &= h_1(r, \phi)dt + H_{11}(r, \phi)dw_1 + H_{12}(r, \phi)dw_2 \\ d\phi &= h_2(r, \phi)dt + H_{21}(r, \phi)dw_1 + H_{22}(r, \phi)dw_2. \end{aligned}$$

Substituting this into (4.95), and remembering that under ensemble averaging,  $dw_i dw_j = \delta_{ij} dt$  and all higher order terms such as  $dt dw_i$  and  $(dt)^2$  vanish, leads to

$$\begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} (H_{21}^2 + H_{22}^2)r \\ 0 \end{pmatrix} dt.$$

This will be satisfied if  $H_{12} = H_{21} = h_2 = 0$  and  $H_{11} = 1$ ,  $H_{22} = 1/r$ , and  $h_1 = 1/(2r)$ . In other words, an Itô equation in polar coordinates that produces sample paths equivalent under ensemble averaging to those generated by the Cartesian Itô SDE  $d\mathbf{x} = d\mathbf{w}$  is

$$\begin{pmatrix} dr \\ d\phi \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1/r \\ 0 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & 1/r \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}. \quad (4.96)$$

**Coordinate Conversion and the Itô SDE (Approach 2)**

This same problem can be approached in a different way. Inverting the transformation of coordinates so that polar coordinates are written in terms of Cartesian coordinates,

$$r = [x_1^2 + x_2^2]^{\frac{1}{2}} \quad \text{and} \quad \phi = \tan^{-1} \left( \frac{x_2}{x_1} \right).$$

It follows that

$$dr = [(x_1 + dx_1)^2 + (x_2 + dx_2)^2]^{\frac{1}{2}} - [x_1^2 + x_2^2]^{\frac{1}{2}}$$

and

$$d\phi = \tan^{-1} \left( \frac{x_2 + dx_2}{x_1 + dx_1} \right) - \tan^{-1} \left( \frac{x_2}{x_1} \right).$$

Expanding the above in a Taylor series to second order in  $dx_i$  (knowing that higher order terms will vanish) gives

$$dr = \frac{1}{2} \frac{[2x_1 dx_1 + (dx_1)^2 + 2x_2 dx_2 + (dx_2)^2]}{[x_1^2 + x_2^2]^{\frac{1}{2}}} - \frac{1}{8} \frac{[4x_1^2(dx_1)^2 + 4x_2^2(dx_2)^2]}{[x_1^2 + x_2^2]^{\frac{3}{2}}}$$

and

$$d\phi = \frac{x_1 dx_2 - x_2 dx_1 + \frac{x_2}{x_1}(dx_1)^2}{x_1^2 + x_2^2} - \frac{x_2 x_1^3 (x_1^{-1}(dx_2)^2 + x_2^2 x_1^{-4}(dx_1)^2)}{(x_1^2 + x_2^2)^2}.$$

Now making the substitutions  $x_1 = r \cos \phi$ ,  $x_2 = r \sin \phi$ ,  $dx_1 = dw_1$ ,  $dx_2 = dw_2$ , and using the usual properties of the Wiener process, this reduces (after some trigonometric simplifications) to

$$dr = \frac{1}{2} r^{-1} dt + \cos \phi dw_1 + \sin \phi dw_2 \tag{4.97}$$

$$d\phi = -r^{-1} \sin \phi dw_1 + r^{-1} \cos \phi dw_2.$$

While (4.96) and (4.97) are not exactly *equal*, they are *equivalent* in the sense that the ensemble of paths generated by both will have the same statistics.

**It Doesn't Matter that These Equations are Different**

At first glance, it may be a source of concern that (4.97) and (4.96) are not the same. After all, it is reasonable to assume that they should be! But referring back to the Fokker–Planck equation in Cartesian coordinates (both in this case, and in the general case in (4.61)), it becomes clear that any two Itô SDEs are *equivalent* if  $d\mathbf{w} \rightarrow R d\mathbf{w}$  where  $R$  is an orthogonal matrix<sup>10</sup> that can be time dependent, and even dependent on the stochastic process defined by the SDE itself. This is exactly the case here, since the substitution  $d\mathbf{w} \rightarrow R d\mathbf{w}$  in (4.97) with

$$R = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

---

<sup>10</sup>Recall that  $R \in \mathbb{R}^{n \times n}$  is called orthogonal if  $RR^T = I$ , and an orthogonal matrix with the additional condition  $\det R = +1$  is called a rotation, or special orthogonal, matrix. The set of all  $n \times n$  rotation matrices is denoted as  $SO(n)$ .

will convert (4.97) to (4.96). Clearly  $RR^T = I$ , and so these two Itô SDEs are equivalent.

Now things can get a little confusing, because the Itô equation (4.97), which is the same as (4.96), and the Itô equation (4.95) are equivalent in the sense that they produce the same Fokker–Planck equation. Moreover, the Stratonovich equation (4.94) is equivalent to these because it too produces the same Fokker–Planck equation.

#### 4.8.2 General Conversion Rules

Formulas were given in Section 4.6 for converting between Itô and Stratonovich versions of the same underlying process described in Cartesian coordinates. The same rules hold for this conversion in curvilinear coordinates.

In general if  $\{q_1, \dots, q_d\}$  is a set of generalized coordinates, given the Stratonovich equation

$$dq_i = h_i^s(\mathbf{q}, t)dt + \sum_{j=1}^m H_{ij}^s(\mathbf{q}, t) \circ dw_j$$

for  $i = 1, \dots, d$  the corresponding Itô equation will be

$$dq_i = h_i(\mathbf{q}, t)dt + \sum_{j=1}^m H_{ij}(\mathbf{q}, t)dw_j$$

where

$$h_i(\mathbf{q}, t) = h_i^s(\mathbf{q}, t) + \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^d \frac{\partial H_{ij}^s}{\partial q_k} H_{kj}^s. \quad (4.98)$$

In the above example of Brownian motion in the plane, the Stratonovich equation (4.94) has no drift, and the corresponding Itô equation (4.97) does have a drift, which is consistent with  $h_i(\mathbf{q}, t) \neq h_i^s(\mathbf{q}, t)$ .

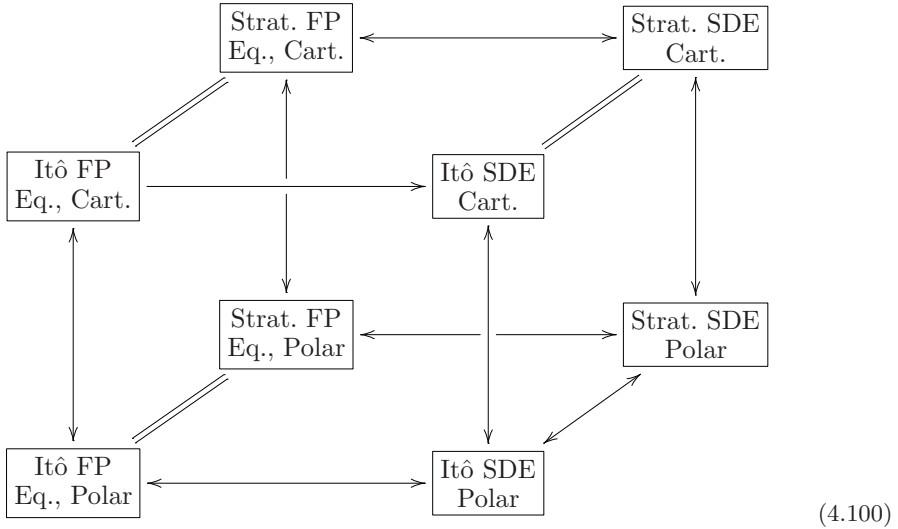
Now consider the Stratonovich equivalent of the Itô equation (4.96). Using (4.98), it becomes clear that

$$\begin{pmatrix} dr \\ d\phi \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1/r \\ 0 \end{pmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & 1/r \end{pmatrix} \circ \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}. \quad (4.99)$$

An important observation can be made from this example: *If for any Itô equation the transformation  $H(\mathbf{q}, t) \rightarrow H(\mathbf{q}, t)R(\mathbf{q}, t)$  is made for any  $R \in SO(m)$  while leaving the drift term the same, the resulting Fokker–Planck equation computed with  $H(\mathbf{q}, t)$  and  $H'(\mathbf{q}, t) = H(\mathbf{q}, t)R(\mathbf{q}, t)$  will be the same. However, this is generally not true for Stratonovich equations.* This is observed in the context of the current example because the coloring matrices,  $H$ , in (4.99) and (4.94) are related by an orthogonal transformation, and in order for them to be produce the same Fokker–Planck equation, they necessarily have different drift terms. The three-dimensional diagram in (4.100) illustrates which equations in this example are equal, and which are equivalent under the conversion rules established in Section 4.6.

#### 4.8.3 Coordinate Changes and Fokker–Planck Equations

The coordinate changes addressed above are for SDEs. When performing coordinate changes for Fokker–Planck equations, the usual calculus is used. Using Cartesian coordinates as the baseline, small changes in  $\mathbf{x}$  are related to small changes in  $\mathbf{q}$  by the



Jacobian matrix:  $d\mathbf{x} = J(\mathbf{q})d\mathbf{q}$ . Differential volume elements described in the two coordinate systems are related by the expression  $d(\mathbf{x}) = |J(\mathbf{q})|d(\mathbf{q})$ , as discussed in Chapter 1. In Chapter 5, it will be shown that the matrix  $G(\mathbf{q}) \doteq J^T(\mathbf{q})J(\mathbf{q})$  (called the metric tensor) contains all of the information needed to measure distances, areas, volumes, etc. Since  $|J(\mathbf{q})| = |G(\mathbf{q})|^{\frac{1}{2}}$ , the volume element in curvilinear coordinates can be expressed as  $d(\mathbf{x}) = |G(\mathbf{q})|^{\frac{1}{2}}d(\mathbf{q})$ . This has an impact on the form of the Fokker–Planck equation in curvilinear coordinates because every “ $d\mathbf{x}$ ” in the derivation in Section 4.5.6 (which in that context is shorthand for  $d(\mathbf{x})$ ) becomes a  $|G(\mathbf{q})|^{\frac{1}{2}}d(\mathbf{q})$ . And when performing integration by parts and localizing, the factor  $|G(\mathbf{q})|^{\frac{1}{2}}$  is introduced into the curvilinear version of the Fokker–Planck equation. The result is presented without proof below. A proof for the more general case of Fokker–Planck equations on manifolds will be provided in Chapter 8. That proof covers the case of curvilinear coordinates in Euclidean space as well. For the reader who wishes to prove these formulas, another route would be to simply start with the Cartesian forms of the general Fokker–Planck equations and work through the change of coordinates as was done early in this section for the specific example of Brownian motion in the plane.

**Itô Version**

The Itô version of the Fokker–Planck equation in generalized coordinates is

$$\frac{\partial f}{\partial t} = -|G|^{-\frac{1}{2}} \sum_i \frac{\partial}{\partial q_i} \left( a_i |G|^{\frac{1}{2}} f \right) + \frac{1}{2} |G|^{-\frac{1}{2}} \sum_{i,j} \frac{\partial^2}{\partial q_i \partial q_j} \left[ (BB^T)_{ij} |G|^{\frac{1}{2}} f \right]. \quad (4.101)$$

Given  $f(\mathbf{q}, 0)$  this generates  $f(\mathbf{q}, t)$  for the Itô SDE

$$d\mathbf{q} = \mathbf{a}(\mathbf{q}, t) + B(\mathbf{q}, t)dw.$$

As illustrated in Exercise 4.10, when  $B(\mathbf{q}, t) = [J(\mathbf{q})]^{-1}$ , (4.101) will be the heat equation under special conditions on  $\mathbf{a}(\mathbf{q}, t)$ .



**Stratonovich Version**

$$\frac{\partial f}{\partial t} = -|G|^{-\frac{1}{2}} \sum_i \frac{\partial}{\partial q_i} \left( a_i^s |G|^{\frac{1}{2}} f \right) + \frac{1}{2} |G|^{-\frac{1}{2}} \sum_{i,j,k} \frac{\partial}{\partial q_i} \left[ B_{ik}^s \frac{\partial}{\partial q_j} (B_{jk}^s |G|^{\frac{1}{2}} f) \right]. \quad (4.102)$$

Given  $f(\mathbf{q}, 0)$  this generates  $f(\mathbf{q}, t)$  for the Stratonovich SDE

$$d\mathbf{q} = \mathbf{a}^s(\mathbf{q}, t) + B^s(\mathbf{q}, t) \circ d\mathbf{w}.$$

As illustrated in Exercise 4.11, when  $B^s(\mathbf{q}, t) = [J(\mathbf{q})]^{-1}$ , (4.102) will be the heat equation under special conditions on  $\mathbf{a}^s(\mathbf{q}, t)$  (which are in general different than the conditions in the Itô case).

**4.9 Chapter Summary**

This chapter introduced concepts from the theory of random (stochastic) processes. Two interpretations of the stochastic integral were reviewed: Itô and Stratonovich. Each has advantages and disadvantages. The Itô calculus is convenient for taking expectations, but does not follow the rules of classical calculus. The Stratonovich calculus follows the rules of classical calculus, but is very difficult to work with when taking expectations.

It is the stochastic integral that can be viewed as the solution of a stochastic differential equation. These two different interpretations mean that, in general, it must be stated up front which kind of SDE is being considered. Rules for converting between Itô and Stratonovich forms were reviewed, as well as the conversion of an SDE of one type into an SDE of the same type, but in curvilinear rather than Cartesian coordinates. In addition, each kind of SDE has a corresponding Fokker–Planck equation. The relationship between all of these concepts is summarized in the cubic diagram (4.100) presented in this chapter for the special case of Cartesian and polar coordinates in the plane.

Many books on stochastic processes exist. These either focus on modeling of physical systems, such as [15, 24, 38], or rigorous mathematical analysis [5, 14, 17, 21, 34, 36, 41]. Several works address the middle ground between applications and theory, including [10, 11, 12, 27].

In later chapters, SDEs that evolve on more exotic spaces than  $\mathbb{R}^d$  will be explored. These include manifolds. In order to understand these concepts, it is important to have sufficient geometric background. This is provided in the next two chapters.

**4.10 Exercises**

4.1. Following up on the last paragraph in Section 4.3, determine the explicit conditions on the covariance and mean of a Gaussian process of the form (4.11) to be a Markov process.

4.2. Prove for all values of  $t$  that  $\rho(\mathbf{x}, t)$  in (4.9) is: (a) a pdf; (b) it has mean  $\mu$ ; (c) it has variance  $\sigma^2$ .

4.3. Given the one-dimensional Ornstein–Uhlenbeck SDE  $dx = -\gamma x dt + c dw$ , write and solve the corresponding Fokker–Planck equation analytically.

4.4. Using the programs provided in [12], simulate 1000 sample paths of the one-dimensional Ornstein–Uhlenbeck SDE in Exercise 4.3. Let each path consist of 100

steps with  $dt = 0.01$ , and let  $\gamma = c = 1$ . Record each  $x(t)$  for  $t = 0.2, 0.5, 1.0$ . Create a histogram for each of these times, and compare it with the analytical solution from the previous problem.

4.5. Prove that a substitution of the form  $d\mathbf{w} \rightarrow R(\mathbf{x}, t)d\mathbf{w}$  into an Itô SDE  $d\mathbf{x} = \mathbf{h}(\mathbf{x}, t)dt + H(\mathbf{x}, t)d\mathbf{w}$  will yield the same Fokker–Planck equation as without this substitution when  $RR^T = I$ .

4.6. Let  $R$  and  $R_0$  denote rotation matrices. Prove that a substitution of the form  $H^s(\mathbf{x}, t) \rightarrow H^s(\mathbf{x}, t)R_0(t)$  into a Stratonovich SDE  $d\mathbf{x} = \mathbf{h}^s(\mathbf{x}, t)dt + H^s(\mathbf{x}, t) \circ d\mathbf{w}$  will yield the same Fokker–Planck equation as without this substitution. Will this statement still be true if  $H^s(\mathbf{x}, t) \rightarrow H^s(\mathbf{x}, t)R(\mathbf{x}, t)$ ? Explain.

4.7. Let  $R$  denote a rotation matrix. If a substitution of the form  $H^s(\mathbf{x}, t) \rightarrow H^s(\mathbf{x}, t)R(\mathbf{x}, t)$  is made in a Stratonovich SDE  $d\mathbf{x} = \mathbf{h}^s(\mathbf{x}, t)dt + H^s(\mathbf{x}, t) \circ d\mathbf{w}$ , how must  $\mathbf{h}^s(\mathbf{x}, t)$  be modified in order to yield the same Fokker–Planck equation as without this substitution?

4.8. Using (4.98) show that the Itô equation (4.97) is equivalent to the Stratonovich equation (4.99), and so in this case it does not matter in which way the SDE is interpreted.

4.9. Starting with the SDE in (4.99) in polar coordinates, and using the rules of Stratonovich stochastic calculus, convert this to an SDE in Cartesian coordinates. Is it equivalent to the Cartesian SDE that yielded (4.94)?

4.10. Show that (4.101) will become the heat equation if  $B(\mathbf{q}, t) = [J(\mathbf{q})]^{-1}$  and

$$a_i(\mathbf{q}, t) = \frac{1}{2}|G|^{-\frac{1}{2}} \sum_j \frac{\partial}{\partial q_j} \left( |G|^{\frac{1}{2}} (BB^T)_{ij} \right).$$

4.11. Show that (4.102) will become the heat equation if  $B^s(\mathbf{q}, t) = [J(\mathbf{q})]^{-1}$  and

$$a_i^s(\mathbf{q}, t) = \frac{1}{2}|G|^{-\frac{1}{2}} \sum_{jk} B_{ik}^s \frac{\partial}{\partial q_j} \left( |G|^{\frac{1}{2}} B_{jk}^s \right).$$

4.12. Show that if

$$\sum_{jk} B_{ik} \frac{\partial}{\partial q_j} (|G|^{\frac{1}{2}} B_{jk}) = \sum_j \frac{\partial}{\partial q_j} (|G|^{\frac{1}{2}} BB^T)_{ij}. \quad (4.103)$$

then the Itô and Stratonovich forms of the Fokker–Planck equation will be the same (and hence the corresponding SDE can be taken as Itô or Stratonovich in this special case without having to specify).

4.13. List three specific examples of when (4.103) will hold. Hint: What if  $B$  is independent of  $\mathbf{q}$ ? Or if  $\mathbf{q}$  is partitioned as  $\mathbf{q} = [\mathbf{q}_1^T, \mathbf{q}_2^T]^T$ , and  $B(\mathbf{q}) = B(\mathbf{q}_2) \oplus B(\mathbf{q}_1)$  (where  $\oplus$  denotes the direct sum reviewed in the appendix), what happens?

4.14. Can the general solution of the Fokker–Planck equation for the Ornstein–Uhlenbeck process in (4.74) be solved in the form of a Gaussian:  $f(\mathbf{x}, t) = \rho(\mathbf{x}; \boldsymbol{\mu}(t), \Sigma(t))$ ? If so, what are the forms of  $\boldsymbol{\mu}(t)$  and  $\Sigma(t)$ ?

## References

1. Bouleau, N., Lépingle, D., *Numerical Methods for Stochastic Processes*, John Wiley & Sons, New York, 1994.
2. Casimir, H.B.G., “On onsager’s principle of microscopic reversibility,” *Rev. Mod. Phys.*, 17(2-3), pp. 343–350, 1945.
3. Chirikjian, G.S., Kyatkin, A.B., *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, Boca Raton, FL, 2001.
4. Doob, J.L., *Stochastic Processes*, John Wiley & Sons, New York, 1953.
5. Durrett, R., *Brownian Motion and Martingales in Analysis*, Wadsworth, Belmont, CA, 1984.
6. Einstein, A., *Investigations on the Theory of the Brownian Movement*, Dover, New York, 1956.
7. Fokker, A.D., “Die Mittlere Energie rotierender elektrischer Dipole in Strahlungs Feld,” *Ann. Phys.*, 43, pp. 810–820, 1914.
8. Fowler, R.H., *Statistical Mechanics*, Cambridge University Press, London, 1929.
9. Fowler, R.H., *Philos. Mag.*, 47, p. 264, 1924.
10. Gard, T.C., *Introduction to Stochastic Differential Equations*, Marcel Dekker, New York, 1988.
11. Gardiner, C.W., *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*, 3rd ed., Springer-Verlag, Berlin, 2004.
12. Higham, D.J., “An algorithmic introduction to numerical simulation of stochastic differential equations,” *SIAM Rev.*, 43, pp. 525–546, 2001.
13. Itô, K., McKean, H.P., Jr. *Diffusion Processes and their Sample Paths*, Springer-Verlag, Berlin, 1974.
14. Karatzas, I., Shreve, S.E., *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer, New York, 1991.
15. Karlin, S., Taylor, H.M., *An Introduction to Stochastic Modeling*, 3rd ed., Academic Press, San Diego, 1998.
16. Kloedon, P.E., Platen, E., *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992.
17. Knight, F.B., *Essentials of Brownian Motion*, Math. Survey 18, American Mathematical Society, Providence, RI, 1981.
18. Kolmogorov, A.N., “Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung,” *Math. Ann.*, 104, pp. 415–458, 1931.
19. Kuo, H.-H., *Introduction to Stochastic Integration*, Springer, New York, 2006.
20. Langevin, P., “Sur la théorie du mouvement brownien,” *C. R. Acad. Sci. Paris*, 146, pp. 530–533, 1908.
21. Lévy, P., *Processus stochastiques et mouvement brownien*, Gauthiers-Villars, Paris, 1948 (and 1965).
22. Maruyama, G., “Continuous Markov processes and stochastic equations,” *Rend. Circ. Mat. Palermo*, 4, pp. 48–90, 1955.
23. McKean, H.P., Jr., *Stochastic Integrals*, Academic Press, New York, 1969.
24. McShane, E.J., *Stochastic Calculus and Stochastic Models*, Academic Press, New York, 1974.
25. Millstein, G.N., “A method of second order accuracy of stochastic differential equations,” *Theory of Probability and Its Applications (USSR)*, 23, pp. 396–401, 1976.
26. Millstein, G.N., Tretyakov, M.V., *Stochastic Numerics for Mathematical Physics*, Springer-Verlag, Berlin, 2004.
27. Øksendal, B., *Stochastic Differential Equations, An Introduction with Applications*, 5th ed., Springer, Berlin, 1998.
28. Onsager, L., “Reciprocal relations in irreversible processes, I, II,” *Phys. Rev.*, 37, pp. 405–426, 38, pp. 2265–2280, 1931.
29. Planck, M., “Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie,” *Sitzungsber. Berlin Akad. Wiss.*, pp. 324–341, 1917.

30. Protter, P., *Stochastic Integration and Differential Equations*, Springer, Berlin, 1990.
31. Rényi, A., *Probability Theory*, North-Holland, Amsterdam, 1970.
32. Ripley, B.D., *Stochastic Simulation*, John Wiley & Sons, New York, 1987.
33. Risken, H., *The Fokker–Planck Equation, Methods of Solution and Applications*, 2nd ed., Springer-Verlag, Berlin, 1989.
34. Rogers, L.C.G., Williams, D., *Diffusion, Markov Processes, and Martingales*, Vols. 1 and 2, John Wiley & Sons, New York, 1987.
35. Stratonovich, R.L., *Topics in the Theory of Random Noise*, Vols. I and II, (translated by R.A. Silverman), Gordon and Breach, New York, 1963.
36. Stroock, D., Varadhan, S.R.S., *Multidimensional Diffusion Processes*, Grundlehren Series #233, Springer-Verlag, Berlin, 1979 (and 1998).
37. Uhlenbeck, G.E., Ornstein, L.S., “On the theory of Brownian motion,” *Phys. Rev.*, 36, pp. 823–841, 1930.
38. van Kampen, N.G., *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam, 1981.
39. van Kampen, N.G., “Derivation of the phenomenological equations from the master equation: I. Even variables only; II. Even and odd variables,” *Physica*, 23, pp. 707–719, pp. 816–824, 1957.
40. Wang, M.C., Uhlenbeck, G.E., “On the theory of Brownian motion II,” *Rev. Mod. Phys.*, 7, pp. 323–342, 1945.
41. Watanabe, S., *Stochastic Differential Equations and Malliavin Calculus*, Tata Institute, 1984.
42. Wiener, N., “Differential space,” *J. Math. Phys.*, 2, pp. 131–174, 1923.

## Geometry of Curves and Surfaces

This chapter consists of a variety of topics in geometry. The approach to geometry that is taken in this chapter and throughout this book is one in which the objects of interest are described as being *embedded*<sup>1</sup> in Euclidean space. There are two natural ways to describe such embedded objects: (1) parametrically and (2) implicitly.

The vector-valued functions  $\mathbf{x} = \mathbf{x}(t)$  and  $\mathbf{x} = \mathbf{x}(u, v)$  are respectively parametric descriptions of curves and surfaces when  $\mathbf{x} \in \mathbb{R}^3$ . For example,  $\mathbf{x}(\psi) = [\cos \psi, \sin \psi, 0]^T$  for  $\psi \in [0, 2\pi)$  is a parametric description of a unit circle in  $\mathbb{R}^3$ , and  $\mathbf{x}(\phi, \theta) = [\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta]^T$  for  $\phi \in [0, 2\pi)$  and  $\theta \in [0, \pi]$  is a parametric description of a unit sphere in  $\mathbb{R}^3$ . Parametric descriptions are not unique. For example,  $\mathbf{x}(t) = [2t/(1+t^2), (1-t^2)/(1+t^2), 0]^T$  for  $t \in \mathbb{R}$  describes the same unit circle as the one mentioned above.<sup>2</sup>

Implicit descriptions of curves and surfaces involve constraint equations in their Cartesian coordinates. For example, the circle in  $\mathbb{R}^3$  can be described as simultaneously satisfying the equation  $x_1^2 + x_2^2 = 1$  (which describes a right-circular cylinder) and  $x_3 = 0$  (which describes the  $x_1$ - $x_2$  plane). An implicit equation for the unit sphere in  $\mathbb{R}^3$  is  $\mathbf{x} \cdot \mathbf{x} = 1$ . Implicit descriptions are generally not unique. For example, the unit circle in the  $x_1$ - $x_2$  plane in  $\mathbb{R}^3$  could have been described as the intersection of the unit sphere with the  $x_3 = 0$  plane rather than the intersection of a cylinder and that plane. Or it could have been described as the intersection of the cylinder and the sphere.

Most of the calculations performed in later chapters involve parametric descriptions. However, it is important to realize that this is not the only approach, and sometimes the implicit approach can result in simpler calculations than when using parametric descriptions. An example of such a situation is described later in this chapter.

<sup>1</sup>A geometrical object that is contained inside of another is said to be immersed in the larger object. If in addition certain properties hold, it is said to be embedded. In this case the mapping that defines the contained object is called an *embedding*. In general, there are many ways to embed one geometrical object inside another. If  $X$  is embedded in  $Y$  then there is an injective mapping  $m : X \rightarrow Y$  that describes the embedded object.

<sup>2</sup>In a strict sense, a curve or surface that differs from another by the removal of a single point is a different mathematical object. For example, the point  $\mathbf{x} = [0, -1, 0]^T$  is on the unit circle, but the parametric description  $\mathbf{x}(t)$  breaks down at that point. From the perspective of computing lengths, areas, volumes, etc., two geometrical objects can be considered equivalent if one coincides with the other except at a locus of points that is lower than the dimension of the object. Therefore, the curve that  $\mathbf{x}(t)$  traces out will be satisfactory proxy for the circle in the context of many applications, and the distinction between the circle and the circle missing one point will be deemphasized.

This chapter is organized as follows. Section 5.1 begins this chapter with an introduction to some basic geometric concepts originating from an application involving robotic arms. Section 5.2 presents a case study in geometry originating from a medical imaging problem. In the context of this one problem, several basic ideas of parametric and implicitly defined curves and surfaces are illustrated in a concrete way. Also some very basic ideas of projective and algebraic geometry are introduced. Section 5.3 reviews the local and global geometry of curves in three-dimensional space. Section 5.4 reviews the differential geometry of two-dimensional surfaces in three-dimensional space. This includes discussions of local and global surface geometry, the divergence theorem, and includes explicit calculations of geometric quantities for the sphere, ellipsoid of revolution, and torus. Section 5.5 introduces Weyl's tube theorem, which is a classical topic not often covered in introductory differential geometry texts. Section 5.6 reviews the concept of the Euler characteristic for surfaces and bodies in two- and three-dimensional Euclidean space. Section 5.7 describes curves and surfaces implicitly, and expresses Stokes' theorem and the divergence theorem in this notation.

The main points to take away from this chapter are:

- Analytical tools exist to compute arc length, area, and volume.
- Curves and surfaces in two- and three-dimensional space can be described parametrically or implicitly, and the local geometry is determined by intrinsic quantities that are independent of the particular description.
- The global topological features of these geometric objects can be related to integrals of curvature. In particular, the Euler characteristic describes how many “holes” there are in an object, and the integrals of certain kinds of curvature over a tubular surface can help to determine whether it is knotted or not.
- The concepts of gradient, divergence, and Laplacian that were defined in Chapter 1 for Cartesian coordinates in Euclidean space apply equally well to curved surfaces.

## 5.1 An Introduction to Geometry Through Robotic Manipulator Kinematics

A *robotic manipulator* (or *robot arm*) is a mechanical device used to move objects around in space. A simple robot arm is shown in Figure 5.1. A few fundamental geometric ideas are introduced in the following subsections in the context of the concrete problem of forward and reverse kinematics of this robot arm.

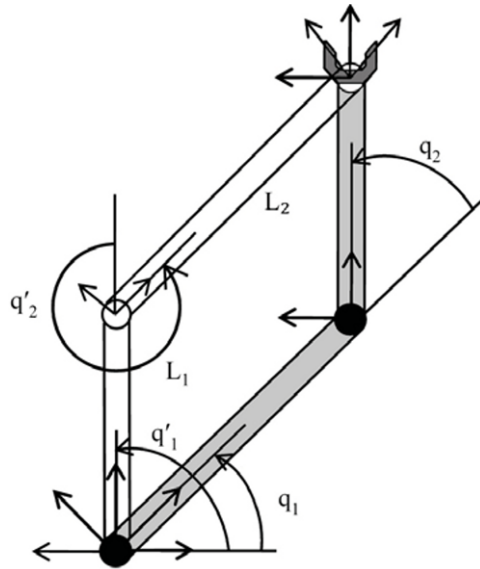
### 5.1.1 Forward (or Direct) Kinematics

A robot arm consisting of two rigid links, of length  $L_1$  and  $L_2$ , and two rotational joints, that turn through angles  $q_1$  and  $q_2$ , is shown in Figure 5.1. The position of the hand of this robot arm is given by the following equations, which can be derived using basic trigonometry and geometrical constructions:

$$\begin{aligned}x_1 &= L_1 \cos q_1 + L_2 \cos(q_1 + q_2) \\x_2 &= L_1 \sin q_1 + L_2 \sin(q_1 + q_2).\end{aligned}\tag{5.1}$$

This can be written in the compact form

$$\mathbf{x} = \mathbf{f}(\mathbf{q}).$$



**Fig. 5.1.** A Robot Arm with Two Rotational Joints

If this arm is treated as a “phantom” that is allowed to pass through itself, then the joints can take the values  $-\pi \leq q_1, q_2 < \pi$ , with the understanding that  $q_i = +\pi$  gives the same *conformation* (or shape) of the arm as  $q_i = -\pi$ . In fact, the joints can spin around and take any real values, but the shape of the arm will be the same for any  $q_i$  and  $q_i + 2n\pi$ , and so it is sufficient to describe all conformations attainable by the arm by limiting things in this way.

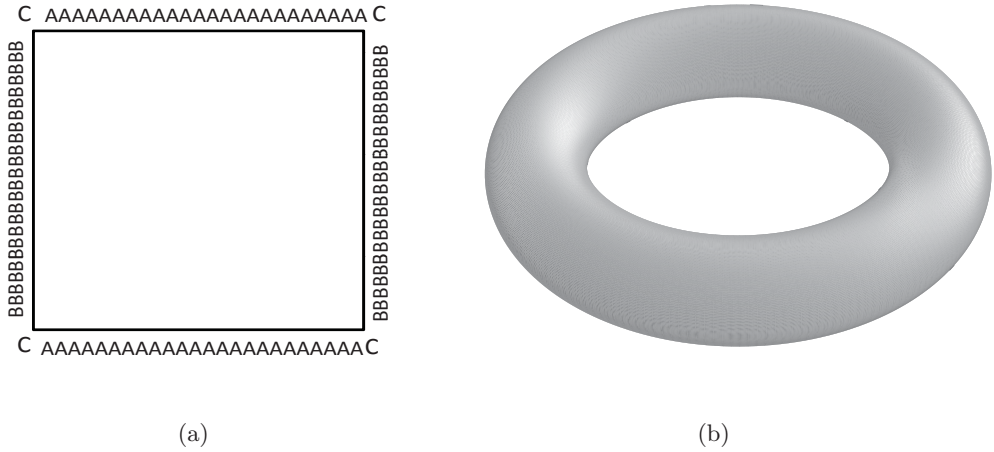
For almost all conformations of the arm, the values  $(q_1, q_2)$  can be perturbed to result in an arbitrary infinitesimal change in position of the hand,  $d\mathbf{x}$ . When  $q_2 = 0$  the arm is fully outstretched, and when  $q_2 = -\pi$  it is folded back on itself. In both of these cases, the hand becomes limited in the directions that it can move, since in both of these cases the hand cannot move instantaneously in the direction tangent to the links. Such a condition is called a *singularity*. Since  $d\mathbf{x} = D\mathbf{f}d\mathbf{q}$  where the Jacobian is

$$D\mathbf{f} = \begin{pmatrix} \frac{\partial f_1}{\partial q_1} & \frac{\partial f_1}{\partial q_2} \\ \frac{\partial f_2}{\partial q_1} & \frac{\partial f_2}{\partial q_2} \end{pmatrix},$$

singularities can be identified by setting  $\det D\mathbf{f} = 0$ . It is common to denote the Jacobian matrix simply as  $J$ , and the absolute value of  $\det D\mathbf{f}$  as  $|J|$ .

The loci of points defined by  $\mathbf{x}(q_1, 0)$  and  $\mathbf{x}(q_1, -\pi)$  are circles of radius  $|L_1 + L_2|$  and  $|L_1 - L_2|$ , respectively. When the hand reaches a specific point on either of these circles, a unique value of  $q_1$  is specified. In the open annular region bounded by these circles, there are two conformations of the arm that reach each end position. This region of the plane is where the robot hand (also called an end-effector or gripper) can operate by moving parts around, and it is called the *workspace*. The two conformations of the arm can be called “elbow up” and “elbow down.” The space of all joint values that the arm can take can be identified with the two-torus. The opposing edges of the square region in the  $q_1$ - $q_2$  plane ranging from  $-\pi$  to  $\pi$  can be “pasted together” by the rule

that each point on these opposing edges corresponds to the same conformation. This is shown in Figure 5.2.



**Fig. 5.2.** Making a Torus from a Square: (a) A “topological torus” in which the directly opposing As and Bs are respectively identified with each other, and C becomes a single point; (b) the “geometric torus,” which is embedded in  $\mathbb{R}^3$

The resulting torus can be visualized as the “donut” surface in three spatial dimensions,  $x_1, x_2, x_3$ . The size of the torus is unimportant, and it can be scaled so that its radii are  $L_1$  and  $L_2$ . From the perspective of the two-dimensional robot arm,  $x_3$  is not a real spatial direction. Rather,  $x_3$  is introduced here only for the convenience of visualization. When the donut is sliced through the  $x_1$ - $x_2$  plane, the result is the boundary of the annular workspace of the arm.

The forward kinematic function in (5.1) can be thought of as a mapping from the torus into this annular workspace,  $\mathbf{f} : T^2 \rightarrow W$ . The workspace can be broken into two parts: the interior  $I(W)$  and the boundary  $\partial W$ . The torus can also be broken into two sets of points: those that map to  $I(W)$  and those that map to  $\partial W$ . Call these sets  $Q(I(W))$  and  $Q(\partial W)$ . From the discussion above, it follows that  $\mathbf{f} : Q(I(W)) \rightarrow I(W)$  is a two-to-one function and  $\mathbf{f} : Q(\partial W) \rightarrow \partial W$  is one-to-one (or *injective*). Both functions are onto (or *surjective*). In general a function that is both injective and surjective is called *bijective*. A bijective function establishes a unique correspondence between elements of two sets.

This can be viewed geometrically as points on the upper and lower halves of the torus being mapped by a projection onto the workspace. The projection is not simply along the vertical (which would correspond to a fixed value of  $q_1$  and two different values of  $q_2$ ) because the value of  $q_1$  is different in up- and down-elbow conformations.

### 5.1.2 Reverse (or Inverse) Kinematics

In practical robotics applications a desired trajectory of the hand,  $\mathbf{x}(t)$ , is given and the goal is to find a trajectory in the joint space of the form  $\mathbf{q}(t)$  such that  $\mathbf{x}(t) = \mathbf{f}(\mathbf{q}(t))$ . This is the reverse, or inverse, problem from that described in the previous



subsection. The three most common ways that this problem is solved are: (1) incremental linearization; (2) analytical solution for the inverse function  $\mathbf{f}^{-1}$ ; and (3) polynomial elimination methods. All three relate to concepts in geometry, and are described below.

### Incremental Linearization

In incremental linearization (which is also called resolved rate motion control), the relationship between an initial set of joint angles and the hand position is assumed to be known. For example, a random value of  $\mathbf{q}(0)$  can be chosen at time  $t = 0$ , and the resulting hand position at that time will be  $\mathbf{x}(0)$ , which can be calculated by  $\mathbf{x}(0) = \mathbf{f}(\mathbf{q}(0))$ . The instantaneous kinematics is described by the equation  $d\mathbf{x} = D\mathbf{f}d\mathbf{q}$ , which means that if the hand is to move from  $\mathbf{x}(0)$  to  $\mathbf{x}(0) + d\mathbf{x}(0)$ , then it had better be the case that  $d\mathbf{x}(0) = D\mathbf{f}(\mathbf{q}(0))d\mathbf{q}(0)$ . If the Jacobian  $D\mathbf{f}(\mathbf{q}(0))$  is invertible, then

$$d\mathbf{q}(0) = [D\mathbf{f}(\mathbf{q}(0))]^{-1}d\mathbf{x}(0)$$

will provide the desired increment. Then the value of  $\mathbf{q}$  can be updated as  $\mathbf{q}(\Delta t) = \mathbf{q}(0) + \Delta t d\mathbf{q}(0)$ . Now a set of joint angles  $\mathbf{q}(\Delta t)$  is known that satisfies  $\mathbf{x}(\Delta t) = \mathbf{f}(\mathbf{q}(\Delta t))$ . The procedure can then be performed again with  $\mathbf{q}(\Delta t)$  taking the place of  $\mathbf{q}(0)$  and  $\mathbf{x}(\Delta t)$  taking the place of  $\mathbf{x}(0)$ . From the starting value  $\mathbf{x}(0)$  that is on a trajectory of the hand, the whole trajectory can be followed by breaking it up into little steps  $d\mathbf{x}(t) = \mathbf{x}(t + \Delta t) - \mathbf{x}(t)$  for any specific end-effector trajectory,  $\mathbf{x}(t)$ .

### Analytical Solution

In the case of the simple two-link arm described by the forward kinematic equations in (5.1), it is possible to obtain closed-form solutions for  $q_1$  and  $q_2$  as a function of any given  $x_1$  and  $x_2$ , provided the position  $\mathbf{x}$  that is specified lies in the workspace (set of reachable positions of the hand).

To start, square and add the equations for  $x_1$  and  $x_2$ :

$$\begin{aligned} x_1^2 + x_2^2 &= [L_1 \cos q_1 + L_2 \cos(q_1 + q_2)]^2 + [L_1 \sin q_1 + L_2 \sin(q_1 + q_2)]^2 \\ &= L_1^2 + L_2^2 + 2L_1L_2 \cos q_2. \end{aligned} \quad (5.2)$$

From this, a solution for  $q_2$  is obtained as

$$q_2(x_1, x_2) = \cos^{-1} \left( \frac{x_1^2 + x_2^2 - L_1^2 - L_2^2}{2L_1L_2} \right). \quad (5.3)$$

Since  $\cos(-\phi) = \cos \phi$ , the above expression represents two solutions: one with the elbow up and the other with the elbow down. Choosing either solution, substituting back into the forward-kinematic expression (5.1), and expanding out gives

$$\begin{aligned} x_1 &= L_1 \cos q_1 + L_2 [\cos q_1 \cos q_2(x_1, x_2) - \sin q_1 \sin q_2(x_1, x_2)] \\ x_2 &= L_1 \sin q_1 + L_2 [\cos q_1 \sin q_2(x_1, x_2) + \sin q_1 \cos q_2(x_1, x_2)]. \end{aligned}$$

Writing the above as a matrix-vector expression and isolating the unknowns  $c = \cos q_1$  and  $s = \sin q_1$  on one side of the equation,

$$\begin{pmatrix} c(x_1, x_2) \\ s(x_1, x_2) \end{pmatrix} = \frac{1}{L_1} \begin{pmatrix} L_1 + L_2 \cos q_2(x_1, x_2) & -L_2 \sin q_2(x_1, x_2) \\ L_2 \sin q_2(x_1, x_2) & L_1 + L_2 \cos q_2(x_1, x_2) \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \quad (5.4)$$

Then

$$q_2(x_1, x_2) = \text{Atan2}[c(x_1, x_2), s(x_1, x_2)] \quad (5.5)$$

where the two-argument tangent function  $\text{Atan2}[\cdot, \cdot]$  takes values in the full range of angular values rather than values restricted to the open interval  $(-\pi/2, \pi/2)$  where  $\tan^{-1}(s/c)$  and  $\text{Atan2}[c, s]$  coincide.

### Polynomial Elimination

While a closed-form analytical solution exists for the simple arm depicted in Figure 5.1, this is not always the case for more complicated manipulators with six joints used to position and orient a hand in three-dimensional space. However, it was shown by Raghavan and Roth [58] that it is always possible to reduce these more complicated cases to algebraic problems, where powerful tools of elimination theory [65] can be applied. This method is illustrated in the context of the robot arm in Figure 5.1 to illustrate how a geometric problem can be reduced to an algebraic one.

Making the substitution  $t_i = \tan(q_i/2)$ , and using trigonometric identities,  $\cos q_i$  and  $\sin q_i$  can be written as

$$\cos q_i = \frac{1 - t_i^2}{1 + t_i^2} \quad \text{and} \quad \sin q_i = \frac{2t_i}{1 + t_i^2}. \quad (5.6)$$

It follows from this that

$$\sin q_i - t_i \cos q_i = t_i \quad \text{and} \quad t_i \sin q_i + \cos q_i = 1. \quad (5.7)$$

Expanding out (5.1) into products of sines and cosines of individual joint angles converts the transcendental forward kinematic expression into one involving rational polynomials in two variables:

$$\begin{aligned} x_1 &= L_1 \frac{1 - t_1^2}{1 + t_1^2} + L_2 \frac{(1 - t_1^2)(1 - t_2^2) - 4t_1 t_2}{(1 + t_1^2)(1 + t_2^2)} \\ x_2 &= L_1 \frac{2t_1}{1 + t_1^2} + L_2 \frac{(1 - t_1^2)t_2 + (1 - t_2^2)t_1}{(1 + t_1^2)(1 + t_2^2)}. \end{aligned}$$

In the inverse kinematics problem,  $x_1$  and  $x_2$  are given and can be treated as inputs, and the goal is to find the inverse kinematic function that returns  $q_1$  and  $q_2$ . This is equivalent to finding  $t_1$  and  $t_2$  as a function of  $x_1$  and  $x_2$  since (5.6) can then be used to obtain  $q_1$  and  $q_2$ .

Multiplying both sides of the above equation by  $(1 + t_1^2)(1 + t_2^2)$  yields

$$\begin{aligned} (1 + t_1^2)(1 + t_2^2)x_1 &= L_1(1 - t_1^2)(1 + t_2^2) + L_2[(1 - t_1^2)(1 - t_2^2) - 4t_1 t_2] \\ (1 + t_1^2)(1 + t_2^2)x_2 &= L_1(2t_1)(1 + t_2^2) + L_2[(1 - t_1^2)t_2 + (1 - t_2^2)t_1]. \end{aligned}$$

These are two polynomial equations in two unknowns. Therefore, the problem has been converted to one of elementary algebraic geometry [38]. For example, in either one of the above equations,  $t_1$  can be solved for in terms of  $t_2$  and substituted into the other equation. The roots of the resulting single equation in  $t_2$  will provide the inverse kinematics solutions.

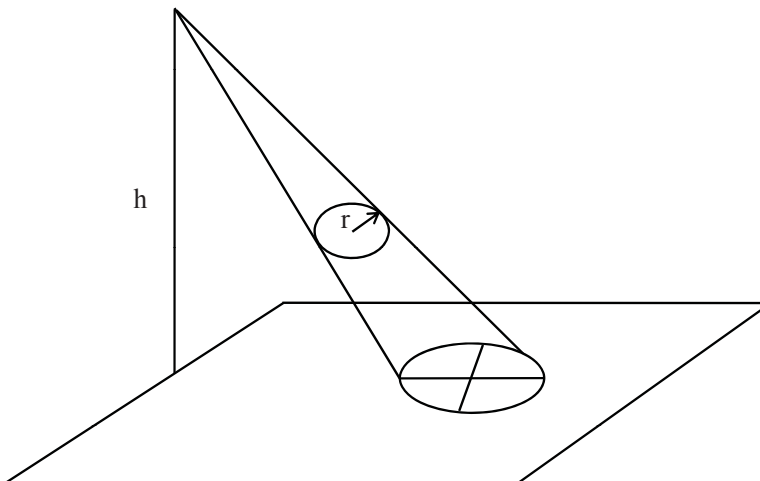
In the particular problem at hand, since (5.2) is already known, the algebraic solution for  $t_2$  is obtained by substituting (5.6) into (5.2). The result is a quadratic equation in

$t_2$ . The roots correspond to up- and down-elbow conformations of the arm. Substituting these into (5.8) reduces them to two quadratic equations in  $t_1$ . The quadratic terms can then be eliminated, resulting in a linear equation giving  $t_1$ . Since  $\tan^{-1}(\cdot)$  is well-defined over the range  $(-\pi/2, \pi/2)$ , the values  $q_i = 2 \tan^{-1} t_i$  are obtained in all cases except when  $q_1 = \pm\pi$ .

The approach described here has been generalized to spatial manipulators and other mechanisms [58, 65]. As an alternative to root-finding approaches, Kohli and Osvatic [44], Chazvini [13], and Manocha and Canny [48] converted the more complicated spatial version of this problem to an eigenvalue–eigenvector problem.

## 5.2 A Case Study in Medical Imaging

Recently, a surgical resident (Dr. Tarun Bhargava) working under the supervision of Drs. Marc Hungerford and Lynn Jones in the Department of Orthopaedics at the Johns Hopkins School of Medicine approached the author with the following problem. An artificial hip implant with a circular metallic rim of radius  $r$  needs to be inserted in the patient at the correct position and orientation. One source of information about this position and orientation is the aspect ratio of the elliptical shape that the metallic circle makes when it is observed in planar x-ray images. These images are obtained by the projection of the circle from a point source of x-rays. Since the circle is located at an arbitrary position in space at some unknown orientation relative to the patient, there is a skewed cone (with vertex at the x-ray source) that contains the metallic circle. The projection of the circle observed in the x-ray image can then be thought of as the intersection of this skewed cone with the imaging plane, which is taken as the  $x_1$ - $x_2$  plane. The problem to be solved is: “How does the aspect ratio of the ellipse observed in the planar projection relate to the spatial position and orientation of the circular rim?” This is depicted in Figure 5.3.



**Fig. 5.3.** Projection of a Circle at Specified Orientation onto a Plane Using a Point Source

This problem can be addressed either parametrically or implicitly. Both approaches are described in the following subsections. First, some preliminary notation that is

relevant to both is reviewed. Let

$$R_1(\theta) \doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\theta & -s\theta \\ 0 & s\theta & c\theta \end{pmatrix}; \quad R_2(\theta) \doteq \begin{pmatrix} c\theta & 0 & s\theta \\ 0 & 1 & 0 \\ -s\theta & 0 & c\theta \end{pmatrix}; \quad R_3(\theta) \doteq \begin{pmatrix} c\theta & -s\theta & 0 \\ s\theta & c\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.8)$$

where  $c\theta \doteq \cos \theta$  and  $s\theta \doteq \sin \theta$ .

Let  $\mathbf{e}_3 = [0, 0, 1]^T$  be the vector that points along the  $x_3$ -axis. Then the unit vector

$$\mathbf{u}(\alpha, \beta) = R_3(\alpha)R_2(\beta)\mathbf{e}_3 = \begin{pmatrix} \cos \alpha \sin \beta \\ \sin \alpha \sin \beta \\ \cos \beta \end{pmatrix} \quad (5.9)$$

is in the standard form of spherical coordinates. This will be used to describe the orientation of the circle in space. The two degrees of freedom  $\alpha$  and  $\beta$  are sufficient for this purpose because of the axial symmetry of the circle.

### 5.2.1 A Parametric Approach

Suppose that we start with a circle of radius  $r$  in the  $x_1$ - $x_2$  plane. This can be parameterized as

$$\mathbf{c}(\theta) = \begin{pmatrix} r \cos \theta \\ r \sin \theta \\ 0 \end{pmatrix} \quad \text{for } \theta \in [0, 2\pi). \quad (5.10)$$

Now suppose that this circle is translated and rotated by a fixed amount in space. The result will be

$$\mathbf{x}(\theta) = \mathbf{m} + R_3(\alpha)R_2(\beta)\mathbf{c}(\theta) \quad \text{where } \mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}. \quad (5.11)$$

The vector  $\mathbf{u}(\alpha, \beta)$  is normal to the plane containing this circle.

Now suppose that there is a point light source located at  $\mathbf{p} = h\mathbf{e}_3$ . We can define a line that connects the point source to each point on the circle. For each fixed value of  $\theta$ , this will be the line parameterized as

$$\mathbf{L}(t; \theta) = \mathbf{p} + [\mathbf{x}(\theta) - \mathbf{p}]t$$

where  $t$  can take any value from  $-\infty$  to  $\infty$ . The above expression for  $\mathbf{L}(t; \theta)$  holds because a line can be defined by the position of a point on it, together with a pointing direction.

Now, suppose that the image plane is the  $x_1$ - $x_2$  plane. A model for the shape of the projected image is the locus of points where each line intersects this plane. This condition is stated as

$$\mathbf{L}(t; \theta) \cdot \mathbf{e}_3 = 0 \quad \implies \quad h + [\mathbf{x}(\theta) \cdot \mathbf{e}_3 - h]t^* = 0,$$

where  $t^*$  denotes the specific value of  $t$  marking the point on the line of projection that connects  $\mathbf{x}(\theta)$  and  $\mathbf{p}$  and intersects the image plane.

But from (5.11), after working out the matrix multiplications, we have that

$$\mathbf{x}(\theta) = \begin{pmatrix} m_1 + r \cos \theta \cos \alpha \cos \beta - r \sin \theta \sin \alpha \\ m_2 + r \cos \theta \sin \alpha \cos \beta + r \sin \theta \cos \alpha \\ m_3 - r \sin \beta \cos \theta \end{pmatrix}.$$

Therefore

$$\mathbf{x}(\theta) \cdot \mathbf{e}_3 = m_3 - r \sin \beta \cos \theta,$$

and so

$$t^* = \frac{h}{h - m_3 + r \sin \beta \cos \theta}.$$

The projected curve observed in the image plane then becomes

$$\boldsymbol{\pi}(\theta) = \begin{pmatrix} \pi_1(\theta) \\ \pi_2(\theta) \end{pmatrix} = \begin{pmatrix} L_1(t^*; \theta) \\ L_2(t^*; \theta) \end{pmatrix} = \begin{pmatrix} \frac{h \cdot m_1 + (h \cdot r)(\cos \theta \cos \alpha \cos \beta - \sin \theta \sin \alpha)}{h - m_3 + r \sin \beta \cos \theta} \\ \frac{h \cdot m_2 + (h \cdot r)(\cos \theta \sin \alpha \cos \beta + \sin \theta \cos \alpha)}{h - m_3 + r \sin \beta \cos \theta} \end{pmatrix} \quad (5.12)$$

where  $x_i = \pi_i$  are the coordinates of the points on the curve of intersection in the image plane.

The most fundamental question is: is this an ellipse? It is not obvious from this parametric description whether it is or not. If it is not, the surgeons will still want to know the maximal distance between antipodal points in the curve, and the minimal distance corresponding to the perpendicular bisector. If the curve is not an ellipse, this becomes a messier problem. The trouble is, just by looking at the parametric form in (5.12), it is difficult to determine if it is an ellipse or not. As it turns out, the projected curve is an ellipse. Proof of this will come later.

Any ellipse must satisfy the implicit equation

$$(\boldsymbol{\pi} - \mathbf{b})^T A (\boldsymbol{\pi} - \mathbf{b}) = 1 \quad (5.13)$$

for some  $A = A^T$  with positive eigenvalues, and some position vector  $\mathbf{b}$ . If (5.12) satisfies this equation, then the planar projection will be an ellipse, and moreover, by finding the eigenvalues of  $A$  the aspect ratio of the ellipse can be calculated. But how can  $(A, \mathbf{b})$  be found from (5.12)?

Since  $\boldsymbol{\pi} = \boldsymbol{\chi}/y$  where  $y = h - m_3 + r \sin \beta \cos \theta$ , (5.13) can be rewritten after substitution of (5.12) as

$$(\boldsymbol{\chi} - y\mathbf{b})^T A (\boldsymbol{\chi} - y\mathbf{b}) = y^2.$$

Expanding this out gives

$$\begin{aligned} & a_{11} [hm_1 + hr(c\theta c\alpha c\beta - s\theta s\alpha) - b_1(h - m_3 + rs\beta c\theta)]^2 \\ & + a_{22} [hm_2 + hr(c\theta s\alpha c\beta + s\theta c\alpha) - b_2(h - m_3 + rs\beta c\theta)]^2 \\ & + 2a_{12} [hm_1 + hr(c\theta c\alpha c\beta - s\theta s\alpha) \\ & - b_1(h - m_3 + rs\beta c\theta)] \cdot [hm_2 + hr(c\theta s\alpha c\beta + s\theta c\alpha) \\ & - b_2(h - m_3 + rs\beta c\theta)] \\ & = [h - m_3 + r \sin \beta \cos \theta]^2. \end{aligned}$$

The above terms can be grouped together by their dependence on  $\theta$  as

$$\begin{aligned}
& a_{11} [x_0 + x_1 c\theta + x_2 s\theta]^2 + a_{22} [y_0 + y_1 c\theta + y_2 s\theta]^2 \\
& + 2a_{12} [x_0 + x_1 c\theta + x_2 s\theta] \cdot [y_0 + y_1 c\theta + y_2 s\theta] \\
& = [z_0 + z_1 c\theta]^2.
\end{aligned}$$

Expanding out, using the fact that  $s^2\theta = 1 - c^2\theta$ , and collecting the coefficients in front of each independent function, the following coefficients must be zero for the above equalities to hold:

$$a_{11}(x_1^2 - x_2^2) + a_{22}(y_1^2 - y_2^2) + 2a_{12}(x_1 y_1 - x_2 y_2) = z_1^2$$

(for the  $c^2\theta$  term);

$$2a_{11}x_0x_1 + 2a_{22}y_0y_1 + 2a_{12}(x_0y_1 + x_1y_0) = 2z_0z_1$$

(for the  $c\theta$  term);

$$2a_{11}x_0x_2 + 2a_{22}y_0y_2 + 2a_{12}(y_0x_2 + x_0y_2) = 0$$

(for the  $s\theta$  term);

$$2a_{11}x_1x_2 + 2a_{22}y_1y_2 + 2a_{12}(x_2y_1 + x_1y_2) = 0$$

(for the  $c\theta s\theta$  term);

$$a_{11}(x_0^2 + x_2^2) + a_{22}(y_0^2 + y_2^2) + 2a_{12}(x_0y_0 + x_2y_2) = z_0^2$$

(for the constant term).

Altogether, this represents five equations in the five unknowns  $a_{11}$ ,  $a_{12}$ ,  $a_{22}$ ,  $b_1$ , and  $b_2$  (the latter two of which are hidden in  $x_i$  and  $y_i$ ), and in the five given quantities:  $\alpha, \beta, m_1, m_2, m_3$ . If the unknowns are labeled as  $\mathbf{q}$  and the known quantities as  $\mathbf{k}$ , then the above equations can be written as

$$\mathbf{F}(\mathbf{q}, \mathbf{k}) = \mathbf{0}. \quad (5.14)$$

These equations are not linear in the unknowns because  $b_i$ 's multiply  $a_{ik}$ 's. In principle, this set of algebraic (i.e., polynomial) equations can be reduced to a single higher order polynomial equation, and root finding methods can be employed. Or, if a solution is known when the circle is at a particular position and orientation, then an artificial trajectory in the five parameters  $\mathbf{k} = [\alpha, \beta, m_1, m_2, m_3]^T$  can be generated from that known starting state to the desired one. Viewing  $\mathbf{k}$  and  $\mathbf{q}$  as functions of an artificial time parameter, (5.14) can be differentiated with respect to this time parameter to give the following "rate-linearized" equations:

$$J_q \dot{\mathbf{q}} + J_k \dot{\mathbf{k}} = \mathbf{0}$$

where

$$J_q = \frac{\partial \mathbf{F}}{\partial \mathbf{q}^T} \quad \text{and} \quad J_k = \frac{\partial \mathbf{F}}{\partial \mathbf{k}^T},$$

and in general these will both be functions of  $\mathbf{q}$  and  $\mathbf{k}$ :  $J_q = J_q(\mathbf{q}, \mathbf{k})$  and  $J_k = J_k(\mathbf{q}, \mathbf{k})$ . If  $\mathbf{k}(t)$  is any trajectory from the initial state (where  $\mathbf{q}(0) = \mathbf{q}_0$  and  $\mathbf{k}(0) = \mathbf{k}_0$  are known and satisfy (5.14)) to the desired state where  $\mathbf{k}(1) = \mathbf{k}_{goal}$ , then a simple iterative

procedure can be used to update the values of  $\mathbf{q}(t)$  from  $t = 0$  to  $t = 1$ . For example, a simple rule such as  $\mathbf{k}(t) = t(\mathbf{k}_{goal} - \mathbf{k}_0)$  can be used, or equivalently  $\dot{\mathbf{k}}(t) = \mathbf{k}_{goal} - \mathbf{k}_0$ . Then, much like the robot arm inverse kinematics problem, iterations using the equations

$$\dot{\mathbf{q}}(t) = -J_q^{-1} J_k \dot{\mathbf{k}}(t) \quad \text{and} \quad \mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \Delta t \dot{\mathbf{q}}(t) \quad (5.15)$$

can be used to steer  $\mathbf{q}(t)$  toward the value that satisfies (5.14) with  $\mathbf{k} = \mathbf{k}_{goal}$ .

However, if the accumulation of numerical errors after many iterations of the above procedure causes (5.14) not to hold, then a correction is in order. In particular, if  $\|\mathbf{k}_{goal} - \mathbf{k}(t)\|$  is small, then

$$\mathbf{F}(\mathbf{q}, \mathbf{k}_{goal}) - \mathbf{F}(\mathbf{q}, \mathbf{k}) \approx \frac{\partial \mathbf{F}}{\partial \mathbf{k}^T} (\mathbf{k}_{goal} - \mathbf{k})$$

and so a variation in the strategy in (5.15) is the update rule

$$\Delta \mathbf{q} = -\alpha \Delta t J_q^{-1}(\mathbf{q}, \mathbf{k}) [\mathbf{F}(\mathbf{q}, \mathbf{k}_{goal}) - \mathbf{F}(\mathbf{q}, \mathbf{k})] \quad (5.16)$$

where  $\alpha$  is a positive weighting scalar, or gain, that regulates the speed of convergence to the goal.

In the special case when (5.14) is of the form

$$\mathbf{f}(\mathbf{q}) - \mathbf{k} = \mathbf{0},$$

$J_k$  is the negative of the identity,  $J_q = J_q(\mathbf{q})$ , and a rule of the form  $\dot{\mathbf{k}} = \alpha(\mathbf{k}_{goal} - \mathbf{k})$  can be used to make the velocity vector point from the current state to the desired one. In this case, (5.16) reduces to

$$\Delta \mathbf{q} = \alpha \Delta t J_q^{-1}(\mathbf{q}) [\mathbf{k}_{goal} - \mathbf{f}(\mathbf{q})]. \quad (5.17)$$

As discussed in Section 5.1, this sort of update rule is very popular in Robotics for guiding the motion of robot arms. In that context,  $\mathbf{q}$  would represent the internal joint angles of the arm, and  $\mathbf{k}$  would denote the variables describing the position and orientation of the hand in space.

If in (5.17) or (5.16) it is the case that the Jacobian matrix is well behaved, i.e., all of the singular values of  $\det J_q$  are close to unity, then these methods tend to work very well. If  $\det J_q \approx 0$ , this is a singularity, and the above algorithms will fail. However, in this case it is possible to modify the algorithms by replacing the inverse of the Jacobian with a generalized (or pseudo-) inverse, such as those described in [4].

### 5.2.2 An Implicit Approach

As an alternative to parameterizing the cone as was done in the previous subsection, it is possible to describe the same geometrical problem using an implicit approach from the start. In other words, instead of describing the cone in the parametric form  $\mathbf{x} = \mathbf{L}(t, \theta)$ , it is possible to describe it implicitly as

$$\phi(\mathbf{x}) = 0. \quad (5.18)$$

Of course, the trick is deriving the form of  $\phi(\mathbf{x})$ . In principle this can be done by starting with the parametric equation  $\mathbf{x} = \mathbf{L}(t, \theta)$  and eliminating the variables  $t$  and  $\theta$  by using appropriate trigonometric and algebraic operations. As an alternative, the approach taken here will be to construct  $\phi(\mathbf{x})$  geometrically.

To begin, consider the equation for a right-circular cone with vertex of height  $d$  above the  $x_1$ - $x_2$  plane, and with base circle of radius  $r$ . This has the implicit equation

$$\phi_0(\mathbf{x}) = \frac{d^2}{r^2}(x_1^2 + x_2^2) - (x_3 - d)^2 = 0.$$

It is easy to check that  $\phi_0(d\mathbf{e}_3) = 0$  and  $\phi_0([r \cos \theta, r \sin \theta, 0]^T) = 0$ , indicating that both the vertex and the base circle satisfy the implicit expression.

Now consider applying particular *affine transformations* (i.e., deformations of the form  $\mathbf{y} = \mathbf{a}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ ) that preserve the circle shape. These will have the effect of rigidly moving the plane containing the circle, but need not be rigid-body motions acting on other planes in  $\mathbb{R}^3$ . These transformations are invertible, and  $\mathbf{x} = \mathbf{a}^{-1}(\mathbf{y}) = A^{-1}(\mathbf{y} - \mathbf{b})$ . This means that if (5.18) holds, then in the new coordinates the condition

$$\phi(\mathbf{a}^{-1}(\mathbf{y})) = 0 \tag{5.19}$$

must hold.

For example, consider the transformation of the form

$$\mathbf{y} = \mathbf{s}(\mathbf{x}) = \begin{pmatrix} x_1 - s_1 x_3/d \\ x_2 - s_2 x_3/d \\ x_3 \end{pmatrix} \iff \mathbf{x} = \mathbf{s}^{-1}(\mathbf{y}) = \begin{pmatrix} y_1 + s_1 y_3/d \\ y_2 + s_2 y_3/d \\ y_3 \end{pmatrix}. \tag{5.20}$$

This can be interpreted geometrically as a linear shear of all the planes parallel to the  $x_1$ - $x_2$  plane. This transformation leaves the  $x_1$ - $x_2$  plane fixed, and moves the point  $\mathbf{x} = d\mathbf{e}_3$  to  $\mathbf{y} = \mathbf{s}(d\mathbf{e}_3) = [-s_1, -s_2, d]^T$ . If this transformation is applied to every point in the right circular cone described by  $\phi_0(\mathbf{x})$ , then it will uniformly *shear* the cone, making it into a slanted cone with base at the same location and vertex at  $\mathbf{s}(d\mathbf{e}_3)$ . Following the rule in (5.19), the equation for this slanted cone would be  $\phi_1(\mathbf{y}) = \phi_0(\mathbf{s}^{-1}(\mathbf{y})) = 0$ , where  $\mathbf{y}$  is the vector of Cartesian coordinates for the space after the transformation has been applied.

Now suppose that we want to rigidly translate this slanted cone so that its vertex is moved from  $\mathbf{s}(d\mathbf{e}_3) = [-s_1, -s_2, d]^T$  to the point  $h\mathbf{e}_3$ . This can be done by applying an affine transformation of the following form to every point in the slanted cone:

$$\mathbf{z} = \mathbf{t}(\mathbf{y}) = \begin{pmatrix} y_1 + s_1 \\ y_1 + s_2 \\ y_3 + h - d \end{pmatrix} \iff \mathbf{y} = \mathbf{t}^{-1}(\mathbf{z}) = \begin{pmatrix} z_1 - s_1 \\ z_1 - s_2 \\ z_3 - (h - d) \end{pmatrix}. \tag{5.21}$$

This gives a translated and sheared cone defined by

$$\phi_2(\mathbf{z}) = \phi_1(\mathbf{t}^{-1}(\mathbf{z})) = \phi_0(\mathbf{s}^{-1}(\mathbf{t}^{-1}(\mathbf{z}))) = 0 \tag{5.22}$$

in the new set of Cartesian coordinates  $\mathbf{z}$ .

Now if  $\phi_2(\mathbf{z}) = 0$  defines the cone that is expected, it should be the case that  $\phi_2(h\mathbf{e}_3) = 0$ , which is in fact easy to verify. Note that  $\mathbf{s}(\mathbf{t}(\cdot)) \neq \mathbf{t}(\mathbf{s}(\cdot))$ . That is, the order in which these transformations are applied is important. And so  $\phi_0(\mathbf{s}^{-1}(\mathbf{t}^{-1}(\cdot))) \neq \phi_0(\mathbf{t}^{-1}(\mathbf{s}^{-1}(\cdot)))$ . These are two different surfaces.

Now suppose that it is desired to rotate the skewed cone defined by  $\phi_2(\mathbf{z}) = 0$  around the point  $\mathbf{z} = h\mathbf{e}_3$ . Such a rotation around a point that is not the origin is actually a rigid-body transformation of the form  $\mathbf{w} = \mathbf{r}(\mathbf{z}) = R\mathbf{z} + \mathbf{r}_0$ . For a general rigid-body transformation the conditions  $RR^T = \mathbb{I}$  and  $\det R = 1$  will hold, and  $\mathbf{r}_0$  will



be an arbitrary translation vector. However, for a particular rigid-body displacement that leaves the point  $\mathbf{z} = h\mathbf{e}_3$  fixed,<sup>3</sup>

$$\mathbf{r}(h\mathbf{e}_3) = R(h\mathbf{e}_3) + \mathbf{r}_0 = h\mathbf{e}_3,$$

which gives

$$\mathbf{r}_0 = h\mathbf{e}_3 - hR\mathbf{e}_3.$$

Therefore, the rotational transformation of interest is

$$\mathbf{w} = \mathbf{r}(\mathbf{z}) = R(\mathbf{z} - h\mathbf{e}_3) + h\mathbf{e}_3 \iff \mathbf{z} = R^T(\mathbf{w} - h\mathbf{e}_3) + h\mathbf{e}_3. \quad (5.23)$$

Using (5.22), the surface defined by  $\phi_3(\mathbf{w}) = \phi_2(\mathbf{r}^{-1}(\mathbf{w}))$  is then one of the form

$$\phi_0(\mathbf{s}^{-1}(\mathbf{t}^{-1}(\mathbf{r}^{-1}(\mathbf{w})))) = 0. \quad (5.24)$$

The problem becomes one of finding the free parameters in the transformations  $\mathbf{s}(\cdot)$ ,  $\mathbf{t}(\cdot)$ , and  $\mathbf{r}(\cdot)$  such that  $\mathbf{r}(\mathbf{t}(\mathbf{s}(\mathbf{c}(\theta))))$  is a copy of the circle  $\mathbf{c}(\theta)$  defined in (5.10) with particular properties. Namely, its center position should be specified as  $\mathbf{m} = [m_1, m_2, m_3]^T$  and the normal to the plane of the circle should be given by  $\mathbf{u}(\alpha, \beta)$  in (5.9). Since the transformations  $\mathbf{s}(\cdot)$  and  $\mathbf{t}(\cdot)$  do not change the orientation of the plane containing the circle (which is the  $x_1$ - $x_2$  plane, with normal  $\mathbf{e}_3$ ) it follows that whatever orientation the matrix  $R$  imparts to  $\mathbf{e}_3$  will be the orientation of the normal to the circle  $\mathbf{r}(\mathbf{t}(\mathbf{s}(\mathbf{c}(\theta))))$ . Therefore, we can set  $R = R_3(\alpha)R_2(\beta)$  and the result will be that  $R\mathbf{e}_3 = \mathbf{u}(\alpha, \beta)$ .

While the height of the point source of x-rays from the image plane is the constant value  $h$ , the parameters  $s_1$ ,  $s_2$ , and  $d$  can be adjusted so that

$$\mathbf{m} = \mathbf{r}(\mathbf{t}(\mathbf{s}(\mathbf{0}))),$$

which is the condition that the center of the circle is in the specified location. Explicitly composing the transformations gives

$$\mathbf{w} = R(\mathbf{x} + (s_1 - s_1x_3/d)\mathbf{e}_1 + (s_2 - s_2x_3/d)\mathbf{e}_2 - d\mathbf{e}_3) + h\mathbf{e}_3 = B\mathbf{x} + \mathbf{b}. \quad (5.25)$$

Evaluating at  $\mathbf{x} = \mathbf{0}$  gives

$$\mathbf{m} = R[s_1, s_2, -d]^T + h\mathbf{e}_3.$$

Therefore,

$$[s_1, s_2, -d]^T = R^T(\mathbf{m} - h\mathbf{e}_3)$$

and so  $B$  and  $\mathbf{b}$  in (5.25) can be written in terms of  $\mathbf{m}$  as

$$B = [R\mathbf{e}_1, R\mathbf{e}_2, (h\mathbf{e}_3 - \mathbf{m})/\mathbf{e}_3^T R^T(h\mathbf{e}_3 - \mathbf{m})] \quad \text{and} \quad \mathbf{b} = \mathbf{m} - h\mathbf{e}_3.$$

Now the elliptical projection of the circle of radius  $r$  with center at  $\mathbf{m}$  and normal given by  $\mathbf{u} = R\mathbf{e}_3$  is the intersection of the  $w_1$ - $w_2$  plane and the skewed cone (5.24) that has its vertex at  $\mathbf{w} = h\mathbf{e}_3$ .

---

<sup>3</sup>Every finite displacement of a rigid body in the plane leaves one point fixed, called a *pole*, and every spatial rigid-body displacement leaves a whole line fixed, called the *screw axis* [9]. By fixing the location of a point in space, the resulting motion is a pure rotation around that point.

In other words, the equation for the elliptical projection is

$$\phi_0(\mathbf{s}^{-1}(\mathbf{t}^{-1}(\mathbf{r}^{-1}(\mathbf{w}')))) = 0$$

where  $\mathbf{w}' = [w_1, w_2, 0]^T$ .

Since  $\phi_0(\mathbf{x})$  is the quadratic form

$$(\mathbf{x} - \mathbf{a})^T A (\mathbf{x} - \mathbf{a}) = 0 \quad \text{where} \quad \mathbf{a} = \begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} d^2/r^2 & 0 & 0 \\ 0 & d^2/r^2 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

with  $d = \mathbf{e}_3^T R^T (h\mathbf{e}_3 - \mathbf{m})$ , and since from (5.25),  $\mathbf{s}^{-1}(\mathbf{t}^{-1}(\mathbf{r}^{-1}(\mathbf{w}))) = B^{-1}(\mathbf{w} - \mathbf{b})$ , it follows that the equation for the projected ellipse is

$$(B^{-1}(\mathbf{w}' - \mathbf{b}) - \mathbf{a})^T A (B^{-1}(\mathbf{w}' - \mathbf{b}) - \mathbf{a}) = 0.$$

After some manipulations, this can be written as

$$\begin{pmatrix} w_1 - c_1 \\ w_2 - c_2 \end{pmatrix}^T \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix} \begin{pmatrix} w_1 - c_1 \\ w_2 - c_2 \end{pmatrix} = 1$$

where

$$C = [\mathbf{e}_1, \mathbf{e}_2]^T B^{-T} A B^{-1} [\mathbf{e}_1, \mathbf{e}_2].$$

This is simply the upper  $2 \times 2$  block of  $B^{-T} A B^{-1}$ .

The eigenvalues of  $C$  are obtained by solving the quadratic equation

$$\det(\lambda I - C) = \lambda^2 - \text{tr}(C)\lambda + \det(C) = 0. \quad (5.26)$$

The aspect ratio of the elliptical projection of the circle is then the square root of

$$\lambda_{min}/\lambda_{max} = \frac{c_{11} + c_{22} - \sqrt{(c_{11} + c_{22})^2 - 4c_{12}^2}}{c_{11} + c_{22} + \sqrt{(c_{11} + c_{22})^2 - 4c_{12}^2}}.$$

Note that unlike in the parametric approach, it is immediately clear when using this implicit formulation that this projection is an ellipse because affine transformations and sections of quadratic forms both result in quadratic forms [38]. And as long as the eigenvalues in the matrix  $C$  are positive, the resulting quadratic form must be an ellipse.

The problem in this section was one of analytic geometry with some ideas from elementary projective geometry. There are many other topics in geometry, including algebraic geometry, stochastic geometry, etc. The remainder of this chapter is devoted to the *differential geometry* of curves and surfaces in  $\mathbb{R}^3$ .

### 5.3 Differential Geometry of Curves

*Differential geometry* is concerned with characterizing the local shape of curves and surfaces using the tools of differential calculus, and relating these local shape properties at each point on the object of interest to its global characteristics. The following subsections respectively address the basic local and global differential geometry of curves in the plane and in three-dimensional space.

### 5.3.1 Local Theory of Curves

The *arc length* of a differentiable space curve,  $\mathbf{x}(t)$ , between points  $\mathbf{x}(t_1)$  and  $\mathbf{x}(t_2)$  is computed as

$$s(t_2) - s(t_1) = \int_{t_1}^{t_2} (\mathbf{x}'(t), \mathbf{x}'(t))^{\frac{1}{2}} dt \quad (5.27)$$

where  $\mathbf{x}' = d\mathbf{x}/dt$ . If the point from which arc length is measured is identified with  $t = 0$ , then  $s(0) = 0$ . Furthermore, if  $t$  and  $s$  are defined to increase in the same direction along the curve, then function  $s(t)$  will be non-decreasing. In most practical situations, this is a strictly increasing function, and so the inverse function  $t = t(s)$  can be defined. This means that given  $\mathbf{x}(t)$ , we can re-parameterize the curve in terms of arc length as  $\tilde{\mathbf{x}}(s) = \mathbf{x}(t(s))$ . When it is clear from the context that the curve is parameterized by arc length, the tilde can be dropped. A *unit tangent* vector can be assigned to each point on a parametric curve  $\mathbf{x}(t)$  by calculating

$$\mathbf{u}(t) \doteq \frac{1}{\left\| \frac{d\mathbf{x}}{dt} \right\|} \frac{d\mathbf{x}}{dt}.$$

When  $t = s$ , this reduces to

$$\mathbf{u}(s) = \frac{d\mathbf{x}}{ds}.$$

Since  $\mathbf{u}(s)$  is a unit vector  $\mathbf{u}(s) \cdot \mathbf{u}(s) = 1$ , and so

$$\frac{d}{ds} (\mathbf{u} \cdot \mathbf{u}) = 0 \quad \implies \quad \mathbf{u} \cdot \frac{d\mathbf{u}}{ds} = 0. \quad (5.28)$$

The (unsigned) *curvature* of an arc-length-parameterized curve (planar or spatial) is defined as

$$\kappa(s) \doteq \left( \frac{d\mathbf{u}}{ds} \cdot \frac{d\mathbf{u}}{ds} \right)^{\frac{1}{2}} = \left( \frac{d^2\mathbf{x}}{ds^2} \cdot \frac{d^2\mathbf{x}}{ds^2} \right)^{\frac{1}{2}}, \quad (5.29)$$

which is a measure of the amount of change in tangent direction at each value of arc length.

In the case of curves that are confined to the plane, it is also possible to give a sign to this curvature depending on whether the tangent to the curve bends clockwise or counterclockwise relative to its location at a prior value of arc length. The *signed curvature* of a planar curve is denoted as  $k(s)$ . By defining the (*principal*) *normal* vector as

$$\mathbf{n}_1(s) \doteq \frac{1}{\kappa(s)} \frac{d\mathbf{u}}{ds} \quad (5.30)$$

when  $\kappa(s) = \|d\mathbf{u}/ds\| \neq 0$ , it follows from (5.28) that

$$\mathbf{u}(s) \cdot \mathbf{n}_1(s) = 0.$$

Thus the tangent and normal vectors define two orthonormal vectors that move along with a point on the curve as  $s$  increases.

The geometry of a planar curve is completely specified by signed curvature (up to the way that the curve is embedded in space by arbitrary rigid-body motions). Or, stated a different way, the *intrinsic geometry* of a planar curve is completely specified by signed curvature, without regard to the way it is situated in space.

For a curve in three-dimensional space, at each  $s$  for which  $\kappa(s) \neq 0$ , three orthonormal vectors can be defined. This requires the definition of a second normal vector (called the binormal):

$$\mathbf{n}_2(s) \doteq \mathbf{u}(s) \times \mathbf{n}_1(s). \quad (5.31)$$

The frames of reference given by the positions  $\mathbf{x}(s)$  and orientations<sup>4</sup>  $Q_{FS} \doteq [\mathbf{u}(s), \mathbf{n}_1(s), \mathbf{n}_2(s)]$  for all values of  $s$  parameterizing the curve are called the *Frenet frames* attached to the curve.

The *torsion* of the curve is defined as

$$\tau(s) \doteq -\frac{d\mathbf{n}_2(s)}{ds} \cdot \mathbf{n}_1(s)$$

and is a measure of how much the curve bends out of the  $(\mathbf{u}, \mathbf{n}_1)$ -plane at each  $s$ .

The information contained in the collection of Frenet frames, the curvature, and the torsion, is termed the Frenet–Serret apparatus. It was published independently by Frenet (1852) and Serret (1851). It can be shown that [49, 53]

$$\frac{d}{ds} \begin{pmatrix} \mathbf{u}(s) \\ \mathbf{n}_1(s) \\ \mathbf{n}_2(s) \end{pmatrix} = \begin{pmatrix} 0 & \kappa(s) & 0 \\ -\kappa(s) & 0 & \tau(s) \\ 0 & -\tau(s) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u}(s) \\ \mathbf{n}_1(s) \\ \mathbf{n}_2(s) \end{pmatrix}. \quad (5.32)$$

The vectors  $\mathbf{u}$ ,  $\mathbf{n}_1$ , and  $\mathbf{n}_2$  are treated like scalars when performing the matrix–vector multiplication on the right-hand side of (5.32). This can be written in the different form

$$\begin{aligned} \frac{d}{ds}[\mathbf{u}(s), \mathbf{n}_1(s), \mathbf{n}_2(s)] &= [\kappa(s)\mathbf{n}_1(s), -\kappa(s)\mathbf{u}(s) + \tau(s)\mathbf{n}_2(s), -\tau(s)\mathbf{n}_1(s)] \\ &= -[\mathbf{u}(s), \mathbf{n}_1(s), \mathbf{n}_2(s)]\Lambda, \end{aligned}$$

or

$$\frac{dQ_{FS}}{ds} = -Q_{FS}\Lambda$$

where  $\Lambda$  is the skew-symmetric matrix in (5.32). The Frenet frame at each value of arc length is then given by a pair consisting of the rotation  $Q_{FS}(s)$  and position  $\mathbf{x}(s)$ . This pair is denoted as  $(Q_{FS}(s), \mathbf{x}(s))$ .

Given a space curve  $\mathbf{x}(t) \in \mathbb{R}^3$  where  $t$  is not necessarily arc length, the (unsigned) curvature is computed as

$$\kappa(t) = \frac{\|\mathbf{x}'(t) \times \mathbf{x}''(t)\|}{\|\mathbf{x}'(t)\|^3} \quad (5.33)$$

and the torsion is

$$\tau(t) = \frac{\det[\mathbf{x}'(t), \mathbf{x}''(t), \mathbf{x}'''(t)]}{\|\mathbf{x}'(t) \times \mathbf{x}''(t)\|^2}. \quad (5.34)$$

---

<sup>4</sup> $Q_{FS} = [\mathbf{u}(s), \mathbf{n}_1(s), \mathbf{n}_2(s)]$  is a rotation matrix (i.e., a  $3 \times 3$  orthogonal matrix with unit determinant).

### 5.3.2 Global Theory of Curves

While curvature and torsion are defined locally at each point on a space curve, some features of the global geometry of closed curves are dictated by curvature and torsion. In particular, the following two theorems should be noted, where  $\oint$  symbolizes an integral around a closed curve, which is invariant under the parametrization used.

**Theorem 5.1.** (Fenchel [27]): For smooth closed curves,

$$\oint \kappa(s) ds \geq 2\pi \quad (5.35)$$

with equality holding only for some kinds of planar ( $\tau(s) = 0$ ) curves.

**Theorem 5.2.** (Fary–Milnor [25, 50]): For closed space curves forming a knot

$$\oint \kappa(s) ds \geq 4\pi. \quad (5.36)$$

Many different kinds of knots exist, and their complexity can be assessed by counting the number of over-crossings that occur in planar projections. This is called the *bridging number*. This number can change for knots of the same type depending on the particular shape of the knot and the projection direction. However, if the minimum number of over-crossings is counted over all possible shapes and projection directions, then the result is called the *bridging index*,  $B$ . For any given knot,  $B$  is the smallest possible number of unknotted arcs that cross over others in any planar projection [2, 59]. This knot invariant was introduced by Schubert [62]. For the “unknot”  $B \doteq 1$ , and for all non-trivial knots  $B \geq 2$ .

The inequalities in (5.35) and (5.36) can be combined and tightened by replacing  $4\pi$  with  $2\pi B$  where  $B$  is the bridging index of the knot [28, 45].

In contrast to the above theorems, for any closed smooth planar curve that does not cross itself<sup>5</sup> and is parameterized such that its interior is to the left side of the tangent,

$$\oint k(s) ds = 2\pi, \quad (5.37)$$

where  $k(s)$  is the *signed curvature* of the curve. The sign is given such that  $|k(s)| = \kappa(s)$  with  $k(s) > 0$  for counterclockwise bending and  $k(s) < 0$  for clockwise bending.

The famous *Jordan curve theorem* states that such a curve divides the plane into two parts: one that is enclosed by the curve (called the interior), and one that is exterior to it. The part of the plane enclosed by a simple closed curve is sometimes called a *simple planar body* or *simply connected planar region*. For any closed curve that is contained inside of a simple planar body, transformations can be constructed that shrink the curve to a single point while the series of shrinking curves are confined to remain within the planar body. In contrast, if a (non-simple) planar body has an interior hole, and a closed curve in the body circumscribes the hole, then such a curve cannot be shrunk to a point and still remain in the body.

A non-simply connected planar region can be constructed by starting with a simple planar body and introducing internal boundaries by “cutting out holes” that are bounded by curves. The *genus*,  $\gamma(B)$ , of a planar body  $B$  can be defined by counting

<sup>5</sup>A closed curve that does not cross itself is called a *simple curve*.

the number of “holes” that it has. The *Euler characteristic* of  $B$ , denoted as  $\chi(B)$ , is obtained by subdividing, or tessellating, the body into disjoint polygonal regions, the union of which is the body, counting the number of polygonal faces,  $f$ , edges,  $e$ , and vertices,  $v$ , and using the formula

$$\boxed{\chi(B) = v(B) - e(B) + f(B)}. \quad (5.38)$$

Interestingly, for a planar body

$$\chi(B) = 1 - \gamma(B). \quad (5.39)$$

Whereas  $\gamma(B)$  is the number of holes in the body,  $\chi(B)$  is the number of closed curves on the exterior boundary (of which there is only one) minus the number of interior boundary curves (of which there are  $\gamma(B)$ ).

If a planar body that initially has one simple boundary curve is deformed so as to enclose some area, and points on the boundary are “glued” to each other, then the result is two bounding curves with opposing orientation (i.e., one turns clockwise and the other counterclockwise). It follows from (5.37) that in this case

$$\int_{\partial B} k(s) ds = 2\pi\chi(B) \quad (5.40)$$

where  $\partial B$  denotes the union of all boundary curves of  $B$ . This is shown in Figure 5.4.

Even though  $\kappa(s)$  (and hence  $k(s)$ ) was defined for curves that are at least twice continuously differentiable, (5.40) can be easily modified to be valid for jagged bodies with boundaries that are only piecewise differentiable. This is because, as Euler observed, the signed curvature of a planar curve can be viewed as the rate of change with respect to arc length of the counterclockwise angle made by the tangent and any fixed line (such as the  $x$ -axis). Written in a different way, this is

$$\theta(s_2) - \theta(s_1) = \int_{s_1}^{s_2} k(s) ds.$$

Therefore when  $\partial B$  is piecewise differentiable, the total angle traversed still can be made to equal  $2\pi\chi(B)$  if at each point where there is a jump in direction the curvature is viewed as a Dirac delta function scaled by the amount of angular change between the tangents to the curve just before and just after each jump.

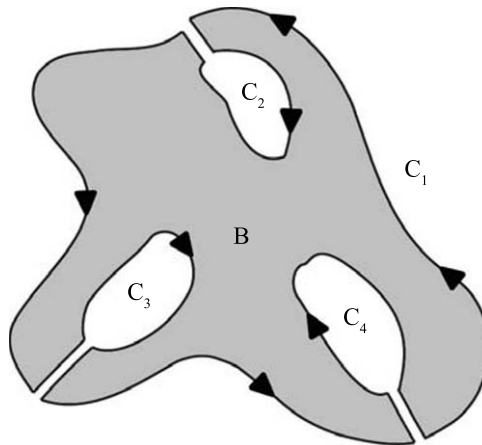
To illustrate this concept, consider the rectangular array consisting of  $L \times W$  unit squares. For this body  $v = (L + 1)(W + 1)$ ,  $e = L(W + 1) + W(L + 1)$ , and  $f = L \cdot W$ . Therefore by (5.38),  $\chi(B) = 1$ . The boundary curve consists of four straight line segments (each having zero curvature) joined by four corners representing four discrete jumps of  $\pi/2$  radians. If  $s = 0$  denotes the lower right corner of the rectangle, and setting  $k(s) = \pi/2[\delta(s - L) + \delta(s - L - W) + \delta(s - 2L - W) + \delta(s - 2L - 2W)]$ , it follows that (5.40) still holds.

Now suppose that certain “sculpting” operations are performed on this rectangle. If one square is removed from the periphery of the rectangle, and the boundary is restored, then the change in Euler characteristic is zero. There are two cases: removal of a square from a straight section of the boundary, and removal of one of the corner squares. Removal of a corner square means  $\Delta f = -1$ ,  $\Delta e = -2$ , and  $\Delta v = -1$ , and so  $\Delta\chi(B) = \Delta v - \Delta e + \Delta f = 0$ . Removal of a square from the straight section gives  $\Delta f = -1$ ,  $\Delta e = -1$ , and  $\Delta v = 0$ , again resulting in  $\Delta\chi(B) = 0$ , indicating that

$\chi(B) = 1$ . In this way the exterior of the rectangle can be “chipped away” without changing its Euler characteristic. For example, it is easy to verify that Figures 5.5(a) and 5.5(b) have the same Euler characteristic.

In contrast, introducing a void into the body changes its Euler characteristic. If one square is removed from deep without the body  $\Delta f = -1$ ,  $\Delta e = 0$ , and  $\Delta v = 0$ , and so  $\Delta\chi(B) = -1$ , or  $\chi(B) = 0$ . If an  $L$ -shaped void is introduced as shown in Figure 5.5(c), then  $\Delta f = -4$ ,  $\Delta e = -3$ , and  $\Delta v = 0$  and again  $\Delta\chi(B) = -1$ , or  $\chi(B) = 0$ .

Finally, if a whole vertical line of squares are removed, splitting the original rectangular body into two disjoint pieces, then  $\Delta f = -L$ ,  $\Delta e = -(L + 1)$  and  $\Delta v = 0$  and  $\Delta\chi(B) = +1$ , or  $\chi(B) = +2$ , as illustrated in Figure 5.5(d). The same effect on the Euler characteristic would result from splitting the body by removal of a horizontal line of squares. In each of the above examples, it can be verified that (5.40) holds.



**Fig. 5.4.** Global Topological Features of a Planar Body are Dictated by Signed Curvature

For a planar object, it is also possible to define the Euler characteristic of the boundary as

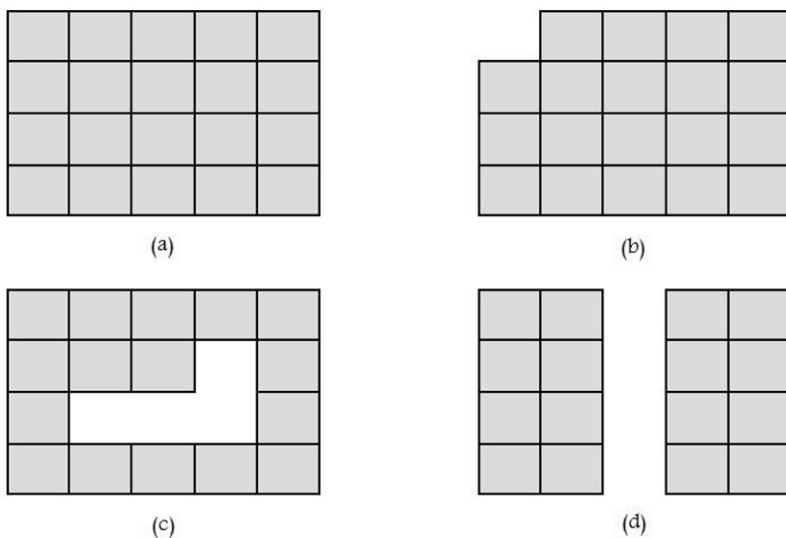
$$\chi(\partial B) = v(\partial B) - e(\partial B). \tag{5.41}$$

It is important to note that  $v$  and  $e$  in (5.41) are *not* the same as in (5.38). Returning to the  $L \times W$  array of squares, the number of boundary vertices can be counted as  $v = 2(L + 1) + 2(w - 1)$  and edges are  $e = 2(L + W)$ . Therefore  $\chi(\partial B) = 0$ . And this will remain true if the initial array of squares is sculpted, voids are introduced, and even if it is split into pieces. If the boundary consists of the union of several simple closed curves  $\{\partial_i B\}$  that are disjoint, i.e.,  $\partial_i B \cap \partial_j B$  over all  $i$  and  $j$ , then  $\chi(\partial B) = 0$  regardless of the topological features of the body.

### 5.4 Differential Geometry of Surfaces in $\mathbb{R}^3$

This section addresses the differential geometry of surface. Closed surfaces and the solid bodies that they enclose are of particular interest. For a *spatial body*  $B$  (i.e., a region in  $\mathbb{R}^3$  with finite non-zero volume), the surface area over the boundary of  $B$ ,

$$F = \int_{\partial B} dS,$$



**Fig. 5.5.** Topological Operations on Body Divided into Squares: (a) An Initial Rectangular Grid; (b) Removal of One Square from the Perimeter; (c) Creation of an L-Shaped Void; (d) Cutting the Body into Two Disjoint Pieces

and volume,

$$V = \int_B dV,$$

are concepts with which engineers and physical scientists are very familiar. *Gauss' divergence theorem*,

$$\int_B \operatorname{div}(\mathbf{f}) dV = \int_{\partial B} \mathbf{f} \cdot \mathbf{n} dS, \quad (5.42)$$

says that the divergence of a differentiable vector field  $\mathbf{f}(\mathbf{x})$ , defined as  $\operatorname{div}(\mathbf{f}) = \sum_{i=1}^3 \partial f / \partial x_i$ , integrated over the volume of  $B$  results in the same answer as integrating  $\mathbf{f} \cdot \mathbf{n}$  over the closed surface containing the body. Here  $\mathbf{n}$  is the outward-pointing normal to the surface.

An immediate consequence of the divergence theorem is that the volume of a body can be computed as a surface integral, by simply constructing a vector field on  $\mathbb{R}^3$  such that  $\operatorname{div}(\mathbf{f}) = 1$ . In particular, letting  $\mathbf{f} = A\mathbf{x}$  for  $A \in \mathbb{R}^{3 \times 3}$ , then  $\operatorname{div}(\mathbf{f}) = \operatorname{tr}(A)$ . Therefore, if  $A$  is any matrix with  $\operatorname{tr}(A) \neq 0$ , then the volume of the body  $B$  can be computed using a surface area integral:

$$V = \frac{1}{\operatorname{tr}(A)} \int_{\partial B} \mathbf{n}^T A \mathbf{x} dS, \quad (5.43)$$

which can be a useful computational tool, since the numerical approximation of a volume involves performing computations over a three-dimensional domain, whereas performing computations of a surface integral is over a two-dimensional domain. The vector field  $\mathbf{f}(\mathbf{x})$  can be chosen so as to make computations as convenient as possible.



### 5.4.1 The First and Second Fundamental Forms

Consider a two-dimensional surface parameterized as  $\mathbf{x}(\mathbf{q})$  where  $\mathbf{x} \in \mathbb{R}^3$  and  $\mathbf{q} \in \mathbb{R}^2$ . The most fundamental quantity from which geometrical properties of the surface can be derived is the metric tensor, which can be viewed in terms of the parameters  $(q_1, q_2)$  as the  $2 \times 2$  real symmetric matrix  $G = [g_{ij}]$ , where

$$g_{ij} = \frac{\partial \mathbf{x}}{\partial q_i} \cdot \frac{\partial \mathbf{x}}{\partial q_j} \quad (5.44)$$

for  $i, j \in \{1, 2\}$ .  $G = [g_{ij}]$  contains all of the information about how the lengths of curves and the area of patches within the surface are calculated.

The *first fundamental form* of the surface is defined as

$$\mathcal{F}^{(1)}(d\mathbf{q}, d\mathbf{q}) \doteq d\mathbf{q}^T G(\mathbf{q}) d\mathbf{q}. \quad (5.45)$$

If we make a change of coordinates to parameterize the surface described as  $\mathbf{q} = \mathbf{q}(\mathbf{s})$ , then by the chain rule,  $d\mathbf{q} = J(\mathbf{s})d\mathbf{s}$ , where

$$J(\mathbf{v}) = \left[ \frac{\partial \mathbf{q}}{\partial s_1}, \frac{\partial \mathbf{q}}{\partial s_2} \right].$$

Therefore, due to the invariance of the fundamental forms under coordinate changes,  $\mathcal{F}_q^{(1)} = \mathcal{F}_s^{(1)}$ , or

$$d\mathbf{s}^T G_s(\mathbf{s}) d\mathbf{s} = d\mathbf{s}^T J^T(\mathbf{s}) G_q(\mathbf{q}(\mathbf{s})) J(\mathbf{s}) d\mathbf{s}.$$

In other words, the metric tensor transforms under coordinate change as

$$G_s(\mathbf{s}) = J^T(\mathbf{s}) G_q(\mathbf{q}(\mathbf{s})) J(\mathbf{s}). \quad (5.46)$$

For the closed surfaces that will be examined later, there will be no need to change between parameterizations, and therefore the subscripts  $q$  and  $s$  on  $G$  need not be stated.

If  $G$  is known, then many important quantities can be computed from  $G$ . For example, the arc length of a curve defined by  $\tilde{\mathbf{x}}(t) = \mathbf{x}(\mathbf{q}(t))$  (which by definition is contained within the surface) for  $t \in [t_1, t_2]$  is

$$L(t_1, t_2) = \int_{t_1}^{t_2} \left( \frac{d\tilde{\mathbf{x}}}{dt} \cdot \frac{d\tilde{\mathbf{x}}}{dt} \right)^{\frac{1}{2}} dt = \int_{t_1}^{t_2} \left( \left[ \frac{d\mathbf{q}}{dt} \right]^T G(\mathbf{q}) \frac{d\mathbf{q}}{dt} \right)^{\frac{1}{2}} dt.$$

An *element of surface area* is calculated as

$$dS = |G(\mathbf{q})|^{\frac{1}{2}} dq_1 dq_2, \quad (5.47)$$

where  $|G(\mathbf{q})|^{\frac{1}{2}} \doteq \sqrt{\det G(\mathbf{q})}$ .

In addition to the direct value of the metric tensor in calculating quantities of interest, the inverse of the metric tensor also arises in applications. In physics, the notation  $g^{ij}$  is used as shorthand for the entries of the inverse of  $G$ . That is,  $G^{-1} = [g^{ij}]$ . This will be useful notation here.

For example, a gradient vector field of a real-valued function on the surface can be defined as

$$\boxed{\text{grad}(f)_i \doteq \sum_j g^{ij} \frac{\partial f}{\partial q_j}}. \quad (5.48)$$

And the divergence of a vector field on the surface can be defined as

$$\boxed{\operatorname{div}(\mathbf{f}) \doteq |G|^{-\frac{1}{2}} \sum_i \frac{\partial}{\partial q_i} (|G|^{\frac{1}{2}} f_i)}. \quad (5.49)$$

Sometimes the parentheses are dropped to lighten the notation a little bit, and these are respectively denoted as  $\operatorname{grad} f$  and  $\operatorname{div} \mathbf{f}$ .

The Laplacian (or *Laplace–Beltrami operator*) of a smooth real-valued function is defined as the divergence of the gradient:

$$\boxed{\operatorname{div}(\operatorname{grad} f) \doteq |G|^{-\frac{1}{2}} \sum_i \frac{\partial}{\partial q_i} \left( |G|^{\frac{1}{2}} \sum_j g^{ij} \frac{\partial f}{\partial q_j} \right)}. \quad (5.50)$$

For 2D surfaces in  $\mathbb{R}^3$ , the above sums are for  $i$  and  $j$  ranging over the set  $\{1, 2\}$ . However, the exact same formulas apply to higher-dimensional surfaces, in which case the sums are taken over  $\{1, 2, \dots, n\}$ .

For two-dimensional surfaces in three-dimensional space,

$$|G(q_1, q_2)| = \left\| \frac{\partial \mathbf{x}}{\partial q_1} \times \frac{\partial \mathbf{x}}{\partial q_2} \right\|^2 \quad (5.51)$$

where  $\times$  denotes the vector cross product.

Furthermore, the unit normal of a two-dimensional surface in  $\mathbb{R}^3$  is computed at each point defined by  $(q_1, q_2)$  as

$$\mathbf{n} = |G|^{-\frac{1}{2}} \left( \frac{\partial \mathbf{x}}{\partial q_1} \times \frac{\partial \mathbf{x}}{\partial q_2} \right).$$

Since  $\mathbf{n} \cdot \mathbf{n} = 1$ , it follows that differentiation of both sides with respect to  $q_1$  or  $q_2$  yields

$$\mathbf{n} \cdot \frac{\partial \mathbf{n}}{\partial q_1} = \mathbf{n} \cdot \frac{\partial \mathbf{n}}{\partial q_2} = 0.$$

In other words,  $\partial \mathbf{n} / \partial q_1$  and  $\partial \mathbf{n} / \partial q_2$  are both in the tangent plane of the surface, which is spanned by the vectors  $\partial \mathbf{x} / \partial q_1$  and  $\partial \mathbf{x} / \partial q_2$ . The *second fundamental form* of a surface is defined as

$$\mathcal{F}^{(2)}(d\mathbf{q}, d\mathbf{q}) = -d\mathbf{x} \cdot d\mathbf{n},$$

where the vectors  $\mathbf{x}$  and  $\mathbf{n}$  are the position and normal at any point on the surface, and  $d\mathbf{x}$  and  $d\mathbf{n}$  are contained in the tangent plane to the surface at that point.

Let the matrix  $L$  be defined by its entries:

$$L_{ij} = \frac{\partial^2 \mathbf{x}}{\partial q_i \partial q_j} \cdot \mathbf{n}. \quad (5.52)$$

The matrix  $L = [L_{ij}]$  contains information about how curved the surface is. For example, for a plane  $L_{ij} = 0$ .

It can be shown that

$$\mathcal{F}^{(2)}(d\mathbf{q}, d\mathbf{q}) = d\mathbf{q}^T L(\mathbf{q}) d\mathbf{q}, \quad (5.53)$$

and that the matrix  $L$  transforms in the same way as  $G$  in (5.46).

### 5.4.2 Curvature

The first and second fundamental forms (5.45) and (5.53) defined in the previous subsection can be used together to compute the curvature of a 2D surface in  $\mathbb{R}^3$ . But before simply providing the equations, some geometrical explanation is in order. Consider a 2D surface parameterized as  $\mathbf{x}(q_1, q_2) \in \mathbb{R}^3$ . Let  $q_i = q_i(s)$  so that  $\mathbf{c}(s) = \mathbf{x}(q_1(s), q_2(s))$  is an arc-length-parameterized curve in  $\mathbb{R}^3$ . The tangent of this curve is computed as the derivative of position with respect to arc length from the chain rule as

$$\mathbf{c}'(s) = \sum_{i=1}^2 \frac{\partial \mathbf{x}}{\partial q_i} q_i'(s)$$

where  $' = d/ds$ . The second derivative of  $\mathbf{c}(s)$  with respect to arc length gives

$$\mathbf{c}''(s) = \sum_{i=1}^2 \frac{\partial \mathbf{x}}{\partial q_i} q_i''(s) + \sum_{i=1}^2 \sum_{j=1}^2 \frac{\partial^2 \mathbf{x}}{\partial q_i \partial q_j} q_i'(s) q_j'(s) = \kappa(s) \mathbf{n}_1(s).$$

The normal to the surface at the point where the curve  $\mathbf{c}(s)$  passes through is generally not the same as the normal  $\mathbf{n}_1(s)$  or binormal  $\mathbf{n}_2(s)$  of the curve. That is,

$$\mathbf{n}_1(s) \neq \mathbf{n}(q_1(s), q_2(s)) \neq \mathbf{n}_2(s).$$

However, all of these normals are perpendicular to the tangent to the curve,  $\mathbf{c}'(s)$ . Furthermore, a new normal to the curve  $\mathbf{c}(s)$  can be defined that is perpendicular to both the tangent to the curve and the normal to the surface as

$$\mathbf{m}(s) = \mathbf{n}(q_1(s), q_2(s)) \times \mathbf{c}'(s).$$

Since by definition it is perpendicular to  $\mathbf{n}(s)$ , this vector must be contained in the tangent plane to the surface that is spanned by  $\partial \mathbf{x} / \partial q_1|_{q_1(s), q_2(s)}$  and  $\partial \mathbf{x} / \partial q_2|_{q_1(s), q_2(s)}$ .

$\mathbf{m}(s)$  and  $\mathbf{n}(s)$  together form an orthonormal basis for describing any vector that is normal to  $\mathbf{c}'(s)$ . This “surface-centric” coordinate system consisting of unit vectors  $\mathbf{u}(s)$ ,  $\mathbf{m}(s)$ , and  $\mathbf{n}(s)$ , which is called the *Darboux frame*, is an alternative basis from the “curve-centric” one defined by  $\mathbf{u}(s)$ ,  $\mathbf{n}_1(s)$ , and  $\mathbf{n}_2(s)$ . Therefore, since  $\mathbf{c}''(s) = \kappa(s) \mathbf{n}_1(s)$  is a vector that is normal to  $\mathbf{c}'(s)$ , it is possible to expand it in the new basis as

$$\kappa(s) \mathbf{n}_1(s) = \kappa_n(s) \mathbf{n}(q_1(s), q_2(s)) + \kappa_g(s) \mathbf{m}(s)$$

where

$$\kappa_n(s) = \kappa(s) \mathbf{n}_1(s) \cdot \mathbf{n}(q_1(s), q_2(s)) \quad \text{and} \quad \kappa_g(s) = \kappa(s) \mathbf{n}_1(s) \cdot \mathbf{m}(s).$$

The quantities  $\kappa_g(s)$  and  $\kappa_n(s)$  are called the *geodesic* and *normal* curvature, respectively. Since  $\mathbf{m}(s) \cdot \mathbf{n}(s) = 0$  and  $\|\mathbf{m}(s)\| = \|\mathbf{n}(s)\| = 1$ , it follows that

$$\kappa^2(s) = \kappa_g^2(s) + \kappa_n^2(s). \quad (5.54)$$

Imagine the set of all possible smooth arc-length-parameterized curves contained within the surface and constrained to pass through the point  $\mathbf{x}(q_1, q_2)$  with a particular tangent direction. Within this set of curves, some will be “wiggling around” a lot inside the surface, and others will be very taut. The most taut ones will have  $\kappa_g(s) = 0$ . These can be used to define how curved the surface is. In particular, it is possible to search

for the minimum and maximum values of  $\kappa_n$  over all taut curves passing through a point on the surface. From these, two tangent directions in a two-dimensional surface can be found at each point (one that maximizes  $\kappa_n$ , and one that minimizes it). These two values of normal curvature provide important information about the local shape of the surface. Of course, the quantities of interest can be obtained without performing an explicit search over all possible taut curves with all possible tangent directions. But this requires some additional notation and observations about the problem at hand.

The first observation that will help is that the normal curvature of every curve  $\mathbf{c}(s) = \mathbf{x}(q_1(s), q_2(s))$  that passes through the point  $\mathbf{c}(s_0) = \mathbf{c}_0$  and has unit tangent vector  $\mathbf{u}(s_0) = d\mathbf{c}/ds|_{s=s_0}$  will have the same normal curvature at that point,  $\kappa_n(s_0)$ . (See Exercise 5.11.) Therefore, it is not necessary to construct the most taut curves passing through a point on a surface to measure how curved the surface is.

The second observation is that the normal curvature from the above definitions is equal to

$$\kappa_n = \mathbf{c}'' \cdot \mathbf{n}, \quad (5.55)$$

and this in turn can be written in the form

$$\kappa_n = \mathcal{F}^{(2)}(d\mathbf{q}, d\mathbf{q}) / \mathcal{F}^{(1)}(d\mathbf{q}, d\mathbf{q}), \quad (5.56)$$

which is independent of the magnitude of  $d\mathbf{q}$ . Therefore, using the following substitution,

$$d\mathbf{q} = G^{-\frac{1}{2}} \mathbf{v} \quad \text{where} \quad \mathbf{v} \cdot \mathbf{v} = 1,$$

and

$$\kappa_n = \mathcal{F}^{(2)}(G^{-\frac{1}{2}} \mathbf{v}, G^{-\frac{1}{2}} \mathbf{v}) = \mathbf{v}^T G^{-\frac{1}{2}} L G^{-\frac{1}{2}} \mathbf{v}. \quad (5.57)$$

The maximal and minimal values of  $\kappa_n$  are attained when  $\mathbf{v}$  points along eigenvectors of  $G^{-\frac{1}{2}} L G^{-\frac{1}{2}}$  corresponding to the maximal and minimal eigenvalues. And so the quantities that would have been obtained by searching over all taut curves with tangent vectors passing through all planar angles can be obtained by performing a simple eigenvalue–eigenvector decomposition of the  $2 \times 2$  matrix  $G^{-1/2} L G^{-1/2}$ .

For a two-dimensional surface in  $\mathbb{R}^3$ , the vector of partial derivatives  $\partial^2 \mathbf{x} / \partial q_i \partial q_j$  can be decomposed into a part that points in the direction of the normal, and a part that lies in the tangent plane. The part that lies in the tangent plane can then be expressed as a projection onto the vectors  $\partial \mathbf{x} / \partial q_k$  for  $k = 1, 2$ . This is written as

$$\frac{\partial^2 \mathbf{x}}{\partial q_i \partial q_j} = L_{ij} \mathbf{n} + \sum_k \Gamma_{ij}^k \frac{\partial \mathbf{x}}{\partial q_k} \quad \text{where} \quad \Gamma_{ij}^k = \sum_l \left( \frac{\partial^2 \mathbf{x}}{\partial q_i \partial q_j} \cdot \frac{\partial \mathbf{x}}{\partial q_l} \right) g^{kl}. \quad (5.58)$$

The *Christoffel symbol* (of the second kind),  $\Gamma_{ij}^k$ , can also be expressed completely in terms of the elements of the metric tensor as

$$\Gamma_{ij}^k = \frac{1}{2} \sum_l \left[ \frac{\partial g_{il}}{\partial q_j} - \frac{\partial g_{ij}}{\partial q_l} + \frac{\partial g_{lj}}{\partial q_i} \right] g^{lk}. \quad (5.59)$$

It can be shown that [49]

$$\kappa_n(s) = \sum_{i,j} L_{ij}(q_1(s), q_2(s)) q_1'(s) q_2'(s)$$

and

$$\kappa_g(s) = \sum_k \left[ q_k'' + \sum_{i,j} \Gamma_{ij}^k q_i' q_j' \right] \frac{\partial \mathbf{x}}{\partial q_k}.$$

The condition  $\kappa_g(s) = 0$  is automatically satisfied when

$$q_k'' + \sum_{i,j} \Gamma_{ij}^k q_i' q_j' = 0. \quad (5.60)$$

Such curves are called *geodesics*.

The *Riemannian curvature* is the four-index tensor given in component form as [49]

$$R_{ijk}^l \doteq \frac{\partial \Gamma_{ik}^l}{\partial q_j} - \frac{\partial \Gamma_{ij}^l}{\partial q_k} + \sum_m (\Gamma_{ik}^m \Gamma_{mj}^l - \Gamma_{ij}^m \Gamma_{mk}^l). \quad (5.61)$$

This can be expressed in terms of the coefficients of the second fundamental form and inverse of the metric tensor as [49]

$$R_{ijk}^l = L_{ik} \sum_m g^{lm} L_{mj} - L_{ij} \sum_m g^{lm} L_{mk}. \quad (5.62)$$

The *principal curvatures* of a 2D surface that minimize and maximize (5.57) can be defined as the roots  $\kappa_1$  and  $\kappa_2$  of the characteristic equation<sup>6</sup>

$$p_1(\kappa) = \det(L - \kappa G) = 0. \quad (5.63)$$

Of course, these roots will be the same as those that solve

$$p_2(\kappa) = \det(G^{-1}L - \kappa I) = \kappa^2 - \text{tr}(G^{-1}L)\kappa + \det(G^{-1}L) = 0.$$

Since  $G$  is a symmetric positive definite matrix in any parametrization, it is possible to write  $G = G^{\frac{1}{2}} G^{\frac{1}{2}}$  and  $G^{-1} = G^{-\frac{1}{2}} G^{-\frac{1}{2}}$ , which means that  $\det(G^{-1}L) = \det(G^{-\frac{1}{2}} L G^{-\frac{1}{2}})$ , and likewise for the trace. This is convenient because  $G^{-\frac{1}{2}} L G^{-\frac{1}{2}}$  is symmetric. If  $\det(L) \neq 0$ , it is also possible to do trace and determinant computations with the symmetric matrix  $L^{\frac{1}{2}} G^{-1} L^{\frac{1}{2}}$ . But  $\det(L) \neq 0$  will not always hold (for example,  $\det(L) \equiv 0$  for a cylinder or plane), whereas  $\det(G) \neq 0$ , except at a set of measure zero where the parametric description breaks down.

From this, the *Gaussian curvature*,  $k(q_1, q_2)$ , is computed as

$$k = \kappa_1 \kappa_2 = \det(G^{-1}L) = |G|^{-\frac{1}{2}} \mathbf{n} \cdot \left( \frac{\partial \mathbf{n}}{\partial q_1} \times \frac{\partial \mathbf{n}}{\partial q_2} \right). \quad (5.64)$$

This quantity, which is computed above in an extrinsic way (i.e., using the normal to the surface in the ambient space  $\mathbb{R}^3$ ), can be computed alternatively as the intrinsic quantity (i.e., only depending on the metric tensor) as

$$k \doteq |G|^{-\frac{1}{2}} R_{1212}, \quad (5.65)$$

where, by definition,  $R_{1212} = \sum_m R_{121m}^m g_{m2}$ . Equation (5.65) can be viewed as a statement of Gauss' *Theorema Egregium* (or, remarkable theorem).

<sup>6</sup>This is because  $G^{-1}L$ ,  $LG^{-1}$ , and  $G^{-\frac{1}{2}} L G^{-\frac{1}{2}}$  all have the same eigenvalues.

The *mean sectional curvature* (or simply *mean curvature*) is defined as

$$m \doteq \frac{1}{2}(\kappa_1 + \kappa_2) = \frac{1}{2}\text{trace}(G^{-1}L). \quad (5.66)$$

This is the average value of  $\kappa_n$  over curves on the surface pointing in all tangent directions and passing through the point  $\mathbf{x}(q_1, q_2)$ . These tangent directions can be generated, for example, by calculating the unit eigenvectors,  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , of  $G^{-\frac{1}{2}}LG^{-\frac{1}{2}}$ . Then letting  $\mathbf{v}(\theta) = \mathbf{v}_1 \cos \theta + \mathbf{v}_2 \sin \theta$  and using (5.57), the normal curvature corresponding to each  $\theta$  is

$$\kappa_n(q_1, q_2; \theta) = \mathcal{F}^{(2)}(G^{-\frac{1}{2}}\mathbf{v}(\theta), G^{-\frac{1}{2}}\mathbf{v}(\theta)) = \kappa_1 \cos^2 \theta + \kappa_2 \sin^2 \theta.$$

Therefore, averaging over  $\theta$  gives

$$\frac{1}{2\pi} \int_0^{2\pi} \kappa_n(q_1, q_2; \theta) d\theta = m(q_1, q_2).$$

This is the same as that resulting from slicing the surface with all planes containing the point  $\mathbf{x}(\mathbf{q})$  and the surface normal, computing the curvature of all curves in these planar sections, and averaging.

The integrals of the Gaussian and mean curvatures over the entirety of a closed surface figure prominently in computations of the probability of intersections of randomly moving bodies, as articulated in the subfield of mathematics known as *integral geometry*, which will be discussed in Volume 2. These integrals of total curvature are defined as

$$K \doteq \int_S k dS \quad (5.67)$$

$$M \doteq \int_S m dS. \quad (5.68)$$

These are respectively called the *total Gaussian curvature* and *total mean curvature*.

These concepts generalize to  $\mathbb{R}^n$ , and even to more abstract submanifolds of intrinsically defined manifolds. One immediate difficulty is that the normal vector can no longer be defined using the vector cross product, which is only defined in  $\mathbb{R}^3$ . However, the concept of the tangent plane still holds, and alternative methods for describing principal curvatures are well known in the differential geometry literature.

The following subsections demonstrate these definitions on concrete examples in three dimensions.

### 5.4.3 The Sphere

A sphere of radius  $R$  can be parameterized as

$$\mathbf{x}(\phi, \theta) = \begin{pmatrix} R \cos \phi \sin \theta \\ R \sin \phi \sin \theta \\ R \cos \theta \end{pmatrix} \quad (5.69)$$

where  $0 \leq \theta \leq \pi$  and  $0 \leq \phi \leq 2\pi$ .

The corresponding metric tensor is

$$G(\phi, \theta) = \begin{pmatrix} g_{\phi, \phi} & g_{\phi, \theta} \\ g_{\theta, \phi} & g_{\theta, \theta} \end{pmatrix} = \begin{pmatrix} R^2 \sin^2 \theta & 0 \\ 0 & R^2 \end{pmatrix}.$$

Clearly,  $\sqrt{\det G(\phi, \theta)} = R^2 \sin \theta$  (there is no need for absolute value signs since  $\sin \theta \geq 0$  for  $\theta \in [0, \pi]$ ). The element of surface area is therefore

$$dS = R^2 \sin \theta \, d\phi \, d\theta.$$

Surface area of the sphere is computed as

$$F = \int_0^\pi \int_0^{2\pi} \sin \theta \, d\phi \, d\theta = 4\pi R^2.$$

The volume of the ball of radius  $R$  can be computed in spherical coordinates in  $\mathbb{R}^3$  (i.e., treating  $\mathbb{R}^3$  as the surface of interest) and restricting the range of parameters defined by the interior of the ball. Alternatively, the divergence theorem can be used in the form of (5.43). If we let the matrix  $A = \mathbb{I}/3$ , then  $\mathbf{n}^T A \mathbf{x} = R/3$  and

$$V = \frac{R}{3} \int_{\partial B} dS = \frac{R}{3} \cdot 4\pi R^2 = \frac{4}{3}\pi R^3.$$

The volume of the ball enclosed by the sphere of radius  $R$ , and surface area of the sphere are summarized, respectively, as

$$V = \frac{4}{3}\pi R^3; \quad F = 4\pi R^2. \quad (5.70)$$

The inward-pointing normal for the sphere is simply  $\mathbf{n} = -\mathbf{x}/R$ , and

$$L(\phi, \theta) = \begin{pmatrix} R \sin^2 \theta & 0 \\ 0 & R \end{pmatrix}.$$

Therefore,

$$G^{-1}L = \begin{pmatrix} 1/R & 0 \\ 0 & 1/R \end{pmatrix}.$$

It follows that

$$m = \frac{1}{2} \operatorname{tr}(G^{-1}L) = 1/R$$

and

$$k = \det(G^{-1}L) = 1/R^2.$$

Since these are both constant, it follows that integrating each of them over the sphere of radius  $R$  is the same as their product with the surface area:

$$M = 4\pi R; \quad K = 4\pi.$$

#### 5.4.4 The Ellipsoid of Revolution

Consider an *ellipsoid of revolution* parameterized as

$$\mathbf{x}(\phi, \theta) = \begin{pmatrix} a \cos \phi \sin \theta \\ a \sin \phi \sin \theta \\ b \cos \theta \end{pmatrix} \quad (5.71)$$

where  $0 \leq \theta \leq \pi$  and  $0 \leq \phi \leq 2\pi$ , and  $a, b$  are positive constants.

The corresponding metric tensor is

$$G(\phi, \theta) = \begin{pmatrix} a^2 \sin^2 \theta & 0 \\ 0 & a^2 \cos^2 \theta + b^2 \sin^2 \theta \end{pmatrix}.$$

The inward-pointing unit normal is

$$\mathbf{n}(\phi, \theta) = -|G(\phi, \theta)|^{-\frac{1}{2}} \begin{pmatrix} ab \cos \phi \sin^2 \theta \\ ab \sin \phi \sin^2 \theta \\ a^2 \sin \theta \cos \theta \end{pmatrix}.$$

Therefore,

$$L(\phi, \theta) = |G(\phi, \theta)|^{-\frac{1}{2}} \begin{pmatrix} a^2 b \sin^3 \theta & 0 \\ 0 & a^2 b \sin \theta \end{pmatrix} \doteq |G(\phi, \theta)|^{-\frac{1}{2}} \tilde{L}(\phi, \theta).$$

Since

$$\det(G^{-1}L) = \det(G^{-1})\det(L) = \det(L)/\det(G),$$

and since for  $A \in \mathbb{R}^{n \times n}$  and  $\det(cA) = c^n \det(A)$ , it follows that

$$\det(G^{-1}L) = \det(\tilde{L})/|\det(G)|^2.$$

Therefore,

$$K = \int_S k \, dS = \int_0^\pi \int_0^{2\pi} \frac{\det(\tilde{L})}{|\det(G)|^2} |\det(G)|^{\frac{1}{2}} d\phi d\theta = \int_0^\pi \int_0^{2\pi} \frac{\det(\tilde{L})}{|\det(G)|^{\frac{3}{2}}} d\phi d\theta,$$

or

$$K = \int_0^\pi \int_0^{2\pi} \frac{ab^2 \sin \theta}{(a^2 \cos^2 \theta + b^2 \sin^2 \theta)^{\frac{3}{2}}} d\phi d\theta = 2\pi ab^2 \int_0^\pi \frac{\sin \theta}{(a^2 \cos^2 \theta + b^2 \sin^2 \theta)^{\frac{3}{2}}} d\theta.$$

Evaluating the remaining integral from tables of closed-form integrals yields  $K = 4\pi$ .

The volume can be computed either using the divergence theorem, or by directly parameterizing the interior of the ellipsoid and integrating to yield

$$V = \frac{4}{3}\pi a^2 b.$$

The values of  $F$  and  $M$  for prolate and oblate ellipsoids have been reported in [37], along with a variety of other solids of revolution. In particular, if  $a = R$  and  $b = \lambda R$  with  $0 < \lambda < 1$ , then

$$F = 2\pi R^2 \left[ 1 + \frac{\lambda^2}{\sqrt{1-\lambda^2}} \log \left( \frac{1 + \sqrt{1-\lambda^2}}{\lambda} \right) \right]; \quad M = 2\pi R \left[ \lambda + \frac{\arccos \lambda}{\sqrt{1-\lambda^2}} \right].$$

In contrast, when  $\lambda > 1$ ,

$$F = 2\pi R^2 \left[ 1 + \frac{\lambda^2 \arccos(1/\lambda)}{\sqrt{\lambda^2-1}} \right]; \quad M = 2\pi R \left[ \lambda + \frac{\log(\lambda + \sqrt{\lambda^2-1})}{\sqrt{\lambda^2-1}} \right].$$



### 5.4.5 The Torus

The 2-torus can be parameterized as

$$\mathbf{x}(\theta, \phi) = \begin{pmatrix} (R + r \cos \theta) \cos \phi \\ (R + r \cos \theta) \sin \phi \\ r \sin \theta \end{pmatrix} \quad (5.72)$$

where  $R > 2r$  and  $0 \leq \theta, \phi \leq 2\pi$ .

The metric tensor for the torus is written in this parametrization as

$$G(\phi, \theta) = \begin{pmatrix} (R + r \cos \theta)^2 & 0 \\ 0 & r^2 \end{pmatrix}.$$

The surface area is computed directly as

$$F = \int_0^{2\pi} \int_0^{2\pi} r(R + r \cos \theta) d\phi d\theta = 2\pi \int_0^{2\pi} r(R + r \cos \theta) d\theta = 4\pi^2 r R.$$

This is the product of the circumference of the *medial* (or “backbone”) circle of radius  $R$ , which has a value of  $2\pi R$ , and the circumference of the circle resulting from the cross section of the torus in the plane normal to the tangent of the medial circle, which has a value of  $2\pi r$ .

Direct calculation yields the inward-pointing normal

$$\mathbf{n}(\theta, \phi) = - \begin{pmatrix} \cos \theta \cos \phi \\ \cos \theta \sin \phi \\ \sin \theta \end{pmatrix}.$$

By defining a vector field

$$\mathbf{f}(\mathbf{x}) = A\mathbf{x} = x_3 \mathbf{e}_3$$

where

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

it is clear that  $\text{div}(\mathbf{f}) = 1$ , and therefore the volume of the torus can be computed via the divergence theorem (in the form of (5.43)) as

$$V = \int_0^{2\pi} \int_0^{2\pi} (r \sin^2 \theta) r (R + r \cos \theta) d\phi d\theta = 2\pi^2 r^2 R.$$

This can be written as  $V = (\pi r^2)(2\pi R)$ , which is the product of the circumference of the medial axis and the cross-sectional area of the interior of the circle of cross section.

The matrix  $L$  is

$$L(\phi, \theta) = \begin{pmatrix} (R + r \cos \theta) \cos \theta & 0 \\ 0 & r \end{pmatrix},$$

and

$$G^{-1}L = \begin{pmatrix} (R + r \cos \theta)^{-1} \cos \theta & 0 \\ 0 & 1/r \end{pmatrix}.$$

The total Gaussian curvature is then computed as

$$K = \int_0^{2\pi} \int_0^{2\pi} [(R + r \cos \theta)^{-1} \cos \theta / r] [r(R + r \cos \theta)] d\phi d\theta = \int_0^{2\pi} \int_0^{2\pi} \cos \theta d\phi d\theta = 0.$$

The mean curvature is  $m = (R + r \cos \theta)^{-1} \cos \theta + 1/r$ . The total mean curvature results from integrating this over the surface. From the above computation of total Gaussian curvature, it is clear that the first term in the expression for  $m$  will integrate to zero, and so

$$M = \frac{F}{2r} = 2\pi^2 R.$$

#### 5.4.6 The Gauss–Bonnet Theorem and Related Inequalities

It is no coincidence that the total Gaussian curvature,  $K$ , is equal to  $4\pi$  for the sphere and ellipsoid, and equal to zero for a torus. The following fundamental theorem relates geometric and topological properties of surfaces.

**Theorem 5.3.** (*Gauss–Bonnet*) *Let  $k$  be the Gaussian curvature of a closed surface  $S$ . Then*

$$\boxed{\int_S k dS = 2\pi\chi(S)}, \quad (5.73)$$

where  $\chi(S)$  is the Euler characteristic of the closed surface  $S$ .

The Euler characteristic of a two-dimensional surface is equal to

$$\boxed{\chi(S) = 2(1 - \gamma(S))} \quad (5.74)$$

where  $\gamma(S)$  is the genus (or “number of donut holes”) of the surface. A sphere and ellipsoid have a genus of zero. The torus has a genus of one. If a polygonal mesh is established on the surface (such as the triangulated meshes used in computer graphics), the Euler characteristic can be computed by the same formula as for a planar body:

$$\chi(S) = v - e + f$$

where  $v$  is the number of vertices,  $e$  is the number of edges, and  $f$  is the number of faces of the polygons.

While the Gauss–Bonnet theorem is by far the most famous theorem involving total curvature, there are a number of other theorems on total curvature. For example, the following result due to K. Voss applies to closed surfaces that are at least twice continuously differentiable [68]:

$$\int_S \max(k, 0) dS \geq 4\pi. \quad (5.75)$$

This inequality holds for all surface topologies. In addition, since  $|k| \geq \max(k, 0)$ , it follows trivially that

$$\int_S |k| dS \geq \int_S \max(k, 0) dS \geq 4\pi. \quad (5.76)$$

Moreover, B.-Y. Chen [14] states the *Chern–Lashof* inequality

$$\int_S |k| dS \geq 4\pi(4 - \chi(S)) = 8\pi(1 + \gamma(S)). \quad (5.77)$$

Integrals of the square of mean curvature have resulted in several inequalities. For example, Willmore (see, e.g., [70, 71] and references therein) proved that

$$\int_S m^2 dS \geq 4\pi \quad (5.78)$$

with equality holding only for the usual sphere in  $\mathbb{R}^3$ , i.e., the undeformed sphere in Section 5.4.3. Shiohama and Takagi [64] proved that no matter how they are smoothly distorted, all 2-tori satisfy

$$\int_{T^2} m^2 dS \geq 2\pi^2, \quad (5.79)$$

with equality holding only for the case of the 2-torus with the undistorted shape given in Section 5.4.5 with the specific relationship between the radii of  $R = \sqrt{2}r > 0$ .

As a final example of a theorem involving an integral of a function of curvature, consider *Ros' theorem* as reported in [53, 56, 60]: Let  $D$  be a bounded domain in  $\mathbb{R}^3$  with finite volume and compact boundary  $\partial D$ . If  $m > 0$  everywhere on this boundary, then

$$\int_{\partial D} \frac{1}{m} dS \geq 3 \cdot \text{Vol}(D). \quad (5.80)$$

## 5.5 Tubes

### 5.5.1 Offset Curves in $\mathbb{R}^2$

The concept of parallel lines is as old as geometry itself. The concept of parallel curves, while somewhat newer, is also fundamental. Given a parametric curve in the plane,  $\mathbf{x}(s)$ , a parallel curve can be constructed as

$$\mathbf{o}(s; r) = \mathbf{x}(s) + r\mathbf{n}(s) \quad (5.81)$$

where  $\mathbf{n}(s)$  is the planar unit normal to the curve at each point, and  $r$  is a constant with value less than the maximal radius of curvature of  $\mathbf{x}(s)$ :  $-1/\kappa_{max} < r < 1/\kappa_{max}$ .

For each fixed value of  $r$ , the curve  $\mathbf{o}(s; r)$  is called an *offset curve*. A non-parametric description of the offset curve defined in (5.81) is

$$\hat{\mathbf{o}}(\mathbf{x}; r) = \mathbf{x} + r\mathbf{n} \quad \implies \quad \hat{\mathbf{o}}(\mathbf{x}(s); r) = \mathbf{o}(s; r). \quad (5.82)$$

The collection of all offset curves for  $-r_0 \leq r \leq r_0$  defines a strip in the plane.

For convenience, the curve parameter  $s$  is usually chosen to be arc length. Then from the Frenet–Serret apparatus, the tangent and normal to the curve are related by the equations

$$\frac{d\mathbf{u}}{ds} = \kappa(s)\mathbf{n} \quad \text{and} \quad \frac{d\mathbf{n}}{ds} = -\kappa(s)\mathbf{u}$$

where  $\kappa(s)$  is the unsigned curvature. Since  $r$  is treated as a constant, this means that

$$\frac{d\mathbf{o}(s; r)}{ds} = [1 - r\kappa(s)]\mathbf{u}(s).$$

Even though  $s$  is not the arc length of the offset curve  $\mathbf{o}(s; r)$ , the unit tangent to  $\mathbf{o}(s; r)$  is  $\mathbf{u}(s)$ . And the unit normal to  $\mathbf{o}(s; r)$  is  $\mathbf{n}(s)$ . Therefore, since  $\mathbf{x}(s)$  and  $\mathbf{o}(s; r)$  have

the same normal and tangent, it follows that taking the offset curve of an offset curve results in an offset curve of the original curve:

$$\hat{\mathbf{o}}(\hat{\mathbf{o}}(\mathbf{x}(s); r_1); r_2) = \hat{\mathbf{o}}(\mathbf{x}(s); r_1 + r_2) \quad (5.83)$$

where the notation in (5.82) is used here.

For fixed  $r$ , the arc length of each  $\mathbf{o}(s; r)$  in (5.81) is computed using the above expressions and the fact that  $\mathbf{u} \cdot \mathbf{n} = 0$  and  $\mathbf{u} \cdot \mathbf{u} = \mathbf{n} \cdot \mathbf{n} = 1$  as

$$\int_0^L \|d\mathbf{o}/ds\| ds = \int_0^L [1 - \kappa(s)r] ds = L - r\Delta\theta.$$

When the curve has no inflection points, and so the unsigned curvature equals the signed curvature, then  $\Delta\theta$  is the angle between the tangents to the curve at  $s = 0$  and  $s = L$ , measured positive when  $\mathbf{u}(0)$  turns counterclockwise to coincide with  $\mathbf{u}(L)$ . For a smooth closed convex curve,  $\mathbf{x}(s)$ , the total angle swept will be  $\Delta\theta = 2\pi$ . Note that for closed curves parameterized such that  $s$  increases with counterclockwise traversal,  $\mathbf{n}$  points to the interior, and so the total length of the offset curve  $\mathbf{o}(t; r)$  will be less than  $\mathbf{x}(s)$  when  $r > 0$ , and greater than  $\mathbf{x}(s)$  when  $r < 0$ . This leads to the following observation explained in [32]:

$$\begin{aligned} \int_{-r_0}^{r_0} \int_0^L \|d\mathbf{o}/ds\| ds dr &= \int_{-r_0}^{r_0} \int_0^L [1 - \kappa(s)r] ds dr \\ &= 2r_0L - \int_0^L \kappa(s) \left( \int_{-r_0}^{r_0} r dr \right) ds \\ &= 2r_0L. \end{aligned} \quad (5.84)$$

The reason for this is that the integral in parentheses vanishes due to the fact that  $f(r) = r$  is an odd function of  $r$  and the range of integration is symmetric around  $r = 0$ .

The analytical and algebraic properties of offset curves have been studied extensively (see, e.g., [23, 24]). Their many applications include path planning in numerically controlled machines [6, 57] and closed-form locally volume preserving deformations of solid models [17].

### 5.5.2 Parallel Fibers, Ribbons, and Tubes of Curves in $\mathbb{R}^3$

For an arc-length-parameterized curve in three-dimensional space,  $\mathbf{x}(s)$ , an offset curve can be defined as in (5.81), but there is more freedom in the way  $\mathbf{n}(s)$  can be chosen. That is, in the three-dimensional case  $\mathbf{n}(s)$  in (5.81) need not be the normal from the Frenet–Serret apparatus. It could be some linear combination of the normal,  $\mathbf{n}_1(s)$ , and bi-normal,  $\mathbf{n}_2(s)$ :

$$\mathbf{n}(s) = \mathbf{n}_1(s) \cos \phi(s) + \mathbf{n}_2(s) \sin \phi(s). \quad (5.85)$$

In this spatial case, using the Frenet formulas  $d\mathbf{n}_1/ds = -\kappa\mathbf{u} + \tau\mathbf{n}_2$ ,  $d\mathbf{n}_2/ds = -\tau\mathbf{n}_1$ ,  $\mathbf{u} \cdot \mathbf{n}_i = 0$ , and  $\mathbf{n}_i \cdot \mathbf{n}_j = \delta_{ij}$ , it is easy to verify that

$$\left\| \frac{\partial \mathbf{o}}{\partial s} \right\|^2 = [1 - r\kappa(s) \cos \phi(s)]^2 + [d\phi/ds + \tau(s)]^2 r^2.$$

In the special case when  $d\phi/ds = -\tau(s)$ , or equivalently,

$$\phi(s) = \theta_0 - \int_0^s \tau(\sigma) d\sigma \quad (5.86)$$

for a constant value of  $\theta_0$ , the curves that result will be the offset curves that evolve with minimal length from a specified starting position. These can be considered to be the spatial curves that are parallel to the curve  $\mathbf{x}(s)$ , and are called *fibers*. The length of a fiber for  $s \in [0, L]$  is

$$\hat{L}(\theta_0, L) = \int_0^L \left\| \frac{\partial \mathbf{o}}{\partial s} \right\| ds = L - r \int_0^L \kappa(s) \cos \left( \theta_0 - \int_0^s \tau(\sigma) d\sigma \right) ds. \quad (5.87)$$

Note that for any values of  $\theta_0$  and  $L$ ,

$$\hat{L}(\theta_0, L) + \hat{L}(\theta_0 + \pi, L) = 2L.$$

Returning to the more general setting of a spatial offset curve (5.81) with the definition of  $\mathbf{n}(s)$  in (5.85), a *ribbon* can be defined as the locus of points with  $s \in [0, L]$  and  $r \in [-r_0, r_0]$ . This can be thought of as a strip of width  $2r_0$  that twists in space with a backbone  $\mathbf{x}(s)$ . The area of such a strip is computed as

$$A(r_0, L) = \int_{-r_0}^{r_0} \int_0^L \left\| \frac{\partial \mathbf{o}}{\partial s} \times \frac{\partial \mathbf{o}}{\partial r} \right\| ds dr$$

where

$$\left\| \frac{\partial \mathbf{o}}{\partial s} \times \frac{\partial \mathbf{o}}{\partial r} \right\|^2 = [1 - r\kappa(s) \cos \phi(s)]^2 + r^2 [d\phi/ds + \tau(s)]^2 r^2. \quad (5.88)$$

When the ribbon has minimal twist (i.e.,  $d\phi/ds = -\tau(s)$ ), the second term in the above expression disappears and

$$A(r_0, L) = \int_{-r_0}^{r_0} \int_0^L [1 - r\kappa(s) \cos \phi(s)] ds dr = 2r_0L$$

for exactly the same reasons as in (5.84).

A constant-radius *tubular body* can be defined around a space curve using (5.81) and (5.85) as the following locus of points parameterized with  $s \in [0, L]$ ,  $r \in [0, r_0]$ , and  $\theta \in [0, 2\pi]$ :

$$\mathbf{T}(s, \alpha; r) = \mathbf{x}(s) + rR[\mathbf{u}(s), \alpha]\mathbf{n}(s) = \mathbf{x}(s) + r[\mathbf{n}_1(s) \cos(\phi(s) + \alpha) + \mathbf{n}_2(s) \sin(\phi(s) + \alpha)]. \quad (5.89)$$

Here  $R[\mathbf{u}, \alpha]$  denotes the rotation matrix describing counterclockwise rotation by angle  $\alpha$  around the vector  $\mathbf{u}$ . A tubular surface (or simply, a tube) is described by (5.89) for fixed value of  $r$ . This is a deformed version of a cylinder with  $\mathbf{x}(s)$  tracing its backbone.

The area of the surface  $\mathbf{T}(s, \alpha; r = r_0)$  enclosing this tube is

$$\begin{aligned} S(r_0, L) &= \int_0^{2\pi} \int_0^L \left\| \frac{\partial \mathbf{T}}{\partial s} \times \frac{\partial \mathbf{T}}{\partial \alpha} \right\| ds d\alpha \\ &= \int_0^{2\pi} \int_0^L r_0 [1 - r_0 \kappa(s) \cos(\phi(s) + \alpha)] ds d\alpha \\ &= 2\pi r_0 L - r_0^2 \int_0^L \kappa(s) \left( \int_0^{2\pi} \cos(\phi(s) + \alpha) d\alpha \right) ds \\ &= 2\pi r_0 L. \end{aligned} \quad (5.91)$$

The reason for the simplification in the last line above is that the integral over  $\alpha$  of  $\cos(\phi(s) + \alpha) = \cos \phi(s) \cos \alpha - \sin \phi(s) \sin \alpha$  is zero.

The volume of a tube defined in (5.89) is computed as the integral of the following triple product:

$$\begin{aligned} V(r_0, L) &= \int_0^{2\pi} \int_0^{r_0} \int_0^L \left[ \frac{\partial T}{\partial s}, \frac{\partial T}{\partial r}, \frac{\partial T}{\partial \alpha} \right] ds dr d\theta \\ &= \int_0^{2\pi} \int_0^{r_0} \int_0^L r ds dr d\theta \\ &= \pi r_0^2 L. \end{aligned} \tag{5.92}$$

Note that the results (5.91) and (5.92) do not depend on  $\phi(s)$  being defined as in (5.86).

A tubular surface around a smooth closed space curve will be a kind of torus. This torus can be embedded in  $\mathbb{R}^3$  in the standard way, or it can be knotted. If  $\phi(s) = 0$ , the element of surface area of a tube and curvature respectively are given as [36]

$$dS = r_0(1 - r_0\kappa(s) \cos \alpha) ds d\alpha \quad \text{and} \quad k(s, \alpha) = \frac{-\kappa(s) \cos \alpha}{r_0(1 - r_0\kappa(s) \cos \alpha)}$$

where  $\kappa(s)$  is the curvature of the backbone curve,  $\mathbf{x}(s)$ . Then

$$\int_T \max(k, 0) dS = - \int_{s=0}^{2\pi} \int_{\alpha=\frac{\pi}{2}}^{\frac{3\pi}{2}} \kappa(s) \cos \alpha d\alpha ds = 2 \int_{s=0}^{2\pi} \kappa(s) ds.$$

From (5.75), it follows that this must be greater than or equal to  $4\pi$ , which is consistent with Fenchel's result that  $\int_0^{2\pi} \kappa(s) ds \geq 2\pi$ . In contrast,

$$\int_T |k| dS = \int_0^{2\pi} \int_0^{2\pi} \kappa(s) |\cos \alpha| d\alpha ds = 4 \int_0^{2\pi} \kappa(s) ds \geq 8\pi$$

when  $\mathbf{x}(s)$  is any smooth closed space curve.

In analogy with the Fary–Milnor theorem that considers necessary conditions on the total curvature of a closed space curve for it to be knotted, Langevin and Rosenberg established the following necessary conditions for a tube to be knotted [46]:

$$\int_T |k| dS \geq 16\pi \tag{5.93}$$

where  $|k|$  is the absolute value of the Gaussian curvature,  $T$  is the surface of the knotted tube, and  $dS$  is the same differential element of surface area used in (5.90). This result was extended by Kuiper and Meeks [45] by replacing  $16\pi$  with  $8\pi B$  where  $B$  is the bridging index of the backbone curve of the tube.

B.-Y. Chen [14] reviews total mean curvature formulas for smooth *knotted closed tubes* such as

$$\int_T m^2 dS \geq 8\pi \tag{5.94}$$

that were initially derived in the early 1970s [15, 16]. This too can be stated in terms of bridging numbers of the backbone curve.

### 5.5.3 Tubes of Surfaces in $\mathbb{R}^3$

Given a smooth parameterized surface,  $\mathbf{x}(t_1, t_2)$ , a unit normal can be defined to the surface at each point as

$$\mathbf{u}(t_1, t_2) = \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} / \left\| \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} \right\|.$$

Then, an offset surface can be defined for a fixed value of  $r$  less than  $1/\max\{|\kappa_1|, |\kappa_2|\}$  (where  $\kappa_1$  and  $\kappa_2$  are the principal curvatures obtained by solving (5.63)) as

$$\mathbf{o}(t_1, t_2; r) = \mathbf{x}(t_1, t_2) + r\mathbf{u}(t_1, t_2). \quad (5.95)$$

The element of surface area for this *offset surface* can be shown to be of the form

$$dS = \left\| \frac{\partial \mathbf{o}}{\partial t_1} \times \frac{\partial \mathbf{o}}{\partial t_2} \right\| dt_1 dt_2$$

where [32]

$$\left\| \frac{\partial \mathbf{o}}{\partial t_1} \times \frac{\partial \mathbf{o}}{\partial t_2} \right\| = [1 - 2rm(t_1, t_2) + r^2k(t_1, t_2)] \left\| \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} \right\|. \quad (5.96)$$

Here  $m(t_1, t_2)$  is the mean curvature and  $k(t_1, t_2)$  is the Gaussian curvature at the point  $\mathbf{x}(t_1, t_2)$ . Therefore, the area of the offset surface will be

$$A = F - 2rM + r^2K \quad (5.97)$$

where  $F$ ,  $M$ , and  $K$  are respectively the area, total mean curvature, and total Gaussian curvature of the original surface,  $S$ .

For a finite body  $B$  with volume  $V(B)$  enclosed by a compact surface  $\partial B$ , an equation similar to (5.97) is *Steiner's formula* for the volume enclosed by the surface offset by an amount  $r$  from  $\partial B$  [67]:

$$\boxed{V(B_r) = V(B) + rF(\partial B) + \frac{r^2}{2}M(\partial B) + \frac{r^3}{3}K(\partial B).} \quad (5.98)$$

It generalizes to higher dimensions and to cases where the surfaces are not smooth.

The volume contained within the two offset surfaces defined by  $r \in [-r_0, r_0]$  can be computed by allowing  $r$  in (5.95) to vary. Then

$$V_o = \int_{-r_0}^{r_0} \int_S \left[ \frac{\partial \mathbf{o}}{\partial r}, \frac{\partial \mathbf{o}}{\partial t_1}, \frac{\partial \mathbf{o}}{\partial t_2} \right] dt_1 dt_2 dr = 2rF + \frac{2r^3}{3} \int_S k dS = 2rF + \frac{4}{3}\pi r^3 \chi(S). \quad (5.99)$$

This formulas generalize to hyper-surfaces in high-dimensional Euclidean spaces. The generalized formulas were proven by Weyl [69]. Although the word “tube” was used in this section for the cylindrical surface around a curve, and an “offset” curve/surface was used more broadly, the word “tube” is used in the mathematics literature to denote both concepts, and “Weyl’s tube theorem” addresses this generalized concept in higher dimensions.

## 5.6 The Euler Characteristic: From One Dimension to $N$ Dimensions

In preceding sections in this chapter the Euler characteristics of finite planar regions and closed surfaces in three-dimensional space played prominent roles in characterizing the topological properties of these two-dimensional objects. In this section, the concept of the Euler characteristic of geometrical objects with dimensions one and three are defined, and rules for defining the Euler characteristic of geometrical objects in higher dimensions are provided. These rules will provide background for understanding the generalized results presented in the next two chapters.

### 5.6.1 The Euler Characteristic of Zero-, One-, and Three-Dimensional Bodies

In (5.40) and (5.73) total curvatures were related to the Euler characteristic of a planar body and a closed surface in space. In both instances, the same formula was used. Here this formula is generalized to other dimensions.

First consider the trivial example of a connected “one-dimensional body.” This is nothing more than a closed interval  $B_1 = [a, b] \subset \mathbb{R}$ . The boundary of the body is the vertices given by the points  $a, b \in \mathbb{R}$ . In this case there are no “faces” to this object. There are only vertices and a single edge (which is the body itself). The Euler characteristic of this can be defined as  $\chi(B_1) = v - e = 2 - 1 = 1$ . Now if there is another interval  $B_2 = [c, d] \subset \mathbb{R}$ , then again  $\chi(B_2) = 1$ . If  $B_1 \cap B_2 = \emptyset$ , then the body  $B_3 = B_1 \cup B_2$  is not connected. A simple count gives  $\chi(B_3) = v - e = 4 - 2 = 2$ . In fact, for this case as well as the 2D cases described earlier, the following rule holds:

$$\chi(B_1 \cup B_2) = \chi(B_1) + \chi(B_2) \quad \text{when} \quad B_1 \cap B_2 = \emptyset. \quad (5.100)$$

The Euler characteristic of a simple closed curve in the plane (which is topologically equivalent to the unit circle,  $S^1$ ) can also be calculated using the formula

$$\chi(S^1) = v - e = 0.$$

From this it is clear that the Euler characteristic, which was initially defined for two-dimensional geometrical objects, can be applied equally well for one-dimensional objects.

A zero-dimensional object can be defined as a set of disconnected points. This can either be viewed as the boundary of a one-dimensional body on the line, or as a zero-dimensional body in any space. The Euler characteristic of a zero-dimensional object is simply the number of points (or vertices) that define the body:  $\chi(B) = v$ .

What about for three-dimensional objects? Given a simple body in three dimensions, it can be tessellated (subdivided) into simple polyhedral cells such as cubes or tetrahedra. For such a body, the definition of the Euler characteristic must be modified as

$$\boxed{\chi(B) = v - e + f - c} \quad (5.101)$$

where again  $v$ ,  $e$ , and  $f$  are the total number of vertices, edges, and faces in the tessellation, but now the total number of spatial *cells*,  $c$ , must also be counted.

For example, consider a body consisting of a single tetrahedral cell. A simple count gives  $\chi(\text{tetcell}) = 4 - 6 + 4 - 1 = 1$ . Now consider a body consisting of a single block defined by a cube and all points on the interior of the cube. This is analogous to what



was done in the planar case in Section 5.3.2. Again, a simple count gives  $\chi(\text{block}) = 8 - 12 + 6 - 1 = 1$ . Suppose that this block is subdivided into an  $L \times W \times H$  array of  $L \cdot W \cdot H$  small blocks, or cells. A careful counting then gives

$$\begin{aligned} c(\text{array}) &= L \cdot W \cdot H \\ f(\text{array}) &= H \cdot W \cdot (L + 1) + L \cdot W \cdot (H + 1) + L \cdot H \cdot (W + 1) \\ e(\text{array}) &= H \cdot (W + 1)(L + 1) + L \cdot (W + 1)(H + 1) + W \cdot (H + 1)(L + 1) \\ v(\text{array}) &= (H + 1)(W + 1)(L + 1). \end{aligned}$$

Therefore,

$$\chi(\text{array}) = 1.$$

If this initial array consisting of  $L \cdot W \cdot H$  blocks is then “sculpted” by removing individual blocks on its exterior, the effect on the Euler characteristic can be examined. Here  $L, W, H$  are all taken to be greater than one. There are three different kinds of blocks that can be removed from the surface of the array: corners, edges, and others. Let  $\Delta\chi$  denote the change in the Euler characteristic of the body that results from this sculpting operation. This means that

$$\Delta\chi(B) = \chi(B_{\text{after}}) - \chi(B_{\text{before}}) = \Delta v - \Delta e + \Delta f - \Delta c$$

where, for example,  $\Delta v$  is the difference between the number of vertices after the sculpting operation and before. If a corner block is removed from the array, one vertex is removed, three edges are removed, three faces are removed, and one cell is removed. Therefore, in this case  $\Delta\chi(B) = (-1) - (-3) + (-3) - (-1) = 0$ . If an edge block is removed,  $\Delta\chi(B) = 0 - (-1) + (-2) - (-1) = 0$ . Similarly, if a surface block that is neither an edge nor a corner is removed,  $\Delta\chi(B) = 0 - 0 + (-1) - (-1) = 0$ . Therefore, it can be concluded that “chipping away” blocks from the surface does not affect the Euler characteristic. A body sculpted in this way will have  $\chi = 1$ .

In contrast, consider the operation of removing a block from deep inside the array. In analogy with a medical treatment that remotely destroys a tumor, call the removal of an interior block an *ablation*. By ablating a block, no vertices, edges, or faces are removed. Only the volumetric cell is removed. This gives  $\Delta\chi = -(-1) = 1$ . If two adjacent blocks are ablated, then two cells and the adjoining face are removed, and again  $\Delta\chi = 1$ . In fact, if this ablation procedure is used to create any simply connected void on the interior of the array, then  $\Delta\chi = 1$ . If rather than forming a simply connected void on the interior, a toroidal void is formed,  $\Delta\chi = 0$  because in that case the number of cells removed is the same as the number of faces, while leaving the number of vertices and edges unchanged. Evidently, the Euler characteristic of a body with a toroidal void is the same as the Euler characteristic for a simply connected body. And this is true regardless of whether the toroidal void is knotted or not.

Consider the following operation. Given the initial array of blocks, if a hole is “drilled” through its center, the result will be to remove  $L$  cells and  $L + 1$  faces, and so  $\Delta\chi = 0 - 0 + (-L - 1) - (-L) = -1$ . (If instead the drilling was through one of the other directions,  $L$  would be replaced with  $W$  or  $H$  but the end result is the same.) In this case the surface that encloses the resulting array will be the same (topologically) as a torus. The Euler characteristic of this toroidal array of blocks will be  $\chi(B) = 1 + \Delta\chi = 0$ . Recall that the Euler characteristic of the surface of a torus is also zero.

If two such holes are drilled parallel to each other in such a way that they do not share any edges, faces or vertices, then the  $\Delta\chi$ 's for each will add together, and the resulting

body will have Euler characteristic of  $\chi(B) = 1 + \Delta\chi(B) = 1 - 2 = -1$ . In contrast, since  $\gamma = 2$ , the surface will have Euler characteristic  $\chi(\partial B) = 2(1 - \gamma) = 2(1 - 2) = -2$ . If two holes are drilled at two orthogonal directions and share one cell in common, the result will be  $\Delta\chi = -3$ , and the resulting surface will have a genus of 3. Therefore,  $\chi(B) = 1 + \Delta\chi(B) = 1 - 3 = -2$ , and  $\chi(\partial B) = 2(1 - \gamma) = 2(1 - 3) = -4$ .

It is also easy to show that if a whole two-dimensional planar array of cells is removed, which cleaves the initial body into two disjoint pieces, then  $\Delta\chi = +1$ .

### 5.6.2 Relationship Between the Euler Characteristic of a Body and Its Boundary

From the examples at the end of the previous section, it appears that in three-dimensional Euclidean space, the Euler characteristic of a body and that of its bounding surface are related as

$$\chi(\partial B) = 2 \cdot \chi(B) = 2[1 - \gamma(B)]. \quad (5.102)$$

Does this formula hold if voids are introduced into the body?

Returning to the  $L \times W \times H$  array of blocks discussed in the previous section, the Euler characteristic of the surface will be

$$\chi(\partial array) = v(\partial array) - e(\partial array) + f(\partial array) = 2$$

where

$$\begin{aligned} v(\partial array) &= 2(H \cdot W + L \cdot W + L \cdot H + 1) \\ e(\partial array) &= 4(H \cdot W + L \cdot W + L \cdot H) \\ f(\partial array) &= 2(H \cdot W + L \cdot W + L \cdot H). \end{aligned}$$

It is easy to verify that the sculpting operations described earlier will not affect the value. Introducing a simple void in an otherwise simple body increases the Euler characteristic of the body from 1 to 2. At the same time, the sum of Euler characteristics of the internal and external bounding surfaces becomes  $2 + 2 = 4$ . And so, (5.102) appears to be correct in three dimensions if the definition of the overall Euler characteristic of the bounding surface is defined as

$$\chi(\cup_k \partial_k B) = \sum_{k=1}^n \chi(\partial_k B) \quad \text{where} \quad \partial_i B \cap \partial_j B = \emptyset \quad \forall i, j \in \{1, \dots, n\}, \quad (5.103)$$

and  $\partial_i B$  is the  $i$ th bounding surface of a complicated boundary consisting of  $n$  disjoint components.

Does (5.102) hold in other dimensions? Consider a one-dimensional simple body on the real line. The Euler characteristics of its boundary and the body itself are respectively  $\chi(\partial B) = v = 2$  and  $\chi(B) = v - e = 2 - 1 = 1$ . For a one-dimensional body on the line consisting of two disjoint components,  $\chi(\partial B) = v = 4$  and  $\chi(B) = v - e = 4 - 2 = 2$ . Clearly if there are  $n$  disjoint parts to the body,  $\chi(\partial B) = 2n$ ,  $\chi(B) = n$ , and there are  $\gamma = n - 1$  holes on the “interior” of the body. And so, on the line  $\chi(B) = \gamma + 1 = \chi(\partial B)/2$ , which means that in the 1D case the first equality in (5.102) holds, but the second does not.

Now consider a 2D body with simple bounding curves. Since for a closed curve in the plane  $\chi(\partial B) = v - e = 0$ , it appears that the first equality in (5.102) cannot hold. However, working out a few examples, it becomes clear that  $\chi(B) = 1 - \gamma(B)$ , which is the second equality in (5.102). Therefore, it seems like a mixed bag, but generally speaking both equalities in (5.102) only hold for the 3D case.

### 5.6.3 The Euler Characteristic of Cartesian Products of Objects

Recall from Chapter 1 that the *Cartesian product* of two sets  $S_1$  and  $S_2$  is the set consisting of all pairs  $(x, y)$  with  $x \in S_1$  and  $y \in S_2$ . This product set is denoted as  $S_1 \times S_2$ . For example, the 2-torus can be viewed as a product of two circles:  $T_2 = S^1 \times S^1$ . A cylinder with closed boundary can be viewed as the product of a closed interval and a circle:  $[0, 1] \times S^1$ . And a unit cube can be thought of as a two-fold product of closed intervals:  $[0, 1] \times [0, 1] \times [0, 1]$ .

It turns out that if a geometric object can be described as a Cartesian product, then the Euler characteristic of the object can be computed as the product of the Euler characteristics of the lower-dimensional objects forming the product:

$$\chi(B_1 \times B_2) = \chi(B_1) \cdot \chi(B_2). \quad (5.104)$$

For example, the two-dimensional toroidal surface has an Euler characteristic of zero, which is the same as the product of the Euler characteristics for two circles, each of which is zero. The Euler characteristic of a toroidal body is zero, which is the same as the product of the Euler characteristics for the circle and a closed circular disk. The Euler characteristic of a cubic body is equal to unity, which is the same as the product of Euler characteristics of each closed interval.

## 5.7 Implicit Surfaces, Level Set Methods, and Curvature Flows

The parametric approach to differential geometry has been augmented in recent years by the use of implicit surface descriptions of the form  $\phi(\mathbf{x}) = 0$ . This single scalar constraint on a vector  $\mathbf{x} \in \mathbb{R}^3$  defines a two-dimensional surface. The implicit and parametric approaches are complementary, each having its benefits. With a parametric surface description of the form  $\mathbf{x} = \mathbf{x}(u_1, u_2)$  it is easy to generate points on the surface by directly evaluating  $\mathbf{x}(u_1, u_2)$  with allowable values of  $u_1$  and  $u_2$ . However, given a point  $\mathbf{y} \in \mathbb{R}^3$  it is difficult to determine directly from the parametric description whether or not this point lies on the surface. In contrast, by evaluating  $\phi(\mathbf{y})$  and comparing this value with zero, it is easy to determine if  $\mathbf{y}$  is on the surface or not. These two approaches are related by the fact that  $\phi(\mathbf{x}(u_1, u_2)) = 0$ .

All of the formulas in the classical parametric differential geometry of surfaces can be recast in terms of implicit surface descriptions. An implicit surface, which is also referred to in the literature as a level set, can be more natural than the parametric approach in some settings. So-called “level set methods” have become popular in recent years in image processing and mechanics. A particular problem that is handled with these methods is the evolution in shape of an initial surface into a new surface. In such “curvature flow” problems, shape changes are defined locally at each point on the surface based on mean or Gaussian curvature. Unlike the Fokker–Planck equations discussed in Chapter 4, which are linear partial differential equations, curvature flow equations are non-linear PDEs. In some applications these PDEs have stochastic coefficients.

### 5.7.1 Implicit Surfaces

An implicit two-dimensional surface is defined by a scalar constraint  $\phi(\mathbf{x}) = 0$  for  $\mathbf{x} \in \mathbb{R}^3$ . When this describes a closed surface,  $\phi(\mathbf{x})$  can be defined such that  $\phi(\mathbf{x}) < 0$  corresponds to the finite body bounded by the surface, and  $\phi(\mathbf{x}) > 0$  corresponds to the

outside world. For example, an ellipsoid with principal axes aligned with the Cartesian coordinate system and centered at the origin has the implicit description

$$\phi(\mathbf{x}) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} - 1 = 0. \quad (5.105)$$

If the corresponding parametric description of the implicit surface  $\phi(\mathbf{x}) = 0$  is  $\mathbf{x} = \mathbf{x}(u_1, u_2)$ , then it must be that  $\phi(\mathbf{x}(u_1, u_2)) = 0$ . This equation provides the key to calculating intrinsic quantities of the surface such as Gaussian and mean curvature directly from the implicit description. This is because the chain rule gives

$$\frac{\partial}{\partial u_i} [\phi(\mathbf{x}(u_1, u_2))] = \frac{\partial \phi}{\partial x_1} \frac{\partial x_1}{\partial u_i} + \frac{\partial \phi}{\partial x_2} \frac{\partial x_2}{\partial u_i} + \frac{\partial \phi}{\partial x_3} \frac{\partial x_3}{\partial u_i} = 0$$

for  $i = 1, 2$ . A more compact way to write this is

$$(\nabla \phi) \cdot \frac{\partial \mathbf{x}}{\partial u_i} = 0 \quad (5.106)$$

where  $\nabla \phi = [\partial \phi / \partial x_1, \partial \phi / \partial x_2, \partial \phi / \partial x_3]^T$  is the *gradient* of  $\phi$  and  $\partial \mathbf{x} / \partial u_i$  is tangent to the surface. This implies immediately that a unit surface normal at  $\mathbf{x}(u_1, u_2)$  is

$$\mathbf{n}(u_1, u_2) = \frac{\nabla \phi|_{\mathbf{x}=\mathbf{x}(u_1, u_2)}}{\left\| \nabla \phi|_{\mathbf{x}=\mathbf{x}(u_1, u_2)} \right\|} = \frac{\nabla \phi|_{\phi(\mathbf{x})=0}}{\left\| \nabla \phi|_{\phi(\mathbf{x})=0} \right\|}. \quad (5.107)$$

Note that while the rightmost quantity in the second equality is not intrinsic (since it is defined relative to the ambient space), it is independent of any parametrization of the surface.

By observing the constraint equations that result from calculating the partial derivatives  $\partial \phi / \partial u_i$  and  $\partial^2 \phi / \partial u_i \partial u_j$  for  $i, j \in \{1, 2\}$ , all of the quantities that appear in the first and second fundamental forms of the surface can be restated in terms of derivatives of  $\phi(\mathbf{x})$  with respect to Cartesian coordinates in  $\mathbb{R}^3$ , followed by the constraint that  $\phi(\mathbf{x}) = 0$ .

## Convolution and Implicit Tubes

A tube is defined by its backbone curve and a radius. If the backbone curve is defined in terms of an arc-length parametrization as  $\mathbf{x} = \hat{\mathbf{x}}(s)$ , then “sweeping” a Gaussian distribution along  $\hat{\mathbf{x}}(s)$  will produce an implicit description of a tube. This is written as a convolution [7, 8]:

$$\phi(\mathbf{x}) \doteq c + \int_{\mathbb{R}^n} \delta_C(\boldsymbol{\xi}) \rho(\mathbf{x} - \boldsymbol{\xi}, \mathbf{0}, \sigma^2 I) d\boldsymbol{\xi} = c + \frac{1}{(2\pi)^{n/2} \sigma^n} \int_0^L \exp\left(-\frac{1}{2} \|\mathbf{x} - \hat{\mathbf{x}}(s)\|^2 / \sigma^2\right) ds \quad (5.108)$$

where  $c$  is a constant that, together with  $\sigma$ , determines the tube radius, and  $\delta_C(\mathbf{x})$  is defined by the second equality above. That is,  $\delta_C(\mathbf{x})$  localizes an integral over the plane or in three dimensions to the backbone curve,  $C$ .

Given a tree-like structure rather than a backbone curve, the same approach can be used to generate branched structures [7, 8]. If the backbone curve is replaced with a surface, an implicit description of an offset surface can be formulated in a similar way.

### Curvature of Implicit Surfaces

By manipulating the relationships discussed previously, the Gaussian and mean curvature of a surface are written in terms of the implicit description as [30]<sup>7</sup>

$$k = \frac{1}{\|\nabla\phi\|^4} \det \begin{bmatrix} \nabla\nabla^T\phi & \nabla\phi \\ \nabla^T\phi & 0 \end{bmatrix} \quad (5.109)$$

and

$$m = \frac{\|\nabla\phi\|^2 \operatorname{tr}(\nabla\nabla^T\phi) - (\nabla^T\phi)(\nabla\nabla^T\phi)(\nabla\phi)}{2\|\nabla\phi\|^3} = \nabla \cdot \left( \frac{\nabla\phi}{\|\nabla\phi\|} \right) \quad (5.110)$$

where it is understood that these only hold subject to the constraint that  $\phi(\mathbf{x}) = 0$ . These formulas have a number of applications in level set methods.

#### 5.7.2 Integration on Implicitly Defined Surfaces and Curves in $\mathbb{R}^3$

##### Integrating Functions on Implicit Surfaces

The integral of a function  $f \in \mathcal{N}(\mathbb{R}^3)$  over the whole surface  $S \subset \mathbb{R}^3$  is expressed in parametric form as

$$\int_S f(\mathbf{x}) dS = \int_{u_1} \int_{u_2} f(\mathbf{x}(u_1, u_2)) |G(u_1, u_2)| du_2 du_1.$$

This alternatively can be thought of as an integral over  $\mathbb{R}^3$  that is localized to the surface implicitly by using the Dirac delta function as

$$\int_S f(\mathbf{x}) dS = \int_{\mathbb{R}^3} f(\mathbf{x}) \delta(\phi(\mathbf{x})) c(\mathbf{x}) d\mathbf{x}. \quad (5.111)$$

The reason why the weighting function  $c(\mathbf{x})$  is needed is analogous to why  $|G(u_1, u_2)|$  is needed in the parametric expression. The exact form of  $c(\mathbf{x})$  is now derived.

Imagine that  $\mathbb{R}^3$  is broken up into an infinite number of concentric parametric surfaces  $\mathbf{x}(u_1, u_2; u_3)$ , where each fixed value of  $u_3$  defines a surface. Take the coordinate  $u_3$  to be orthogonal to  $u_1$  and  $u_2$ . Then the metric tensor for the three-dimensional region parameterized by  $u_1, u_2, u_3$  will be of the form

$$G(u_1, u_2, u_3) = \begin{pmatrix} g_{11} & g_{12} & 0 \\ g_{12} & g_{22} & 0 \\ 0 & 0 & g_{33} \end{pmatrix},$$

and so

$$|G|^{\frac{1}{2}} = g_{33} \cdot (g_{11}g_{22} - g_{12}^2) = \left\| \frac{\partial \mathbf{x}}{\partial u_3} \right\| \cdot \left\| \frac{\partial \mathbf{x}}{\partial u_1} \times \frac{\partial \mathbf{x}}{\partial u_2} \right\|.$$

Let  $\phi(\mathbf{x}) = 0$  denote the particular surface that results for the constant value  $u_3 = u_3^0$ . Then

$$\phi(\mathbf{x}(u_1, u_2; u_3)) = u_3 - u_3^0. \quad (5.112)$$

The evaluation of (5.111) in the coordinates  $u_1, u_2, u_3$  then becomes

<sup>7</sup>The sign convention used here is changed so as to be consistent with the rest of the text.

$$\int_{u_1} \int_{u_2} \int_{u_3} f(\mathbf{x}(u_1, u_2; u_3)) \delta(u_3 - u_3^0) c(\mathbf{x}(u_1, u_2; u_3)) \left\| \frac{\partial \mathbf{x}}{\partial u_3} \right\| \cdot \left\| \frac{\partial \mathbf{x}}{\partial u_1} \times \frac{\partial \mathbf{x}}{\partial u_2} \right\| du_3 du_2 du_1.$$

Due to the definition of the Dirac delta function, this reduces to

$$\int_{u_1} \int_{u_2} f(\mathbf{x}(u_1, u_2; u_3^0)) c(\mathbf{x}(u_1, u_2; u_3^0)) \left\| \frac{\partial \mathbf{x}}{\partial u_3} \right\|_{u_3=u_3^0} \cdot \left\| \frac{\partial \mathbf{x}}{\partial u_1} \times \frac{\partial \mathbf{x}}{\partial u_2} \right\|_{u_3=u_3^0} du_2 du_1. \quad (5.113)$$

This integral becomes the usual surface integral when

$$c(\mathbf{x}(u_1, u_2; u_3^0)) \left\| \frac{\partial \mathbf{x}}{\partial u_3} \right\|_{u_3=u_3^0} = 1. \quad (5.114)$$

Therefore, determining the form of  $c(\mathbf{x})$  such that this will be true is required.

To address this, observe from the chain rule that partial differentiation of (5.112) with respect to each  $u_i$  yields

$$\frac{\partial \mathbf{x}}{\partial u_1} \cdot \text{grad } \phi = 0; \quad \frac{\partial \mathbf{x}}{\partial u_2} \cdot \text{grad } \phi = 0; \quad \frac{\partial \mathbf{x}}{\partial u_3} \cdot \text{grad } \phi = 1. \quad (5.115)$$

The first two of these indicate that  $\text{grad } \phi$  must be parallel to  $\partial \mathbf{x} / \partial u_3$ , which must be parallel to  $\partial \mathbf{x} / \partial u_1 \times \partial \mathbf{x} / \partial u_2$ . Furthermore, the magnitude of  $\text{grad } \phi$  is set by the third equality in (5.115). Altogether, these equalities dictate that

$$\text{grad } \phi = \frac{1}{\|\partial \mathbf{x} / \partial u_3\|^2} \frac{\partial \mathbf{x}}{\partial u_3} \quad \text{and} \quad \frac{\partial \mathbf{x}}{\partial u_3} = \|\text{grad } \phi\|^2, \quad (5.116)$$

neither of which are unit vectors.

From (5.116) it is therefore clear that (5.114) holds when

$$c(\mathbf{x}) = \|\text{grad } \phi(\mathbf{x})\|.$$

Therefore (5.111) becomes

$$\boxed{\int_S f(\mathbf{x}) dS = \int_{\mathbb{R}^3} f(\mathbf{x}) \delta(\phi(\mathbf{x})) \|\text{grad } \phi(\mathbf{x})\| d\mathbf{x}.} \quad (5.117)$$

### Integrating Functions on Implicit Curves in $\mathbb{R}^3$

The implicit description of a curve in  $\mathbb{R}^3$  is defined by the intersection of two implicitly defined surfaces. This is expressed as the simultaneous constraints

$$\phi_1(\mathbf{x}) = \phi_2(\mathbf{x}) = 0. \quad (5.118)$$

Two kinds of integrals over curves are common: (a) integrals of scalar functions over the curve, akin to total curvature; (b) work-like integrals in which the dot product of a vector field with the tangent is integrated over the curve. These are expressed in parametric form as

$$I_C^{(1)} \doteq \int_C f(\mathbf{x}(t)) \left\| \frac{d\mathbf{x}}{dt} \right\| dt \quad \text{and} \quad I_C^{(2)} \doteq \int_C \mathbf{f}(\mathbf{x}(t)) \cdot \frac{d\mathbf{x}}{dt} dt. \quad (5.119)$$

Below, an expression for  $I_C^{(2)}$  in implicit form is derived. The case of  $I_C^{(1)}$  is left as an exercise.

If  $\mathbf{x}(t)$  is a parametric description of the curve  $C$ , then substitution into (5.118) and taking the derivative with respect to  $t$  gives

$$\frac{d\mathbf{x}}{dt} \cdot \text{grad } \phi_1 = \frac{d\mathbf{x}}{dt} \cdot \text{grad } \phi_2 = 0. \quad (5.120)$$

This means that  $d\mathbf{x}/dt$  and  $\text{grad}\phi_1 \times \text{grad}\phi_2$  must be parallel. Therefore, there must be some scalar function  $\alpha(\mathbf{x})$  (that is yet to be determined) such that

$$I_C^{(2)} = \int_{\mathbb{R}^3} \mathbf{f}(\mathbf{x}) \cdot [\text{grad } \phi_1 \times \text{grad } \phi_2] \alpha(\mathbf{x}) \delta(\phi_1(\mathbf{x})) \delta(\phi_2(\mathbf{x})) d\mathbf{x}. \quad (5.121)$$

In order to derive the form of  $\alpha(\mathbf{x})$ , a parametrization of the tubular region that envelops the curve can be established. Any point in this region can be expressed as  $\mathbf{x}(t; u_1, u_2)$ , where  $u_1$  and  $u_2$  are Cartesian coordinates in the plane normal to the curve at  $t$ , and the curve itself can be defined by  $u_1 = u_2 \equiv 0$ . And the constraint (5.118) can be expressed in these coordinates as

$$\phi_1(\mathbf{x}(t; u_1, u_2)) = u_1 \quad \text{and} \quad \phi_2(\mathbf{x}(t; u_1, u_2)) = u_2. \quad (5.122)$$

Taking the partial derivatives of the above two equations with respect to  $u_1$  and  $u_2$  gives

$$\frac{\partial \mathbf{x}}{\partial u_1} \cdot \text{grad } \phi_1 = \frac{\partial \mathbf{x}}{\partial u_2} \cdot \text{grad } \phi_2 = 1 \quad (5.123)$$

and

$$\frac{\partial \mathbf{x}}{\partial u_2} \cdot \text{grad } \phi_1 = \frac{\partial \mathbf{x}}{\partial u_1} \cdot \text{grad } \phi_2 = 0. \quad (5.124)$$

These imply that

$$\text{grad } \phi_i = \frac{1}{\|\partial \mathbf{x} / \partial u_i\|^2} \frac{\partial \mathbf{x}}{\partial u_i} \quad \text{for } i = 1, 2. \quad (5.125)$$

Since  $(t, u_1, u_2)$  is an orthogonal curvilinear coordinate system, each of the vectors  $\partial \mathbf{x} / \partial t$ ,  $\partial \mathbf{x} / \partial u_1$ , and  $\partial \mathbf{x} / \partial u_2$  is orthogonal to each other, and so  $|G(t, u_1, u_2)|^{\frac{1}{2}}$  is expressed as the product of the magnitudes of these partials. This means that (5.121) can be expressed in parametric form as

$$\begin{aligned} I_C^{(2)} &= \int_t \int_{u_1} \int_{u_2} \mathbf{f}(\mathbf{x}(t; u_1, u_2)) \cdot \frac{\partial \mathbf{x} / \partial u_1 \times \partial \mathbf{x} / \partial u_2}{\|\partial \mathbf{x} / \partial u_1\|^2 \cdot \|\partial \mathbf{x} / \partial u_2\|^2} \alpha(\mathbf{x}(t; u_1, u_2)) \\ &\quad \times \delta(u_1) \delta(u_2) \|\partial \mathbf{x} / \partial t\| \|\partial \mathbf{x} / \partial u_1\| \|\partial \mathbf{x} / \partial u_2\| du_2 du_1 dt. \end{aligned}$$

This simplifies to

$$I_C^{(2)} = \int_t \mathbf{f}(\mathbf{x}(t; 0, 0)) \cdot \frac{\partial \mathbf{x} / \partial u_1 \times \partial \mathbf{x} / \partial u_2}{\|\partial \mathbf{x} / \partial u_1\| \cdot \|\partial \mathbf{x} / \partial u_2\|} \alpha(\mathbf{x}(t; 0, 0)) \|\partial \mathbf{x} / \partial t\| dt. \quad (5.126)$$

But due to the orthogonality of the tubular coordinate system,

$$\frac{\partial \mathbf{x} / \partial u_1 \times \partial \mathbf{x} / \partial u_1}{\|\partial \mathbf{x} / \partial u_1\| \cdot \|\partial \mathbf{x} / \partial u_2\|} = \frac{\partial \mathbf{x} / \partial t}{\|\partial \mathbf{x} / \partial t\|}.$$

And so (5.126) becomes the expression in (5.119) when  $\alpha(\mathbf{x}) = 1$ .

The line integral  $I_C^{(2)}$  can be expressed in two forms that are both independent of parameters:

$$\int_C \mathbf{f}(\mathbf{x}) \cdot d\mathbf{x} = \int_{\mathbb{R}^3} \mathbf{f}(\mathbf{x}) \cdot [\text{grad } \phi_1 \times \text{grad } \phi_2] \delta(\phi_1(\mathbf{x})) \delta(\phi_2(\mathbf{x})) d(\mathbf{x}). \quad (5.127)$$

This is one of the rare instances in which  $d\mathbf{x} = \mathbf{x}(t + dt) - \mathbf{x}(t)$  and  $d(\mathbf{x}) = dx_1 dx_2 dx_3$  appear in the same expression. And therefore the shorthand  $d\mathbf{x}$  cannot be used for  $d(\mathbf{x})$ .

### 5.7.3 Integral Theorems for Implicit Surfaces

The classical theorems of multivariate integral calculus can all be recast in implicit function notation. For a smooth vector field  $\mathbf{f}(\mathbf{x})$ , the divergence theorem (A.120) becomes

$$\int_{\mathbb{R}^3} (\nabla \cdot \mathbf{f}) [1 - H(\phi(\mathbf{x}))] d\mathbf{x} = \int_{\mathbb{R}^3} (\mathbf{f} \cdot \nabla \phi) \delta(\phi(\mathbf{x})) d\mathbf{x} \quad (5.128)$$

where  $\delta(\cdot)$  is the Dirac delta function on the real line, which is used to enforce the constraint  $\phi(\mathbf{x}) = 0$ , and  $H(\cdot)$  is the Heaviside step function in (2.8) that makes  $H(\phi(\mathbf{x})) = 0$  on the solid body bound by the surface since by definition  $\phi(\mathbf{x}) < 0$  on the interior. Note that  $\delta(y) = dH/dy$ .

Stokes' theorem (5.42) becomes

$$\int_{\mathbb{R}^3} (\nabla \times \mathbf{f}) \delta(\phi_1(\mathbf{x})) d\mathbf{x} = \int_{\mathbb{R}^3} \frac{[\mathbf{f}, \nabla \phi_1, \nabla \phi_2]}{\|\nabla \phi_1 \times \nabla \phi_2\|} \delta(\phi_1(\mathbf{x})) \delta(\phi_2(\mathbf{x})) d\mathbf{x} \quad (5.129)$$

(where  $[\cdot, \cdot, \cdot]$  is the triple product and the boundary curve is described as the intersection of two implicit surfaces  $\phi_1(\mathbf{x}) = 0$  and  $\phi_2(\mathbf{x}) = 0$ , the tangent to which is found by crossing the normals of the two surfaces), and the Gauss–Bonnet theorem (5.73) becomes

$$\int_{\mathbb{R}^3} \frac{1}{\|\nabla \phi\|^3} \det \begin{bmatrix} \nabla \nabla^T \phi & \nabla \phi \\ \nabla^T \phi & 0 \end{bmatrix} \delta(\phi(\mathbf{x})) d\mathbf{x} = 2\pi \chi(S) \quad (5.130)$$

where  $S$  is the closed surface defined by the constraint  $\phi(\mathbf{x}) = 0$ .

### 5.7.4 Level Sets and Curvature Flows

#### Level Set and Fast Marching Methods

Given an implicit surface defined by the equation  $\phi(\mathbf{x}) = 0$ , a fundamental question that arises is, “How can points on the surface be traced out rapidly?” This is one of the subjects discussed in the area of research known as *level set methods* [54, 55, 63]. The basic idea is that once one point on the surface is known, other neighboring points can be approximated well since the local shape of the surface can be computed.

If an initial set of surface points are known, and then the initial function is perturbed and becomes  $\phi(\mathbf{x}) + \epsilon(\mathbf{x}) = 0$  where  $\epsilon(\mathbf{x})$  is a function that takes small values, how should the initially determined points update their positions so as to sit on the new surface? Another related problem is, given an implicitly defined surface, how can the tube/offset of that surface be computed rapidly [43]? These dynamic problems that involve change of a surface, and in which a corresponding change in the positions of points is sought, are addressed well using *fast marching methods*.

Level set and fast marching methods have found applications in many areas including computer vision, image processing [41], mechanics, and control. A particular application, which involves the evolution of a surface so as to minimize the mean curvature of the resulting surface, is discussed in the following subsection.



## Evolving Surfaces and Curvature Flows

Two different kinds of stochastic models involving surfaces are common in the literature. The first kind involves surfaces that change shape in a stochastic way. This can be written in parametric form as  $\mathbf{x}(u_1, u_2; t)$ , or in implicit form as  $\phi(\mathbf{x}; t) = 0$ . A simple example of this would be a sphere with radius that varies with time:  $r = r(t)$ , where  $r(t)$  might consist of a deterministic and a noise part. The second sort of stochastic model involving surfaces is the evolution of stochastic paths on a fixed surface. This can be written in parametric form as  $\mathbf{x}(u_1(t), u_2(t))$  or in implicit form as  $\phi(\mathbf{x}(t)) = 0$ . This second kind of problem, where the “surface” can actually be a more-than-two-dimensional extension of the concept of a surface (called a manifold), is the subject of Chapter 8, and the intervening chapters lay the foundations required to fully address this sort of problem. However, as a side note, stochastic surface models of the first kind are briefly reviewed here. In recent years, such models have found applications in mechanics [42, 72], computer vision [26, 51, 61], and image processing [40, 66]. In these applications a surface is evolved by a simple rule (which can be either deterministic or stochastic) so as to capture some physical or visual feature. The deterministic version of the problem has been applied widely in image processing [12, 73], and to the approximation of molecular surfaces [3].

A popular problem is to evolve a surface by its own local features. In particular, the implicit function  $\phi(\mathbf{x}; t) = 0$  that defines the surface at any particular time is forced to obey the following non-linear evolution equation:

$$\frac{\partial \phi}{\partial t} = f(m) \|\nabla \phi\| \quad \text{with} \quad \phi(\mathbf{x}; 0) = \phi_0(\mathbf{x}) \quad (5.131)$$

where  $m$  is the mean curvature given in (5.110). Both  $m$  and  $\|\nabla \phi\|$  are non-linear operators that act on  $\phi$ , and  $f : \mathbb{R} \rightarrow \mathbb{R}$  can also be a non-linear function that is either deterministic or stochastic. A surface that evolves deterministically according to mean curvature flow is one for which  $f(m) = m$ . This sort of formulation is used to evolve surfaces into ones that are smoother (i.e., have reduced curvature). When boundary conditions are imposed, this can be used to evolve an initial surface into one with minimal total mean curvature [10, 18, 39]. Interestingly, in the case when a closed convex planar curve is used as the initial shape, then the shape transformation process can be viewed as one in which the “curvature entropy”

$$S(\kappa) \doteq - \oint \kappa(s, t) \log \kappa(s, t) ds$$

increases as a function of time until the maximum entropy curvature function results (which occurs when the curvature flow converges to a circle). This sort of argument has been used in [29, 33] (though their definition of entropy has an opposite sign).

Both the deterministic [21, 22, 34] and stochastic [40, 66] versions of curvature flow problem have been studied extensively. For more pointers to the literature, see [11, 55, 63].

While the focus of this section has been on surfaces in  $\mathbb{R}^3$ , many of the formulas extend naturally to include  $(n - 1)$ -dimensional hyper-surfaces in  $\mathbb{R}^n$  [20, 30, 35]. In contrast, the next chapter discusses the intrinsic coordinate-free geometry of “manifolds” (which includes hyper-surfaces as a special case).

## 5.8 Chapter Summary

This chapter focused on the geometry of curves and surfaces in two- and three-dimensional space. It began with elementary (and practical) examples involving a robot arm and a medical imaging problem. These examples illustrate how problems in analytic and projective geometry can be posed either parametrically or implicitly. Then the mathematical machinery required to compute arc length, area, volume, curvature, etc., was introduced. The relationships between locally defined geometric quantities and global topological features were examined. Geometric quantities that are usually described parametrically were recast implicitly. The divergence theorem and Stokes' theorem were written in implicit form (which is something that the author has not seen elsewhere).

In addition to those works cited throughout this chapter, accessible introductions to differential geometry of curves and surfaces with many examples include [19, 31, 47]. Alternative ways to attach reference frames to curves is an area that has received attention in both the kinematics [9] and mathematics literature [5]. This will be discussed in the context of variational problems in Volume 2. It is important to note that differential geometry is not the only kind of geometry. Accessible introductions to algebraic aspects of geometry include [1, 52].

The next two chapters extend the differential geometric concepts presented here to higher dimensions. The concept of differential forms is very useful when it comes to describing integrals on high-dimensional manifolds. After a detailed introduction to differential forms in Chapter 6, the concepts of volume, curvature, Euler characteristic, etc., are defined for high-dimensional geometric objects in Chapter 7. The material presented in these chapters will be important for understanding how to define SDEs and Fokker–Planck equations on manifolds and Lie groups.

## 5.9 Exercises

5.1. Compute the Jacobian determinant for (5.1) and verify that the singularities occur when  $q_2 = 0$  or  $\pi$ .

5.2. For the robot arm described in Section 5.1, write computer programs to implement all three of the inverse kinematics routines described in Section 5.1.2. Let  $L_1 = L_2 = 1$ , and take as the starting conformation  $\mathbf{q}(0) = [\pi/2, -\pi/2]^T$ , which corresponds to  $\mathbf{x}(0) = [1, 1]^T$ . Simulate the arm following a circle of the form  $\mathbf{x}(t) = [1 - r + r \cos 2\pi t, 1 + r \sin 2\pi t]^T$  for  $0 \leq t \leq 1$ . Try values of  $r = 0.2, 0.5, 1.1$ . What happens? Why does this happen?

5.3. Pick several values for  $h, r, \alpha, \beta, m_1, m_2, m_3$  and plot the parametric curve  $\boldsymbol{\pi}(\theta)$  in (5.12) for  $\theta \in [0, 2\pi)$ .

5.4. Write two computer programs, one to implement the iterative Jacobian-inversion scheme in (5.15), the other to use (5.16) for updating using the rule  $\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \Delta \mathbf{q}(t)$ . Using the same parameters as those chosen in Exercise 3 above, evaluate the convergence of these iterative methods.

5.5. Compare the performance of the iterative numerical methods in Exercise 4 above, with the implicit algebraic approach in Section 5.2.2.

5.6. Show that a planar arc-length, parameterized curve  $\mathbf{x}(s)$  with  $\mathbf{x}(0) = \mathbf{0}$  and tangent vector  $\mathbf{x}'(0) = \mathbf{e}_1$  can be completely characterized in terms of signed curvature  $k(s)$  as

$$\mathbf{x}(s) = \begin{pmatrix} \int_0^s \cos \theta(\sigma) d\sigma \\ \int_0^s \sin \theta(\sigma) d\sigma \end{pmatrix} \quad \text{where} \quad \theta(s) = \int_0^s k(\sigma) d\sigma.$$

5.7. Using the chain rule, show that (5.29) and (5.33) are equivalent.

5.8. Calculate the curvature and torsion of the right-handed circular helix

$$\mathbf{x}(s) = (r \cos as, r \sin as, has)^T \quad (5.132)$$

where  $a = (r^2 + h^2)^{-1/2}$ , and  $r$  and  $h$  are constants. (Before doing so, verify that  $s$  is in fact the arc length.)

5.9. Prove that a curve confined to the surface of a sphere of radius  $R$  in  $\mathbb{R}^3$  can have curvature no less than  $1/R$ .

5.10. Prove that the volume of the region in  $\mathbb{R}^3$  defined by all vectors of the form  $\mathbf{x}(u, v, w) = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$  for  $(u, v, w) \in [0, 1] \times [0, 1] \times [0, 1]$  is given by  $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ .

5.11. Consider the torus in Section 5.4.5 with  $R = 1$  and  $r = 0.2$ . Use the Langevin–Rosenburg theorem to verify that the torus is not knotted by numerically calculating the integral of absolute Gaussian curvature.

5.12. Apply a shear transformation of the form  $\mathbf{s}(\mathbf{x})$  with  $s_1 = 0.1$  and  $s_2 = 0.5$  in (5.20) to the torus in the previous problem. Either show analytically, or write computer programs to numerically verify that: (a) the volume enclosed by the deformed torus is not changed by this transformation; (b) the integral of Gaussian curvature over the whole surface is not changed by this deformation. Hint: When computing volume, use the divergence theorem to calculate it as a surface integral.

5.13. Prove that every curve  $\mathbf{c}(s) = \mathbf{x}(q_1(s), q_2(s))$  that passes through the point  $\mathbf{c}(s_0) = \mathbf{c}_0$  and has tangent vector  $\mathbf{u}(s_0) = d\mathbf{c}/ds|_{s=s_0}$  has the same normal curvature at that point.

5.14. Prove that (5.55) and (5.56) are equivalent.

5.15. Prove that the Gaussian and mean curvature can be written in terms of the coefficient matrices of the first and fundamental forms as

$$m = \frac{g_{11}L_{22} + g_{22}L_{11} - 2g_{12}L_{12}}{2(g_{11}g_{21} - g_{12}^2)} \quad \text{and} \quad k = \frac{L_{11}L_{22} - L_{12}^2}{g_{11}g_{22} - g_{12}^2}. \quad (5.133)$$

5.16. Using facts about eigenvectors of symmetric matrices, what can be concluded about the tangent vectors to a surface that point along the directions of principal curvature?

5.17. Using the definition in (5.58) and the fact that  $GG^{-1} = G^{-1}G = \mathbb{I}$ , prove (5.59).

5.18. Prove the *Codazzi–Mainardi equations*:

$$\frac{\partial L_{ij}}{\partial u_k} - \frac{\partial L_{ik}}{\partial u_j} = \sum_l (\Gamma_{ik}^l L_{lj} - \Gamma_{ij}^l L_{lk}).$$

5.19. The simple closed space curve (which can be found in the MATLAB™ Demos, and is due to Professor Rouben Rastamian)

$$\mathbf{x}(t) = [x(t), x'(t), x''(t)]^T \quad \text{where} \quad x(t) = \sin(t) + 2 \sin(2t) - \frac{3}{2} \cos(2t)/2 + \frac{3}{2} \sin(3t)$$

for  $t \in [0, 2\pi)$  forms a knot. Calculate the curvature of this knot as a closed form expression, and write a computer program to numerically calculate the total curvature, and verify that the conditions in the Fary–Milnor theorem are satisfied.

5.20. Construct a tube of radius 0.5 around the curve in the previous exercise. Write a computer program that numerically calculates the Gaussian curvature on a fine grid of points on that tubular surface (e.g., increments in  $t$  of  $\Delta t = 2\pi/100$  and in the circumferential variable  $\theta$  of  $\Delta\theta = \delta t$ ). Use these values to (approximately) verify the Gauss–Bonnet and Langevin–Rosenburg theorems.

5.21. Prove (5.96).

5.22. Prove (5.99).

5.23. The concept of a tube (or offset) is not limited to curves and surfaces in Euclidean space. Let  $\mathbf{u}(t)$  be a smooth closed simple curve contained in the unit sphere,  $S^2$ . Within the unit sphere, the distance between two points is calculated as  $d(\mathbf{u}_1, \mathbf{u}_2) = \cos^{-1}(\mathbf{u}_1 \cdot \mathbf{u}_2)$ . A tube around  $\mathbf{u}(t)$  can then be defined as the set of points on the surface of the sphere:  $T_u = \{\mathbf{x} \in S^2 \mid d(\mathbf{u}, \mathbf{x}) < r\}$ , where  $r$  is smaller than the minimal radius of curvature of  $\mathbf{u}(t)$ . What will the tube formulas for the length of offset curves and the area of the strip on the sphere be in this case?

5.24. A surface of revolution in  $\mathbb{R}^3$  can be parameterized as

$$\mathbf{x}(\phi, \theta) = [r(z) \cos \phi, r(z) \sin \phi, z]^T$$

where  $r(z)$  is a specified function. (a) Under what conditions on  $r(z)$  will the surface of revolution be a simple, closed, and differentiable surface? (b) Transform the original surface to  $\mathbf{x}'(\phi, \theta) = R_3(\theta(z))\mathbf{x}(\phi, \theta)$  where  $\theta(z)$  is a smooth function. What will the new surface look like? Compute the Gaussian curvature of  $\mathbf{x}'(\phi, \theta)$  and  $\mathbf{x}(\phi, \theta)$  and compare.

5.25. Write a computer program to numerically verify for the ellipsoid in Section 5.4.4 that  $K = 4\pi$  and  $F$  is given by the provided formulas.

5.26. For the ellipsoid in (5.105) calculate the Gaussian and mean curvature using (5.109) and (5.110), and compare with the values obtained for the parameterized ellipsoid of revolution in Section 5.4.4.

5.27. A *ruled surface* is one that can be parameterized as

$$\mathbf{x}(u_1, u_2) = \mathbf{c}(u_1) + u_2\mathbf{v}(u_1) \tag{5.134}$$

where  $\mathbf{c}(u_1)$  and  $\mathbf{v}(u_1)$  are arbitrary differentiable vector-valued functions. A ribbon is a special case of a ruled surface. Do the following: (a) Calculate the mean and Gaussian curvatures for the ruled surface  $\mathbf{x}(u_1, u_2)$  in (5.134) (b) What conclusions can be drawn about the general properties of curvature of ruled surfaces? (c) Show that a hyperboloid of one sheet given by the implicit equation

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{a^2} - \frac{x_3^2}{c^2} = 1$$

is a ruled surface by finding a parametrization for it of the form in (5.134).

5.28. Obtain a closed-form implicit formula of the form  $\phi(\mathbf{x}) = 0$  for the torus parameterized in (5.72).

5.29. Prove (5.109) and (5.110) by using the parametric formulas for Gaussian and mean curvature, and substituting in the corresponding quantities defined in terms of the implicit surface constraint  $\phi(\mathbf{x}) = 0$ . Verify that (5.110) is written explicitly in component form as

$$m = \frac{1}{2\|\nabla\phi\|^3} [\phi_{x_1,x_1}(\phi_{x_2}^2 + \phi_{x_3}^2) + \phi_{x_2,x_2}(\phi_{x_1}^2 + \phi_{x_3}^2) + \phi_{x_3,x_3}(\phi_{x_1}^2 + \phi_{x_2}^2) - 2\phi_{x_1,x_2}\phi_{x_1}\phi_{x_2} - 2\phi_{x_1,x_3}\phi_{x_1}\phi_{x_3} - 2\phi_{x_2,x_3}\phi_{x_2}\phi_{x_3}]$$

where  $\phi_{x_i} \doteq \partial\phi/\partial x_i$  and  $\phi_{x_i,x_j} \doteq \partial^2\phi/\partial x_i\partial x_j$ . What is (5.109) explicitly in component form?

5.30 Show that the signed curvature of an implicitly defined planar curve  $\psi(x_1, x_2) = 0$  can be written as

$$k = (\psi_{x_1,x_1}\psi_{x_2}^2 - 2\psi_{x_1,x_2}\psi_{x_1}\psi_{x_2} + \psi_{x_2,x_2}\psi_{x_1}^2)/(\psi_{x_1}^2 + \psi_{x_2}^2)^{3/2} \quad (5.135)$$

where  $\psi_{x_i} \doteq \partial\psi/\partial x_i$  and  $\psi_{x_i,x_j} \doteq \partial^2\psi/\partial x_i\partial x_j$ . Use this fact to derive (5.110) by performing the following steps: (1) slice the surface  $\phi(\mathbf{x}) = 0$  with all planes passing through the point  $\mathbf{x} = \mathbf{x}(u_1, u_2)$  and containing the normal; (2) compute the curvature of the resulting plane curve; and (3) average this curvature over all of the slicing planes.

5.31. Verify that (5.135) can be computed as

$$k = -\frac{1}{\|\nabla\psi\|^3} \det \begin{bmatrix} \nabla\nabla^T\psi & \nabla\psi \\ \nabla^T\psi & 0 \end{bmatrix} = \frac{\|\nabla\psi\|^2 \text{tr}(\nabla\nabla^T\psi) - (\nabla^T\psi)(\nabla\nabla^T\psi)(\nabla\psi)}{\|\nabla\psi\|^3} \quad (5.136)$$

$$= \nabla \cdot \left( \frac{\nabla\psi}{\|\nabla\psi\|} \right) \quad (5.137)$$

where  $\nabla\psi = [\psi_{x_1}, \psi_{x_2}]^T$ . That is, in this planar case the formulas for mean and Gaussian curvature collapse into the same expression. Note the slight differences in (5.136)–(5.137) relative to (5.109) and (5.110). In general, for an implicit  $n$ -dimensional “hyper-surface”  $\phi(\mathbf{x}) = 0$  for  $\mathbf{x} \in \mathbb{R}^{n+1}$ , it can be shown that [20, 30, 35]

$$k = (-1)^n \frac{1}{\|\nabla\phi\|^{n+2}} \det \begin{bmatrix} \nabla\nabla^T\phi & \nabla\phi \\ \nabla^T\phi & 0 \end{bmatrix} \quad (5.138)$$

and

$$m = \frac{\|\nabla\phi\|^2 \text{tr}(\nabla\nabla^T\phi) - (\nabla^T\phi)(\nabla\nabla^T\phi)(\nabla\phi)}{n\|\nabla\phi\|^3} = \nabla \cdot \left( \frac{\nabla\phi}{\|\nabla\phi\|} \right) \quad (5.139)$$

where  $\nabla = \partial/\partial\mathbf{x} = [\partial/\partial x_1, \dots, \partial/\partial x_{n+1}]^T$ .

Evaluate these formulas for the sphere of radius  $r$  in  $\mathbb{R}^{n+1}$  defined by  $x_1^2 + x_2^2 + \dots + x_{n+1}^2 = r^2$ .

5.32. Derive the implicit version of the integral  $I_C^{(1)}$  in (5.119).

5.33. By extending the pattern observed in the plane and in three-dimensional space, the Euler characteristic can be extended to bodies in four dimensions as

$$\chi(B) = f_0(B) - f_1(B) + f_2(B) - f_3(B) + f_4(B) \quad (5.140)$$

where  $f_0$  denotes zero-dimensional vertices,  $f_1$  denotes one-dimensional edges, etc. Given an array consisting of  $L \times W \times H \times D$  of four-dimensional cubes, what would the formulas for  $f_i(B)$  analogous to those given in Section 5.6.1 be? Hint: The formulas should be symmetric in  $L, W, H, D$ , in the same way that the formulas in lower dimensions were.

## References

1. Abhyankar, S.S., *Algebraic Geometry for Scientists and Engineers*, Mathematical Surveys and Monographs, 35, American Mathematical Society, Providence, RI, 1990.
2. Adams, C.C., *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*, W.H. Freeman, New York, 1994.
3. Bates, P.W., Wei, G.W., Zhao, S., "Minimal molecular surfaces and their applications," *J. Comput. Chem.*, 29, pp. 380–391, 2008.
4. Ben-Israel, A., Greville, T.N.E., *Generalized Inverses: Theory and Applications*, 2nd ed., Canadian Mathematical Society Books in Mathematics, Springer, New York, 2003.
5. Bishop, R., "There is more than one way to frame a curve," *Amer. Math. Month.*, 82, pp. 246–251, 1975.
6. Blackmore, D., Leu, M.C., Wang, L.P., "The sweep-envelope differential equation algorithm and its application to NC machining verification," *Computer-Aided Design*, 29, pp. 629–637, 1997.
7. Bloomenthal, J., (ed.), *Introduction to Implicit Surfaces*, Morgan Kaufmann, San Francisco, 1997.
8. Bloomenthal, J., Shoemake, K., "Convolution surfaces," *Computer Graphics*, 25, pp. 251–256, 1991 (Proc. SIGGRAPH'91).
9. Bottema, O., Roth, B., *Theoretical Kinematics*, Dover, New York, 1990.
10. Brakke, K.A., *The Motion of a Surface by its Mean Curvature*, Princeton University Press, Princeton, NJ, 1978.
11. Buttazzo, G., Visintin, A., eds., *Motion by Mean Curvature and Related Topics*, Proceedings of the international conference held at Trento, July 20–24, 1992. de Gruyter, Berlin, 1994.
12. Chan, T.F., Vese, L.A., "Active contours without edges," *IEEE Trans. Image Process.*, 10, pp. 266–277, 2001.
13. Chazvini, M., "Reducing the inverse kinematics of manipulators to the solution of a generalized eigenproblem," *Computational Kinematics*, pp. 15–26, 1993.
14. Chen, B.-Y., *Total Mean Curvature and Submanifolds of Finite Type*, World Scientific, Singapore, 1984.
15. Chen, B.-Y., "On the total curvature of immersed manifolds," *Amer. J. Math.*, 93, pp. 148–162, 1971; 94, pp. 899–907, 1972; 95, pp. 636–642, 1973.
16. Chen, B.-Y., "On an inequality of mean curvature," *J. London Math. Soc.*, 4, pp. 647–650, 1972.
17. Chirikjian, G.S., "Closed-form primitives for generating locally volume preserving deformations," *ASME J. Mech. Des.*, 117, pp. 347–354, 1995.

18. Chopp, D.L., "Computing minimal surfaces via level set curvature flow," *J. Comput. Phys.*, 106, pp. 77–91, 1993.
19. do Carmo, M., *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, NJ, 1976.
20. Dombrowski, P., "Krümmungsgrossen gleichungsdefinierter Untermannigfaltigkeiten Riemannscher Mannigfaltigkeiten," *Math. Nachr.*, 38, pp. 133–180, 1968.
21. Evans, L.C., Spruck, J., "Motion of level sets by mean curvature," *J. Diff. Geom.*, 33, pp. 635–681, 1991.
22. Evans, L.C., Spruck, J., "Motion of level sets by mean curvature II," *Trans. Amer. Math. Soc.*, 330, pp. 321–332, 1992.
23. Farouki, R.T., Neff, C.A., "Analytical properties of plane offset curves," *Computer Aided Geometric Design*, 7, pp. 83–99, 1990.
24. Farouki, R.T., Neff, C.A., "Algebraic properties of plane offset curves," *Computer Aided Geometric Design*, 7, pp. 101–127, 1990.
25. Fary, I., "Sur la courbure totale d'une courbe gauche faisant un noeud," *Bull. Soc. Math. Fr.*, 77, pp. 128–138, 1949.
26. Faugeras, O., Keriven, R., "Variational principles, surface evolution, PDE's, level set methods and the stereo problem," *IEEE Trans. Image Process.*, 7, pp. 336–344, 1998.
27. Fenchel, W., "Über Krümmung und Windung geschlossenen Raumkurven," *Math. Ann.*, 101, pp. 238–252, 1929.
28. Fox, R.H., "On the total curvature of some tame knots," *Ann. Math.*, 52, pp. 258–261, 1950.
29. Gage, M., Hamilton, R.S., "The heat equation shrinking convex plane curves," *J. Diff. Geom.*, 23, pp. 69–96, 1986.
30. Goldman, R., "Curvature formulas for implicit curves and surfaces," *Computer Aided Geometric Design*, 22, pp. 632–658, 2005.
31. Gray, A., Abbena, E., Salamon, S., *Modern Differential Geometry of Curves and Surfaces with MATHEMATICA*, Chapman & Hall/CRC, Boca Raton, FL, 2006.
32. Gray, A., *Tubes*, 2nd ed., Birkhäuser, Boston, 2004.
33. Grayson, M., "The heat equation shrinks embedded plane curves to round points," *J. Diff. Geom.*, 26, pp. 285–314, 1987.
34. Grayson, M., "A short note on the evolution of a surface by its mean curvature," *Duke Math. J.*, 58, pp. 555–558, 1989.
35. Gromoll, D., Klingenberg, W., Meyer, W., *Riemannsche Geometrie im Grossen*. Lecture Notes in Mathematics, Vol. 55. Springer, Berlin, 1975.
36. Guggenheimer, H.W., *Differential Geometry*, Dover, New York, 1977.
37. Hadwiger, H., *Altes und Neues über Konvexe Körper*, Birkhäuser Verlag, Basel, 1955.
38. Hodge, W.V.D., Pedoe, D., *Methods of Algebraic Geometry*, Vols. 1–3, Cambridge University Press, London, 1952, (reissued 1994).
39. Huisken, G., "Flow by mean curvature of convex surfaces into spheres," *J. Diff. Geom.*, 20, p. 237, 1984.
40. Juan, O., Keriven, R., Postelnicu, G., "Stochastic motion and the level set method in computer vision: Stochastic active contours," *Int. J. Comput. Vision*, 69, pp. 7–25, 2006.
41. Kass, M., Witkin, A., Terzopoulos, D., "Snakes: Active contour models," *Int. J. Comput. Vision*, 1, pp. 321–331, 1988.
42. Katsoulakis, M.A., Kho, A.T., "Stochastic curvature flows: Asymptotic derivation, level set formulation and numerical experiments," *J. Interfaces Free Boundaries*, 3, pp. 265–290, 2001.
43. Kimmel, R., Bruckstein, A.M., "Shape offsets via level sets," *Computer-Aided Design*, 25, pp. 154–162, 1993.
44. Kohli, D., Osvatic, M., "Inverse kinematics of the general 6R and 5R; P serial manipulators," *ASME J. Mech. Des.*, 115, pp. 922–931, 1993.
45. Kuiper, N.H., Meeks, W.H., "The total curvature of a knotted torus," *J. Diff. Geom.*, 26, pp. 371–384, 1987.

46. Langevin, R., Rosenburg, H., "On curvature integrals and knots," *Topology*, 15, pp. 405–416, 1976.
47. Lipschutz, M.M., *Differential Geometry*, Schaum's Outline Series, McGraw-Hill, New York, 1969.
48. Manocha, D., Canny, J., "Efficient inverse kinematics for general 6R manipulators," *IEEE Trans. Robot. Automat.*, 10, pp. 648–657, 1994.
49. Millman, R.S., Parker, G.D., *Elements of Differential Geometry*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
50. Milnor, J., "On the total curvature of knots," *Ann. Math.*, 52, pp. 248–257, 1950.
51. Mumford, D., Shah, J., "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, 42, p. 577, 1989.
52. Olver, P.J., *Classical Invariant Theory*, Cambridge University Press, London, 1999.
53. Oprea, J., *Differential Geometry and Its Applications*, 2nd ed., The Mathematical Association of America, Washington, DC, 2007.
54. Osher, S., Sethian, J.A., "Fronts propagating with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations," *J. Comput. Phys.*, 79, pp. 12–49, 1988.
55. Osher, S.J., Fedkiw, R.P., *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York, 2002.
56. Osserman, R., "Curvature in the eighties," *Amer. Math. Month.*, 97, pp. 731–756, 1990.
57. Pham, B., "Offset curves and surfaces: A brief survey," *Computer-Aided Design*, 24, pp. 223–229, 1992.
58. Raghavan, M., Roth, B., "Inverse kinematics of the general 6R manipulator and related linkages," *ASME J. Mech. Des.*, 115, pp. 502–508, 1993.
59. Rolfsen, D., *Knots and Links*, Publish or Perish Press, Wilmington, DE, 1976.
60. Ros, A., "Compact surfaces with constant scalar curvature and a congruence theorem," *J. Diff. Geom.*, 27, pp. 215–220, 1988.
61. San Jose Estepar, R., Haker, S., Westin, C.F., "Riemannian mean curvature flow," in *Lecture Notes in Computer Science: ISVC05*, 3804, pp. 613–620, Springer, 2005.
62. Schubert, H., "Über eine Numerische Knoteninvariante," *Math. Z.*, 61, pp. 245–288, 1954.
63. Sethian, J.A., *Level Set Methods and Fast Marching Methods : Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, 2nd ed., Cambridge University Press, London, 1999.
64. Shiohama, K., Takagi, R., "A characterization of a standard torus in  $E^3$ ," *J. Diff. Geom.*, 4, pp. 477–485, 1970.
65. Sommese, A.J., Wampler, C.W., *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*, World Scientific, Singapore, 2005.
66. Soner, H.M., Touzi, N., "A stochastic representation for mean curvature type geometric flows," *Ann. Prob.*, 31, pp. 1145–1165, 2003.
67. Sullivan, J.M., "Curvatures of smooth and discrete surfaces," in *Discrete Differential Geometry*, A.I. Bobenko, P. Schröder, J.M. Sullivan, and G.M. Ziegler, eds., Oberwolfach Seminars, Vol. 38, pp. 175–188, Birkhäuser, Basel, 2008.
68. Voss, K., "Eine Bemerkung über die Totalkrümmung geschlossener Raumkurven," *Arch. Math.*, 6, pp. 259–263, 1955.
69. Weyl, H., "On the volume of tubes," *Amer. J. Math.*, 61, pp. 461–472, 1939.
70. Willmore, T.J., "Mean curvature of Riemannian immersions," *J. London Math. Soc.*, 3, pp. 307–310, 1971.
71. Willmore, T.J., "Tight immersions and total absolute curvature," *Bull. London Math. Soc.*, 3, pp. 129–151, 1971.
72. Yip, N.K., "Stochastic motion by mean curvature," *Arch. Rational Mech. Anal.*, 144, pp. 331–355, 1998.
73. Zhang, S., Younes, L., Zweck, J., Ratnanather, J.T., "Diffeomorphic surface flow: A novel method of surface evolution," *SIAM J. Appl. Math.*, 68, pp. 806–824, 2008.



## Differential Forms

This chapter introduces differential forms, exterior differentiation, and multi-vectors in a concrete and explicit way by restricting the discussion to  $\mathbb{R}^n$ . This is extended to more general settings later. Roughly speaking, differential forms generalize and unify the concepts of the contour integral, curl, element of surface area, divergence, and volume element that are used in statements of Stokes' theorem and the divergence theorem. At first it may seem unnecessary to learn yet another new mathematical construction. The trouble is that without an appropriate extension of the concept of the cross product, it is difficult and messy to try to extend the theorems of vector calculus to higher dimensions, and to non-Euclidean spaces. As was illustrated in Chapter 1 in the context of heat and fluid flow problems, these theorems play a central role. Likewise, in probability flow problems involving stochastic differential equations and their associated Fokker–Planck equations, these theorems play a role in assessing how much probability density flows past a given surface. Since the problems of interest (such as the stochastic cart in Figure 1.1) will involve stochastic flows on Lie groups, understanding how to extend Stokes' theorem and the divergence theorem to these generalized settings will be useful. The first step in achieving this goal is to understand differential forms in  $\mathbb{R}^n$ .

Differential forms were developed by E. Cartan. Much of what is presented in this chapter is stated (in more abstract terminology) in [2, 4, 5]. The presentation here most closely follows that in [3], with the exception that the *subscript–superscript notation*, which is explained in the paragraph below, is *not* used here.

In many books on differential forms and manifolds, notation such as

$$\mathbf{v} = \sum_i v^i \mathbf{e}_i \quad \text{and} \quad df = \sum_i \frac{\partial f}{\partial x^i} dx^i \quad (6.1)$$

is used. This is consistent with the physicists' shorthand that repetition over raised and lowered indices automatically implies summation, and so in summation notation  $\mathbf{v} = v^i \mathbf{e}_i$  and  $df = \frac{\partial f}{\partial x^i} dx^i$ . It is also consistent with the idea that if the Hermitian conjugate of a vector  $\mathbf{w}$  is computed, then the result can be written as  $\mathbf{w}^* = \sum_i w_i^* \mathbf{e}^i$  where  $\mathbf{e}^i = \mathbf{e}_i^*$  is the dual (transpose) of  $\mathbf{e}_i$ ,  $w_i^* = \overline{w_i}$ , and the operation of conjugation flips superscripts and subscripts. Tu [14] points out that this raising and lowering convention has the benefit that expressions are balanced, and this provides a check for accuracy. However, in the current presentation, all summation signs will be explicitly written, and all indices will be subscripts (except in special cases where the usual subscript location is too cluttered, or if particular superscripts have already been widely accepted, e.g., the Christoffel symbol  $\Gamma_{ij}^k$ ). The use of subscript-only notation, while not universal, is

consistent with a number of other works including [2, 8, 9, 12, 15]. It also keeps things simple, and consistent with the vector and matrix notation used in engineering and computer science. Therefore, instead of (6.1), the following notation will be used:

$$\mathbf{v} = \sum_i v_i \mathbf{e}_i \quad \text{and} \quad df = \sum_i \frac{\partial f}{\partial x_i} dx_i. \quad (6.2)$$

The main points to take away from this chapter are:

- To understand the concept and properties of differential forms and multi-vectors on  $\mathbb{R}^n$ ;
- To be able to perform computations with them, including computing exterior derivatives;
- To understand how differential forms on  $\mathbb{R}^n$  behave under coordinate changes;
- To be able to perform computations with differential forms in the context of the generalized Stokes' theorem in  $\mathbb{R}^n$  (at least on box-like domains).

## 6.1 An Informal Introduction to Differential Forms on $\mathbb{R}^n$

In classical vector calculus, the differential of a smooth<sup>1</sup> scalar function  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  is defined as

$$d\phi \doteq \sum_{i=1}^n \frac{\partial \phi}{\partial x_i} dx_i$$

where  $\mathbf{x} \in \mathbb{R}^n$ . The real-valued function  $\phi(\mathbf{x})$  is sometimes called a *0-form*. The formula for the differential given above can be generalized as

$$\omega_1 \doteq \sum_{i=1}^n a_i(\mathbf{x}) dx_i, \quad (6.3)$$

where  $a_i(\mathbf{x}) = \frac{\partial \phi}{\partial x_i}$  can be viewed as a special case. The generalized quantity  $\omega_1$  is called a *differential 1-form* (or *1-form* for short). It can be viewed as a functional  $\omega_1(\mathbf{x}, d\mathbf{x})$  that is linear in  $d\mathbf{x}$  and has no restrictions on its dependence on  $\mathbf{x}$  other than that it is a smooth function in this argument. It is immediate from these definitions that a special kind of 1-form results from differentiating a 0-form.

When  $n = 1$ , the zero-form  $\omega_0(x) = \phi(x)$  and the one-form  $\omega_1(x, dx) = \frac{d\phi}{dx} dx$  are related by the Fundamental Theorem of Calculus as

$$\int_{[a,b]} \omega_1 = \omega_0|_a^b.$$

### 6.1.1 Definitions and Properties of $n$ -Forms and Exterior Derivatives

So far in this discussion, nothing of value has been added by the concept of a form. However, things get more interesting when a *differential 2-form* (or *2-form*, for short) is

---

<sup>1</sup>Infinite differentiability is not required in this definition, but it is often more convenient to restrict the discussion to  $C^\infty(\mathbb{R}^n)$  from the start rather than to  $C^1(\mathbb{R}^n)$ , followed by a restriction to  $C^2(\mathbb{R}^n)$  at the point in the discussion when two derivatives are taken, etc.

defined, and the differential of a 1-form is written as a 2-form. For each fixed  $\mathbf{x} \in \mathbb{R}^n$ , a 2-form is like a *quadratic form*<sup>2</sup> in the variable  $d\mathbf{x}$ , and is denoted as

$$\omega_2 \doteq \sum_{i=1}^n \sum_{j=1}^n b_{ij}(\mathbf{x}) dx_i \wedge dx_j. \tag{6.4}$$

The function  $\omega_2(\mathbf{x}, d\mathbf{x})$  is quadratic in  $d\mathbf{x}$  in the sense that two entries in this vector “multiply” each other with the *wedge product*,  $\wedge$ . However, unlike usual quadratic forms which are symmetric in the variables due to the commutativity of scalar multiplication ( $v_1 v_2 = v_2 v_1$ ), a two-form is defined to be *skew-symmetric*, or *anti-commuting*, due to the following postulated property of the wedge product:

$$\boxed{dx_j \wedge dx_i \doteq -dx_i \wedge dx_j.} \tag{6.5}$$

With the exception of this anti-commutative property, the multiplication of differentials when using the wedge product can be viewed as scalar multiplication, which means that it is distributive and associative. In other words, it is postulated that for any real-valued functions  $f(\mathbf{x})$  and  $g_j(\mathbf{x})$ ,

$$\boxed{dx_i \wedge (f dx_j) \doteq (f dx_i) \wedge dx_j \doteq f \cdot (dx_i \wedge dx_j)} \tag{6.6}$$

and

$$\boxed{dx_i \wedge \left( \sum_j g_j dx_j \right) \doteq \sum_j dx_i \wedge (g_j dx_j)} \tag{6.7}$$

where  $\cdot$  just means scalar multiplication.

It follows immediately from (6.5) that

$$dx_i \wedge dx_i = 0.$$

And furthermore, any 2-form on  $\mathbb{R}^n$  can be written as

$$\sum_{i=1}^n \sum_{j=1}^n f_{ij} dx_i \wedge dx_j = \sum_{i=1}^n \sum_{j=i+1}^n \tilde{f}_{ij} dx_i \wedge dx_j$$

where  $\tilde{f}_{ij} = f_{ij} - f_{ji}$ .

Therefore, if  $n = 1$ , then  $\omega_2 = 0$  because there is no way to avoid differentials with repeated indices multiplying each other under the wedge product in this case. And if  $n = 2$ , then working through the double sum in (6.4) gives

$$\omega_2 = b_{12} dx_1 \wedge dx_2 + b_{21} dx_2 \wedge dx_1 = (b_{12} - b_{21}) dx_1 \wedge dx_2.$$

Note that more generally if  $b_{ij} = b_{ji}$  for all values of  $i, j \in \{1, 2, \dots, n\}$ , then  $\omega_2 \equiv 0$ . Since  $b \doteq b_{12} - b_{21}$  is an arbitrary function when  $b_{ij}$  are arbitrary, a 2-form when  $n = 2$  can always be written as

$$\omega_2 = b(\mathbf{x}) dx_1 \wedge dx_2.$$

---

<sup>2</sup>The term “quadratic form” refers to a function  $a : \mathbb{R}^n \rightarrow \mathbb{R}$  that has the structure  $a(\mathbf{v}) = \mathbf{v}^T A \mathbf{v}$  for some  $A \in \mathbb{R}^{n \times n}$ .

Whereas the usual calculus was used to go from a 0-form to a 1-form, a newer kind of calculus, called *exterior calculus*, is required to take a 1-form into a 2-form. The associated *exterior derivative* is defined by the following rule:<sup>3</sup>

$$\begin{aligned} d\left(\sum_{i=1}^n a_i(\mathbf{x})dx_i\right) &\doteq \sum_{i=1}^n \left(\sum_{j=1}^n \frac{\partial a_i}{\partial x_j} dx_j\right) \wedge dx_i \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial a_i}{\partial x_j} dx_j \wedge dx_i \\ &= -\sum_{i=1}^n \sum_{j=1}^n \frac{\partial a_i}{\partial x_j} dx_i \wedge dx_j. \end{aligned} \quad (6.8)$$

The first equality above is a definition, and the others derive from the fact that the wedge product is distributive, associative, and anti-commuting from (6.5)–(6.7).

In analogy with the way that the usual differential takes an arbitrary differentiable scalar function (or 0-form) into a special kind of 1-form, the exterior derivative defined above takes an arbitrary 1-form into a special kind of 2-form.

A 3-form is defined as

$$\omega_3 \doteq \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n c_{ijk}(\mathbf{x}) dx_i \wedge dx_j \wedge dx_k \quad (6.9)$$

where each  $c_{ijk}(\mathbf{x})$  is a smooth real-valued function. The anti-commuting nature of the wedge product, together with distributivity and associativity, gives

$$-dx_i \wedge dx_j \wedge dx_k = dx_j \wedge dx_i \wedge dx_k = dx_i \wedge dx_k \wedge dx_j.$$

In other words, any pairwise transposition of *adjacent differentials* produces the negative of the original product. But performing two such pairwise transpositions produces two negative signs that cancel:

$$dx_i \wedge dx_j \wedge dx_k = dx_k \wedge dx_i \wedge dx_j = dx_j \wedge dx_k \wedge dx_i.$$

It also means that whenever there is a repeated index, the result will be zero:

$$dx_i \wedge dx_j \wedge dx_i = dx_i \wedge dx_i \wedge dx_j = dx_j \wedge dx_i \wedge dx_i = 0 \quad \forall i, j \in \{1, 2, \dots, n\}.$$

If  $n < 3$ , then this means that  $\omega_3 = 0$  because in this case there is no way to avoid wedge products of differentials with repeated indices. If  $n = 3$ , working through the  $3^3 = 27$  terms in the triple sum in (6.9) gives

$$\omega_3 \doteq c(\mathbf{x}) dx_1 \wedge dx_2 \wedge dx_3$$

where  $c(\mathbf{x})$  is written in terms of  $c_{ijk}(\mathbf{x})$  in a way that is left as an exercise.

In analogy with the way that a special kind of 2-form was generated by exterior differentiation of a 1-form, a special kind of 3-form can be generated by exterior differentiation of a 2-form by using the following rule:

---

<sup>3</sup>The notation  $d(\cdot)$  should *not* be confused with the usual differential.

$$\begin{aligned}
d\left(\sum_{i=1}^n \sum_{j=1}^n b_{ij} dx_i \wedge dx_j\right) &\doteq \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{k=1}^n \frac{\partial b_{ij}}{\partial x_k} dx_k\right) \wedge dx_i \wedge dx_j & (6.10) \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial b_{ij}}{\partial x_k} dx_k \wedge dx_i \wedge dx_j \\
&= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial b_{ij}}{\partial x_k} dx_i \wedge dx_j \wedge dx_k.
\end{aligned}$$

The reason why there is no negative sign in the final equality, whereas there was one in (6.4), is that in (6.4) only a single change in order of adjacent differentials took place. In contrast, in the second and third equality in (6.10) two adjacent swaps are required:  $dx_k \wedge dx_i \wedge dx_j \rightarrow (-1)dx_i \wedge dx_k \wedge dx_j$ , and  $dx_i \wedge dx_k \wedge dx_j \rightarrow (-1)dx_i \wedge dx_j \wedge dx_k$ . Therefore the negative signs cancel.

Some notation to be aware of when reading more theoretical treatments is the following. The set of differential  $k$ -forms on  $\mathbb{R}^n$  is denoted as  $\Omega^k(\mathbb{R}^n)$ . Based on the informal discussion above, the exterior derivative can then be viewed as the mapping

$$d: \Omega^k(\mathbb{R}^n) \rightarrow \Omega^{k+1}(\mathbb{R}^n). \quad (6.11)$$

### 6.1.2 Exterior Derivatives of $(n-1)$ -Forms on $\mathbb{R}^n$ for $n=2,3$

Consider the 1-form in  $\mathbb{R}^2$ :

$$\omega_1 = a_1(x_1, x_2)dx_1 + a_2(x_1, x_2)dx_2.$$

According to the rule given in (6.8), the exterior derivative of this will be

$$d\omega_1 = \frac{\partial a_1}{\partial x_2} dx_2 \wedge dx_1 + \frac{\partial a_2}{\partial x_1} dx_1 \wedge dx_2 = \left(\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2}\right) dx_1 \wedge dx_2. \quad (6.12)$$

Taking the exterior derivative of this will produce a 3-form. But in general an  $(n+1)$ -form will be zero on  $\mathbb{R}^n$  because in this case there is no way to avoid wedge products involving the same differentials. Or stated in a different way, since the exterior derivative of an  $n$ -form produces an  $(n+1)$ -form, and since every  $(n+1)$ -form on  $\mathbb{R}^n$  is zero, the exterior derivative of an  $n$ -form on  $\mathbb{R}^n$  must always be zero. Therefore, in  $\mathbb{R}^2$  it must be that  $d(d\omega_1) = 0$ , and more generally  $d(d\omega_{n-1}) = 0$  when  $\omega_{n-1}$  is an  $(n-1)$ -form on  $\mathbb{R}^n$ .

But this is not the only time that the exterior derivative of an exterior derivative will be zero. For example, starting with the 0-form (scalar function)  $\omega_0 = \phi(x_1, x_2)$ , the following 1-form results:

$$\omega_1 = \frac{\partial \phi}{\partial x_1} dx_1 + \frac{\partial \phi}{\partial x_2} dx_2.$$

Now taking a second exterior derivative of this 1-form, and evaluating the result using (6.12), gives  $d(d\omega_1) = 0$  because

$$\frac{\partial}{\partial x_1} \left(\frac{\partial \phi}{\partial x_2}\right) = \frac{\partial}{\partial x_2} \left(\frac{\partial \phi}{\partial x_1}\right).$$

Now consider the following 2-form in  $\mathbb{R}^3$ :

$$\omega_2 = \tilde{b}_{12}(x_1, x_2, x_3) dx_1 \wedge dx_2 + \tilde{b}_{23}(x_1, x_2, x_3) dx_2 \wedge dx_3 + \tilde{b}_{13}(x_1, x_2, x_3) dx_1 \wedge dx_3. \quad (6.13)$$

At first this may look less general than the definition in (6.4) evaluated at  $n = 3$ . But actually choosing  $\tilde{b}_{ij} \doteq b_{ij} - b_{ji}$  in (6.13) makes it exactly the same. For this reason, (6.4) can be re-written as  $\sum_{1 \leq i < j \leq 3} \tilde{b}_{ij} dx_i \wedge dx_j$ , and more generally any differential  $k$ -form in  $\mathbb{R}^n$  can be written with the summations nested as

$$\omega_k \doteq \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \tilde{a}_{i_1, i_2, \dots, i_k}(\mathbf{x}) dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k} \quad (6.14)$$

where  $1 \leq k \leq n$ . Here the functions  $\tilde{a}_{i_1, i_2, \dots, i_k}(\mathbf{x})$  generalize the skew-symmetric part of the  $b_{ij}$  and  $c_{ijk}$  discussed previously.

The exterior derivative of the above  $k$ -form<sup>4</sup> is defined as

$$d\omega_k \doteq \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \left( \sum_{1 < i_{k+1} \leq n} \frac{\partial \tilde{a}_{i_1, i_2, \dots, i_k}}{\partial x_{i_{k+1}}} dx_{i_{k+1}} \right) \wedge dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k}. \quad (6.15)$$

It follows from using this definition twice that

$$d(d\omega_k) = \sum_{1 \leq i_1 < \dots < i_k \leq n} \left( \sum_{1 \leq i_{k+1}, i_{k+2} \leq n} \frac{\partial^2 \tilde{a}_{i_1, i_2, \dots, i_k}}{\partial x_{i_{k+1}} \partial x_{i_{k+2}}} dx_{i_{k+2}} \wedge dx_{i_{k+1}} \right) \wedge dx_{i_1} \wedge \dots \wedge dx_{i_k}.$$

But since each  $\tilde{a}_{i_1, i_2, \dots, i_k}(\mathbf{x})$  is a smooth function, the order of taking partial derivatives does not matter. This introduces a symmetry. And any symmetry in the coefficients of a differential form means that (6.5) will force the resulting sum to be equal to zero. Therefore, the following general equality for any  $k$ -form on  $\mathbb{R}^n$  is observed:

$$d(d\omega_k) = 0. \quad (6.16)$$

This can be thought of as the natural generalization of the classical rules  $\nabla \times (\nabla \phi) = \mathbf{0}$  and  $\nabla \cdot (\nabla \times \phi) = 0$ .

Returning to the 2-form  $\omega_2$  in  $\mathbb{R}^3$  defined in (6.13), the exterior derivative becomes

$$d\omega_2 = \left( \frac{\partial \tilde{b}_{23}}{\partial x_1} + \frac{\partial \tilde{b}_{13}}{\partial x_2} + \frac{\partial \tilde{b}_{12}}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3. \quad (6.17)$$

Since this is a 3-form, it follows that  $d(d\omega_2) = 0$ , since that would be a 4-form in  $\mathbb{R}^3$ , and would necessarily have wedge products involving differentials with repeated indices.

Now consider what happens to 1-forms in  $\mathbb{R}^3$  when exterior derivatives are applied. Let

$$\omega_1 = a_1(x_1, x_2, x_3) dx_1 + a_2(x_1, x_2, x_3) dx_2 + a_3(x_1, x_2, x_3) dx_3.$$

Taking one exterior derivative using the rule (6.8), and simplifying using the properties of the wedge product,

<sup>4</sup>Here  $k < n$  because the case  $k = n$  results in  $d\omega_n = 0$ .

$$d\omega_1 = \left( \frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) dx_1 \wedge dx_2 + \left( \frac{\partial a_3}{\partial x_2} - \frac{\partial a_2}{\partial x_3} \right) dx_2 \wedge dx_3 + \left( \frac{\partial a_3}{\partial x_1} - \frac{\partial a_1}{\partial x_3} \right) dx_1 \wedge dx_3. \quad (6.18)$$

It can also be verified that  $d(d\omega_1) = 0$ , which is left as an exercise.

For those familiar with vector calculus (see the appendix for a review), (6.17) should look familiar because it resembles the divergence of a vector field and (6.18) looks curiously like the curl operation. Indeed, this will be discussed in Section 6.8.

The properties of differential forms were defined above in terms of the anti-symmetric nature of pairwise transpositions of adjacent differentials. In practice it can be rather tedious to look for the path of pairwise adjacent transpositions that convert a differential form defined in one ordering of the differentials into another. Therefore, the powerful and general language of permutations is useful in re-defining differential forms, and relating them to their “dual object,” which is called a multi-vector. This will be defined shortly. But first some additional properties and notation will be useful.

### 6.1.3 Products of Differential Forms

Exterior differentiation is one way to change a  $k$ -form into a  $(k+1)$ -form. However, it is not the only way. It is also possible to take the product of two forms. This product follows the simple rule: Given forms  $\omega_p$  and  $\alpha_q$  that are respectively  $p$ - and  $q$ -forms defined in a similar way as in (6.14), their product is the  $(p+q)$ -form

$$\omega_p \wedge \alpha_q \doteq \left( \sum_{1 \leq i_1 < i_2 < \dots < i_p \leq n} \tilde{a}_{i_1, i_2, \dots, i_p}(\mathbf{x}) dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_p} \right) \wedge \left( \sum_{1 \leq j_1 < j_2 < \dots < j_q \leq n} \tilde{b}_{j_1, j_2, \dots, j_q}(\mathbf{x}) dx_{j_1} \wedge dx_{j_2} \wedge \dots \wedge dx_{j_q} \right). \quad (6.19)$$

Therefore, if  $p+q > n$ , then  $\omega \wedge \alpha = 0$ , since in that case there is no way to avoid repeated differentials.

The definition in (6.19) together with the properties of differential forms and the properties of the wedge product in (6.5)–(6.7) are sufficient to explicitly compute the product of any two differential forms. For example, consider the product of the following two differential 1-forms on  $\mathbb{R}^3$ :

$$\omega_1 = a_1 dx_1 + a_2 dx_2 + a_3 dx_3 \quad \text{and} \quad \alpha_1 = b_1 dx_1 + b_2 dx_2 + b_3 dx_3.$$

Then

$$\omega_1 \wedge \alpha_1 = (a_1 dx_1 + a_2 dx_2 + a_3 dx_3) \wedge (b_1 dx_1 + b_2 dx_2 + b_3 dx_3).$$

Using the distributive law and anti-symmetry of the wedge product, this reduces to

$$\begin{aligned} \omega_1 \wedge \alpha_1 &= a_1 b_2 dx_1 \wedge dx_2 + a_1 b_3 dx_1 \wedge dx_3 + a_2 b_1 dx_2 \wedge dx_1 \\ &\quad + a_2 b_3 dx_2 \wedge dx_3 + a_3 b_1 dx_3 \wedge dx_1 + a_3 b_2 dx_3 \wedge dx_2 \\ &= (a_1 b_2 - a_2 b_1) dx_1 \wedge dx_2 + (a_1 b_3 - a_3 b_1) dx_1 \wedge dx_3 + (a_2 b_3 - a_3 b_2) dx_2 \wedge dx_3. \end{aligned} \quad (6.20)$$

Other wedge products are left as exercises.

### 6.1.4 Concise Notation for Differential Forms and Exterior Derivatives

Long expressions involving differential forms such as (6.14) and (6.15) can be reduced in size by defining an *index set*,  $I_k^n$ , that takes into account all non-redundant orderings of indices in the expressions for  $k$ -forms on  $\mathbb{R}^n$ . Then a  $k$ -form on  $\mathbb{R}^n$  can be written succinctly as

$$\omega_k = \sum_{s \in I_k^n} a_s dx_s$$

where each  $dx_s$  is a  $k$ -fold wedge product consisting of differentials with indices drawn from  $I_k^n$ . For example, the index set for a 2-form in  $\mathbb{R}^3$  would be

$$I_2^3 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\},$$

and if  $s = \{1, 2\}$  then  $dx_s \doteq dx_1 \wedge dx_2$ . As another example, the index sets for 2-forms and 3-forms in  $\mathbb{R}^4$  respectively would be

$$I_2^4 = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$$

and

$$I_3^4 = \{\{1, 2, 3\}, \{1, 3, 4\}, \{1, 2, 4\}, \{2, 3, 4\}\}.$$

When the value of  $n$  is clear, the superscript will often be dropped, and the index set will be written as  $I_k$ . Furthermore, many books make the notation even more condensed by not even introducing the parameter “ $s$ ” mentioned above. Then

$$\omega_k = \sum_{I_k} a_{I_k} dx_{I_k}.$$

The ultimate in terse notation suppresses  $s$ ,  $n$ , and  $k$ , in which case  $\omega = \sum_I a_I dx_I$ . Notation that is this light can make it very difficult for the uninitiated reader to pick up the concepts, and so the presentation here will not go that far.

The exterior derivative of a  $k$ -form such as (6.15) can be written more succinctly as [2, 3, 14]

$$d\omega_k = \sum_{I_k} da_{I_k} \wedge dx_{I_k} = \sum_{I_k} \left( \sum_j \frac{\partial a_{I_k}}{\partial x_j} dx_j \right) \wedge dx_{I_k}. \quad (6.21)$$

It is a  $(k+1)$ -form as discussed previously.

The product of a  $p$ -form and  $q$ -form is written concisely as

$$\omega_p \wedge \alpha_q = \left( \sum_{I_p} a_{I_p} dx_{I_p} \right) \wedge \left( \sum_{I_q} b_{I_q} dx_{I_q} \right) = \sum_{I_p} \sum_{I_q} (a_{I_p} dx_{I_p}) \wedge (b_{I_q} dx_{I_q}). \quad (6.22)$$

However, the one drawback of the concise notation is that the expression cannot be simplified further without expanding everything out.

The introduction to differential forms presented in this section should be sufficient for understanding the theorems in Section 6.8 and the use of differential forms in subsequent chapters. The other sections in this chapter illustrate the relationship between differential forms and mathematical objects called multi-vectors, which is provided for completeness. In order to understand these concepts fully, a review of permutations is helpful.



## 6.2 Permutations

The set of all permutation operations on  $n$  elements together with the operation of composition is called the *symmetric group*, or *permutation group*, and is denoted here as  $\Pi_n$ . It is a finite group containing  $n!$  elements. The elements of  $\Pi_n$  can be arranged in any order, and any fixed arrangement of the elements of  $\Pi_n$  can be numbered as  $\pi_i$  for  $i = 0, \dots, n! - 1$ . It is convenient to retain the label  $\pi_0$  for the “do nothing” permutation.

An arbitrary element  $\pi \in \Pi_n$  is denoted as

$$\pi = \begin{pmatrix} 1 & 2 & \dots & n \\ \pi(1) & \pi(2) & \dots & \pi(n) \end{pmatrix}.$$

Changing the order of the columns in the above element does not change the element. So in addition to the above expression,

$$\pi = \begin{pmatrix} 2 & 1 & \dots & n \\ \pi(2) & \pi(1) & \dots & \pi(n) \end{pmatrix} = \begin{pmatrix} n & 2 & \dots & 1 \\ \pi(n) & \pi(2) & \dots & \pi(1) \end{pmatrix}$$

where the rows of dots denote those columns not explicitly listed. The above expression should be read as “the number  $i$  goes to  $\pi(i)$  for  $i = 1, \dots, n$ .”

Two permutations can be composed by the rule

$$\begin{aligned} \pi_i \circ \pi_j &= \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_i(1) & \pi_i(2) & \dots & \pi_i(n) \end{pmatrix} \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_j(1) & \pi_j(2) & \dots & \pi_j(n) \end{pmatrix} \\ &\doteq \begin{pmatrix} \pi_j(1) & \pi_j(2) & \dots & \pi_j(n) \\ \pi_i(\pi_j(1)) & \pi_i(\pi_j(2)) & \dots & \pi_i(\pi_j(n)) \end{pmatrix} \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_j(1) & \pi_j(2) & \dots & \pi_j(n) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_i(\pi_j(1)) & \pi_i(\pi_j(2)) & \dots & \pi_i(\pi_j(n)) \end{pmatrix} \in \Pi_n. \end{aligned}$$

This rule is not the only way to define compositions of permutations, but this is the rule that will be used below.

### 6.2.1 Examples of Permutations and Their Products

For example, the product of the permutations

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 5 & 4 & 3 \end{pmatrix} \quad \text{and} \quad \pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 1 & 3 & 4 \end{pmatrix}$$

is

$$\pi_2 \circ \pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 4 & 3 & 1 \end{pmatrix} \quad \text{and} \quad \pi_1 \circ \pi_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 5 & 4 \end{pmatrix}.$$

Since these are not the same, it can be concluded that the product of permutations is not commutative.

An alternative way of denoting permutations is by decomposing them into cycles. For instance in  $\pi_1$  above,  $1 \rightarrow 2 \rightarrow 1$ , and  $3 \rightarrow 5 \rightarrow 3$  and  $4 \rightarrow 4$ . This means that we can decompose  $\pi_1$  as a product of cycles:

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}.$$

In fact, there is no need to explicitly include the permutation corresponding to the cycle of length 1 corresponding to  $4 \rightarrow 4$ , since it is the identity. While in general permutations do not commute under the operation  $\circ$ , permutations corresponding to disjoint cycles do commute. The shorthand notation for the above that reflects this commutativity of cycles is  $\pi_1 = (12)(35)(4) = (4)(12)(35) = (35)(4)(12) = \dots$ . Not every permutation can be broken down into smaller cycles. For example,  $\pi_2 = (12543)$ .

### 6.2.2 The Sign of a Permutation

The final feature of permutations that will be important in the geometric and linear algebraic computations that follow is the *sign* (or *signature*) of a permutation. A permutation that swaps two entries while leaving the rest fixed is called a *transposition*. Any permutation can be broken down into a product of transpositions. If that product consists of an even number of transpositions, the sign of the original permutation is designated as  $+1$ . If the product consists of an odd number of transpositions, the sign of the original permutation is  $-1$ . In other words,  $\text{sign}(\pi) = (-1)^{|\text{trans}(\pi)|}$  where  $|\text{trans}(\pi)|$  denotes the number of transpositions in a decomposition of  $\pi$ . For example, in  $\pi_1$  given above, there are two transpositions corresponding to the two cycles each with two elements. So  $|\text{trans}(\pi_1)| = 2$  and  $\text{sign}(\pi_1) = +1$ . For  $\pi_2$ , it is possible to transform the sequence 2, 5, 1, 3, 4 back to 1, 2, 3, 4, 5 by performing the following transpositions:

$$(2, 5, 1, 3, 4) \rightarrow (1, 5, 2, 3, 4) \rightarrow (1, 2, 5, 3, 4) \rightarrow (1, 2, 3, 5, 4) \rightarrow (1, 2, 3, 4, 5).$$

Alternatively, the following transpositions could be performed:

$$(2, 5, 1, 3, 4) \rightarrow (2, 4, 1, 3, 5) \rightarrow (2, 3, 1, 4, 5) \rightarrow (1, 3, 2, 4, 5) \rightarrow (1, 2, 3, 4, 5).$$

While the number of transpositions used to define a path from one permutation to another is not unique since there are many paths, the *minimal number* of transpositions required to restore the original ordering is unique. But regardless of whether or not that minimal number is realized, the signature of any sequence of transpositions connecting a permutation with the identity will be the same. Counting the number of arrows above, it is clear that  $|\text{trans}(\pi_2)| = 4$  and so  $\text{sign}(\pi_2) = +1$ .

Since the number of transpositions in the product of two transpositions will be

$$|\text{trans}(\pi_1 \circ \pi_2)| = |\text{trans}(\pi_1)| + |\text{trans}(\pi_2)|,$$

it follows that

$$\text{sign}(\pi_1 \circ \pi_2) = \text{sign}(\pi_1) \cdot \text{sign}(\pi_2).$$

More generally the product of even permutations is even, the product of odd permutations is even, and the product of even with odd is odd.

### 6.2.3 Multi-Dimensional Version of the Levi-Civita Symbol

The *Kronecker delta* is defined as the function  $\delta : \mathbb{Z} \times \mathbb{Z} \rightarrow \{0, 1\}$  such that

$$\delta_{i,j} \doteq \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases}$$

One way to extend this to a multi-index version is

$$\delta_{i_1, i_2, \dots, i_n} \doteq \begin{cases} 1 & \text{for } i_1 = i_2 = \dots = i_n \\ 0 & \text{otherwise.} \end{cases}$$

This is the same as the  $(n - 1)$ -fold product

$$\delta_{i_1, i_2, \dots, i_n} = \delta_{i_1, i_2} \delta_{i_2, i_3} \dots \delta_{i_{n-1}, i_n}.$$

If each index is limited to have the  $N$  possible values  $1, \dots, N$ , then only  $N$  out of the  $N^n$  possible combinations of indices will result in the value of unity.

Another extension is [11]

$$\delta_{i_1, i_2, \dots, i_n}^{j_1, j_2, \dots, j_n} \doteq \det \begin{bmatrix} \delta_{i_1, j_1} & \delta_{i_1, j_2} & \dots & \delta_{i_1, j_n} \\ \delta_{i_2, j_1} & \delta_{i_2, j_2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{i_n, j_1} & \delta_{i_n, j_2} & \dots & \delta_{i_n, j_n} \end{bmatrix}. \tag{6.23}$$

The *Levi-Civita symbol* (sometimes called the *alternating tensor*) is defined as

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if } (i, j, k) \in \{(1, 2, 3), (3, 1, 2), (2, 3, 1)\} \\ -1 & \text{if } (i, j, k) \in \{(2, 1, 3), (1, 3, 2), (3, 2, 1)\} \\ 0 & \text{otherwise.} \end{cases}$$

Here the “otherwise” refers to any case where indices are repeated, i.e.,  $\epsilon_{112} = \epsilon_{121} = \epsilon_{211} = 0$  and likewise if 1 or 2 is replaced with 3, etc. Altogether there are  $3^3 = 27$  possibilities, and only six are non-zero.

Note that the first two conditional equalities in the above definition of the Levi-Civita symbol can be written as

$$\epsilon_{\pi(1), \pi(2), \pi(3)} = \text{sgn}(\pi) \quad \text{for } \pi \in \Pi_3.$$

These symbols are related by the equalities

$$\boxed{\sum_{i=1}^3 \epsilon_{ijk} \epsilon_{imn} = \delta_{jm} \delta_{kn} - \delta_{jn} \delta_{km} \quad \text{and} \quad \frac{1}{2} \sum_{i, j=1}^3 \epsilon_{ijk} \epsilon_{ijn} = \delta_{kn}.} \tag{6.24}$$

Letting  $n = k$  in the second equality above makes the right side equal unity. Summing over  $k$  and multiplying by 2 then gives  $\sum_{i, j, k=1}^3 \epsilon_{ijk} \epsilon_{ijk} = 6$ .

The Kronecker delta and Levi-Civita symbols are used both in classical mechanics and throughout mathematics and physics. For example, the identity matrix is the one defined in terms of its entries as  $\mathbb{I} = [\delta_{ij}]$ , and the cross product of three-dimensional vectors can be defined in component form as

$$(\mathbf{a} \times \mathbf{b})_i = \sum_{j, k=1}^3 \epsilon_{ijk} a_j b_k. \tag{6.25}$$

Given the above equation, it should come as no surprise that the extension of the concept of a cross product to higher dimensions will be defined using a multi-dimensional version of the Levi-Civita symbol, which is defined below.

Let  $I_n \doteq (1, \dots, n)$  denote the numbers 1 through  $n$  arranged in this order, and let  $m : I_n \rightarrow I_n$ , where the result of the mapping is stored in the same order as the inputs.

That is,  $m : (1, \dots, n) \rightarrow (m(1), \dots, m(n))$ . If the mapping is not one-to-one, the repeated values will be stored. In the latter case, the result of applying this mapping to  $I$  will be viewed as a *multi-set*<sup>5</sup> with  $n$  entries. For example, if  $n = 3$ ,  $m$  might be a function  $m(1, 2, 3) = (2, 3, 1)$  or  $m(1, 2, 3) = (1, 2, 1)$ . That is,  $m$  will either be a permutation, or it will produce a multi-set in which some entries are repeated and others are omitted. In this context, the multi-dimensional Levi-Civita symbol is defined as

$$\epsilon_{m(1),m(2),\dots,m(n)} \doteq \begin{cases} \text{sgn}(m) & \text{for } m \in \Pi_n \\ 0 & \text{otherwise} \end{cases}.$$

In other words, it again will take the values  $+1$ ,  $-1$ , or  $0$  depending on whether  $m$  is an even or odd permutation, or not a permutation at all. Letting  $i_k$  denote  $m(k)$ , the multi-dimensional versions of (6.24) can be written as [11]

$$\epsilon_{i_1,i_2,\dots,i_n} \epsilon_{j_1,j_2,\dots,j_n} = \delta_{i_1,i_2,\dots,i_n}^{j_1,j_2,\dots,j_n} \quad \text{and} \quad \sum_{i_1,i_2,\dots,i_n=1}^n \epsilon_{i_1,i_2,\dots,i_n} \epsilon_{i_1,i_2,\dots,i_n} = n! \quad (6.26)$$

because out of the  $n^n$  possible combinations of indices, only the subset corresponding to permutations produces a non-zero contribution to the sum. In a similar way, for an  $n$ th order tensor,  $A = [a_{i_1,i_2,\dots,i_n}]$ ,

$$\sum_{i_1,i_2,\dots,i_n=1}^n \epsilon_{i_1,i_2,\dots,i_n} a_{i_1,i_2,\dots,i_n} = \sum_{\pi \in \Pi_n} \text{sgn}(\pi) a_{\pi(1),\pi(2),\dots,\pi(n)}.$$

Knowing this can be useful in relating some of what follows in subsequent sections to definitions given in some older books. Note that sometimes it will be useful to use the shorthand

$$\epsilon(\pi) = \epsilon_{\pi(1),\pi(2),\dots,\pi(n)}.$$

### 6.3 The Hodge Star Operator

Given a differential  $k$ -form on  $\mathbb{R}^n$ ,

$$\omega = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} a_{i_1,\dots,i_k} dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k},$$

the *Hodge star operator* produces from this  $k$ -form the  $(n - k)$ -form on  $\mathbb{R}^n$ , denoted as  $*\omega$ , that results from the substitution

$$dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k} \longrightarrow \text{sgn } \pi dx_{j_1} \wedge dx_{j_2} \wedge \dots \wedge dx_{j_{n-k}}$$

together with an additional summation over  $\{j_1, \dots, j_{n-k}\}$  where

$$\pi = \begin{pmatrix} 1 & \dots & k & k+1 & \dots & n \\ i_1 & \dots & i_k & j_1 & \dots & j_{n-k} \end{pmatrix} \quad \text{and} \quad \{i_1, \dots, i_k\} \cap \{j_1, \dots, j_{n-k}\} = \emptyset.$$

The latter condition ensures that  $\pi$  is in fact a permutation.

---

<sup>5</sup>Unlike a usual set, in which each element appears once, in a multi-set, elements can appear several times.

Explicitly,

$$*\omega \doteq \sum_{\substack{i_1 < i_2 < \dots < i_k \\ j_1 < j_2 < \dots < j_{n-k}}} \operatorname{sgn} \begin{pmatrix} 1 & \dots & k & k+1 & \dots & n \\ i_1 & \dots & i_k & j_1 & \dots & j_{n-k} \end{pmatrix} a_{i_1, \dots, i_k} dx_{j_1} \wedge \dots \wedge dx_{j_{n-k}}. \tag{6.27}$$

Due to the anti-symmetric nature of the wedge product, the condition  $\{i_1, \dots, i_k\} \cap \{j_1, \dots, j_{n-k}\} = \emptyset$  need not be enforced explicitly when  $k > 1$ . If this condition is violated, the wedge products involving terms that are common between the two sets of indices will be equal to zero. Therefore, it does not matter whether or not  $\pi$  (or the sign of  $\pi$ ) is defined when  $\pi$  is not actually a permutation since the result will multiply zero. When  $k = 1$  the condition either needs to be enforced, or  $\operatorname{sgn} \pi$  should be defined to be zero when  $\pi$  is not a permutation.

From the properties of permutations, it can be shown that the Hodge star operator applied twice,  $**\omega = **\omega$ , results in

$$**\omega = (-1)^{k(n-k)}\omega. \tag{6.28}$$

Now consider some examples. If  $\phi(\mathbf{x})$  is a 0-form on  $\mathbb{R}^n$  (i.e., a function  $\mathbb{R}^n \rightarrow \mathbb{R}$ ), then

$$*\phi \doteq \phi dx_1 \wedge \dots \wedge dx_n.$$

If  $\omega = a_1 dx_1 + a_2 dx_2$  is a 1-form on  $\mathbb{R}^2$ , then

$$*\omega = \sum_j \left\{ \operatorname{sgn} \begin{pmatrix} 1 & 2 \\ 1 & j \end{pmatrix} a_1 dx_j + \operatorname{sgn} \begin{pmatrix} 1 & 2 \\ 2 & j \end{pmatrix} a_2 dx_j \right\} = a_1 dx_2 - a_2 dx_1.$$

If  $\omega = a_1 dx_1 + a_2 dx_2 + a_3 dx_3$  is a 1-form on  $\mathbb{R}^3$ , then

$$*\omega = a_1 dx_2 \wedge dx_3 - a_2 dx_1 \wedge dx_3 + a_3 dx_1 \wedge dx_2.$$

If  $\omega = a_{12} dx_1 \wedge dx_2 + a_{13} dx_1 \wedge dx_3 + a_{23} dx_2 \wedge dx_3$  is a 2-form on  $\mathbb{R}^3$ , then

$$*\omega = a_{12} dx_3 - a_{13} dx_2 + a_{23} dx_1.$$

If  $\omega = a dx_1 \wedge dx_2 \wedge dx_3$  is a 3-form on  $\mathbb{R}^3$ , then  $*\omega = a$ .

The beauty of the Hodge star operator will become apparent at the end of this chapter when it is used to restate the integral theorems of vector calculus in a concise form that generalizes nicely to non-Euclidean settings. One reason for this is that if  $\omega = \sum_i a_i dx_i$  is a 1-form on  $\mathbb{R}^n$ , and  $\mathbf{a} = [a_1, \dots, a_n]^T$ , then it can be shown that the exterior derivative of  $*\omega$  is the  $n$ -form

$$d(*\omega) = (\operatorname{div} \mathbf{a}) dx_1 \wedge \dots \wedge dx_n. \tag{6.29}$$

And furthermore, if  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ , then the Hodge star operator of the 1-form  $d\phi$  is

$$d(*d\phi) = \operatorname{div}(\operatorname{grad} \phi) dx_1 \wedge \dots \wedge dx_n. \tag{6.30}$$

## 6.4 Tensor Products and Dual Vectors

A *tensor product* of two vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  can be defined as

$$\mathbf{a} \otimes \mathbf{b} \doteq \mathbf{a}\mathbf{b}^T \in \mathbb{R}^{n \times n}. \quad (6.31)$$

This can be thought of as an operation that produces a two-dimensional matrix from two column vectors. Sometimes this is referred to as an *outer product*, and is related to the inner product by the equality  $\text{tr}(\mathbf{a} \otimes \mathbf{b}) = \mathbf{a} \cdot \mathbf{b}$ .

The elements of the resulting matrix are  $(\mathbf{a} \otimes \mathbf{b})_{ij} = a_i b_j$ . The tensor product of three vectors can be defined to be the  $n \times n \times n$  array indexed by  $i, j, k$  with entries

$$((\mathbf{a} \otimes \mathbf{b}) \otimes \mathbf{c})_{ijk} = (\mathbf{a} \otimes (\mathbf{b} \otimes \mathbf{c}))_{ijk} = (\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c})_{ijk} = a_i b_j c_k.$$

This extends in an obvious way to higher dimensions. Furthermore, the vectors can be allowed to have different dimensions, resulting in a rectangular box of numbers, rather than square or cubic array.

The dual space of  $\mathbb{R}^n$  consists of all linear functions that take in vectors from  $\mathbb{R}^n$  and return real numbers. This dual space can be thought of as being equivalent to all real  $n$ -dimensional row vectors, which, after multiplication with a vector in  $\mathbb{R}^n$ , results in a real scalar. In other words, if  $V = \mathbb{R}^n$  and  $\mathbf{v} \in \mathbb{R}^n$ , then any  $\varphi \in (\mathbb{R}^n)^*$  (the dual of  $\mathbb{R}^n$ ) can be defined as  $\varphi(\mathbf{v}) = \mathbf{w}^T \mathbf{v}$  for some  $\mathbf{w} \in \mathbb{R}^n$ . A function  $\varphi(\mathbf{v})$  is sometimes called a *dual vector*. It contains the same information as  $\mathbf{w}^T$ . Whereas the concept of the transpose makes sense for vectors in  $\mathbb{R}^n$ , the concept of a dual generalizes to more abstract vector spaces.

Tensor products can be defined in more general contexts than (6.31) by using elements of the dual space. These more general tensor products are constructed as follows [7]: Let  $V$  and  $W$  be vector spaces with elements  $\mathbf{v}$  and  $\mathbf{w}$ , respectively. Let  $\varphi_1 \in V^*$  and  $\varphi_2 \in W^*$ . Then for any  $(\mathbf{v}, \mathbf{w}) \in V \times W$ , define

$$(\varphi_1 \otimes \varphi_2)(\mathbf{v}, \mathbf{w}) \doteq \varphi_1(\mathbf{v})\varphi_2(\mathbf{w}). \quad (6.32)$$

Since  $\varphi_1$  and  $\varphi_2$  are both linear functions, the tensor product is a *bilinear function*:

$$\begin{aligned} (\varphi_1 \otimes \varphi_2)(a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2, \mathbf{w}) &= a_1 (\varphi_1 \otimes \varphi_2)(\mathbf{v}_1, \mathbf{w}) + a_2 (\varphi_1 \otimes \varphi_2)(\mathbf{v}_2, \mathbf{w}) \\ (\varphi_1 \otimes \varphi_2)(\mathbf{v}, a_1 \mathbf{w}_1 + a_2 \mathbf{w}_2) &= a_1 (\varphi_1 \otimes \varphi_2)(\mathbf{v}, \mathbf{w}_1) + a_2 (\varphi_1 \otimes \varphi_2)(\mathbf{v}, \mathbf{w}_2). \end{aligned}$$

For example, if  $V = \mathbb{R}^n$ , then for  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  and  $\varphi_1(\mathbf{v}) \doteq \mathbf{a}^T \mathbf{v}$  and  $\varphi_2 \doteq \mathbf{b}^T \mathbf{w}$ ,

$$(\varphi_1 \otimes \varphi_2)(\mathbf{v}, \mathbf{w}) = (\mathbf{a}^T \mathbf{v})(\mathbf{b}^T \mathbf{w}) = \mathbf{v}^T (\mathbf{a}\mathbf{b}^T) \mathbf{w}.$$

At the core of this is the outer product in (6.31), and so when  $V = \mathbb{R}^n$  it can be convenient to blur the distinction between  $\varphi_1 \otimes \varphi_2$  and  $\mathbf{a}\mathbf{b}^T$  while keeping in mind that the  $\varphi_1 \otimes \varphi_2$  construction is more general.

Returning to the more general case, the tensor product of dual vectors can be iterated as  $\varphi_1 \otimes \varphi_2 \otimes \dots \otimes \varphi_n$  due to the associative property

$$(\varphi_1 \otimes \varphi_2) \otimes \varphi_3 = \varphi_1 \otimes (\varphi_2 \otimes \varphi_3).$$

A  $k$ -fold tensor product of dual vectors can take as its domain the  $k$ -fold Cartesian product of a vector space. The dimension of this  $k$ -fold Cartesian product is the same as the dimension of a  $k$ -fold tensor product of vectors. And so, a  $k$ -fold tensor product of dual vectors can be thought of as a function that takes in a  $k$ -fold tensor product of vectors and returns a scalar.

## 6.5 Exterior Products

This subsection begins by presenting the concept of the exterior product of two vectors in a concrete way as an alternating sum of tensor products. From this concrete starting point, the abstract definitions are put into context.

### 6.5.1 A Concrete Introduction to Exterior Products

The cross product in three-dimensional space can be defined using the tensor product of vectors. First let

$$\mathbf{a} \wedge \mathbf{b} \doteq \frac{1}{2} [\mathbf{a} \otimes \mathbf{b} - \mathbf{b} \otimes \mathbf{a}].$$

This is called an *exterior product* (or *wedge product*) of the vectors. Explicitly, this matrix has the form

$$\mathbf{a} \wedge \mathbf{b} \doteq \frac{1}{2} \begin{pmatrix} 0 & a_1b_2 - a_2b_1 & a_1b_3 - a_3b_1 \\ a_2b_1 - a_1b_2 & 0 & a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 & a_3b_2 - b_3a_2 & 0 \end{pmatrix}.$$

Note that since scalar multiplication is commutative, this is a skew-symmetric matrix with entries that are functions of the vectors  $\mathbf{a}$  and  $\mathbf{b}$ :

$$\mathbf{a} \wedge \mathbf{b} = -(\mathbf{b} \wedge \mathbf{a}).$$

This is reminiscent of that which was postulated for differential forms in (6.5).

This  $3 \times 3$  skew-symmetric matrix has three independent pieces of information that can be arranged in a three-dimensional vector. The  $\vee$  operation can be defined to convert  $\mathbf{a} \wedge \mathbf{b}$  into a vector by extracting the three non-zero independent entries and arranging them as

$$(\mathbf{a} \wedge \mathbf{b})^\vee = \frac{1}{2} \begin{pmatrix} a_2b_3 - b_2a_3 \\ a_3b_1 - b_3a_1 \\ a_1b_2 - a_2b_1 \end{pmatrix}. \quad (6.33)$$

When arranged in this way,  $(\mathbf{a} \wedge \mathbf{b})^\vee = \frac{1}{2} \mathbf{a} \times \mathbf{b}$ . While the above definition of the  $\vee$  operation is natural in some sense for  $\mathbb{R}^3$  since it relates to the familiar cross-product operation, it does not generalize well. In higher dimensions a more natural way to define the  $\vee$  of  $\mathbf{a} \wedge \mathbf{b}$  would be to arrange the entries *lexicographically*<sup>6</sup> as  $[a_1b_2 - a_2b_1, a_1b_3 - a_3b_1, a_2b_3 - b_2a_3]^T$ . This is illustrated in Exercises 6.12–6.15. Note that there is also a sign change in this new definition of  $\vee$  relative to  $\vee$  in (6.33), which will no longer be used.

If  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ , then  $\mathbf{a} \wedge \mathbf{b}$  will have  $n(n-1)/2$  independent entries. These too could be arranged in a long column vector, but there is no need to do so at this point. Rather,  $\mathbf{a} \wedge \mathbf{b}$  will itself be referred to as a *2-vector*.

The exterior product of multiple vectors from the same vector space,  $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ , can be defined as

---

<sup>6</sup>In general a lexicographical ordering of a string of characters, each of which has a natural ordering, arranges the first character of each string according to the ordering for that character, followed by ordering of the second character, etc. This is the way a dictionary is organized with the usual alphabetical ordering of characters. The same applies to natural numbers ordered in the usual way.

$$\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_k = \frac{1}{k!} \sum_{\pi \in \Pi_k} \text{sgn}(\pi) \mathbf{v}_{\pi(1)} \otimes \dots \otimes \mathbf{v}_{\pi(k)}. \quad (6.34)$$

The result is called a *multi-vector*. Here  $k$  can be less than or equal to the dimension of  $V$ , and in this particular case  $\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_k$  is called a *k-vector*. The vector space of all  $k$ -vectors is denoted as  $\Lambda^k(V)$  or  $\Lambda^k V$ . A  $k$ -vector can be viewed as a block of numbers in a  $k$ -dimensional array (in which each index can take  $n$  values) that has particular symmetries. Or, by defining a “ $\vee$ ” operator, in analogy with (6.33), the non-redundant entries in the  $k$ -dimensional array can be extracted and arranged in a single column vector. As will be shown below, this  $k$ -vector would have  $\binom{n}{k}$  entries.

Since the concept of tensor products can be defined both for vectors in  $V$  and for dual vectors in  $V^*$ , it follows that  $\Lambda^k(V^*)$  can be constructed as

$$\varphi_1 \wedge \dots \wedge \varphi_k \doteq \frac{1}{k!} \sum_{\pi \in \Pi_k} \text{sgn}(\pi) \varphi_{\pi(1)} \otimes \dots \otimes \varphi_{\pi(k)}. \quad (6.35)$$

Rather than calling this a dual- $k$ -vector or  $k$ -dual-vector, it is called a *k-form*. Substitution of a differential one-form for each  $\varphi_i$  in (6.35) results in the differential  $k$ -forms discussed earlier in this chapter. This establishes an equivalence:

$$\Lambda^k(V^*) \cong \Omega^k(V).$$

With the concrete definition in (6.35) in mind, the modern abstract definition of exterior products can be more easily grasped.

### 6.5.2 Abstract Definition of the Exterior Product of Two Vectors

An alternative (more abstract) definition of the exterior product of vectors to the one given in the previous section is used in many books on differential forms. This is reviewed here in order to make it easier to reconcile the presentations provided in different books.

To begin, let  $V = \mathbb{R}^n$  and  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  be the natural basis, and let  $p = 0, 1, 2, \dots, n$ . For any vectors  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  and real numbers  $a, b \in \mathbb{R}$ , an abstract *wedge product* is defined to take pairs of vectors in  $V$ , and return a vector in a new vector space  $W$ , i.e.,  $\wedge : V \times V \rightarrow W$ , while obeying the following rules:

$$(a\mathbf{u} + b\mathbf{v}) \wedge \mathbf{w} = a(\mathbf{u} \wedge \mathbf{w}) + b(\mathbf{v} \wedge \mathbf{w}) \quad (6.36)$$

$$\mathbf{w} \wedge (a\mathbf{u} + b\mathbf{v}) = a(\mathbf{w} \wedge \mathbf{u}) + b(\mathbf{w} \wedge \mathbf{v}) \quad (6.37)$$

$$\mathbf{v} \wedge \mathbf{v} = \mathbf{0} \quad (6.38)$$

$$\{\mathbf{e}_i \wedge \mathbf{e}_j \mid i, j \in \{1, \dots, n\}\} \text{ spans } W. \quad (6.39)$$

As a consequence of (6.36)–(6.39),

$$\mathbf{w} \wedge \mathbf{v} = -\mathbf{v} \wedge \mathbf{w}. \quad (6.40)$$

$W$  is called  $\Lambda^2 V$ , and products of the form  $\mathbf{v} \wedge \mathbf{v}$  are called *2-vectors*. The definitions  $\Lambda^0 V = \mathbb{R}$  and  $\Lambda^1 V = V$  can also be made.<sup>7</sup>

<sup>7</sup>In some books  $\Lambda$  is denoted as  $\bigwedge$ .



### 6.5.3 The Exterior Product of Several Vectors

The  $p$ th exterior power of  $V$ , denoted  $\Lambda^p V$ , is a vector space with elements that are  $p$ -vectors of the form

$$\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_p \in \Lambda^p V \quad \forall \quad \mathbf{v}_i \in V,$$

where for  $i = 1, \dots, p$  and  $j > i$  the  $p$ -fold wedge product must satisfy<sup>8</sup>

$$\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_{i-1} \wedge (a\mathbf{u} + b\mathbf{v}) \wedge \mathbf{w}_{i+1} \wedge \dots \wedge \mathbf{w}_p = \quad (6.41)$$

$$a(\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_{i-1} \wedge \mathbf{u} \wedge \mathbf{w}_i \wedge \dots \wedge \mathbf{w}_p) + b(\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_{i-1} \wedge \mathbf{v} \wedge \mathbf{w}_i \wedge \dots \wedge \mathbf{w}_p)$$

$$\mathbf{w}_1 \wedge \dots \wedge \mathbf{w}_{i-1} \wedge \mathbf{u} \wedge \mathbf{w}_{i+1} \wedge \dots \wedge \mathbf{w}_{j-1} \wedge \mathbf{u} \wedge \mathbf{w}_{j+1} \wedge \dots \wedge \mathbf{w}_p = \mathbf{0} \quad (6.42)$$

and

$$\{\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \dots \wedge \mathbf{e}_{i_p} \mid 1 \leq i_1 < i_2 < \dots < i_p \leq n\} \text{ spans } \Lambda^p V \quad (6.43)$$

for all  $a, b \in \mathbb{R}$  and  $\mathbf{u}, \mathbf{v}, \mathbf{w}_1, \dots, \mathbf{w}_p \in V$ .

From (6.43) it follows that the dimension of the vector space  $\Lambda^p V$  is

$$\dim(\Lambda^p V) = \binom{n}{p} = \frac{n!}{(n-p)!p!} \quad (6.44)$$

for  $p \in \{0, 1, 2, \dots, n\}$ . Note that  $\Lambda^0 V$  and  $\Lambda^n V$  are both one-dimensional vector spaces (hence equivalent to  $\mathbb{R}$ ), and  $\Lambda^1 V$  and  $\Lambda^{n-1} V$  are equivalent to  $V = \mathbb{R}^n$ . To be more precise, the word “equivalent” here means *isomorphic* as vector spaces, as defined in the appendix.

Furthermore, from the combination of the above rules, it can be shown that [5, 3]

$$\mathbf{v}_{\pi(1)} \wedge \mathbf{v}_{\pi(2)} \wedge \dots \wedge \mathbf{v}_{\pi(p)} = \text{sgn}(\pi) \mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_p \quad (6.45)$$

where  $\pi$  is a permutation on  $n$  letters and  $\text{sgn}(\pi) \in \{-1, +1\}$  with  $+1$  corresponding to an even number of pairwise transpositions and  $-1$  corresponding to the odd case. See Section 6.2 and the appendix for examples.

As an example that demonstrates the calculation of exterior products of vectors, let  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ . Then

$$\mathbf{v} \wedge \mathbf{w} = (v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2) \wedge (w_1 \mathbf{e}_1 + w_2 \mathbf{e}_2) = (v_1 w_2 - v_2 w_1) \mathbf{e}_1 \wedge \mathbf{e}_2.$$

Similarly, if  $A \in \mathbb{R}^{2 \times 2}$ , it is easy to see that  $(A\mathbf{v}) \wedge (A\mathbf{w})$  can be expanded out as

$$\begin{aligned} & [(a_{11}v_1 + a_{12}v_2)\mathbf{e}_1 + (a_{21}v_1 + a_{22}v_2)\mathbf{e}_2] \wedge [(a_{11}w_1 + a_{12}w_2)\mathbf{e}_1 + (a_{21}w_1 + a_{22}w_2)\mathbf{e}_2] \\ &= (a_{11}a_{22} - a_{12}a_{21})(v_1w_2 - v_2w_1) \mathbf{e}_1 \wedge \mathbf{e}_2 \\ &= (\det A) \mathbf{v} \wedge \mathbf{w}. \end{aligned}$$

Generalizations of this observation are explained in the following section.

<sup>8</sup>Of course, when  $i = 1, 2, p-1, p$  these expressions need to be modified so as to make sense, since  $\mathbf{w}_0$  and  $\mathbf{w}_{p+1}$  are not defined.

### 6.5.4 Computing with Exterior Products

Recall from the appendix that permutations enter in the definition of the determinant of a matrix. And there is a connection here as well. Namely, if  $A \in \mathbb{R}^{n \times n}$ , then [5, 3]

$$(A\mathbf{v}_1) \wedge (A\mathbf{v}_2) \wedge \dots \wedge (A\mathbf{v}_n) = (\det A)(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_n) \in \Lambda^n V. \quad (6.46)$$

In fact, using this formula twice (once with  $A$  and once with  $B$ ) gives

$$\begin{aligned} (AB\mathbf{v}_1) \wedge (AB\mathbf{v}_2) \wedge \dots \wedge (AB\mathbf{v}_n) &= (\det A)((B\mathbf{v}_1) \wedge (B\mathbf{v}_2) \wedge \dots \wedge (B\mathbf{v}_n)) \\ &= (\det A \det B)(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_n). \end{aligned}$$

But direct evaluation gives

$$((AB)\mathbf{v}_1) \wedge ((AB)\mathbf{v}_2) \wedge \dots \wedge ((AB)\mathbf{v}_n) = \det(AB)(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_n).$$

Picking off the coefficients reproduces the well-known fact that

$$\det(AB) = \det A \det B. \quad (6.47)$$

As another immediate consequence of (6.46), it is clear that if  $\mathbf{v}_i = \mathbf{e}_i$ , and if  $A$  is an orthogonal matrix, a change of orthogonal basis of  $\Lambda^n V$  can be implemented by making an orthogonal change of basis in  $V$ .

The determinant can be viewed as a special case of a more general set of functions of a matrix generated from exterior products. More specifically, the quantity  $\Lambda^p A$  can be defined so as to satisfy

$$(A\mathbf{v}_1) \wedge (A\mathbf{v}_2) \wedge \dots \wedge (A\mathbf{v}_p) = (\Lambda^p A)(\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_p) \in \Lambda^p V. \quad (6.48)$$

Here  $p \in \{1, 2, \dots, n\}$  whereas (6.46) holds for the special case when  $p = n$ . In the exercises the explicit form of  $\Lambda^p A$  is computed for several concrete cases.

Following the same arguments that led to (6.47), but now using (6.48) in place of (6.46),

$$\boxed{\Lambda^p(AB) = \Lambda^p(A)\Lambda^p(B)}. \quad (6.49)$$

### 6.5.5 The Exterior Product of Two Exterior Products

If  $\mathbf{v} = \mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_p \in \Lambda^p V$  and  $\mathbf{u} = \mathbf{u}_1 \wedge \mathbf{u}_2 \wedge \dots \wedge \mathbf{u}_q \in \Lambda^q V$ , then there exists a unique way to construct  $\mathbf{v} \wedge \mathbf{u} \in \Lambda^{p+q} V$ . Namely,

$$\mathbf{v} \wedge \mathbf{u} = (\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \dots \wedge \mathbf{v}_p) \wedge (\mathbf{u}_1 \wedge \mathbf{u}_2 \wedge \dots \wedge \mathbf{u}_q).$$

Given three such exterior products,  $\mathbf{v} \in \Lambda^p V$ ,  $\mathbf{u} \in \Lambda^q V$ , and  $\mathbf{w} \in \Lambda^r V$ , the following properties follow from this definition [3]:

$$(a\mathbf{v} + b\mathbf{u}) \wedge \mathbf{w} = a(\mathbf{v} \wedge \mathbf{w}) + b(\mathbf{u} \wedge \mathbf{w}) \quad (6.50)$$

$$\mathbf{w} \wedge (a\mathbf{v} + b\mathbf{u}) = a(\mathbf{w} \wedge \mathbf{v}) + b(\mathbf{w} \wedge \mathbf{u}) \quad (6.51)$$

$$(\mathbf{v} \wedge \mathbf{u}) \wedge \mathbf{w} = \mathbf{v} \wedge (\mathbf{u} \wedge \mathbf{w}) \quad (6.52)$$

$$\mathbf{v} \wedge \mathbf{u} = (-1)^{pq}(\mathbf{u} \wedge \mathbf{v}) \quad (6.53)$$

where in (6.50)  $p = q$ , and in (6.51)  $q = r$ . If  $\mathbf{v} \in \Lambda^p V$  and  $\mathbf{u} \in \Lambda^q V$ , then [5, 3]

$$\boxed{(\Lambda^{p+q} A)(\mathbf{v} \wedge \mathbf{u}) = [(\Lambda^p A)(\mathbf{v})] \wedge [(\Lambda^q A)(\mathbf{u})]}. \quad (6.54)$$

### 6.5.6 The Inner Product of Two Exterior Products

If  $V$  is an inner product space (e.g.,  $\mathbb{R}^n$  with inner product  $(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T \mathbf{w}$ ), then an inner product on  $\Lambda^p V$  can be defined relative to the inner product on  $V$  as follows:

$$(\mathbf{u}, \mathbf{v})_p \doteq \det \begin{bmatrix} (\mathbf{u}_1, \mathbf{v}_1) & (\mathbf{u}_1, \mathbf{v}_2) & \cdots & (\mathbf{u}_1, \mathbf{v}_p) \\ (\mathbf{u}_2, \mathbf{v}_1) & (\mathbf{u}_2, \mathbf{v}_2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{u}_p, \mathbf{v}_1) & (\mathbf{u}_p, \mathbf{v}_2) & \cdots & (\mathbf{u}_p, \mathbf{v}_p) \end{bmatrix} \quad (6.55)$$

where  $\mathbf{v} = \mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_p$  and  $\mathbf{u} = \mathbf{u}_1 \wedge \mathbf{u}_2 \wedge \cdots \wedge \mathbf{u}_p$ .

### 6.5.7 The Dual of an Exterior Product

If  $\mathbf{v}_i \in V$  and  $\varphi_j \in V^*$ , then  $\varphi_j(\mathbf{v}_i) \in \mathbb{R}$ . However, since  $\{\varphi_j\}$  forms a vector space (which in the case when  $V = \mathbb{R}^n$  is isomorphic to the vector space consisting of transposed vectors, or the Hermitian conjugate in the complex case) their exterior products can be computed also. The space of all dual  $p$ -vectors is denoted  $\Lambda^p(V^*)$ . That is,

$$\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_p \in \Lambda^p(V^*)$$

and

$$\Lambda^p(V^*) = (\Lambda^p V)^*.$$

The dual exterior product evaluated on the exterior product of the same dimension is computed as

$$\begin{aligned} (\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_p) \cdot [\mathbf{v}_1 \wedge \mathbf{v}_2 \wedge \cdots \wedge \mathbf{v}_p] &= \sum_{\pi \in \Pi_p} \text{sgn}(\pi) \varphi_1(\mathbf{v}_{\pi(1)}) \varphi_2(\mathbf{v}_{\pi(2)}) \cdots \varphi_p(\mathbf{v}_{\pi(p)}) \\ &= \det \begin{bmatrix} \varphi_1(\mathbf{v}_1) & \varphi_1(\mathbf{v}_2) & \cdots & \varphi_1(\mathbf{v}_p) \\ \varphi_2(\mathbf{v}_1) & \varphi_2(\mathbf{v}_2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_p(\mathbf{v}_1) & \varphi_p(\mathbf{v}_2) & \cdots & \varphi_p(\mathbf{v}_p) \end{bmatrix} \in \mathbb{R}. \end{aligned} \quad (6.56)$$

It should not come as a surprise that (6.56) and (6.55) are essentially the same, since associated with each linear function  $\varphi \in V^*$  is a vector  $\mathbf{u} \in V$  such that  $\varphi(\mathbf{v}) = (\mathbf{u}, \mathbf{v})$  for all  $\mathbf{v} \in V$ .

## 6.6 Invariant Description of Vector Fields

Given a differentiable function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , the directional derivative (in the direction  $\mathbf{a}$ , and evaluated at  $\mathbf{x} \in \mathbb{R}^n$ ) is

$$(D_{\mathbf{a}}f)(\mathbf{x}) \doteq \left. \frac{d}{dt} f(\mathbf{x} + t\mathbf{a}) \right|_{t=0}. \quad (6.57)$$

If in addition to being differentiable,  $f(\mathbf{x})$  is analytic, then for fixed  $\mathbf{x}$  and  $\mathbf{a}$  the function  $f_{(\mathbf{x}, \mathbf{a})}(t) = f(\mathbf{x} + t\mathbf{a})$  can be expanded in a one-dimensional Taylor series in  $t$ . Following this by taking the derivative  $d/dt$  and setting  $t = 0$  yields

$$(D_{\mathbf{a}}f)(\mathbf{v}) = a_1 \left. \frac{\partial f}{\partial x_1} \right|_{\mathbf{x}=\mathbf{v}} + \dots + a_n \left. \frac{\partial f}{\partial x_n} \right|_{\mathbf{x}=\mathbf{v}} = \mathbf{a} \cdot (\nabla_{\mathbf{x}}f)|_{\mathbf{x}=\mathbf{v}}. \quad (6.58)$$

Here  $|_{\mathbf{x}=\mathbf{v}}$  means evaluation of the function by substituting  $\mathbf{x}$  with  $\mathbf{v}$ . This will be written in shorthand as  $|_{\mathbf{v}}$ . Equation (6.58) can also be viewed as a direct application of the chain rule.

As  $\mathbf{a}$  is allowed to visit all possible values in  $\mathbb{R}^n$ , the set

$$V = \{(D_{\mathbf{a}}f)(\mathbf{v}) \mid \mathbf{a} \in \mathbb{R}^n\} \quad (6.59)$$

forms a vector space with the operations of addition,  $+$ , and scalar multiplication,  $\cdot$ , following from the linearity property of derivatives:

$$(D_{\alpha \cdot \mathbf{a} + \beta \cdot \mathbf{b}}f)(\mathbf{v}) = \alpha \cdot (D_{\mathbf{a}}f)(\mathbf{v}) + \beta \cdot (D_{\mathbf{b}}f)(\mathbf{v}).$$

The above properties hold for *any* differentiable function  $f \in C^1(\mathbb{R}^n)$ , and so it is convenient to think of

$$B = \left\{ \left. \frac{\partial}{\partial x_1} \right|_{\mathbf{v}}, \dots, \left. \frac{\partial}{\partial x_n} \right|_{\mathbf{v}} \right\}$$

as a basis for the vector space  $(V, +, \cdot)$ . If  $\mathbf{a} = \mathbf{a}(\mathbf{v}) = \sum_{i=1}^n a_i(\mathbf{v})\mathbf{e}_i$  is a vector field on  $\mathbb{R}^n$ , then so too is

$$\mathcal{A}(\mathbf{v}) = \sum_{i=1}^n a_i(\mathbf{v}) \left. \frac{\partial}{\partial x_i} \right|_{\mathbf{v}}, \quad \text{or} \quad \mathcal{A} = \sum_{i=1}^n a_i \frac{\partial}{\partial x_i}. \quad (6.60)$$

The second expression is shorthand for the first. In this notation, the application of a vector field to a function results in a directional derivative:

$$\boxed{\mathcal{A}f = D_{\mathbf{a}}f}. \quad (6.61)$$

The *Lie bracket* of two such vector fields is defined as

$$\boxed{[\mathcal{A}, \mathcal{B}](f) = \mathcal{A}(\mathcal{B}f) - \mathcal{B}(\mathcal{A}f)} \quad (6.62)$$

where each vector field is evaluated at the same value of  $\mathbf{v}$ .

Why go through all of this trouble when  $\{\mathbf{e}_i\}$  is a perfectly good basis for  $\mathbb{R}^n$ ? Two answers to this question are: (a) the form of (6.60) is independent of the basis used; and (b) it generalizes better to the intrinsic study of manifolds.

As a demonstration of point (a), consider the smooth and invertible deformation of space  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , and let  $\mathbf{x} = \mathbf{g}(\mathbf{x}')$  and  $\mathbf{v} = \mathbf{g}(\mathbf{v}')$ . The Jacobian of this transformation is

$$J = \frac{\partial \mathbf{x}}{\partial (\mathbf{x}')^T} \in \mathbb{R}^{n \times n}.$$

In component form this is  $J_{ij} = \partial x_i / \partial x'_j$ , and the elements of the inverse Jacobian are  $J^{ij} = \partial x'_i / \partial x_j$ .

Let  $f'(\mathbf{x}') = f(\mathbf{g}(\mathbf{x}'))$ . Then  $f'(\mathbf{g}^{-1}(\mathbf{x})) = f(\mathbf{x})$ . From the chain rule,

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^n \frac{\partial f'}{\partial x'_j} \frac{\partial x'_j}{\partial x_i} = \sum_{j=1}^n J^{ji} \frac{\partial f'}{\partial x'_j}.$$

Then

$$\mathcal{A}(\mathbf{v})f = \sum_{i=1}^n a_i(\mathbf{v}) \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{v}} = \sum_{j=1}^n \left( \sum_{i=1}^n J^{ji}(\mathbf{x}') a_i(\mathbf{g}(\mathbf{x}')) \right) \left. \frac{\partial f'}{\partial x'_j} \right|_{\mathbf{v}'}$$

Therefore, if

$$\mathbf{a}'(\mathbf{x}') = [J(\mathbf{x}')]^{-1} \mathbf{a}(\mathbf{g}(\mathbf{x}')),$$

or equivalently,

$$\mathbf{a}(\mathbf{x}) = J(\mathbf{g}^{-1}(\mathbf{x})) \mathbf{a}'(\mathbf{g}^{-1}(\mathbf{x})),$$

then

$$\boxed{\mathcal{A}(\mathbf{v})f = \mathcal{A}'(\mathbf{v}')f'} \quad (6.63)$$

In the special case when  $\mathbf{a}(\mathbf{x}) = A_0\mathbf{x}$  and  $\mathbf{g}(\mathbf{y}) = G_0\mathbf{y}$  where  $A_0$  and  $G_0$  are invertible constant matrices, (6.63) holds with  $\mathbf{a}'(\mathbf{x}') = (G_0^{-1}A_0G_0)\mathbf{x}' = A'_0\mathbf{x}'$ .

In the modern view,  $\mathbf{a}(\mathbf{x}) \notin \mathbb{R}^n$ . Rather, the tangent vector  $\mathcal{A}(\mathbf{x})$  as defined in (6.60) belongs to a new space called *the tangent space to  $\mathbb{R}^n$  at the point  $\mathbf{x}$* . This is denoted as  $\mathcal{A} \in T_x\mathbb{R}^n$ . In a similar way, when considering a manifold,  $M$  (i.e., higher-dimensional generalization of a simple curve or surface), a point in the manifold (which is not necessarily described as a vector) is denoted as  $x \in M$ , and a vector in the tangent space to  $M$  at  $x$  is denoted as  $T_xM$ .

Why go through all of this trouble when a perfectly valid definition of vector fields already existed? Detailed answers can be found in [1, 3, 14]. The short answer is that there is a bijective mapping between the set of all  $\mathbf{a}(\mathbf{x})$ 's and the set of all  $\mathcal{A}(\mathbf{x})$ 's, and so for any  $\mathbf{a}(\mathbf{x})$  there is a unique  $\mathcal{A}(\mathbf{x})$ , and vice versa. And it is  $\mathcal{A}(\mathbf{x})$  (rather than  $\mathbf{a}(\mathbf{x})$ ) that has some nice properties that are used in the following section, which in turn are useful in describing vector fields on manifolds in a way that is independent of how they are embedded in a higher-dimensional Euclidean space.

## 6.7 Push-Forwards and Pull-Backs in $\mathbb{R}^n$

### 6.7.1 General Theory

Let  $U$  and  $V$  be two open subsets of  $\mathbb{R}^n$  that are related to each other through a transformation  $\psi : U \rightarrow V$  that is invertible, and both  $\psi$  and  $\psi^{-1}$  are smooth (i.e., infinitely differentiable). Then  $\psi$  is called a *diffeomorphism*.

Now suppose that there is a smooth function  $f : V \rightarrow \mathbb{R}$ . The *differential* of the diffeomorphism  $\psi$  is denoted as  $d\psi(x)$ , and is defined by the equality

$$(d\psi(x)\mathcal{A})f \doteq \mathcal{A}f(\psi(x)) \quad (6.64)$$

for *any* smooth  $f$ , where  $\mathcal{A}$  is a vector field as interpreted in the modern sense in (6.61). The definition of the differential in (6.64) is related to the Jacobian  $D\psi = [\partial\psi_i/\partial x_j]$ , where  $D\psi$  is shorthand for  $(D\psi)(x)$ , by the fact that [3]

$$\mathcal{B} = d\psi(x)\mathcal{A} \quad \iff \quad \mathbf{b} = [D\psi]\mathbf{a}.$$

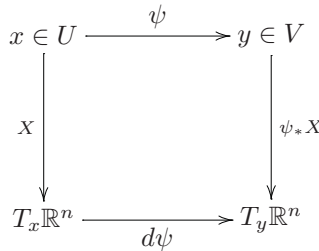
However, (6.64) would *not* work if  $\mathbf{a}$  were substituted for  $\mathcal{A}$ . Henceforth in this section boldface will be dropped and vectors interpreted in the form of  $v$  will be used in order to be consistent with the literature.

Now suppose that a vector field is defined on  $U$ , such that for every  $x \in U$ , we have a vector  $X(x) \in T_x\mathbb{R}^n \cong \mathbb{R}^n$ . (Note that  $X(x)$  need not be confined to  $U$ .) Using the

function  $\psi$ , the vector field  $X$  can be used to define a vector field  $Y$  on  $V$  by assigning to each  $y \in V$  a vector  $Y(y) = \psi_*X(y)$  where  $\psi_*X$  is called a *push-forward* (vector field), and is defined by the expression

$$\psi_*X(y) \doteq d\psi(x)(X(\psi^{-1}(y))). \tag{6.65}$$

As pointed out in [3], the above equation can be expressed in the commutative diagram:



The *pull-back*  $\psi^*\omega$  is defined for any 1-form  $\omega$  by the equality [3, 4, 14]

$$(\psi^*\omega) \cdot X(x) \doteq \omega \cdot (\psi_*X)(y) \tag{6.66}$$

where  $X$  is a vector field on  $U$  and  $\psi_*X$  is the push-forward vector field on  $V$ . The syntax is “ $\psi^*\omega$  is the pull-back of  $\omega$ .” The definition in (6.66) reflects the fact that  $X$  and  $\psi_*X$  are 1-vectors and  $\omega$  and  $\psi^*\omega$  are 1-forms, and the result of dotting a 1-form and a 1-vector is a scalar.

The pull-back of a  $p$ -form  $\omega$  is defined by the expression

$$\boxed{(\psi^*\omega) \cdot [X_1(x) \wedge \dots \wedge X_p(x)] \doteq \omega \cdot [(\psi_*X_1)(y) \wedge \dots \wedge (\psi_*X_p)(y)]} \tag{6.67}$$

where the dot is interpreted as in (6.56).

The pull-back of a form can be defined more generally, i.e., it is not restricted to the case of diffeomorphisms. But this is the case that will be most common in the applications considered later. In the following subsection, examples illustrate calculations with forms in detail.

### 6.7.2 Example Calculations

This subsection demonstrates the definitions in great detail.

#### Example 1: Forms and the Chain Rule

Let  $(x_1, x_2, x_3)$ ,  $(r, \phi, z)$ , and  $(R, \Phi, \Theta)$  respectively denote Cartesian, cylindrical, and spherical coordinates in  $\mathbb{R}^3$ , and the ranges of all parameters are chosen so as to avoid the singularities at  $r = R = 0$  and  $\Theta = 0, \pi$ . Let

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \zeta(r, \phi, z) = \begin{pmatrix} \zeta_1(r, \phi, z) \\ \zeta_2(r, \phi, z) \\ \zeta_3(r, \phi, z) \end{pmatrix} \doteq \begin{pmatrix} r \cos \phi \\ r \sin \phi \\ z \end{pmatrix} \tag{6.68}$$

and

$$\begin{pmatrix} r \\ \phi \\ z \end{pmatrix} = \xi(R, \Phi, \Theta) = \begin{pmatrix} \xi_1(R, \Phi, \Theta) \\ \xi_2(R, \Phi, \Theta) \\ \xi_3(R, \Phi, \Theta) \end{pmatrix} \doteq \begin{pmatrix} R \sin \Theta \\ \Phi \\ R \cos \Theta \end{pmatrix}. \tag{6.69}$$

Then

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \psi(R, \Phi, \Theta) = \begin{pmatrix} \psi_1(R, \Phi, \Theta) \\ \psi_2(R, \Phi, \Theta) \\ \psi_3(R, \Phi, \Theta) \end{pmatrix} \doteq \begin{pmatrix} R \cos \Phi \sin \Theta \\ R \sin \Phi \sin \Theta \\ R \cos \Theta \end{pmatrix} \quad (6.70)$$

where

$$\psi = \zeta \circ \xi \quad (6.71)$$

(that is, the function  $\zeta$  composed with the function  $\xi$ ).

The differentials in Cartesian coordinates can be related to those in polar and spherical coordinates as

$$dx_1 = \cos \phi dr - r \sin \phi d\phi; \quad dx_2 = \sin \phi dr + r \cos \phi d\phi; \quad dx_3 = dz \quad (6.72)$$

and

$$dx_1 = \cos \Phi \sin \Theta dR - R \sin \Phi \sin \Theta d\Phi + R \cos \Phi \cos \Theta d\Theta \quad (6.73)$$

$$dx_2 = \sin \Phi \sin \Theta dR + R \cos \Phi \sin \Theta d\Phi + R \sin \Phi \cos \Theta d\Theta \quad (6.74)$$

$$dx_3 = \cos \Theta dR - R \sin \Theta d\Theta. \quad (6.75)$$

It follows that

$$\begin{aligned} dx_1 \wedge dx_2 &= (\cos \phi dr - r \sin \phi d\phi) \wedge (\sin \phi dr + r \cos \phi d\phi) \\ &= r(\cos^2 \phi + \sin^2 \phi) dr \wedge d\phi \\ &= r dr \wedge d\phi \end{aligned}$$

$$dx_1 \wedge dx_3 = \cos \phi dr \wedge dz - r \sin \phi d\phi \wedge dz$$

$$dx_2 \wedge dx_3 = \sin \phi dr \wedge dz + r \cos \phi d\phi \wedge dz$$

and

$$dx_1 \wedge dx_2 = R \sin^2 \Theta dR \wedge d\Phi - R^2 \sin \Theta \cos \Theta d\Phi \wedge d\Theta$$

$$dx_1 \wedge dx_3 = -R \cos \Phi dR \wedge d\Theta - R \sin \Phi \sin \Theta \cos \Theta d\Phi \wedge dR + R^2 \sin \Phi \sin^2 \Theta d\Phi \wedge d\Theta$$

$$dx_2 \wedge dx_3 = -R \sin \Phi dR \wedge d\Theta + R \cos \Phi \sin \Theta \cos \Theta d\Phi \wedge dR - R^2 \cos \Phi \sin^2 \Theta d\Phi \wedge d\Theta.$$

Furthermore,

$$dx_1 \wedge dx_2 \wedge dx_3 = r dr \wedge d\phi \wedge dz \quad (6.76)$$

$$dx_1 \wedge dx_2 \wedge dx_3 = -R^2 \sin \Theta dR \wedge d\Phi \wedge d\Theta. \quad (6.77)$$

### Example 2: Inverse Mappings and Push-Forward Vector Fields

Each of the mappings in the previous example can be inverted if the domain is restricted to exclude the singularities. Then

$$\begin{pmatrix} r \\ \phi \\ z \end{pmatrix} = \zeta^{-1}(x_1, x_2, x_3) = \begin{pmatrix} \zeta_1^{-1}(x_1, x_2, x_3) \\ \zeta_2^{-1}(x_1, x_2, x_3) \\ \zeta_3^{-1}(x_1, x_2, x_3) \end{pmatrix} \doteq \begin{pmatrix} (x_1^2 + x_2^2)^{\frac{1}{2}} \\ \tan^{-1}(x_2/x_1) \\ x_3 \end{pmatrix} \quad (6.78)$$

and

$$\begin{pmatrix} R \\ \Phi \\ \Theta \end{pmatrix} = \boldsymbol{\xi}^{-1}(r, \phi, z) = \begin{pmatrix} \xi_1^{-1}(r, \phi, z) \\ \xi_2^{-1}(r, \phi, z) \\ \xi_3^{-1}(r, \phi, z) \end{pmatrix} \doteq \begin{pmatrix} (r^2 + z^2)^{\frac{1}{2}} \\ \phi \\ \tan^{-1}(r/z) \end{pmatrix} \quad (6.79)$$

and

$$\boldsymbol{\psi}^{-1} = (\boldsymbol{\zeta} \circ \boldsymbol{\xi})^{-1} = \boldsymbol{\xi}^{-1} \circ \boldsymbol{\zeta}^{-1}.$$

Given a vector field

$$\mathcal{V} = v_1(\mathbf{q}) \frac{\partial}{\partial q_1} + \dots + v_n(\mathbf{q}) \frac{\partial}{\partial q_n} = \sum_{i=1}^n v_i(\mathbf{q}) \frac{\partial}{\partial q_i}$$

(with  $\mathbf{v} \doteq [v_1, \dots, v_n]^T$ ) in coordinates,  $\mathbf{q}$ , the push-forward vector field is what this vector field should appear as in the coordinates  $\mathbf{x} = \boldsymbol{\psi}(\mathbf{q})$ . (Here  $\boldsymbol{\psi}$  is referring to a general transformation, i.e., not necessarily spherical coordinates.) The push-forward can then be computed in terms of the inverse mapping and Jacobian matrix as

$$\boxed{\boldsymbol{\psi}_* \mathcal{V} = \sum_{i=1}^n \mathbf{e}_i^T \mathbf{v}_* \frac{\partial}{\partial x_i} \quad \text{where} \quad \mathbf{v}_*(\mathbf{x}) \doteq \left[ \frac{\partial \boldsymbol{\psi}}{\partial \mathbf{q}^T} \mathbf{v}(\mathbf{q}) \right] \Big|_{\mathbf{q}=\boldsymbol{\psi}^{-1}(\mathbf{x})}} \quad (6.80)$$

and, as always,  $(\mathbf{e}_i)_j = \delta_{ij}$ .

For example, given a vector field in cylindrical coordinates of the form

$$\mathcal{V} = v_1(r, \phi, z) \frac{\partial}{\partial r} + v_2(r, \phi, z) \frac{\partial}{\partial \phi} + v_3(r, \phi, z) \frac{\partial}{\partial z},$$

then with  $\mathbf{q} = [r, \phi, z]^T$  the Jacobian in this case is

$$\frac{\partial \boldsymbol{\zeta}}{\partial \mathbf{q}^T} = \begin{pmatrix} \cos \phi & -r \sin \phi & 0 \\ \sin \phi & r \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\frac{\partial \boldsymbol{\zeta}}{\partial \mathbf{q}^T} \Big|_{\mathbf{q}=\boldsymbol{\zeta}^{-1}(\mathbf{x})} = \begin{pmatrix} x_1/(x_1^2 + x_2^2)^{\frac{1}{2}} & -x_2 & 0 \\ x_2/(x_1^2 + x_2^2)^{\frac{1}{2}} & x_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the corresponding push-forward vector field in Cartesian coordinates is therefore

$$\boldsymbol{\zeta}_* \mathcal{V} = v'_1(\boldsymbol{\zeta}^{-1}(\mathbf{x})) \frac{\partial}{\partial x_1} + v'_2(\boldsymbol{\zeta}^{-1}(\mathbf{x})) \frac{\partial}{\partial x_2} + v'_3(\boldsymbol{\zeta}^{-1}(\mathbf{x})) \frac{\partial}{\partial x_3},$$

where

$$v'_i(\boldsymbol{\zeta}^{-1}(\mathbf{x})) = v'_i((x_1^2 + x_2^2)^{\frac{1}{2}}, \tan^{-1}(x_2/x_1), x_3) \quad \text{and} \quad \mathbf{v}' = \frac{\partial \boldsymbol{\zeta}}{\partial \mathbf{q}^T} \mathbf{v}.$$

Note that while the mappings  $\boldsymbol{\zeta}$ ,  $\boldsymbol{\xi}$ , and  $\boldsymbol{\psi}$  are vector-valued and are therefore denoted in bold above, in order to be consistent with the literature henceforth they are denoted in the “lighter” (non-bold) notation, as was the case in the discussion earlier in this section.



**Example 3: Forms and Composition of Transformations**

Differential one-forms in different curvilinear coordinate systems are obtained from those in Cartesian coordinates via the classical chain rule and composition of transformations and functions as

$$\begin{aligned}\omega_1 &= a_1(\mathbf{x})dx_1 + a_2(\mathbf{x})dx_2 + a_3(\mathbf{x})dx_3 \\ &= a_1(\zeta(r, \phi, z))(dr \cos \phi - d\phi \sin \phi) \\ &\quad + a_2(\zeta(r, \phi, z))(dr \sin \phi + d\phi \cos \phi) + a_3(\zeta(r, \phi, z))dz \\ &\doteq \zeta^* \omega_1\end{aligned}\tag{6.81}$$

and

$$\begin{aligned}\omega_1 &= a_1(\mathbf{x})dx_1 + a_2(\mathbf{x})dx_2 + a_3(\mathbf{x})dx_3 \\ &= a_1(\xi(R, \Phi, \Theta))(\cos \Phi \sin \Theta dR - R \sin \Phi \sin \Theta d\Phi + R \cos \Phi \cos \Theta d\Theta) \\ &\quad + a_2(\xi(R, \Phi, \Theta))(\sin \Phi \sin \Theta dR + R \cos \Phi \sin \Theta d\Phi + R \sin \Phi \cos \Theta d\Theta) \\ &\quad + a_3(\xi(R, \Phi, \Theta))(\cos \Theta dR - R \sin \Theta d\Theta) \\ &\doteq \psi^* \omega_1.\end{aligned}$$

In other words,  $\zeta^* \omega_1$  and  $\psi^* \omega_1$  are simply  $\omega_1$  as it appears in polar and spherical coordinates, respectively.

Furthermore, given a form such as  $\zeta^* \omega_1$  that can be written as

$$\beta_1 = b_1(r, \phi, z)dr + b_2(r, \phi, z)d\phi + b_3(r, \phi, z)dz\tag{6.82}$$

in polar coordinates, it is possible to compute

$$\begin{aligned}\xi^* \beta_1 &= b_1(\xi(R, \Phi, \Theta))(\sin \Theta dR + R \cos \Theta d\Theta) + b_1(\xi(R, \Phi, \Theta))d\Phi \\ &\quad + b_1(\xi(R, \Phi, \Theta))(\cos \Theta dR - R \sin \Theta d\Theta).\end{aligned}$$

Amid this exercise in the chain rule, things become interesting with the observation that

$$\xi^*(\zeta^* \omega_1) = (\zeta \circ \xi)^* \omega_1.\tag{6.83}$$

In other words, pull-backs can either be concatenated, or the transformations can be composed and the corresponding pull-back can be calculated, and the result will be the same!

Now suppose that we are given two generic one-forms,  $\alpha_1$  and  $\omega_1$  in Cartesian coordinates. Then after some straightforward calculations, it is can be verified that<sup>9</sup>

$$\zeta^*(\alpha_1 \wedge \omega_1) = \zeta^*(\alpha_1) \wedge \zeta^*(\omega_1) \quad \text{and} \quad \xi^*(\beta_1 \wedge \eta_1) = \xi^*(\beta_1) \wedge \xi^*(\eta_1)$$

where  $\beta_1$  and  $\eta_1$  are the same kind of form as that defined in (6.82). Also, when performed directly, some tedious (though conceptually not difficult) calculations lead to

$$\psi^*(\alpha_1 \wedge \omega_1) = \psi^*(\alpha_1) \wedge \psi^*(\omega_1).$$

However, this tedium can be avoided by breaking the problem up into two simpler problems and using (6.83) as follows:

<sup>9</sup>The notation  $\zeta^*(\omega_1)$  and  $\zeta^* \omega_1$  mean exactly the same thing, but it is sometimes clearer to use the former in writing expressions such as  $\zeta^*(d\omega_1)$  or  $\zeta^*(\omega_1) \wedge \zeta^*(\alpha_1)$  rather than  $\zeta^* d\omega_1$  or  $\zeta^* \omega_1 \wedge \zeta^* \alpha_1$ .

$$\xi^*(\zeta^*(\alpha_1 \wedge \omega_1)) = (\zeta \circ \xi)^*(\alpha_1 \wedge \omega_1) = \psi^*(\alpha_1 \wedge \omega_1).$$

On the other hand, choosing  $\beta_1 = \zeta^* \alpha_1$  and  $\eta_1 = \zeta^* \omega_1$  gives

$$\xi^*(\beta_1) \wedge \xi^*(\eta_1) = \xi^*(\zeta^* \alpha_1) \wedge \xi^*(\zeta^* \omega_1) = \psi^*(\alpha_1) \wedge \psi^*(\omega_1).$$

Indeed, in this regard there is nothing special about cylindrical and spherical coordinates, and the above hold in general for pull-backs of differential one-forms.

And things become even more interesting when this exercise is attempted for two-forms and three-forms, and it is concluded that for general differentiable transformations  $\xi$  and  $\zeta$ , and general forms  $\omega$  and  $\alpha$  the following hold:

$$\boxed{\xi^*(\zeta^* \omega) = (\zeta \circ \xi)^* \omega} \quad (6.84)$$

and

$$\boxed{\xi^*(\alpha \wedge \omega) = \xi^*(\alpha) \wedge \xi^*(\omega)}. \quad (6.85)$$

Another important property of the pull-back is that it is linear. This is left as an exercise to prove.

#### Example 4: Pull-Backs and Exterior Derivatives of Forms

The exterior derivative of the 1-form  $\omega_1$  in (6.81) was given in (6.18). If this is then converted to cylindrical coordinates, then

$$\begin{aligned} \zeta^*(d\omega_1) &= \left( \frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) \Big|_{\mathbf{x}=\zeta(r,\phi,z)} r dr \wedge d\phi \\ &+ \left( \frac{\partial a_3}{\partial x_2} - \frac{\partial a_2}{\partial x_3} \right) \Big|_{\mathbf{x}=\zeta(r,\phi,z)} (\sin \theta dr \wedge dz + r \cos \theta d\theta \wedge dz) \\ &+ \left( \frac{\partial a_3}{\partial x_1} - \frac{\partial a_1}{\partial x_3} \right) \Big|_{\mathbf{x}=\zeta(r,\phi,z)} (\cos \theta dr \wedge dz - r \sin \theta d\theta \wedge dz). \end{aligned}$$

On the other hand,

$$\begin{aligned} d(\zeta^* \omega_1) &= d[a_1(\zeta(r, \phi, z))(dr \cos \phi - d\phi \sin \phi) + a_2(\zeta(r, \phi, z))(dr \sin \phi + d\phi \cos \phi) \\ &+ a_3(\zeta(r, \phi, z))dz] \\ &= d[a_1(\zeta(r, \phi, z)) \cos \phi + a_2(\zeta(r, \phi, z)) \sin \phi] dr \\ &+ d[a_2(\zeta(r, \phi, z)) \cos \phi - a_1(\zeta(r, \phi, z)) \sin \phi] d\phi \\ &+ d[a_3(\zeta(r, \phi, z))] dz. \end{aligned}$$

The first term can be expanded out as

$$\begin{aligned} &\frac{\partial}{\partial \phi} [a_1(\zeta(r, \phi, z)) \cos \phi + a_2(\zeta(r, \phi, z)) \sin \phi] d\phi \wedge dr + \\ &\frac{\partial}{\partial z} [a_1(\zeta(r, \phi, z)) \cos \phi + a_2(\zeta(r, \phi, z)) \sin \phi] dz \wedge dr = \\ &\left[ \frac{\partial a_1}{\partial \zeta^T} \frac{\partial \zeta}{\partial \phi} \cos \phi + \frac{\partial a_2}{\partial \zeta^T} \frac{\partial \zeta}{\partial \phi} \sin \phi \right] d\phi \wedge dr + \\ &[-a_1(\zeta(r, \phi, z)) \sin \phi + a_2(\zeta(r, \phi, z)) \cos \phi] d\phi \wedge dr + \end{aligned}$$

$$\left[ \frac{\partial a_1}{\partial \zeta^T} \frac{\partial \zeta}{\partial z} \cos \phi + \frac{\partial a_2}{\partial \zeta^T} \frac{\partial \zeta}{\partial z} \sin \phi \right] dz \wedge dr$$

where the chain rule was used to obtain the right-hand side, and partial derivatives of  $\zeta$  are computed easily by referring back to (6.68). Similar expansions of the other terms, together with re-collecting terms, lead to the equality

$$d(\zeta^* \omega_1) = \zeta^*(d\omega_1).$$

This observation generalizes as

$$\boxed{d(\psi^* \omega) = \psi^*(d\omega)} \quad (6.86)$$

where  $\omega$  is any  $k$ -form, and  $\psi^*$  is the pull-back of any smooth mapping  $\psi : U \rightarrow V$  where  $U, V \in \mathbb{R}^n$ . Equipped with the intuition gained by the above example, and the general definition of pull-back and exterior derivative given earlier in this chapter, it is not difficult to prove (6.86) using properties of Jacobians and multilinear algebra. Proofs of this, as well as for (6.84) and (6.85), that circumvent the direct use of Jacobians can be found in [3, 4, 5].

### Example 5: Some Very Concrete Calculations

Consider the one-form

$$\gamma_1 = x_1^2 \sin x_2 dx_1 + x_3 e^{-x_1} dx_2 + \cos x_2 dx_3 \quad (6.87)$$

and the two-form

$$\gamma_2 = x_1^2 \sin x_2 dx_1 \wedge dx_2 + x_3 e^{-x_1} dx_2 \wedge dx_3 + \cos x_2 dx_1 \wedge dx_3. \quad (6.88)$$

Both are forms on  $\mathbb{R}^3$ , where  $\mathbf{x} = [x_1, x_2, x_3]^T \in \mathbb{R}^3$  denotes Cartesian coordinates. These are given the names  $\gamma_k$  to distinguish them from the generic forms denoted as  $\omega, \alpha, \beta$ . The subscript  $k$  denotes that it is a  $k$ -form.

The exterior derivative of  $\gamma_1$  is computed as

$$\begin{aligned} d\gamma_1 &= \left[ \frac{\partial}{\partial x_1} (x_1^2 \sin x_2) dx_1 + \frac{\partial}{\partial x_2} (x_1^2 \sin x_2) dx_2 + \frac{\partial}{\partial x_3} (x_1^2 \sin x_2) dx_3 \right] \wedge dx_1 \\ &\quad + \left[ \frac{\partial}{\partial x_1} (x_3 e^{-x_1}) dx_1 + \frac{\partial}{\partial x_2} (x_3 e^{-x_1}) dx_2 + \frac{\partial}{\partial x_3} (x_3 e^{-x_1}) dx_3 \right] \wedge dx_2 \\ &\quad + \left[ \frac{\partial}{\partial x_1} (\cos x_2) dx_1 + \frac{\partial}{\partial x_2} (\cos x_2) dx_2 + \frac{\partial}{\partial x_3} (\cos x_2) dx_3 \right] \wedge dx_3 \\ &= (x_1^2 \cos x_2 dx_2) \wedge dx_1 + (e^{-x_1} dx_3 - x_3 e^{-x_1} dx_1) \wedge dx_2 + (-\sin x_2 dx_2) \wedge dx_3 \\ &= -(x_1^2 \cos x_2 + x_3 e^{-x_1}) dx_1 \wedge dx_2 - (e^{-x_1} + \sin x_2) dx_2 \wedge dx_3. \end{aligned} \quad (6.89)$$

Some of the simplifications result from partial derivatives being zero, and others are due to the fact that  $dx_1 \wedge dx_1 = dx_2 \wedge dx_2 = dx_3 \wedge dx_3 = 0$  and  $dx_1 \wedge dx_2 = -dx_2 \wedge dx_1$ , etc.

Similarly, the exterior derivative of  $\gamma_2$  is computed as

$$\begin{aligned}
d\gamma_2 &= \left[ \frac{\partial}{\partial x_1}(x_1^2 \sin x_2) dx_1 + \frac{\partial}{\partial x_2}(x_1^2 \sin x_2) dx_2 \right] \wedge dx_1 \wedge dx_2 \\
&+ \left[ \frac{\partial}{\partial x_1}(x_3 e^{-x_1}) dx_1 + \frac{\partial}{\partial x_3}(x_3 e^{-x_1}) dx_3 \right] \wedge dx_2 \wedge dx_3 \\
&+ \left[ \frac{\partial}{\partial x_2}(\cos x_2) dx_2 \right] \wedge dx_1 \wedge dx_3 \\
&= -x_3 e^{-x_1} dx_1 \wedge dx_2 \wedge dx_3 - \sin x_2 dx_2 \wedge dx_1 \wedge dx_3 \\
&= (\sin x_2 - x_3 e^{-x_1}) dx_1 \wedge dx_2 \wedge dx_3.
\end{aligned}$$

Taking the exterior derivative of  $d\gamma_2$  will clearly introduce repeated wedge products of the form  $dx_1 \wedge dx_1 \wedge dx_2 \wedge dx_3$ ,  $dx_2 \wedge dx_1 \wedge dx_2 \wedge dx_3$ , and  $dx_3 \wedge dx_1 \wedge dx_2 \wedge dx_3$  that are all equal to zero. From this it follows that  $d(d\gamma_2) = 0$ .

In the case of the exterior derivative of  $d\gamma_1$ ,

$$\begin{aligned}
d(d\gamma_1) &= -\frac{\partial}{\partial x_3}(x_1^2 \cos x_2 + x_3 e^{-x_1}) dx_3 \wedge dx_1 \wedge dx_2 \\
&\quad - \frac{\partial}{\partial x_1}(e^{-x_1} + \sin x_2) dx_1 \wedge dx_2 \wedge dx_3 \\
&= -e^{-x_1}(dx_3 \wedge dx_1 \wedge dx_2 - dx_1 \wedge dx_2 \wedge dx_3) \\
&= 0,
\end{aligned}$$

which is just a special case of (6.16).

It is easy to verify by inspection that

$$\gamma_1 \wedge \gamma_1 = \gamma_2 \wedge \gamma_2 = 0.$$

It is not difficult to compute

$$\gamma_1 \wedge \gamma_2 = [x_1^2 x_3 e^{-x_1} \sin x_2 - x_3 e^{-x_1} \cos x_2 + x_1^2 \sin x_2 \cos x_2] dx_1 \wedge dx_2 \wedge dx_3$$

and

$$\gamma_1 \wedge d\gamma_1 = -[x_1^2 + e^{-x_1}(x_3 \cos x_2 + x_1^2 \sin x_2)] dx_1 \wedge dx_2 \wedge dx_3.$$

The Hodge star operator gives

$$*\gamma_1 = x_1^2 \sin x_2 dx_2 \wedge dx_3 - x_3 e^{-x_1} dx_1 \wedge dx_3 + \cos x_2 dx_1 \wedge dx_2$$

and

$$*\gamma_2 = x_1^2 \sin x_2 dx_3 + x_3 e^{-x_1} dx_1 - \cos x_2 dx_2.$$

The forms  $\gamma_1$  and  $\gamma_2$  can be described in a curvilinear coordinate system rather than Cartesian coordinates. Substituting (6.68) and (6.72) into (6.87) gives

$$\begin{aligned}
\zeta^* \gamma_1 &= [(r \cos \phi)^2 \sin(r \sin \phi)](\cos \phi dr - r \sin \phi d\phi) \\
&\quad + z e^{-r \cos \phi}(\sin \phi dr + r \cos \phi d\phi) + \cos(r \sin \phi) dz \\
&= [r^2 \cos^3 \phi \sin(r \sin \phi) + z e^{-r \cos \phi} \sin \phi] dr \\
&\quad + [z r e^{-r \cos \phi} \cos \phi - r^3 \cos^2 \phi \sin \phi \sin(r \sin \phi)] d\phi \\
&\quad + \cos(r \sin \phi) dz.
\end{aligned} \tag{6.90}$$

Computation of  $d\gamma_1$  in Cartesian coordinates as in (6.89) followed by substitution of all  $x_i$  and  $dx_i$  represented in cylindrical coordinates as in (6.68) and (6.72) gives

$$\begin{aligned} \zeta^*(d\gamma_1) = & - [(r \cos \phi)^2 \cos(r \sin \phi) + ze^{-r \cos \phi}] r dr \wedge d\phi \\ & - [e^{-r \cos \phi} + \sin(r \sin \phi)] (\sin \phi dr \wedge dz + r \cos \phi d\phi \wedge dz). \end{aligned} \quad (6.91)$$

On the other hand, computing the exterior derivative of  $\gamma_1$  expressed in cylindrical coordinates as in (6.90) gives

$$\begin{aligned} d(\zeta^* \gamma_1) = & \left[ \frac{\partial}{\partial \phi} (r^2 \cos^3 \phi \sin(r \sin \phi) + ze^{-r \cos \phi} \sin \phi) \right] d\phi \wedge dr \\ & + \left[ \frac{\partial}{\partial z} (r^2 \cos^3 \phi \sin(r \sin \phi) + ze^{-r \cos \phi} \sin \phi) \right] dz \wedge dr \\ & + \left[ \frac{\partial}{\partial z} (zre^{-r \cos \phi} \cos \phi - r^3 \cos^2 \phi \sin \phi \sin(r \sin \phi)) \right] dz \wedge d\phi \\ & + \left[ \frac{\partial}{\partial r} (zre^{-r \cos \phi} \cos \phi - r^3 \cos^2 \phi \sin \phi \sin(r \sin \phi)) \right] dr \wedge d\phi \\ & + \frac{\partial}{\partial \phi} [\cos(r \sin \phi)] d\phi \wedge dr \\ & + \frac{\partial}{\partial r} [\cos(r \sin \phi)] dr \wedge dz. \end{aligned}$$

After expanding out these partial derivatives and rearranging terms it can be observed that this is the same as (6.91), as must be the case for the definition of the exterior derivative to be self-consistent.

## 6.8 Generalizing Integral Theorems from Vector Calculus

In this section, the integration of forms, and generalizations of Stokes' theorem are reviewed.

### 6.8.1 Integration of Differential Forms

Consider the linear function  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  defined by

$$\varphi(\mathbf{x}) = \mathbf{a}^T \mathbf{x} = \sum_{i=1}^n a_i x_i \quad (6.92)$$

where each  $a_i \in \mathbb{R}$  is a constant. The differential of this linear map is

$$d\varphi = \sum_{i=1}^n \frac{\partial \varphi}{\partial x_i} dx_i = \sum_{i=1}^n a_i dx_i.$$

Or, stated in another way, if the function  $x_i : \mathbb{R}^n \rightarrow \mathbb{R}$  is defined to extract the  $i$ th coordinate of a vector as  $x_i(\mathbf{v}) = \mathbf{e}_i^T \mathbf{v} = v_i$ , then the collection of all

$$dx_i(\mathbf{v})|_{\mathbf{v}=\mathbf{x}} = dx_i$$

for  $i = 1, \dots, n$  forms a basis for the vector space of all such maps,  $V^*$ .

The quantity  $d\varphi$  is a differential 1-form. As discussed earlier, a differential  $k$ -form can be constructed from the wedge product of  $k$  differential 1-forms. A differential  $k$ -form,  $\omega$ , on  $\mathbb{R}^n$  can be defined with respect to Cartesian coordinates  $\mathbf{x} = [x_1, \dots, x_n]^T$ , and a set of smooth functions  $\{a_{i_1, i_2, \dots, i_k}(\mathbf{x})\}$  as

$$\omega = \sum_{i_1, i_2, \dots, i_k} a_{i_1, i_2, \dots, i_k}(\mathbf{x}) dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k} \quad \text{where } 1 \leq i_1 < i_2 < \dots < i_k \leq n.$$

The above equation can be written more concisely as

$$\omega = \sum_{I_k} a_{I_k} dx_{I_k} \tag{6.93}$$

where  $I_k = \{i_1, i_2, \dots, i_k\}$  is any subset of  $\{1, \dots, n\}$  consisting of  $k$  distinct numbers written in strictly increasing order,  $a_{I_k} = a_{i_1, i_2, \dots, i_k}$  and  $dx_{I_k} = dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k}$ .

When  $k = n$ , and there is a change of coordinates  $\mathbf{x}'(\mathbf{x})$ ,

$$dx'_1 \wedge dx'_2 \wedge \dots \wedge dx'_n = \det \left[ \frac{\partial x'_i}{\partial x_j} \right] dx_1 \wedge dx_2 \wedge \dots \wedge dx_n. \tag{6.94}$$

This is almost the same as the usual change of coordinates for a differential volume element in  $\mathbb{R}^n$ :

$$dx'_1 dx'_2 \dots dx'_n = \left| \det \left[ \frac{\partial x'_i}{\partial x_j} \right] \right| dx_1 dx_2 \dots dx_n.$$

The important differences come from the facts that: (1) the order of multiplication of differentials is unimportant in the latter expression, whereas the wedge product is anti-commuting; (2) there is no absolute value sign on determinant in (6.94). These differences become quite important, for example, when studying non-orientable manifolds, or generalizing Stokes' theorem in high-dimensional spaces. However, for most of the mundane sorts of integration problems that arise in the probabilistic problems that will be discussed in this work it suffices to write

$$\int_{\mathbb{R}^n} f(\mathbf{x}) dx_1 \wedge dx_2 \wedge \dots \wedge dx_n = \int_{\mathbb{R}^n} f(\mathbf{x}) dx_1 dx_2 \dots dx_n \tag{6.95}$$

with the understanding that the order of terms and bounds of integration will not be changed.

### 6.8.2 The Inner Product of Forms

Multi-vectors form a vector space, and the inner product of multi-vectors with the same dimension was defined in a natural way. It is natural to assume that differential forms, which are derived from the dual of multi-vectors, should also lend themselves to a natural definition of inner product. Such a product should take two differential forms of the same dimensionality and return a scalar. In order to do so, it is expected that an integral should be involved to cancel the “differential” aspect of a differential form.

Based on the discussion in the previous section, it really only makes sense to integrate  $n$ -forms over  $\mathbb{R}^n$  (or a body  $B \subset \mathbb{R}^n$ ). Given two  $n$ -forms,  $\alpha_n = a(\mathbf{x}) dx_1 \wedge \dots \wedge dx_n$  and  $\beta_n = b(\mathbf{x}) dx_1 \wedge \dots \wedge dx_n$ , their inner product can be defined as

$$\langle \alpha_n, \beta_n \rangle \doteq \int_{B \subset \mathbb{R}^n} a(\mathbf{x}) b(\mathbf{x}) dx_1 \dots dx_n.$$

This can be written in terms of the Hodge star operator as

$$\langle \alpha_n, \beta_n \rangle = \int_B \alpha_n \wedge * \beta_n = \int_B * \alpha_n \wedge \beta_n = \langle \beta_n, \alpha_n \rangle$$

where the wedge product of a 0-form and an  $n$ -form is interpreted as scalar multiplication of the function defining the 0-form with the  $n$ -form.

The beauty of this approach is that it generalizes. If  $\alpha_k$  and  $\beta_k$  are two  $k$ -forms on  $B \subset \mathbb{R}^n$ , then

$$\langle \alpha_k, \beta_k \rangle \doteq \int_B \alpha_k \wedge * \beta_k \quad (6.96)$$

returns a scalar value and is consistent with all of the properties of an inner product. For example,  $\langle \alpha_k, \beta_k \rangle = \langle \beta_k, \alpha_k \rangle$ , it is bi-linear, etc.

### 6.8.3 Green's Theorem for a Square Region in $\mathbb{R}^2$

Consider the bi-unit square  $B = [-1, 1] \times [-1, 1] \subset \mathbb{R}^2$  that has corners  $(-1, -1)$ ,  $(-1, 1)$ ,  $(1, -1)$ , and  $(1, 1)$ . Let  $\mathbf{x} \in B$  and  $\omega = a_1(\mathbf{x})dx_1 + a_2(\mathbf{x})dx_2$ . Then

$$\int_{\partial B} \omega = \int_{\partial B} a_1(\mathbf{x})dx_1 + a_2(\mathbf{x})dx_2 \quad (6.97)$$

$$= \int_B \left( \frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) dx_1 dx_2 \quad (6.98)$$

$$= \int_B \left( \frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) dx_1 \wedge dx_2 \quad (6.99)$$

$$= \int_B d\omega. \quad (6.100)$$

The equality in (6.98) is from the classical Green's theorem. The rest are from the definitions of forms and exterior derivatives.

In contrast, if  $*\omega = a_1 dx_2 - a_2 dx_1$ , then from the divergence theorem in the plane,

$$\begin{aligned} \int_{\partial B} *\omega &= \int_{\partial B} \mathbf{a} \cdot \mathbf{n} ds \\ &= \int_B \operatorname{div}(\mathbf{a}) dx_1 dx_2 \\ &= \int_B d(*\omega) \end{aligned}$$

where  $ds$  denotes the differential element of arc length along the boundary and (6.29) was used to establish the final equality.

While the above statements are in fact true for any connected region  $B \subset \mathbb{R}^2$  and associated boundary  $\partial B$ , the computation in the general case involves knowing how  $*\omega$  behaves with changes of coordinates. In contrast, restricting the discussion to the bi-unit square allows all calculations to be performed in Cartesian coordinates.

### 6.8.4 Stokes' Theorem for a Cube in $\mathbb{R}^3$

Now consider the cubic volume  $[-1, 1] \times [-1, 1] \times [-1, 1] \subset \mathbb{R}^3$ . All of the six faces of the bounding cube are copies of the square region  $[-1, 1] \times [-1, 1]$ . Let  $B$  denote the union of some number of these faces, and let  $\partial B$  denote the boundary of  $B$ . In classical

terms,  $B$  would be denoted as  $S$  and the boundary would be described by a piecewise smooth parameterized curve,  $C$ . Note that in the present context  $B$  is now a surface in  $\mathbb{R}^3$  rather than a volume.

If  $\omega = a_1 dx_1 + a_2 dx_2 + a_3 dx_3$ , then from (6.18) and the classical version of Stokes' theorem,

$$\begin{aligned} \int_{\partial B} \omega &= \int_C \mathbf{a} \cdot d\mathbf{x} \\ &= \int_S \nabla \times \mathbf{a} dS \\ &= \int_B d(\omega). \end{aligned}$$

### 6.8.5 The Divergence Theorem for a Cube in $\mathbb{R}^3$

Unlike in Stokes' theorem in the previous subsection, now let  $B = [-1, 1] \times [-1, 1] \times [-1, 1]$  be a volume in  $\mathbb{R}^3$ . If  $\omega = a_1 dx_1 + a_2 dx_2 + a_3 dx_3$ , then  $*\omega = a_3 dx_1 \wedge dx_2 - a_2 dx_1 \wedge dx_3 + a_1 dx_2 \wedge dx_3$  and  $d(*\omega) = \operatorname{div}(\mathbf{a}) dx_1 \wedge dx_2 \wedge dx_3$ . Therefore, from the classical divergence theorem in  $\mathbb{R}^3$ ,

$$\begin{aligned} \int_{\partial B} *\omega &= \int_{\partial B} \mathbf{a} \cdot \mathbf{n} dS \\ &= \int_B \operatorname{div}(\mathbf{a}) dV \\ &= \int_B d(*\omega). \end{aligned}$$

The pattern that emerges from the previous three subsections is that given an  $(n-1)$ -form,  $\alpha$ , on the  $(n-1)$ -dimensional boundary,  $\partial B$ , of an  $n$ -dimensional domain,  $B$ ,

$$\boxed{\int_{\partial B} \alpha = \int_B d\alpha.} \quad (6.101)$$

In some contexts  $\alpha$  is defined directly, and in others  $\alpha = *\omega$ , when  $\omega$  is a one-form.

The equality in (6.101) was not proved here, only observed repeatedly for  $n = 2, 3$ . The next chapter will sketch the proof for the more general case, and point to the literature for more complete treatments. In that discussion it will be important to understand how general  $k$ -forms transform under coordinate changes as a generalization of (6.94), which will be addressed at the end of this chapter. But first a connection between forms and diffusion equations is illustrated.

### 6.8.6 Detailed Examples

#### Example 1: Stokes' theorem and the Bi-Unit Cube

As an example of Stokes' theorem, consider the closed bi-unit block  $B = [-1, 1] \times [-1, 1] \times [-1, 1] \subset \mathbb{R}^3$ . The boundary of this body is the bi-unit cube,  $\partial B$ . While  $\partial B$  is



not a smooth surface, it can be viewed as the limit of a series of smooth *superquadric surfaces* of the form

$$x_1^{2n} + x_2^{2n} + x_3^{2n} = 1 \quad \text{where } \mathbb{Z} \ni n \rightarrow \infty.$$

The integral over  $B$  of the form  $d\gamma_2$  (where  $\gamma_2$  is defined in (6.88)) can be computed as

$$\int_B d\gamma_2 = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 (\sin x_2 - x_3 e^{-x_1}) dx_1 dx_2 dx_3 = 0.$$

The value of zero results because the functions  $f_1(x_2) = \sin x_2$  and  $f_2(x_3) = x_3$  are odd over the symmetric domains of integration  $-1 \leq x_2 \leq 1$  and  $-1 \leq x_3 \leq 1$ .

The surface  $\partial B$  consists of six planar faces that appear in pairs. The integral of  $\gamma_2$  over  $\partial B$  then can be written as three parts:

$$\int_{\partial B} \gamma_2 = c_1 + c_2 + c_3$$

where

$$\begin{aligned} c_1 &= \int_{-1}^1 \int_{-1}^1 x_3 e^{+1} (-dx_2 dx_3) + \int_{-1}^1 \int_{-1}^1 x_3 e^{-1} dx_2 dx_3 = 0 \\ c_2 &= \int_{-1}^1 \int_{-1}^1 \cos(-1) (-dx_1 dx_3) + \int_{-1}^1 \int_{-1}^1 \cos(+1) dx_1 dx_3 = 0 \\ c_3 &= \int_{-1}^1 \int_{-1}^1 x_1^2 \sin x_2 (-dx_1 dx_2) + \int_{-1}^1 \int_{-1}^1 x_1^2 \sin x_2 dx_1 dx_2 = 0. \end{aligned}$$

The negative signs on the differential area elements appear when evaluating  $dx_j \wedge dx_i$  as  $-dx_i \wedge dx_j = -dx_i dx_j$  under the integral. This happens for faces with outward normals pointing in negative coordinate directions. Each of the integrals in  $c_1$  and  $c_3$  happens to be zero due to the fact that the integrands are odd functions. However,  $c_2$  vanishes because the two integrands cancel as a result of the signed area elements and the evenness of the cosine function.

This example has demonstrated that  $\int_{\partial B} \gamma_2 = \int_B d\gamma_2$ , where  $B$  is a three-dimensional domain and  $\partial B$  is its two-dimensional boundary. The next example illustrates another case.

### Example 2: Stokes' Theorem in Curvilinear Coordinates

Consider a cylinder in  $\mathbb{R}^3$  defined by  $x_1^2 + x_2^2 = r_0^2$  and  $0 \leq z \leq h_0$  that has an open top and closed base (e.g., a coffee can). Call this surface  $C$  and let  $\partial C$  denote the counterclockwise-oriented circular rim at the top of this surface. The form  $\zeta^* \gamma_1$  is given in cylindrical coordinates in (6.90). If this is written as

$$\zeta^* \gamma_1 = a_r(r, \phi, z) dr + a_\phi(r, \phi, z) d\phi + a_z(r, \phi, z) dz,$$

then

$$\int_{\partial C} \zeta^* \gamma_1 = \int_0^{2\pi} a_\phi(r_0, \phi, h) d\phi$$

because  $r = r_0$  and  $z = h$  are constants, and so  $dr = dz = 0$ .

The evaluation of the other side of the equation in Stokes' theorem is evaluated as

$$\int_C d(\zeta^* \gamma_1) = \int_0^h \int_0^{2\pi} \left( \frac{\partial a_z}{\partial \phi} - \frac{\partial a_\phi}{\partial z} \right) \Big|_{r=r_0} d\phi dz + \int_0^{2\pi} \int_0^{r_0} \left( \frac{\partial a_\phi}{\partial r} - \frac{\partial a_r}{\partial \phi} \right) \Big|_{z=0} dr d\phi.$$

This can be simplified by observing that

$$\int_0^{2\pi} \frac{\partial a_z}{\partial \phi} d\phi = a_z(r, 2\pi, z) - a_z(r, 0, z) = 0$$

due to the continuity of the function  $a_z(\cdot)$  and the topology of the circle. In addition,

$$\int_0^h \frac{\partial a_\phi}{\partial z} dz = a_\phi(r, \phi, h) - a_\phi(r, \phi, 0)$$

and

$$\int_0^{r_0} \frac{\partial a_\phi}{\partial r} dr = a_\phi(r, \phi, z) - a_\phi(0, \phi, z).$$

Note that in the specific case of (6.90), the function  $a_\phi(0, \phi, z) = 0$  because of the factor of  $r$  that resulted from changing from Cartesian to polar coordinates.

Putting all of these facts together leads to the simplification

$$\begin{aligned} \int_C d(\zeta^* \gamma_1) &= \int_0^{2\pi} \{ [a_\phi(r_0, \phi, h) - a_\phi(r_0, \phi, 0)] + [a_\phi(r_0, \phi, 0) - a_\phi(0, \phi, 0)] \} d\phi \\ &= \int_0^{2\pi} a_\phi(r_0, \phi, h) d\phi. \end{aligned}$$

Therefore, Stokes' theorem has been demonstrated in the form

$$\int_C d(\zeta^* \gamma_1) = \int_{\partial C} \zeta^* \gamma_1.$$

## 6.8.7 Closed Forms and Diffusion Equations

Consider the diffusion equation

$$\frac{\partial f}{\partial t} = \Delta_{(K, \mathbf{v})} f \quad (6.102)$$

where

$$\Delta_{(K, \mathbf{v})} f = \frac{1}{2} \sum_{i,j=1}^n k_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} - \sum_{k=1}^n v_k \frac{\partial f}{\partial x_k}$$

where  $K = K^T \in \mathbb{R}^{n \times n}$  and  $\mathbf{v} \in \mathbb{R}^n$  are both constant quantities. Recall that (6.102) was examined extensively in Chapter 2.

Equipped with knowledge of differential forms in Euclidean space, it is possible to construct a form on the  $(n+1)$ -dimensional space-time domain  $D \subset \mathbb{R}^n \times \mathbb{R}_{>0}$ . In particular, let

$$\omega = * \left( \frac{1}{2} \sum_{i,j=1}^n k_{ij} \frac{\partial f}{\partial x_j} dx_j - \sum_{k=1}^n v_k f \right) \wedge dt - (-1)^n f dx_1 \wedge dx_2 \wedge \dots \wedge dx_n, \quad (6.103)$$

where  $*$  is the Hodge star operator for the spatial part of the domain.

Then, as was shown in [5] for the one-dimensional heat equation,

$$d\omega = \left( \Delta_{(K,\mathbf{v})}f - \frac{\partial f}{\partial t} \right) dx_1 \wedge dx_2 \wedge \dots \wedge dx_n \wedge dt = 0.$$

The last equality is due to (6.102). Whenever a form  $\omega$  has the property that  $d\omega = 0$  identically, then  $\omega$  is called a *closed form*. In contrast, if a form  $\alpha = d\beta$  where  $\beta$  is another form, then  $\alpha$  is called an *exact form*. An exact form is always closed because  $d\alpha = d(d\beta) = 0$ . However, not every closed form is exact.

For the particular form defined in (6.103), it follows from Stokes' theorem that for any  $(n + 1)$ -dimensional space-time domain  $D$ ,

$$\int_{\partial D} \omega = \int_D d\omega = 0.$$

Other differential forms can be constructed to elucidate properties of solutions of (6.102) through the use of Stokes' theorem as explained in [5] for the case when  $K = \mathbb{I}$  and  $\mathbf{v} = \mathbf{0}$ .

### 6.9 Differential Forms and Coordinate Changes

Let  $\mathbf{x}$  and  $\mathbf{y}$  denote positions in  $\mathbb{R}^n$ . They can be related by a smooth mapping  $\mathbf{y} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $\mathbf{y} = \mathbf{y}(\mathbf{x})$ . If the values of  $\mathbf{x}$  are restricted so as to sweep out a finite  $n$ -dimensional volume,  $N \subset \mathbb{R}^n$ , then as  $\mathbf{x} \in N$  is evaluated under the function, the result will be the finite volume  $\mathbf{y}(N) \subset \mathbb{R}^n$ . Since  $\mathbf{x}$  and  $\mathbf{y}$  have the same dimensions, the Jacobian matrix  $\partial\mathbf{y}/\partial\mathbf{x}^T$  is square, and the classical inverse function theorem applies.

Given a differential  $k$ -form in one set of coordinates, it is possible to express the same form in the other set of coordinates. To start, let  $A^T = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n] \in \mathbb{R}^{n \times n}$  (i.e., the  $i$ th row of  $A$  is  $\mathbf{a}_i^T$ ). Then each  $\mathbf{a}_i \cdot d\mathbf{x} = \mathbf{a}_i^T d\mathbf{x}$  is a one-form. It can be shown that the form resulting from the substitution  $dx_i \rightarrow \mathbf{a}_i^T d\mathbf{x}$  (which is equivalent to  $\mathbf{x} \rightarrow A\mathbf{x}$ ) is

$$(\mathbf{a}_1^T d\mathbf{x}) \wedge \dots \wedge (\mathbf{a}_k^T d\mathbf{x}) = \sum_{1 \leq i_1 < \dots < i_k \leq n} \begin{vmatrix} \alpha_{1,i_1} & \alpha_{1,i_2} & \dots & \alpha_{1,i_k} \\ \alpha_{2,i_1} & \alpha_{2,i_2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \alpha_{k-1,i_k} \\ \alpha_{k,i_1} & \dots & \alpha_{k,i_{k-1}} & \alpha_{k,i_k} \end{vmatrix} dx_{i_1} \wedge \dots \wedge dx_{i_k} \tag{6.104}$$

where here  $\alpha_{k,i} \doteq \mathbf{a}_k \cdot \mathbf{e}_i = a_{ik}$ . This relationship for the linear transformation  $\mathbf{x} \rightarrow A\mathbf{x} \doteq \mathbf{y}$  can be used to build up the way that a  $k$ -form transforms under coordinate change. In particular, if  $\mathbf{y} = \psi(\mathbf{x})$  is now a non-linear change of coordinates, then the expression analogous to (6.104) is

$$dy_{j_1} \wedge \dots \wedge dy_{j_k} = \sum_{1 \leq i_1 < \dots < i_k \leq n} \begin{vmatrix} \frac{\partial \psi_{j_1}}{\partial x_{i_1}} & \frac{\partial \psi_{j_1}}{\partial x_{i_2}} & \dots & \frac{\partial \psi_{j_1}}{\partial x_{i_k}} \\ \frac{\partial \psi_{j_2}}{\partial x_{i_1}} & \frac{\partial \psi_{j_2}}{\partial x_{i_2}} & \dots & \vdots \\ \vdots & \vdots & \ddots & \frac{\partial \psi_{j_{k-1}}}{\partial x_{i_k}} \\ \frac{\partial \psi_{j_k}}{\partial x_{i_1}} & \dots & \frac{\partial \psi_{j_k}}{\partial x_{i_{k-1}}} & \frac{\partial \psi_{j_k}}{\partial x_{i_k}} \end{vmatrix} dx_{i_1} \wedge \dots \wedge dx_{i_k} \tag{6.105}$$

where of course  $dy_{j_k} = \mathbf{e}_{j_k}^T d\mathbf{y}$ .

Therefore, if

$$\omega_k = \sum_{1 \leq j_1 < \dots < j_k \leq n} a_{j_1, \dots, j_k}(\mathbf{y}) dy_{j_1} \wedge dy_{j_2} \wedge \dots \wedge dy_{j_k},$$

then in the new set of coordinates,

$$\begin{aligned} \psi^* \omega_k &= \sum_{1 \leq j_1 < \dots < j_k \leq n} a_{j_1, \dots, j_k}(\psi(\mathbf{x})) \\ &\sum_{1 \leq i_1 < \dots < i_k \leq n} \begin{vmatrix} \frac{\partial \psi_{j_1}}{\partial x_{i_1}} & \frac{\partial \psi_{j_1}}{\partial x_{i_2}} & \dots & \frac{\partial \psi_{j_1}}{\partial x_{i_k}} \\ \frac{\partial \psi_{j_2}}{\partial x_{i_1}} & \frac{\partial \psi_{j_2}}{\partial x_{i_2}} & \dots & \vdots \\ \vdots & \vdots & \ddots & \frac{\partial \psi_{j_{k-1}}}{\partial x_{i_k}} \\ \frac{\partial \psi_{j_k}}{\partial x_{i_1}} & \dots & \frac{\partial \psi_{j_k}}{\partial x_{i_{k-1}}} & \frac{\partial \psi_{j_k}}{\partial x_{i_k}} \end{vmatrix} dx_{i_1} \wedge \dots \wedge dx_{i_k}. \end{aligned} \quad (6.106)$$

In the special case when  $k = n$ ,

$$\omega_n = a(\mathbf{y}) dy_1 \wedge dy_2 \wedge \dots \wedge dy_n$$

and (6.106) simplifies to

$$\psi^* \omega_n = a(\psi(\mathbf{x})) |D\psi| dx_1 \wedge dx_2 \wedge \dots \wedge dx_n, \quad (6.107)$$

which is the result from Section 6.8.1.

The expression in (6.106) will be particularly useful when it comes to writing the integral theorems discussed in the previous section in different curvilinear coordinate systems. It will also be useful when discussing parameterized  $m$ -dimensional embedded manifolds<sup>10</sup> in  $\mathbb{R}^n$ .

## 6.10 Chapter Summary

This chapter has served as an introduction to differential forms and multi-vectors. Multi-vectors are vectors in a vector space  $\Lambda^p(V)$ , which is generated by performing the  $p$ -fold wedge product of vectors drawn from a vector space  $V$ . The dual space of  $V$ , denoted as  $V^*$ , is the space of forms (i.e., functions  $\varphi : V \rightarrow \mathbb{R}$ ). The wedge product of the space of forms can be defined in such a way that  $\Lambda^p(V^*) = (\Lambda^p V)^*$ . The exterior derivative of a form gives a differential form. Simple rules define how to directly compute the exterior derivatives of these differential forms. And these rules make differential forms an ideal tool for extending classical theorems of multivariable calculus, such as Stokes' theorem.

This chapter has covered the basics of differential forms. Other accessible treatments include Darling [3] and Schreiber [13]. In fact, this chapter was modeled after the presentations in those works. Sometimes it is easier to understand a mathematical concept by seeing it used in practice. For more on applications of differential forms in the “real world” see [10].

The next chapter applies the concept of differential forms beyond how they were used here. In particular, the curvature and torsion of an  $m$ -dimensional “manifold” (i.e., higher-dimensional analog of a simple curve or surface) in  $\mathbb{R}^n$  is defined in a very natural way using differential forms. And Stokes' theorem extends in a very elegant way to manifolds when stated in terms of differential forms.

<sup>10</sup>Think of these as  $m$ -dimensional surfaces.

## 6.11 Exercises

6.1. Using the defining properties in (6.5)–(6.7), show that

$$dx_i \wedge \left( \sum_j g_j dx_j \right) = \sum_j g_j \cdot (dx_i \wedge dx_j)$$

where  $\cdot$  just means scalar multiplication.

6.2. Starting with the definition in (6.9), and using the properties of the wedge product,  $\wedge$ , determine  $c(\mathbf{x})$  as a function of  $c_{ijk}(\mathbf{x})$  such that  $\omega_3 = c(\mathbf{x}) dx_1 \wedge dx_2 \wedge dx_3$  in the special case when  $n = 3$ .

6.3. Prove the following: (a) that (6.17) holds; (b) given an arbitrary 1-form in  $\mathbb{R}^3$ , denoted as  $\omega_1$ , verify (6.18) and that  $d(d\omega_1) = 0$ .

6.4. Using the defining properties in (6.5)–(6.7), show that

$$\left( \sum_{i=1}^n f_i dx_i \right) \wedge \left( \sum_{j=1}^n g_j dx_j \right) = \sum_{1 \leq i < j \leq n} (f_i g_j - f_j g_i) dx_i \wedge dx_j.$$

6.5. Show that

$$\mathbf{v} \wedge \mathbf{w} = \sum_{i,j=1}^n v_i w_j \mathbf{e}_i \wedge \mathbf{e}_j = \sum_{i < j} (v_i w_j - v_j w_i) \mathbf{e}_i \wedge \mathbf{e}_j,$$

and from (A.14) the magnitude of  $\mathbf{v} \wedge \mathbf{w}$  (viewed as a column vector of dimension  $n(n-1)/2$ ) satisfies

$$\|\mathbf{v} \wedge \mathbf{w}\|^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 - (\mathbf{v} \cdot \mathbf{w})^2.$$

6.6. Prove both equalities in (6.24).

6.7. Verify (6.26) when  $n = 2, 3, 4$ .

6.8. Verify (6.49) when  $n = 2, 3, 4$ .

6.9. Let  $\omega_p$  and  $\alpha_q$  respectively be differential  $p$ - and  $q$ -forms in  $\mathbb{R}^n$ . Work out  $\omega_p \wedge \alpha_q$  for the following cases: (a)  $p = 1, q = 1, n = 2$ ; (b)  $p = 1, q = 2, n = 3$ ; (c)  $p = 1, q = 2, n = 4$ ; (d)  $p = 2, q = 2, n = 5$ .

6.10. Again let  $\omega_p$  and  $\alpha_q$  respectively be differential  $p$ - and  $q$ -forms in  $\mathbb{R}^n$ . Prove that

$$\omega_p \wedge \alpha_q = (-1)^{pq} \alpha_q \wedge \omega_p. \quad (6.108)$$

6.11. Show that any permutation  $\pi \in \Pi_n$  has an inverse, and that the associative law holds for permutations, i.e.,  $(\pi_1 \circ \pi_2) \circ \pi_3 = \pi_1 \circ (\pi_2 \circ \pi_3)$ .

6.12. Let  $V = \mathbb{R}^3$  and  $A \in \mathbb{R}^{3 \times 3}$ . Verify that  $A^2(AB) = A^2(A)A^2(B)$  and  $A^3(AB) = A^3(A)A^3(B)$ . Hint: For the first part of the problem order the basis elements of  $A^2(\mathbb{R}^3)$  as  $\mathbf{e}_1 \wedge \mathbf{e}_2, \mathbf{e}_1 \wedge \mathbf{e}_3, \mathbf{e}_2 \wedge \mathbf{e}_3$  and identify  $\mathbf{x} \wedge \mathbf{y}$  with the column vector via a  $\vee$  operation defined by

$$(\mathbf{x} \wedge \mathbf{y})^\vee \doteq [x_1y_2 - x_2y_1, x_1y_3 - x_3y_1, x_2y_3 - x_3y_2]^T.$$

Then the unique matrix  $\Lambda^2(A)$  that satisfies

$$\Lambda^2(A)(\mathbf{x} \wedge \mathbf{y})^\vee = ((A\mathbf{x}) \wedge (A\mathbf{y}))^\vee$$

is

$$\Lambda^2(A) = \begin{pmatrix} a_{11}a_{22} - a_{21}a_{12} & a_{11}a_{23} - a_{21}a_{13} & a_{12}a_{23} - a_{22}a_{13} \\ a_{11}a_{32} - a_{31}a_{12} & a_{11}a_{33} - a_{31}a_{13} & a_{12}a_{33} - a_{32}a_{13} \\ a_{21}a_{32} - a_{31}a_{22} & a_{21}a_{33} - a_{31}a_{23} & a_{22}a_{33} - a_{32}a_{23} \end{pmatrix}. \quad (6.109)$$

6.13. Let  $V = \mathbb{R}^4$  and  $A \in \mathbb{R}^{4 \times 4}$ . Using the lexicographical ordering of basis elements of  $\Lambda^2(\mathbb{R}^4)$ :  $\mathbf{e}_1 \wedge \mathbf{e}_2$ ,  $\mathbf{e}_1 \wedge \mathbf{e}_3$ ,  $\mathbf{e}_1 \wedge \mathbf{e}_4$ ,  $\mathbf{e}_2 \wedge \mathbf{e}_3$ ,  $\mathbf{e}_2 \wedge \mathbf{e}_4$ ,  $\mathbf{e}_3 \wedge \mathbf{e}_4$  and the  $\vee$  operation defined by

$$(\mathbf{x} \wedge \mathbf{y})^\vee \doteq \begin{pmatrix} x_1y_2 - x_2y_1 \\ x_1y_3 - x_3y_1 \\ x_1y_4 - x_4y_1 \\ x_2y_3 - x_3y_2 \\ x_2y_4 - x_4y_2 \\ x_3y_4 - x_4y_3 \end{pmatrix},$$

show that the resulting matrix  $\Lambda^2(A)$  has entries:

$$\begin{aligned} \Lambda^2(A)_{11} &= a_{11}a_{22} - a_{12}a_{21}; & \Lambda^2(A)_{12} &= a_{11}a_{23} - a_{13}a_{21}; & \Lambda^2(A)_{13} &= a_{11}a_{24} - a_{14}a_{21}; \\ \Lambda^2(A)_{14} &= a_{12}a_{23} - a_{13}a_{22}; & \Lambda^2(A)_{15} &= a_{12}a_{24} - a_{14}a_{22}; & \Lambda^2(A)_{16} &= a_{13}a_{24} - a_{14}a_{23}; \\ \Lambda^2(A)_{21} &= a_{11}a_{32} - a_{12}a_{31}; & \Lambda^2(A)_{22} &= a_{11}a_{33} - a_{13}a_{31}; & \Lambda^2(A)_{23} &= a_{11}a_{34} - a_{14}a_{31}; \\ \Lambda^2(A)_{24} &= a_{12}a_{33} - a_{13}a_{32}; & \Lambda^2(A)_{25} &= a_{12}a_{34} - a_{14}a_{32}; & \Lambda^2(A)_{26} &= a_{13}a_{34} - a_{14}a_{33}; \\ \Lambda^2(A)_{31} &= a_{11}a_{42} - a_{12}a_{41}; & \Lambda^2(A)_{32} &= a_{11}a_{43} - a_{13}a_{41}; & \Lambda^2(A)_{33} &= a_{11}a_{44} - a_{14}a_{41}; \\ \Lambda^2(A)_{34} &= a_{12}a_{43} - a_{13}a_{42}; & \Lambda^2(A)_{35} &= a_{12}a_{44} - a_{14}a_{42}; & \Lambda^2(A)_{36} &= a_{13}a_{44} - a_{14}a_{43}; \\ \Lambda^2(A)_{41} &= a_{21}a_{32} - a_{22}a_{31}; & \Lambda^2(A)_{42} &= a_{21}a_{33} - a_{23}a_{31}; & \Lambda^2(A)_{43} &= a_{21}a_{34} - a_{24}a_{31}; \\ \Lambda^2(A)_{44} &= a_{22}a_{33} - a_{23}a_{32}; & \Lambda^2(A)_{45} &= a_{22}a_{34} - a_{24}a_{32}; & \Lambda^2(A)_{46} &= a_{23}a_{34} - a_{24}a_{33}; \\ \Lambda^2(A)_{51} &= a_{21}a_{42} - a_{22}a_{41}; & \Lambda^2(A)_{52} &= a_{21}a_{43} - a_{23}a_{41}; & \Lambda^2(A)_{53} &= a_{21}a_{44} - a_{24}a_{41}; \\ \Lambda^2(A)_{54} &= a_{22}a_{43} - a_{23}a_{42}; & \Lambda^2(A)_{55} &= a_{22}a_{44} - a_{24}a_{42}; & \Lambda^2(A)_{56} &= a_{23}a_{44} - a_{24}a_{43}; \\ \Lambda^2(A)_{61} &= a_{31}a_{42} - a_{32}a_{41}; & \Lambda^2(A)_{62} &= a_{31}a_{43} - a_{33}a_{41}; & \Lambda^2(A)_{63} &= a_{31}a_{44} - a_{34}a_{41}; \\ \Lambda^2(A)_{64} &= a_{32}a_{43} - a_{33}a_{42}; & \Lambda^2(A)_{65} &= a_{32}a_{44} - a_{34}a_{42}; & \Lambda^2(A)_{66} &= a_{33}a_{44} - a_{34}a_{43}. \end{aligned}$$

6.14. Again let  $V = \mathbb{R}^4$  and  $A \in \mathbb{R}^{4 \times 4}$ . This time compute  $\Lambda^3(A)$ . Using the lexicographical ordering of basis elements of  $\Lambda^3(\mathbb{R}^4)$ :  $\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$ ,  $\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_4$ ,  $\mathbf{e}_1 \wedge \mathbf{e}_3 \wedge \mathbf{e}_4$ ,  $\mathbf{e}_2 \wedge \mathbf{e}_3 \wedge \mathbf{e}_4$ , and the  $\vee$  operation defined by

$$(\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z})^\vee \doteq \begin{pmatrix} (x_2y_3 - x_3y_2)z_1 - (x_1y_3 - x_3y_1)z_2 + (x_1y_2 - x_2y_1)z_3 \\ (x_2y_4 - x_4y_2)z_1 - (x_1y_4 - x_4y_1)z_2 + (x_1y_2 - x_2y_1)z_4 \\ (x_3y_4 - x_4y_3)z_1 - (x_1y_4 - x_4y_1)z_3 + (x_1y_3 - x_3y_1)z_4 \\ (x_3y_4 - x_4y_3)z_2 - (x_2y_4 - x_4y_2)z_3 + (x_2y_3 - x_3y_2)z_4 \end{pmatrix},$$

show that the unique matrix  $\Lambda^2(A)$  that satisfies

$$\Lambda^3(A)(\mathbf{x} \wedge \mathbf{y} \wedge \mathbf{z})^\vee = ((A\mathbf{x}) \wedge (A\mathbf{y}) \wedge (A\mathbf{z}))^\vee$$

has entries:

$$\Lambda^3(A)_{11} = (a_{22}a_{33} - a_{32}a_{23})a_{11} - (a_{12}a_{33} - a_{32}a_{13})a_{21} + (a_{12}a_{23} - a_{22}a_{13})a_{31};$$

$$\Lambda^3(A)_{12} = (a_{22}a_{34} - a_{32}a_{24})a_{11} - (a_{12}a_{34} - a_{32}a_{14})a_{21} + (a_{12}a_{24} - a_{22}a_{14})a_{31};$$

$$\Lambda^3(A)_{13} = (a_{23}a_{34} - a_{33}a_{24})a_{11} - (a_{13}a_{34} - a_{33}a_{14})a_{21} + (a_{13}a_{24} - a_{23}a_{14})a_{31};$$

$$\Lambda^3(A)_{14} = (a_{23}a_{34} - a_{33}a_{24})a_{12} - (a_{13}a_{34} - a_{33}a_{14})a_{22} + (a_{13}a_{24} - a_{23}a_{14})a_{32};$$

$$\Lambda^3(A)_{21} = (a_{22}a_{43} - a_{42}a_{23})a_{11} - (a_{12}a_{43} - a_{42}a_{13})a_{21} + (a_{12}a_{23} - a_{22}a_{13})a_{41};$$

$$\Lambda^3(A)_{22} = (a_{22}a_{44} - a_{42}a_{24})a_{11} - (a_{12}a_{44} - a_{42}a_{14})a_{21} + (a_{12}a_{24} - a_{22}a_{14})a_{41};$$

$$\Lambda^3(A)_{23} = (a_{23}a_{44} - a_{43}a_{24})a_{11} - (a_{13}a_{44} - a_{43}a_{14})a_{21} + (a_{13}a_{24} - a_{23}a_{14})a_{41};$$

$$\Lambda^3(A)_{24} = (a_{23}a_{44} - a_{43}a_{24})a_{12} - (a_{13}a_{44} - a_{43}a_{14})a_{22} + (a_{13}a_{24} - a_{23}a_{14})a_{42};$$

$$\Lambda^3(A)_{31} = (a_{32}a_{43} - a_{42}a_{33})a_{11} - (a_{12}a_{43} - a_{42}a_{13})a_{31} + (a_{12}a_{33} - a_{32}a_{13})a_{41};$$

$$\Lambda^3(A)_{32} = (a_{32}a_{44} - a_{42}a_{34})a_{11} - (a_{12}a_{44} - a_{42}a_{14})a_{31} + (a_{12}a_{34} - a_{32}a_{14})a_{41};$$

$$\Lambda^3(A)_{33} = (a_{33}a_{44} - a_{43}a_{34})a_{11} - (a_{13}a_{44} - a_{43}a_{14})a_{31} + (a_{13}a_{34} - a_{33}a_{14})a_{41};$$

$$\Lambda^3(A)_{34} = (a_{33}a_{44} - a_{43}a_{34})a_{12} - (a_{13}a_{44} - a_{43}a_{14})a_{32} + (a_{13}a_{34} - a_{33}a_{14})a_{42};$$

$$\Lambda^3(A)_{41} = (a_{32}a_{43} - a_{42}a_{33})a_{21} - (a_{22}a_{43} - a_{42}a_{23})a_{31} + (a_{22}a_{33} - a_{32}a_{23})a_{41};$$

$$\Lambda^3(A)_{42} = (a_{32}a_{44} - a_{42}a_{34})a_{21} - (a_{22}a_{44} - a_{42}a_{24})a_{31} + (a_{22}a_{34} - a_{32}a_{24})a_{41};$$

$$\Lambda^3(A)_{43} = (a_{33}a_{44} - a_{43}a_{34})a_{21} - (a_{23}a_{44} - a_{43}a_{24})a_{31} + (a_{23}a_{34} - a_{33}a_{24})a_{41};$$

$$\Lambda^3(A)_{44} = (a_{33}a_{44} - a_{43}a_{34})a_{22} - (a_{23}a_{44} - a_{43}a_{24})a_{32} + (a_{23}a_{34} - a_{33}a_{24})a_{42}.$$

6.15. Compare the determinants of  $A$ ,  $\Lambda^2(A)$ , and  $\Lambda^3(A)$  in the previous three problems. Is there a general pattern for the determinant of  $\Lambda^p(A)$  for  $A \in \mathbb{R}^{n \times n}$  where  $n \geq p$ ?

6.16. Using only the defining properties of a wedge product, show that (6.40) holds. How is  $W$  related to a Lie algebra?

6.17. Let  $\mathbf{v}_i, \mathbf{w}_i \in \mathbb{R}^n$  for  $i = 1, \dots, p \leq n$  and  $\pi \in \Pi_n$ . Using the definition in (6.56) where  $\varphi_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x}$ , together with the multi-linearity of the wedge product, prove the *Lagrange identity* [6]

$$\det \begin{bmatrix} \mathbf{v}_1 \cdot \mathbf{w}_1 & \mathbf{v}_2 \cdot \mathbf{w}_1 & \cdots & \mathbf{v}_p \cdot \mathbf{w}_1 \\ \mathbf{v}_1 \cdot \mathbf{w}_2 & \mathbf{v}_2 \cdot \mathbf{w}_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_1 \cdot \mathbf{w}_p & \mathbf{v}_2 \cdot \mathbf{w}_p & \cdots & \mathbf{v}_p \cdot \mathbf{w}_p \end{bmatrix} = \sum_{\pi \in \Pi_n \mid \pi(1) < \cdots < \pi(p)} \det[\mathbf{w}_i \cdot \mathbf{e}_{\pi(j)}] \det[\mathbf{v}_i \cdot \mathbf{e}_{\pi(j)}] \quad (6.110)$$

where  $[a_{ij}]$  denotes the  $p \times p$  matrix with entries  $a_{ij}$ .

6.18. Rewrite the “light” expressions in (6.64) and (6.65) in the “heavy” (coordinate-dependent) way that they would appear using classical multivariable calculus and Jacobian matrices.

6.19. Let  $\psi(\mathbf{x})$  be the non-linear shear transformation

$$\psi(\mathbf{x}) = \begin{pmatrix} x_1 + s_1(x_2, x_3) \\ x_2 + s_2(x_3) \\ x_3 \end{pmatrix}$$

where  $s_1(x_2, x_3)$  and  $s_2(x_3)$  are both smooth functions. Let  $\mathbf{y} = \psi(\mathbf{x})$  and define the vector field

$$\mathbf{X}(\mathbf{y}) = \begin{pmatrix} y_2^2 + y_3^2 \\ y_1^2 + y_3^2 \\ y_1^2 + y_3^2 \end{pmatrix}.$$

Explicitly, what are  $\psi_*X$  and  $\psi^*\omega$  where  $\omega$  is the 1-form  $\omega = d\varphi$  and  $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}$  is defined as  $\varphi(\mathbf{x}) = x_1^2 + 2x_1x_2 + x_3$ ?

6.20. Verify for  $n = 1$ ,  $n = 2$ , and  $n = 3$  that  $\omega$  in (6.103) is a closed form when  $f(\mathbf{x}, t)$  solves the diffusion equation in (6.102).

6.21. Prove that the pull-back is linear. That is, given two generic  $k$ -forms on  $\mathbb{R}^n$ , denoted as  $\omega$  and  $\alpha$  in curvilinear coordinates,  $\mathbf{q}$ , and if  $\mathbf{q} = \psi(\mathbf{s})$ , then the pull-back of the linear combination is the linear combination of the pull-backs:

$$\boxed{\psi^*(c_1\omega + c_2\alpha) = c_1\psi^*\omega + c_2\psi^*\alpha} \quad (6.111)$$

for any  $c_1, c_2 \in \mathbb{R}$ .

## References

1. Abraham, R., Marsden, J.E., Ratiu, T., *Manifolds, Tensor Analysis, and Applications*, 2nd ed., Applied Mathematical Sciences 75, Springer, New York, 1988.
2. Cartan, H., *Differential Forms*, Hermann, Paris; Houghton Mifflin, Boston, 1970.
3. Darling, R.W.R., *Differential Forms and Connections*, Cambridge University Press, London, 1994.
4. do Carmo, M.P., *Differential Forms and Applications*, Springer-Verlag, Berlin, 1994.
5. Flanders, H., *Differential Forms with Applications to the Physical Sciences*, Dover, New York, 1989.
6. Greub, W.H., *Multilinear Algebra*, 2nd ed., Universitext, Springer-Verlag, Berlin, 1987.
7. Guggenheimer, H.W., *Differential Geometry*, Dover, New York, 1977.
8. Guillemin, V., Pollack, A., *Differential Topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
9. Lang, S., *Fundamentals of Differential Geometry*, Springer, New York, 1999.
10. Lindell, I.V., *Differential Forms in Electromagnetics*, IEEE Press/Wiley-Interscience, New York, 2004.
11. Lovelock, D., Rund, H., *Tensors, Differential Forms, and Variational Principles*, Dover, New York, 1989.
12. Mukherjee, A., *Topics in Differential Topology*, Hindustan Book Agency, New Delhi, 2005.
13. Schreiber, M., *Differential Forms: A Heuristic Introduction*, Universitext, Springer-Verlag, New York, 1977.
14. Tu, L.W., *An Introduction to Manifolds*, Springer, New York, 2008.
15. Warner, F.W., *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, New York, 1983.



---

## Polytopes and Manifolds

This chapter extends the review of geometrical ideas from the previous chapters to include geometrical objects in higher dimensions. These include hyper-surfaces and “hyper-polyhedra” (or *polytopes*) in  $\mathbb{R}^n$ . A parametric description of an  $m$ -dimensional *embedded manifold*<sup>1</sup> in an  $n$ -dimensional Euclidean space is of the form  $\mathbf{x} = \mathbf{x}(\mathbf{q})$  where  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{q} \in \mathbb{R}^m$  with  $m \leq n$ . If  $m = n - 1$ , then this is called a *hyper-surface*. An implicit description of an  $m$ -dimensional embedded manifold in  $\mathbb{R}^n$  is a system of constraint equations of the form  $\phi_i(\mathbf{x}) = 0$  for  $i = 1, \dots, n - m$ . In the context of engineering applications, the two most important differences between the study of two-dimensional surfaces in  $\mathbb{R}^3$  and  $m$ -dimensional embedded manifolds in  $\mathbb{R}^n$  are: (1) there is no cross-product operation for  $\mathbb{R}^n$ ; and (2) if  $m \ll n$ , it can be more convenient to leave behind  $\mathbb{R}^n$  and describe the manifold intrinsically. For these reasons, modern mathematical concepts such as differential forms and coordinate-free differential geometry can be quite powerful.

Section 7.1 discusses some properties of non-differentiable geometric objects such as polyhedra in three-dimensional space, and extends these ideas to higher dimensions. Section 7.2 discusses several examples of manifolds that arise in applications. Section 7.3 extends concepts from the parametric treatment of differential geometry in three dimensions to  $n$ -dimensional Euclidean space. Section 7.5 illustrates how differential forms can be used to simplify calculations associated with embedded manifolds. Section 7.6 applies differential forms to the coordinate-free treatment of manifolds, including the generalized definition of curvature and extensions of the Gauss–Bonnet theorem. Section 7.7 provides a brief introduction to fiber bundles and connections. Section 7.8 discusses the heat equation on a manifold. some exercises involving calculations on manifolds.

The main points to take away from this chapter are:

- Higher-dimensional versions of polygons and polyhedra are called polytopes. A product of polytopes, called the Minkowski sum, produces new polytopes from old ones.
- The concepts of simple planar or spatial curves and simply connected surfaces in  $\mathbb{R}^3$  extend to higher dimensions and are examples of more general mathematical structures called manifolds.
- Sometimes it is natural to treat these geometric objects as “living in” a higher dimensional Euclidean space, and sometimes it is more natural to use purely intrinsic approaches.

---

<sup>1</sup>For now, think of this as a smooth simple surface that does not self-intersect. The word “embedded” means that the manifold is “seated in”  $\mathbb{R}^n$  in a way that will be made more precise later.

- Tools for handling intrinsic geometry exist, including formulas for the curvature of a manifold.
- Integral theorems, such as the extension of the Gauss–Bonnet theorem and Stokes’ theorem, hold for manifolds and polytopes, and explicit computations can be performed using the methods of this chapter.

## 7.1 Properties and Operations on Convex Polytopes in $\mathbb{R}^n$

In the context of this discussion, a *polytope* is a closed and bounded geometric hyper-surface that encapsulates a finite volume, and the shape of which is defined by a finite number of intersecting hyper-planes. In three-dimensional space, this amounts to a surface with flat faces and straight edges (a polyhedron), and in the plane this becomes a polygon.

For polyhedra in three-dimensional space the quantities  $F$ ,  $V$ , and  $M$  can be computed regardless of the fact that polyhedra do not satisfy the differentiability requirements assumed in Section 5.4. This extends to polytopes in higher dimensions.

The simplest polytopes are those that are *convex*. A convex region  $C \subset \mathbb{R}^n$  is one in which every pair of points contained in the region can be connected with a line segment, every point of which is contained within the region. In other words, if  $\mathbf{x}, \mathbf{y} \in C \subset \mathbb{R}^n$ , then  $(1 - \lambda)\mathbf{x} + \lambda\mathbf{y} \in C$  for all  $\lambda \in [0, 1]$ .

The concept of a convex polytope is closely related to that of a multi-dimensional convex function, that extends the definition in (3.21). Namely, a convex function  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$  is one for which

$$\boxed{\Phi(t\mathbf{x} + (1 - t)\mathbf{y}) \leq t\Phi(\mathbf{x}) + (1 - t)\Phi(\mathbf{y}) \quad \forall t \in [0, 1]} \quad (7.1)$$

for any  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ . This is a direct generalization of the one-dimensional case in (3.21). The graph of such a function in  $\mathbb{R}^{n+1}$ ,  $z = \Phi(\mathbf{x})$  is a convex (semi-infinite) body. The intersection of a convex body with a hyper-plane results in a convex body. As a special case, when  $z$  is fixed as a constant,  $c$ , the hyper-surface in  $\mathbb{R}^n$  defined by  $\Phi(\mathbf{x}) = c$  is a convex hyper-surface, and the implicitly defined body  $\Phi(\mathbf{x}) \leq c$  is convex. However, such a body need not be a polytope. For example, it could be the  $n$ -dimensional volume bounded by a hyper-ellipsoid.

In contrast, a polytope  $C \subset \mathbb{R}^n$  can be constructed by intersecting many half spaces, which are rotated and translated copies of the space  $\mathbf{x} \cdot \mathbf{e}_n \geq 0$ . Doing this  $m$  times generates an inequality of the form

$$A\mathbf{x} \leq \mathbf{b},$$

which is interpreted as  $m$  scalar inequality constraints where  $A \in \mathbb{R}^{m \times n}$  and  $\mathbf{b} \in \mathbb{R}^m$ . Computer algorithms are available that compute the vertices of these polytopes given the equations of the half spaces that define them [24]. In some cases the result is not a finite body, in which case it is sometimes called a “polyhedral cone” where the word “polyhedron” is used in a generalized sense in place of the word “polytope.”

In what follows, it is assumed that the polytopes that are provided bound bodies of finite volume.

### 7.1.1 Computing the Volume and Surface Area of Polyhedra

Consider a polygon in the plane defined by vertices  $\mathbf{x}_i \in \mathbb{R}^2$  for  $i = 1, \dots, k$ . These vertices are connected pairwise with line segments (or edges) from  $\mathbf{x}_j$  to  $\mathbf{x}_{j+1}$  for  $j = 1, \dots, k-1$ , and  $\mathbf{x}_k$  to  $\mathbf{x}_1$ . Let us assume that the points are arranged such that edges only intersect at vertices.

The perimeter of the polygon can then be calculated by simply summing up the length of all of the edges:

$$L = \sum_{i=1}^k \|\mathbf{x}_{i+1} - \mathbf{x}_i\|$$

with  $\mathbf{x}_{k+1} = \mathbf{x}_1$ . The area of a polygon can be computed by subdividing it into triangles, using the area formula for each triangle, and summing over all of the triangles.

The surface area of a polyhedron can be computed if the positions of all of the vertices are known. This amounts to nothing more than adding up the areas of each of the polygonal faces.

The volume of a polyhedron can be computed using a discrete version of the divergence theorem.

### 7.1.2 Properties of Minkowski Sums

Given a convex body  $C \subset \mathbb{R}^n$ , let  $V(C)$  denote its volume. Re-scaling  $C$  as

$$\lambda \cdot C \doteq \{\lambda \mathbf{x} \mid \mathbf{x} \in C \subset \mathbb{R}^n, \lambda \in \mathbb{R}_{>0}\}$$

clearly does not change its convexity. However, the volume of the re-scaled body will be  $V(\lambda \cdot C) = \lambda^n V(C)$ .

Let us denote a translated copy of the body  $C$  by vector  $\mathbf{t} \in \mathbb{R}^n$  as

$$\mathbf{t} + C \doteq \{\mathbf{x} + \mathbf{t} \mid \mathbf{x} \in C\}.$$

Then  $V(\mathbf{t} + C) = V(C)$ . Likewise, if  $A \in \mathbb{R}^{n \times n}$  with  $\det A > 0$ , define

$$A \cdot C \doteq \{A\mathbf{x} \mid \mathbf{x} \in C\}.$$

It follows that  $V(A \cdot C) = |\det A|^n V(C)$ .

As is the case always, rigid-body motions do not change the volume, surface area, or total curvatures (Gaussian and mean) of a body.

Given two convex bodies,  $C_1$  and  $C_2$ , the *Minkowski sum* of the bodies is defined as

$$\boxed{C_1 + C_2 \doteq \{\mathbf{x}_1 + \mathbf{x}_2 \mid \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2\}}. \quad (7.2)$$

This results in a new convex body. Clearly, since vector addition is commutative and associative,

$$C_1 + C_2 = C_2 + C_1 \quad \text{and} \quad (C_1 + C_2) + C_3 = C_1 + (C_2 + C_3),$$

and so we can write the Minkowski sum of  $k$  bodies simply as  $C_1 + C_2 + \dots + C_k$ .

An interesting and important result is the *Brunn–Minkowski* inequality [25, 26]

$$V(C_1 + C_2)^{\frac{1}{n}} \geq V(C_1)^{\frac{1}{n}} + V(C_2)^{\frac{1}{n}}. \quad (7.3)$$

Equality only holds when  $C_1$  and  $C_2$  are scaled versions of each other, in which case they are called *homothetic*.

A very important fact is that for  $\lambda_i \in \mathbb{R}_{>0}$  [25, 26],

$$V(\lambda_1 \cdot C_1 + \dots + \lambda_k \cdot C_k) = \sum_{1 \leq i_1, \dots, i_n \leq k} \lambda_{i_1} \cdots \lambda_{i_n} \nu(C_{i_1}, \dots, C_{i_n}) \tag{7.4}$$

is a homogeneous polynomial in  $\lambda_1, \dots, \lambda_k$ . The functions  $\nu(C_{i_1}, \dots, C_{i_n})$  are called *mixed volumes*, and have the following properties [30, 70]:

$$\nu(C_{i_1}, \dots, C_{i_r}, \dots, C_{i_s}, \dots, C_{i_r}, \dots, C_{i_n}) = \nu(C_{i_1}, \dots, C_{i_s}, \dots, C_{i_r}, \dots, C_{i_n}) \tag{7.5}$$

$$\nu(\lambda_{i_1} \cdot C_{i_1}, \dots, \lambda_{i_n} \cdot C_{i_n}) = \lambda_{i_1} \cdots \lambda_{i_n} \nu(C_{i_1}, \dots, C_{i_n}) \tag{7.6}$$

$$\nu(\mathbf{x}_{i_1} + C_{i_1}, \dots, \mathbf{x}_{i_n} + C_{i_n}) = \nu(C_{i_1}, \dots, C_{i_n}). \tag{7.7}$$

The *Aleksandrov–Fenchel* inequality states that [30, 70]

$$\nu(C_{i_1}, C_{i_1}, C_{i_3}, \dots, C_{i_n}) \nu(C_{i_2}, C_{i_2}, C_{i_3}, \dots, C_{i_n}) \leq |\nu(C_{i_1}, C_{i_2}, C_{i_3}, \dots, C_{i_n})|^2.$$

The convex bodies  $C_1, \dots, C_k$  in the above equations can be polytopes, but they need not be. For example, they could be the volumes bounded by hyper-spheres or hyper-ellipsoids. The volume of the Minkowski sum of a convex body  $C \subset \mathbb{R}^n$  with a solid ball of radius  $r \geq 0$ ,  $B_r^n \doteq r \cdot B^n \subset \mathbb{R}^n$ , is computed as [25]

$$V(C + r \cdot B^n) = W_0(C) + \sum_{k=1}^n \binom{n}{k} W_k(C) r^k \tag{7.8}$$

where the weighting coefficients  $W_k(C)$  are defined in terms of mixed volumes as

$$W_k(C) = \nu(\underbrace{C, \dots, C}_{n-k}, \underbrace{B^n, \dots, B^n}_k).$$

Here  $C$  is repeated  $n - k$  times and  $B^n$  (the ball with unit radius) is repeated  $k$  times. Equation (7.8) is Steiner’s multi-dimensional version of formula (5.98), extended to nonsmooth surfaces.

The coefficients  $W_k(C)$  are called *quermassintegrals*. A version of these quantities normalized and reordered as [25]

$$\mu_{n-k}(C) = \frac{k}{\mathcal{O}_k} \binom{n}{k} W_k(C) \tag{7.9}$$

are called *intrinsic volumes*. Recall that  $\mathcal{O}_k$  was defined in (2.37). The intrinsic volumes  $\mu_i$  for  $i = 0, 1, 2, \dots, n$  take the place of  $V, F, M$ , and  $K$  that appeared in the three-dimensional version of Steiner’s formula.

The values of  $V, F$ , and  $M$  for the Platonic solids with vertices lying on a sphere of unit radius have been reported in [30] as follows.

Tetrahedron:

$$V = \frac{8\sqrt{3}}{27}; \quad F = \frac{8\sqrt{3}}{3}; \quad M = 2\sqrt{6} \arccos(-1/3).$$

Cube:

$$V = \frac{8\sqrt{3}}{9}; \quad F = 8; \quad M = 2\sqrt{3}\pi.$$

Octahedron:

$$V = \frac{4}{3}; \quad F = 4\sqrt{3}; \quad M = 6\sqrt{2} \arccos(1/3).$$

Dodecahedron:

$$V = \frac{2}{9}\sqrt{15}(\sqrt{5} + 1); \quad F = \sqrt{200 - 40\sqrt{5}}; \quad M = 5\sqrt{3}(\sqrt{5} - 1) \arctan(2).$$

Icosahedron:

$$V = \frac{1}{3}\sqrt{40 + 8\sqrt{5}}; \quad F = \sqrt{3}(10 - 2\sqrt{5}); \quad M = \sqrt{450 - 90\sqrt{5}} \arcsin(2/3).$$

From the discretized version of the Gauss–Bonnet theorem,  $K = 4\pi$  for all of these polyhedra, since they are all topologically equivalent to the sphere.

### 7.1.3 Convolution of Bodies

Any convex body,  $C$ , of finite volume in  $\mathbb{R}^n$  can be viewed as the support of a function  $f_C : \mathbb{R}^n \rightarrow \{0, 1\}$  where the value of 0 occurs when  $\mathbf{x} \notin C$ , and the value of 1 occurs when  $\mathbf{x} \in C$ . The integral of such a function over  $\mathbb{R}^n$  will then be the volume of  $C$ .

The function  $f_C$  is the *indicator function* for the body  $C$ . In general, the *support* of a function is the set of values of the argument for which the value of the function is not zero. The support of the indicator function is the body itself. Or stated another way, the support of an indicator function is the pre-image  $f_C^{-1}(1)$ .

Given two bodies,  $C_1$  and  $C_2$ , two indicator functions,  $f_{C_1}(\mathbf{x})$  and  $f_{C_2}(\mathbf{x})$ , can be defined, one for each body. Then, the indicator function for the two bodies can be convolved:

$$(f_{C_1} * f_{C_2})(\mathbf{x}) = \int_{\mathbb{R}^n} f_{C_1}(\boldsymbol{\xi}) f_{C_2}(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi}.$$

Convolution is an operation that was used extensively in Chapters 2 and 3, and will be generalized to the group-theoretic setting in Volume 2.

In analogy with the way that  $C_1 + C_2 = C_2 + C_1$ , it is the case that  $f_{C_1} * f_{C_2} = f_{C_2} * f_{C_1}$ . Furthermore, an interesting thing to observe is that  $(f_{C_1} * f_{C_2})(\mathbf{x}) > 0$  on the interior of  $C_1 + C_2$ . And since the boundary  $\partial(C_1 + C_2)$  is a set of measure zero in  $\mathbb{R}^n$ , the support of  $(f_{C_1} * f_{C_2})$  and the body  $C_1 + C_2$  are indistinguishable in terms of the  $n$ -dimensional volumes that they occupy. In this sense they are equal. And if the Heaviside step function  $H : \mathbb{R} \rightarrow \{0, 1\}$  defined in (2.8) is composed with  $(f_{C_1} * f_{C_2})(\mathbf{x})$ , then an object that is indistinguishable from the Minkowski sum will be produced:

$$f_{C_1+C_2}(\mathbf{x}) = H((f_{C_1} * f_{C_2})(\mathbf{x})). \quad (7.10)$$

The relationship between the Minkowski sum and convolution was observed in [28]. This observation has been used in robotics applications together with using the fast Fourier transform (FFT) to rapidly evaluate Minkowski sums [36].

## 7.2 Examples of Manifolds

A manifold can be thought of as a higher dimensional generalization of the concept of a surface that locally “looks like”  $\mathbb{R}^n$  at every point. A necessary condition for this intuitive property to hold is that self intersections such as those that occur in a figure-eight pattern do not occur in manifolds.

The general discussion of manifolds in this chapter will be easier to follow with a few concrete examples in mind. In applications, manifolds with more than two dimensions mostly arise in the following five ways: (1) as the solution to a set of constraint equations in a large number of variables; (2) as the product manifold of two or more lower dimensional manifolds; (3) as the interior of a polytope or other region in  $\mathbb{R}^n$  with opposing faces (or antipodal points) “pasted” together according to some rule; (4) as the space of all values of a matrix Lie group; (5) as the quotient of a Lie group relative to one of its subgroups. These categories are not mutually exclusive, as will be demonstrated shortly. And the last category will not be explicitly addressed in this volume, though it will be in Volume 2.

### Example 1: The Sphere $S^3$ Embedded in $\mathbb{R}^4$

A simple example of a manifold resulting from a constraint equation is the unit sphere in  $\mathbb{R}^4$ , which is denoted as  $S^3$ , and is described in terms of Cartesian coordinates as

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1.$$

Since  $\mathbb{R}^4$  is a four-dimensional space, and this is a single constraint equation, we conclude that  $S^3$  is a  $4 - 1 = 3$ -dimensional manifold. Parametric equations that satisfy this constraint and “reach” every point on  $S^3$  (as well as  $S^n$ ) were given in Section 2.3.

### Example 2: The Product of a Torus and Ellipsoid Embedded in $\mathbb{R}^6$

The torus and ellipsoid are both two-dimensional surfaces in  $\mathbb{R}^3$ . Let  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ . If

$$\phi_1(\mathbf{x}) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} - 1 = 0$$

is the equation of an ellipsoid and  $\phi_2(\mathbf{y}) = 0$  is the extrinsic description of the torus derived in Exercise 5.28, then these two two-dimensional manifolds can be combined to define a manifold in  $\mathbb{R}^6$ . For example, let  $A_1$  and  $A_2$  be invertible  $3 \times 3$  real matrices and let  $A$  be an invertible  $2 \times 2$  matrix. Also let  $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^3$  and  $\mathbf{b} \in \mathbb{R}^2$ . Then for appropriate choices of  $A, A_i, b$ , and  $b_i$  the following system of two equations in the six unknowns  $\mathbf{z} = [\mathbf{x}^T, \mathbf{y}^T]^T \in \mathbb{R}^6$  will define a four-dimensional product manifold:

$$A \begin{pmatrix} \phi_1(A_1^{-1}(\mathbf{x} - \mathbf{b}_1)) \\ \phi_2(A_2^{-1}(\mathbf{x} - \mathbf{b}_2)) \end{pmatrix} = \mathbf{b}.$$

The way that  $A, A_1, A_2$  and  $\mathbf{b}, \mathbf{b}_1, \mathbf{b}_2$  are chosen defines a particular embedding. If the ellipsoid and torus are described using the parametric equations provided in Chapter 5, which can be denoted respectively as  $\mathbf{x}_1(\phi_1)$  and  $\mathbf{x}_2(\phi_2)$ , then for any real invertible  $6 \times 6$  matrix  $C$  and  $\mathbf{d} \in \mathbb{R}^6$ , the equation

$$\mathbf{z} = C \begin{pmatrix} \mathbf{x}_1(\phi_1) \\ \mathbf{x}_2(\phi_2) \end{pmatrix} + \mathbf{d}$$

would be one way to embed the parameterized four-dimensional product manifold in  $\mathbb{R}^6$ . A not-so-difficult exercise would be to work out the relationship between  $C$ ,  $A$ ,  $A_i$ , etc., in these two different approaches.

### Example 3: The Group of Motions of the Euclidean Plane

The group of planar rigid-body motions has been encountered several times earlier in this volume. Elements of this group are described using matrices of the form

$$g = \begin{pmatrix} \cos \theta & -\sin \theta & x \\ \sin \theta & \cos \theta & y \\ 0 & 0 & 1 \end{pmatrix} \quad \text{with } x, y \in \mathbb{R} \quad \text{and } \theta \in [0, 2\pi). \quad (7.11)$$

The set of all such matrices is called the *special Euclidean group* of the plane, and is denoted as  $SE(2)$ , where the “2” corresponds to the dimension of the plane. The group operation is matrix multiplication. In fact, any Lie group with elements that are matrices and which has a group operation of matrix multiplication is called a *matrix Lie group*. Therefore, when referring to a matrix Lie group, there is no need to mention the group operation, since it is understood in advance to be matrix multiplication.

The manifold of the group<sup>2</sup> can be embedded in  $\mathbb{R}^n$  in several ways. The standard way to embed any matrix Lie group with elements that are  $m \times m$  matrices is to stack the columns into one long vector of dimension  $m^2$ . This then defines an embedding of the matrix Lie group in  $\mathbb{R}^{m^2}$ . In the case of (7.11) this is somewhat “inefficient” in the sense that  $SE(2)$  can be embedded in a much smaller space due to the fact that the last row in the matrix  $g$  consists of constants. For example,  $SE(2)$  can be embedded in  $\mathbb{R}^6$  by parameterizing a vector of the form<sup>3</sup>

$$[\cos \theta, -\sin \theta, x, \sin \theta, \cos \theta, y]^T \in \mathbb{R}^6.$$

This parametric description is equivalent to the implicit description

$$[w, -z, x, z', w', y]^T \in \mathbb{R}^6 \quad \text{with } w^2 + z^2 = 1, z = z', w = w'.$$

Or, since there is repetition, it could be embedded parametrically as

$$[\cos \theta, x, \sin \theta, y]^T \in \mathbb{R}^4$$

or implicitly as

$$[w, x, z, y]^T \in \mathbb{R}^4 \quad \text{with } w^2 + z^2 = 1.$$

Furthermore, in the same way that the topological circle can be thought of as the interval  $[0, 2\pi)$  with the ends “glued” together, a useful way to visualize  $SE(2)$  is as the space  $\mathbb{R}^2 \times [0, 2\pi)$  with the  $\theta$  values glued.

The point is, there is more than one way to embed a manifold in  $\mathbb{R}^n$  for some value of  $n$  sufficiently large. Each different way that it is embedded will induce an associated curvature and twisting/torsion of the embedded manifold that is in addition to the natural intrinsic nature of the manifold.

<sup>2</sup>The manifold of a Lie group is called a *group manifold*

<sup>3</sup>As should be clear from Example 2, embeddings are not unique.

**Example 4: The Group of Rotations of Three-Dimensional Euclidean Space**

The group of rotations in three-dimensional space, or *special orthogonal group*,  $SO(3)$ , is the matrix Lie group with elements that are  $3 \times 3$  real orthogonal matrices with unit determinant. Since it is a matrix group, the group operation is matrix multiplication. If each element of this group is written in terms of columns as

$$R = [\mathbf{a}, \mathbf{b}, \mathbf{c}],$$

then the orthogonality constraint implies that

$$\mathbf{a} \cdot \mathbf{a} = \mathbf{b} \cdot \mathbf{b} = \mathbf{c} \cdot \mathbf{c} = 1$$

and

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} = 0.$$

Altogether this constitutes six constraints imposed on the nine degrees of freedom inherent in a  $3 \times 3$  matrix. The additional constraint that  $\det R = +1$  can be obtained without any further loss of degrees of freedom by observing that when

$$\mathbf{a} \cdot \mathbf{a} = \mathbf{b} \cdot \mathbf{b} = 1 \quad \text{and} \quad \mathbf{a} \cdot \mathbf{b} = 0 \tag{7.12}$$

then setting  $\mathbf{c} = \mathbf{a} \times \mathbf{b}$  will simultaneously satisfy  $\det R = +1$  and the remaining constraints on  $\mathbf{c}$ . This means that for  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$  together with (7.12) describes rotation matrices of the form

$$R = [\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}].$$

The mapping  $\mathbf{v} : SO(3) \rightarrow \mathbb{R}^6$  defined by  $\mathbf{v}(R) = [\mathbf{a}^T, \mathbf{b}^T]^T$  constitutes an embedding of  $SO(3)$  in  $\mathbb{R}^6$ .

Is this efficient or can  $SO(3)$  be embedded in a lower dimensional Euclidean space? Well, it turns out that any  $3 \times 3$  rotation matrix can be parameterized using the *Euler parameters* as

$$R(u_1, u_2, u_3, u_4) = \begin{pmatrix} u_1^2 - u_2^2 - u_3^2 + u_4^2 & 2(u_1 u_2 - u_3 u_4) & 2(u_3 u_1 + u_2 u_4) \\ 2(u_1 u_2 + u_3 u_4) & u_2^2 - u_3^2 - u_1^2 + u_4^2 & 2(u_2 u_3 - u_1 u_4) \\ 2(u_3 u_1 - u_2 u_4) & 2(u_2 u_3 + u_1 u_4) & u_3^2 - u_1^2 - u_2^2 + u_4^2 \end{pmatrix} \tag{7.13}$$

where

$$u_1^2 + u_2^2 + u_3^2 + u_4^2 = 1.$$

There is, however, some redundancy in this because making the substitution  $u_i \rightarrow -u_i$  gives exactly the same rotation matrix (since all of the  $u_i$ 's appear as quadratic terms in  $R(u_1, u_2, u_3, u_4)$ ). This means that as a manifold, the group of rotations can be visualized as the sphere  $S^3$  with antipodal points identified with each other. While this does not mean that  $SO(3)$  can be realized as a non-self-intersecting hyper-surface in  $\mathbb{R}^4$ , it does mean that any parametrization of the sphere  $S^3$  can be used to parameterize the rotation group  $SO(3)$ . Several such parameterizations are provided in [18], and several more are provided in Volume 2.

**Example 5: Real Projective Space**

The *real projective space*, denoted as  $\mathbb{R}P^n$ , is a compact  $n$ -dimensional manifold, the points of which can be identified with lines in  $\mathbb{R}^{n+1}$ , that pass through the origin. Since lines through the origin are defined by a unit direction that can point along the



line in either direction, another way to view  $\mathbb{R}P^n$  is as the sphere  $S^n$  with antipodal points glued together. This means that  $\mathbb{R}P^3$  resembles  $SO(3)$ , and in fact a differentiable and invertible mapping (called a *diffeomorphism*) can be established between these two spaces.

The *real projective plane*,  $\mathbb{R}P^2$ , can be thought of as the usual sphere  $S^2$  with antipodal points glued, or as the set of lines in  $\mathbb{R}^3$  that pass through the origin. It can also be visualized as the unit disk with antipodal points glued [4, 65]. The manifold  $\mathbb{R}P^2$  itself cannot be realized as a non-self-intersecting surface in  $\mathbb{R}^3$ . However, it can be realized as such a surface in  $\mathbb{R}^4$  by the mapping  $\mathbf{m} : S^2 \rightarrow \mathbb{R}^4$  defined by  $\mathbf{m}(\mathbf{u}) = [u_1u_2, u_1u_3, u_2u_3, u_2^2 - u_3^2]^T$ . There are many possible such mappings [37]. Since this mapping is a quadratic form in the entries  $u_i$ , it means that  $+\mathbf{u}$  and  $-\mathbf{u}$  map to the same point. And given a point  $\mathbf{m}$  on this embedded manifold, it is possible to recover (up to the sign that has been lost) the pair of unit vectors  $\pm\mathbf{u}$ . For example, if  $m_3 \neq 0$ , then since  $m_1m_2/m_3 = u_1^2$  it is possible to recover  $u_1$  up to an unknown sign. Then this can be used to find  $u_2 = m_1/u_1$  and  $u_3 = m_3/u_1$ . If  $m_3 = 0$ , then either  $u_2 = 0$  or  $u_3 = 0$ , or both. Exactly which case it is will be evident from examining the other entries of  $\mathbf{m}$ .

Furthermore, many ways exist to map the manifold  $\mathbf{m}(\mathbf{u})$  into a surface in  $\mathbb{R}^3$  that intersects itself. Two of these self-intersecting surfaces are called the *Roman surface* (which was discovered by Steiner while he was visiting Rome) and *Boy's surface* [4]. Whereas an *embedding* is a way to map a manifold into  $\mathbb{R}^n$  in such a way that it does not intersect itself, a mapping from a manifold into  $\mathbb{R}^n$  that results in self-intersections is called an *immersion*, and the resulting geometric object is called an *immersed manifold* (even though it is not actually a manifold). For example, a figure-eight pattern can be thought of as an immersed circle.

### Example 6: Polytopes with a Twist and Crystallography

Crystals appear in nature in many forms: salt, diamonds, man-made silicon wafers, etc. In protein crystallography, many copies of a protein molecule are coaxed into forming a crystal in the laboratory. Then x-ray diffraction experiments can be performed to gain information about the shape of these molecules. Atomic models are then fit to these shapes. Readable introductions to the subject of protein crystallography, including discussions of experimental methods and the mathematics of crystal structure, include [32, 41, 46, 53]. In such crystals, the basic unit that is translated to replicate the whole crystal is called the *unit cell*. This can be constructed from several translated and rotated copies of the same protein molecule, and/or it can share the unit cell with copies of other protein molecules (in which case it is called a co-crystal). If only one copy of a protein inhabits a unit cell, then the whole crystal might look like Figure 7.1, where the letter R is used to denote a shape without rotational or mirror symmetries. (Most capital Roman letters, including A, B, C, D, E, H, I, M, N, and others, have some kind of symmetry, but R is one of the few that does not.) The use of the letter R in this presentation (rather than the other asymmetric letters F, P, L, etc.) follows [21]. The unit cell can be broken up into so-called *asymmetric units* containing exactly one object. The union of these asymmetric units reconstitutes the unit cell, and translated copies of the unit cell completely tile space. In the case of Figure 7.2, one object and its mirror image inhabit one unit cell which is then repeated. In Figure 7.3, four variants on the same object constitute one unit cell.

A relationship exists between crystallography, polytopes, and the theory of manifolds. The same sort of construction that was used in Figure 5.2 to form a torus from

a rectangle can be used to construct more exotic manifolds. For example, consider the two-dimensional lattices as shown in Figures 7.1–7.3. As mentioned above, the letter R is used to illustrate the basic crystallographic asymmetric unit for each tiling of the plane described in the lower left of each figure as a “gluing” of arrows on opposing edges of the rectangle. This gluing of asymmetric units produces the unit cell shown in the upper left of each figure.<sup>4</sup> By copying and translating the contents of this unit cell, an infinite lattice can be formed (truncated versions of which are illustrated on the right-hand side of Figures 7.1 and 7.2). A lattice can also be formed by translating unit cells like those on the right side of Figure 7.3.

The set of translations of unit cells to form a lattice in  $\mathbb{R}^n$  is a discrete subset of the group of rigid-body motions,  $SE(n)$ . This subset is closed under composition. In the planar case, such operations can be expressed as rigid-body motions of the form in (7.11). But in the context of planar crystallographic motions,  $(x, y, \theta)$  are restricted to very specific values. For example, if the dimensions of the cells in Figures 7.1–7.3 are  $w \times h$  (i.e., width by height) with  $w \neq h$ , then the only allowable translations that will place unit cells in the correct locations are of the form  $(x, y, \theta) = (m \cdot w, n \cdot h, 0)$  for  $m, n \in \mathbb{Z}$  for Figure 7.1,  $(x, y, \theta) = (2m \cdot w, n \cdot h, 0)$  for Figure 7.2, and  $(x, y, \theta) = (2m \cdot w, 2n \cdot h, 0)$  for Figure 7.3.

Now even though the dimensions of the asymmetric units in the lattices generated by the gluings in Figures 7.1–7.3 are exactly the same, the unit cells are clearly different, and the lattices have different symmetry groups. In the case of Figure 7.1 all crystallographic motions are of the form in (7.11) with  $(x, y, \theta) = (m \cdot w, n \cdot h, 0)$ . That is, they are purely translated copies of each other. For the other two cases, combinations of rotations and mirror reflections are required to generate unit cells from asymmetric units.<sup>5</sup>

The lattice in Figure 7.1 can be generated by repeated application of transformations of the form

$$t_1 = \begin{pmatrix} 1 & 0 & w \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad t_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix}$$

(and their inverses) to the unit cell. The set of all such translations forms a group  $G_0 \cong (\mathbb{Z}^2, +)$ .

Now consider the transformations that can be applied to the contents of the asymmetric unit at the lower left of Figure 7.2 to produce the lattice on the right side of that figure. The basic motion when translating horizontally from one cell to an adjacent one is to flip the orientation by  $\pi$  radians (or 180 degrees). Each such transformation can be viewed as acting on a reference frame attached to the lower left corner of each asymmetric unit. In other words, for the lattice in Figure 7.2 the basic motions between asymmetric units are of the form

$$b_1 = \begin{pmatrix} 1 & 0 & w \\ 0 & -1 & h \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad b_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix}.$$

These act on the contents of asymmetric units by multiplication of the position  $(x, y)$  represented as a vector  $[x, y, 1]^T$  to produce  $[x', y', 1]^T$  from which the new position

<sup>4</sup>A unit cell is the smallest unit of a crystal lattice from which the whole lattice can be constructed by the application of a set of rigid-body translations drawn from a discrete subgroup of  $SE(n)$ .

<sup>5</sup>If spatial rigid-body motions are allowed, then the mirror image of a planar figure could be produced by a 180-degree spatial rotation around the line in the plane representing the mirror.

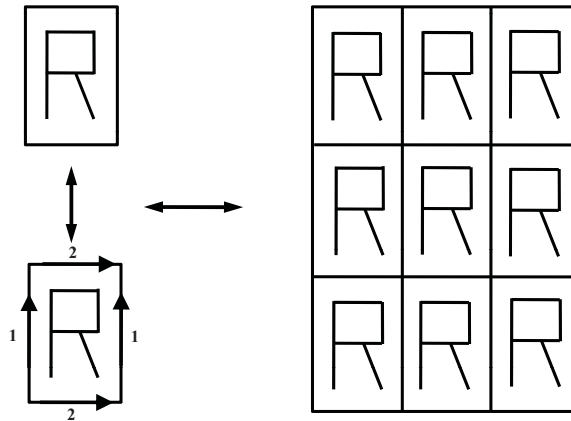
$(x', y')$  can be extracted. The transformation  $b_1$  produces the neighbor to the right of a given asymmetric unit, and  $b_2$  produces the neighbor directly below. The set of crystallographic operations for this lattice is obtained by all possible repeated applications of the basic operations  $b_1$  and  $b_2$  and their inverses, which is implemented as a matrix multiplication  $b_1 \circ b_1, b_1 \circ b_2, b_2 \circ b_2, b_1 \circ b_1 \circ b_2$ , etc. This generates a discrete (though infinite) set of transformations that we can call  $G_1$ .

The asymmetric unit at the lower left of Figure 7.2 can be thought of as an object which, when multiplied by all elements of  $G_1$ , covers the plane  $\mathbb{R}^2$ . Therefore, a good notation for this asymmetric unit is  $G_1 \backslash \mathbb{R}^2$  since, in a sense,  $G_1 \times (G_1 \backslash \mathbb{R}^2) = \mathbb{R}^2$ .

The set of basic transformations that produces the lattice generated by the asymmetric unit and unit cells shown in Figure 7.3 are

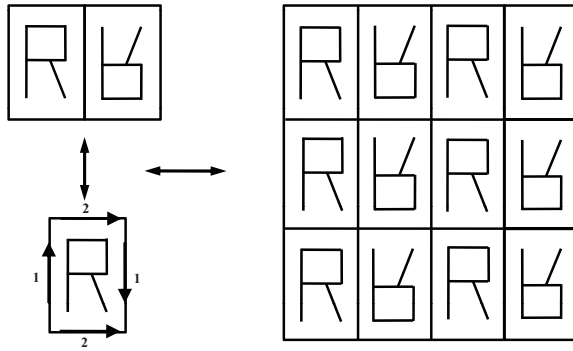
$$b'_1 = b_1 = \begin{pmatrix} 1 & 0 & w \\ 0 & -1 & h \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad b'_2 = \begin{pmatrix} 1 & 0 & w \\ 0 & -1 & -h \\ 0 & 0 & 1 \end{pmatrix}.$$

The group generated by repeated application of these transformations and their inverses can be called  $G_2$ . Note that while the asymmetric units in these cases have the same shape and dimensions,  $G_0 \backslash \mathbb{R}^2 \neq G_1 \backslash \mathbb{R}^2 \neq G_2 \backslash \mathbb{R}^2$  because the gluings are different.

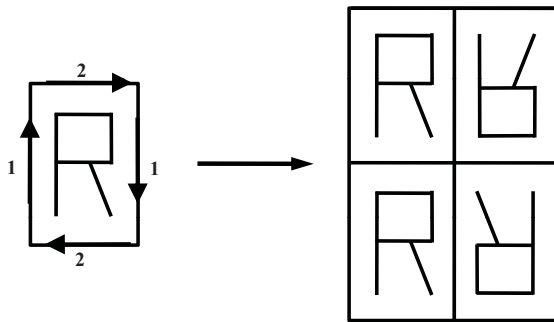


**Fig. 7.1.** A Pattern on the Torus Transferred to the Euclidean Plane

There are 17 classes of planar tilings, or regular tessellations (also called “wallpaper patterns”), and there are 230 such patterns in the spatial case. These are described by groups of rigid-body motions and reflections that can be applied to each unit cell to either transform the contents of the cell back into itself, or translate one unit cell to an adjacent one. These are the *crystallographic space groups*. If such a group is denoted as  $G$ , then we can think of  $\mathbb{R}^n$  as being “divided up” by  $G$ . That is, each asymmetric unit can be viewed as an element of  $G \backslash \mathbb{R}^n$ , and when elements of  $G$  are applied to elements of  $G \backslash \mathbb{R}^n$ , the result is a tiling of  $\mathbb{R}^n$ . These crystallographic space groups have been studied extensively, and were completely classified by the late nineteenth and early twentieth century [21, 39, 44]. More recently, the classification of 3-manifolds constructed as quotients of  $\mathbb{R}^3$  by space groups (called *orbifolds*) was initiated by Thurston [61]. Other treatments can be found in [14, 35, 47, 57, 65]. A precursor to the orbifold



**Fig. 7.2.** A Pattern on the Klein Bottle Transferred to the Euclidean Plane



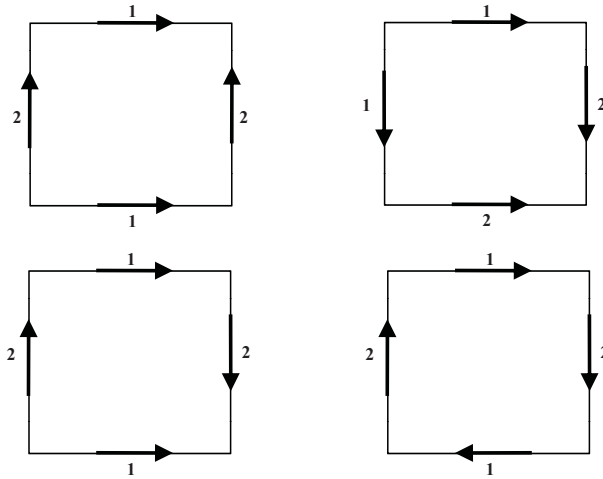
**Fig. 7.3.** A Pattern on the Real Projective Plane Transferred to the Euclidean Plane

concept is that of the *V-manifold* introduced by Satake [55]. The relationship between these manifolds and mathematical objects called groupoids has also been studied [66].

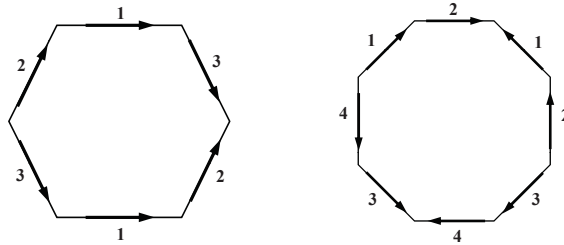
The Klein bottle and real projective plane depicted as gluings in Figures 7.2, 7.3, and 7.4 are both non-orientable two-dimensional surfaces that cannot be embedded in  $\mathbb{R}^3$ . They can be displayed as planar gluings, but this should not be confused with planar embeddings.

Figures 7.1–7.5 represent only a few of the two-dimensional orbifolds that can be constructed from crystallographic space groups. And in addition, these concepts apply in higher dimensions. For example, if  $\mathbb{R}^3$  is broken up into cubes by the group  $(\mathbb{Z}^3, +)$ , then the opposing faces of a cube are glued in a natural way, and the result is the 3-torus,  $T^3$ . Other three-manifolds can be obtained by using other of the 230 crystallographic space groups to define unit cells and associated gluings [14, 35, 47].

In addition, it is possible to define other manifolds by appropriate gluing of the edges of polygons, faces of polygons, or more generally polytopes. Figure 7.4 shows how some of the most studied two-dimensional manifolds can be thought of as gluings of squares. And there is no reason to limit the discussion to squares or rectangles. Figure 7.5 shows how tori (with one or two holes) can be described as gluings of hexagons and octagons. Whereas the hexagon can be used to tile the plane without leaving any empty spaces, the octagon cannot. In three dimensions, various gluings of the opposing faces of a cube lead to different manifolds. The simplest of these is the three-torus,  $T^3$ . As more exotic



**Fig. 7.4.** Various Squares with Glued Edges: (upper left) The Torus; (upper right) The Sphere; (lower left) The Klein Bottle; (lower right) The Real Projective Plane



**Fig. 7.5.** Two-Dimensional Manifolds Represented as Polygons with Glued Edges: (left) The Torus as a Glued Hexagon; (right) The Two-Holed Torus as a Glued Octagon

examples, if the opposing pentagonal faces of a dodecahedron are twisted by  $1/5$  or  $3/5$  of a 360-degree turn and glued, the results will be two different 3-manifolds [61].

**Counterexamples**

When presented with such a wide variety of examples, it can be tempting to think that everything is a manifold. But this certainly is not true, and counterexamples that illustrate spaces that are not manifolds are in order. As a first class of examples, two manifolds that are glued together at a single point result in an object that is not a manifold. This includes two kissing spheres, a double-sided cone, and two cubes offset by some distance and connected with an umbilical curve. These are not manifolds because the neighborhoods containing the points of contact are unlike any open neighborhoods in  $\mathbb{R}^n$ . Another example that is not a manifold is the closed unit square  $[0, 1] \times [0, 1]$  because the neighborhoods containing points from the boundary are unlike any open neighborhood in  $\mathbb{R}^n$  since they are partially closed. (However, the open unit square  $(0, 1) \times (0, 1)$  is a manifold, and the closed unit square, while not a manifold, is an example of a *manifold with boundary*.) And finally, it should be noted that not every twisted and glued polytope results in a manifold. It can be that a gluing results in too many points that accumulate at one edge or vertex and not enough at others.

Again, the test in this case is to assess whether or not each neighborhood at each set of glued vertices and edges has the same properties as a neighborhood in  $\mathbb{R}^n$ . This can be approached intuitively when  $n = 2$  or  $3$ , but becomes very difficult for  $n > 3$ .

### 7.3 Embedded Manifolds, Part I: Using Vector Calculus

In this section, the concepts of tangent, normal, surface area integral, curvature, integration by parts, and the divergence theorem are described in the context of  $m$ -dimensional manifolds embedded in  $n$ -dimensional Euclidean space. The reason for doing this is to show that it can be done (even though it is rather messy and notationally heavy). This is then followed by the lighter modern approach involving differential forms in Section 7.5 and the coordinate-free approach in Section 7.6.

#### 7.3.1 The Inner Product of Vector Fields on Manifolds Embedded in $\mathbb{R}^n$

Let  $\mathbf{F} \in \mathbb{R}^n$  and  $\mathbf{H} \in \mathbb{R}^n$  be two real vector fields on the manifold  $M \subset \mathbb{R}^n$ . That is, for each point on the manifold a single vector is assigned to the tangent hyper-plane at that point. Since that hyper-plane is a subspace of  $\mathbb{R}^n$ , the vector that is assigned can be viewed as being in  $\mathbb{R}^n$ . More concretely, if  $M$  is parameterized as  $\mathbf{x} = \mathbf{x}(\mathbf{q})$  with  $\mathbf{x} \in \mathbb{R}^n$  and  $\mathbf{q} \in \mathbb{R}^m$ , then

$$\mathbf{F} = \sum_{i=1}^m f_i \mathbf{T}_i, \quad \text{where } \mathbf{T}_i \doteq \frac{\partial \mathbf{x}}{\partial q_i}. \quad (7.14)$$

In this way, the vector  $\mathbf{F} \in \mathbb{R}^n$  has  $m$  independent components  $\{f_i\}$ . These can be stacked to form a column array of the form  $\mathbf{f} = [f_1, f_2, \dots, f_m]^T$ . Note that in the context of this definition,  $\mathbf{F} \cdot \mathbf{e}_j = F_j \neq f_j$ .

It then makes sense to talk about the inner products of vectors of the form of  $\mathbf{F}$  in  $\mathbb{R}^n$  (which is just the usual dot product of vectors), and this induces an inner product in the  $m$ -dimensional tangent hyper-plane since

$$\mathbf{F} \cdot \mathbf{H} = \left( \sum_{i=1}^m f_i \mathbf{T}_i \right) \cdot \left( \sum_{j=1}^m h_j \mathbf{T}_j \right) = \sum_{i,j=1}^m g_{ij} f_i h_j \doteq \langle \mathbf{f}, \mathbf{h} \rangle. \quad (7.15)$$

Furthermore, by taking the dot product on both sides of (7.14) with  $\mathbf{T}_j$ , it follows that

$$\mathbf{F} \cdot \mathbf{T}_j = \sum_{i=1}^m g_{ji} f_i \quad \implies \quad \mathbf{f} = G^{-1} \frac{\partial \mathbf{x}^T}{\partial \mathbf{q}} \mathbf{F}. \quad (7.16)$$

The conversion from  $\mathbf{f}$  to  $\mathbf{F}$  is trivial, since  $f_i = \mathbf{f} \cdot \mathbf{e}_i$  for  $i = 1, \dots, m$ , which can then be substituted into (7.14).

The calculations in (7.15), which are at a single point on the manifold, can be integrated over the whole manifold, leading to the definition<sup>6</sup>

---

<sup>6</sup>Here  $\langle \mathbf{f}, \mathbf{h} \rangle$  does *not* denote expected value as in Chapter 3. Rather, angle brackets are used here in place of parentheses to distinguish the combination of dot product and integral from the pointwise inner product of vector fields  $(\mathbf{f}, \mathbf{h})$  defined in (7.15).

$$\langle \mathbf{f}, \mathbf{h} \rangle \doteq \int_M \mathbf{F} \cdot \mathbf{H} dV = \int_M (\mathbf{f}, \mathbf{h}) dV \quad (7.17)$$

$$= \int_D (\mathbf{f}(\mathbf{q}), \mathbf{h}(\mathbf{q})) |G(\mathbf{q})|^{\frac{1}{2}} dq_1 \dots dq_m \quad (7.18)$$

$$= \sum_{i,j=1}^m \int_D f_i g_{ij} h_j |G|^{\frac{1}{2}} dq_1 dq_2 \dots dq_m. \quad (7.19)$$

Here  $D \subset \mathbb{R}^m$  is the range of coordinate values that define the manifold, and the Riemannian metric tensor is computed as in Chapter 5:

$$G = [g_{ij}] \quad \text{where} \quad g_{ij} = \frac{\partial \mathbf{x}}{\partial q_i} \cdot \frac{\partial \mathbf{x}}{\partial q_j}, \quad \text{for } i, j = 1, \dots, m.$$

Clearly,  $\langle \mathbf{f}, \mathbf{h} \rangle = \langle \mathbf{h}, \mathbf{f} \rangle$ .

Furthermore, given a smooth function  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  and the definition of the gradient in (5.48),

$$\langle \mathbf{f}, \text{grad } \phi \rangle = \sum_{i,j,k=1}^m \int_D g^{ki} \frac{\partial \phi}{\partial q_i} g_{kj} f_j |G(\mathbf{q})|^{\frac{1}{2}} dq_1 \dots dq_m.$$

This quantity will appear on one side of the divergence theorem for embedded manifolds. In order to write the other side of the equation, an appropriate concept of surface normal is required.

### 7.3.2 An Example: A Hyper-Spherical Cap

Now a concrete example of a manifold,  $M$ , is examined in detail. The manifold is the upper hemisphere of  $S^3$  embedded in  $\mathbb{R}^4$  with open boundary where the hemisphere meets the hyper-plane  $\mathbb{R}^3 \subset \mathbb{R}^4$ . Here two different parameterizations of  $M$  are used to illustrate concrete geometric calculations, and conversions between coordinate systems.

#### Cartesian Coordinates

The part of the unit sphere  $S^3 \subset \mathbb{R}^4$  defined by the Cartesian coordinates  $\mathbf{q} \in \mathbb{R}^3$  as

$$\mathbf{x}(\mathbf{q}) = \begin{pmatrix} x_1(q_1, q_2, q_3) \\ x_2(q_1, q_2, q_3) \\ x_3(q_1, q_2, q_3) \\ x_4(q_1, q_2, q_3) \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ \sqrt{1 - q_1^2 - q_2^2 - q_3^2} \end{pmatrix} \quad (7.20)$$

where  $x_4 > 0$ , is a manifold.

The vectors  $\mathbf{T}_i = \partial \mathbf{x} / \partial q_i$  are computed from  $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$  defined in (7.20) as

$$\mathbf{T}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \frac{-q_1}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} \end{pmatrix}; \quad \mathbf{T}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \frac{-q_2}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} \end{pmatrix}; \quad \mathbf{T}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \frac{-q_3}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} \end{pmatrix}. \quad (7.21)$$

An arbitrary vector tangent to the manifold in  $\mathbb{R}^4$  is then

$$\mathbf{V} = \sum_{i=1}^3 v_i \mathbf{T}_i. \quad (7.22)$$

The notation  $\mathbf{v} = [v_1, v_2, v_3]^T$  will be convenient. However, it is important to not confuse  $\mathbf{V}$  and  $\mathbf{v}$ , and the associated definitions of inner products. For a vector in the tangent space of  $M$  to be a unit vector,  $\mathbf{V} \cdot \mathbf{V} = \mathbf{v}^T G \mathbf{v} \doteq (\mathbf{v}, \mathbf{v}) = 1$ . In general, when viewing a vector in the tangent space of an embedded manifold as a vector in the Euclidean space in which the manifold itself is embedded, capital bold Roman letters will be used. The lower case bold versions of these same letters will denote the column arrays that extract the relevant information from these large vectors. The lower case versions will always be of lower dimension than the capital versions because the Euclidean space in which the manifold is embedded is always larger than the dimension of the manifold. The conversion between these two descriptions in the general case is given in (7.16).

Since this orientable manifold has dimension  $n = 3$  and it is embedded in  $\mathbb{R}^{n+1}$  (i.e., it has *co-dimension* of one), it makes sense to define a single outward-pointing unit normal at each point of  $M$ . In particular, in this example the normal to  $M$  is

$$\mathbf{N}_M(\mathbf{q}) = \mathbf{x}(\mathbf{q}).$$

It is easy to verify that  $\mathbf{N}_M \cdot \mathbf{N}_M = 1$  and  $\mathbf{N}_M \cdot \mathbf{T}_i = 0$  for  $i = 1, 2, 3$ .

The metric tensor for  $S^3$  in this coordinate patch is

$$G(\mathbf{q}) = \frac{1}{1 - q_1^2 - q_2^2 - q_3^2} \begin{pmatrix} 1 - q_2^2 - q_3^2 & q_1 q_2 & q_1 q_3 \\ q_1 q_2 & 1 - q_1^2 - q_3^2 & q_2 q_3 \\ q_1 q_3 & q_2 q_3 & 1 - q_1^2 - q_2^2 \end{pmatrix} \quad (7.23)$$

and it follows that

$$|G(\mathbf{q})| = \frac{1}{1 - q_1^2 - q_2^2 - q_3^2} \quad \text{and} \quad G^{-1} = \begin{pmatrix} 1 - q_1^2 & -q_1 q_2 & -q_1 q_3 \\ -q_1 q_2 & 1 - q_2^2 & -q_2 q_3 \\ -q_1 q_3 & -q_2 q_3 & 1 - q_3^2 \end{pmatrix}.$$

where  $|G(q)| = \det G(q)$ .

### Hyper-Spherical Coordinates

Now consider the completely different way to parameterize the upper hemisphere of  $S^3$  embedded in  $\mathbb{R}^4$ :

$$\mathbf{x}(\phi) = \begin{pmatrix} x_1(\phi, \theta, \psi) \\ x_2(\phi, \theta, \psi) \\ x_3(\phi, \theta, \psi) \\ x_4(\phi, \theta, \psi) \end{pmatrix} = \begin{pmatrix} \sin \psi \sin \theta \cos \phi \\ \sin \psi \sin \theta \sin \phi \\ \sin \psi \cos \theta \\ \cos \psi \end{pmatrix} \quad \text{where } (\phi, \theta, \psi) \in [0, 2\pi) \times [0, \pi] \times [0, \pi/2). \quad (7.24)$$

As a vector-valued function,  $\mathbf{x}(\phi)$  is *not* simply the same as that given in (7.20) with  $\mathbf{q}$  replaced by  $\phi$ ; rather the shorthand  $\mathbf{x}(\phi) = \mathbf{x}(\mathbf{q}(\phi))$  is being used. It is easy to see from matching the first three components in the vectors in (7.20) and (7.24) that  $\mathbf{q}(\phi)$  can be expressed explicitly in terms of components as

$$\begin{aligned} q_1 &= \sin \psi \sin \theta \cos \phi \\ q_2 &= \sin \psi \sin \theta \sin \phi \\ q_3 &= \sin \psi \cos \theta. \end{aligned} \quad (7.25)$$

The inverse mapping can be computed from this as



$$\begin{aligned}
\phi &= \tan^{-1} \frac{q_2}{q_1} \\
\theta &= \tan^{-1} \frac{\sqrt{q_1^2 + q_2^2}}{q_3} \\
\psi &= \sin^{-1} \sqrt{q_1^2 + q_2^2 + q_3^2}.
\end{aligned} \tag{7.26}$$

Vectors tangent to  $M$  can be obtained using these coordinates as

$$\frac{\partial \mathbf{x}}{\partial \phi} = \begin{pmatrix} -\sin \phi \sin \theta \sin \psi \\ \cos \phi \sin \theta \sin \psi \\ 0 \\ 0 \end{pmatrix}; \quad \frac{\partial \mathbf{x}}{\partial \theta} = \begin{pmatrix} \cos \phi \cos \theta \sin \psi \\ \sin \phi \cos \theta \sin \psi \\ -\sin \theta \sin \psi \\ 0 \end{pmatrix}; \quad \frac{\partial \mathbf{x}}{\partial \psi} = \begin{pmatrix} \cos \psi \sin \theta \cos \phi \\ \cos \psi \sin \theta \sin \phi \\ \cos \psi \cos \theta \\ -\sin \psi \end{pmatrix}.$$

These vectors span the same hyper-plane in  $\mathbb{R}^4$  as the vectors  $\mathbf{T}_i$  given in (7.21). This hyper-plane is an embedded version of the tangent space of  $M$  at the specified value of  $\phi$ .

Unlike the coordinate system  $\{q_1, q_2, q_3\}$ , this one is orthogonal, meaning that the metric tensor is diagonal:

$$G(\phi) = \begin{pmatrix} \sin^2 \theta \sin^2 \psi & 0 & 0 \\ 0 & \sin^2 \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad |G(\phi)|^{\frac{1}{2}} = \sin^2 \psi \sin \theta. \tag{7.27}$$

As with the vector function  $\mathbf{x}(\phi)$ , the shorthand  $G(\phi)$  used here is to denote “the metric tensor obtained by undergoing calculations analogous to those used to obtain  $G(\mathbf{q})$  with  $\phi$  taking the place of  $\mathbf{q}$  at every step of the calculation.” However, a major difference between the relationships  $\mathbf{x}(\phi) \leftrightarrow \mathbf{x}(\mathbf{q})$  and  $G(\phi) \leftrightarrow G(\mathbf{q})$  is that partial derivatives with respect to coordinates were required when computing  $G$ . This means that in general  $G(\phi) \neq G(\mathbf{q}(\phi))$ . Rather,

$$G(\phi) = \frac{\partial \mathbf{q}^T}{\partial \phi} G(\mathbf{q}(\phi)) \frac{\partial \mathbf{q}}{\partial \phi^T}. \tag{7.28}$$

This can be verified by computing the Jacobian corresponding to the equations in (7.25). Explicitly,

$$\frac{\partial \mathbf{q}}{\partial \phi^T} = \begin{pmatrix} -\sin \phi \sin \theta \sin \psi & \cos \phi \cos \theta \sin \psi & \cos \phi \sin \theta \cos \psi \\ \cos \phi \sin \theta \sin \psi & \sin \phi \cos \theta \sin \psi & \sin \phi \sin \theta \cos \psi \\ 0 & -\sin \theta \sin \psi & \cos \theta \cos \psi \end{pmatrix}$$

and

$$\left| \frac{\partial \mathbf{q}}{\partial \phi^T} \right| = -\sin^2 \psi \sin \theta \cos \psi.$$

The transpose of this matrix is  $\partial \mathbf{q}^T / \partial \phi$ . The inverse of  $\partial \mathbf{q} / \partial \phi^T$  can be computed explicitly as

$$\left[ \frac{\partial \mathbf{q}}{\partial \phi^T} \right]^{-1} = \begin{pmatrix} -\sin \phi / \sin \theta \sin \psi & \cos \phi \cos \theta / \sin \psi & \cos \phi \sin \theta / \cos \psi \\ \cos \phi / \sin \theta \sin \psi & \sin \phi \cos \theta / \sin \psi & \sin \phi \sin \theta / \cos \psi \\ 0 & -\sin \theta / \sin \psi & \cos \theta / \cos \psi \end{pmatrix}^T.$$

This same result can be obtained by computing the Jacobian of the transformation defined by the equations in (7.26), followed by substitution of  $\mathbf{q} = \mathbf{q}(\phi)$ . Explicitly,

$$\frac{\partial \phi}{\partial \mathbf{q}^T} = \begin{pmatrix} -q_2/(q_1^2 + q_2^2) & q_1/(q_1^2 + q_2^2) & 0 \\ \frac{q_1 q_3 (q_1^2 + q_2^2)^{-\frac{1}{2}}}{q_1^2 + q_2^2 + q_3^2} & \frac{q_2 q_3 (q_1^2 + q_2^2)^{-\frac{1}{2}}}{q_1^2 + q_2^2 + q_3^2} & -\frac{(q_1^2 + q_2^2)^{\frac{1}{2}}}{q_1^2 + q_2^2 + q_3^2} \\ \frac{(q_1^2 + q_2^2 + q_3^2)^{-\frac{1}{2}} q_1}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} & \frac{(q_1^2 + q_2^2 + q_3^2)^{-\frac{1}{2}} q_2}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} & \frac{(q_1^2 + q_2^2 + q_3^2)^{-\frac{1}{2}} q_3}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} \end{pmatrix},$$

from which it is easy to verify that

$$\left[ \frac{\partial \mathbf{q}}{\partial \phi^T} \right]^{-1} = \frac{\partial \phi}{\partial \mathbf{q}^T} \Big|_{\mathbf{q}=\mathbf{q}(\phi)}. \quad (7.29)$$

### An Example of a Manifold with Boundary

Consider the part of the upper hemisphere of  $S^3$  embedded in  $\mathbb{R}^4$  for which  $x_4 > h$  for some constant  $h > 0$ . This means that  $\|\mathbf{q}\| < (1 - h^2)^{\frac{1}{2}}$  and the vector  $\mathbf{x}(\mathbf{q}) \in \mathbb{R}^4$  points to all locations in this space, which is denoted as  $M$ . Or, in the  $\phi$  parametrization,  $\psi < \cos^{-1} h$ . The volume of this manifold can be computed in either coordinate system. In particular,

$$\begin{aligned} V_M &= \int_0^{\cos^{-1} h} \int_0^\pi \int_0^{2\pi} \sin^2 \psi \sin \theta d\phi d\theta d\psi = 4\pi \int_0^{\cos^{-1} h} \sin^2 \psi d\psi \\ &= 2\pi [\psi - \sin \psi \cos \psi]_0^{\cos^{-1} h} = 2\pi [\cos^{-1} h - (1 - h^2)^{\frac{1}{2}} h]. \end{aligned}$$

The boundary,  $\partial M$ , is the intersection of  $S^3$  with a copy of  $\mathbb{R}^3$  that has been translated  $h$  units along the  $x_4$  axis. This boundary manifold (the dimension of which is lower by one than the dimension of  $M$ ) can also be described as a parameterized embedding of  $\mathbb{R}^4$  as

$$\mathbf{x}'(\mathbf{s}) = \mathbf{x}(\mathbf{q}(\mathbf{s})) = \begin{pmatrix} r_h \cos s_1 \sin s_2 \\ r_h \sin s_1 \sin s_2 \\ r_h \cos s_2 \\ h \end{pmatrix} \quad \text{where } r_h \doteq \sqrt{1 - h^2}. \quad (7.30)$$

The metric tensor for this two-dimensional manifold is obtained by computing

$$\begin{pmatrix} \frac{\partial \mathbf{x}'}{\partial s_1} \cdot \frac{\partial \mathbf{x}'}{\partial s_1} & \frac{\partial \mathbf{x}'}{\partial s_1} \cdot \frac{\partial \mathbf{x}'}{\partial s_2} \\ \frac{\partial \mathbf{x}'}{\partial s_1} \cdot \frac{\partial \mathbf{x}'}{\partial s_2} & \frac{\partial \mathbf{x}'}{\partial s_2} \cdot \frac{\partial \mathbf{x}'}{\partial s_2} \end{pmatrix} = \begin{pmatrix} (r_h)^2 \sin^2 \theta & 0 \\ 0 & (r_h)^2 \end{pmatrix}.$$

This is just the metric for the sphere  $S_{r_h}^2 \subset \mathbb{R}^3$ , and its surface area can be computed in the usual way as

$$V_{\partial M} = \int_0^\pi \int_0^{2\pi} (r_h)^2 \sin \theta d\phi d\theta = 4\pi (r_h)^2.$$

### 7.3.3 Computing Normals Extrinsically Without the Cross Product

In  $\mathbb{R}^n$  for  $n \neq 3$  the vector cross product is not defined. However, the concept of the normal to an  $m$ -dimensional manifold in  $\mathbb{R}^n$  is still valid. (Recall, for example, how

normals were constructed in the  $n = 2$  case.) In the general case, each point on an  $m$ -dimensional embedded manifold in  $\mathbb{R}^n$  will have an associated subspace of normals in  $\mathbb{R}^n$  defined by  $n - m$  directions. This subsection addresses how to construct this “normal subspace” explicitly without having the convenience of a cross product.

Take as the standard orthonormal basis for  $\mathbb{R}^n$  the set of unit vectors  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ . The subspace of  $\mathbb{R}^n$  spanned by the tangent vectors to the embedded manifold at point  $\mathbf{x}(\mathbf{q})$  is defined by  $\mathbf{T}_i = \partial\mathbf{x}/\partial q_i$  for  $i = 1, \dots, m$ .

A basis for the subspace of normals to a smooth embedded manifold can be constructed by projection. Namely, start with vector  $\mathbf{e}_1$ , and subtract away all components of it that are in the tangent plane:

$$\mathbf{N}'_1 \doteq \mathbf{e}_1 - \sum_{i=1}^m \frac{\mathbf{e}_1 \cdot \mathbf{T}_i}{\mathbf{T}_i \cdot \mathbf{T}_i} \mathbf{T}_i.$$

Define this to be the first column of the matrix  $N$ . Do the same operation on  $\mathbf{e}_2, \mathbf{e}_3$ , etc. The result will be a matrix with  $n$  columns  $N = [\mathbf{N}'_1, \mathbf{N}'_2, \dots, \mathbf{N}'_n]$ , each of which is normal to the embedded manifold. However, only  $n - m$  of these vectors will be linearly independent. To find a basis for the space of normals, row reduce  $N^T$ , and pick off the  $n - m$  independent (non-zero) rows after row reduction is complete. The transpose of these rows will be the normals  $\mathbf{N}_1, \dots, \mathbf{N}_{n-m}$ .

In general, this basis for the space of normals to the embedded manifold at the point  $\mathbf{x}(\mathbf{q})$  will not be orthonormal, but it does not need to be to perform calculations. And the *Gram–Schmidt orthogonalization process* (see Section A.1.4) can be used if an orthogonal basis is desired.

Explicit calculations used to compute normals to submanifolds of embedded manifolds are now demonstrated on the hyper-spherical cap,  $M$ , and its bounding sphere,  $\partial M$ . The tangent vectors to this boundary manifold  $\partial M$ , as seen in  $\mathbb{R}^4$ , are

$$\mathbf{T}'_j = \frac{\partial\mathbf{x}'}{\partial s_j} = \left. \frac{\partial\mathbf{x}}{\partial \mathbf{s}^T} \right|_{\mathbf{q}=\mathbf{q}(\mathbf{s})} \frac{\partial\mathbf{q}}{\partial s_j}.$$

For this example,

$$\mathbf{T}'_1(\mathbf{s}) = \begin{pmatrix} -r_h \sin s_1 \sin s_2 \\ r_h \cos s_1 \sin s_2 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{T}'_2(\mathbf{s}) = \begin{pmatrix} r_h \cos s_1 \cos s_2 \\ r_h \sin s_1 \cos s_2 \\ -r_h \sin s_2 \\ 0 \end{pmatrix}. \quad (7.31)$$

An arbitrary tangent vector to  $\partial M$  is then of the form  $\mathbf{V}' = \sum_{j=1}^2 v'_j \mathbf{T}'_j$ . Now if we want to obtain a unit vector normal to  $\partial M$  that is in the tangent space of  $M$ , then we seek a vector

$$\mathbf{N}_{\partial M}(\mathbf{q}) = \sum_{i=1}^3 n_i(\mathbf{q}) \mathbf{T}_i(\mathbf{q}) \quad \text{such that} \quad \mathbf{N}_{\partial M} \cdot \mathbf{T}'_j = 0 \quad \text{and} \quad \mathbf{N}_{\partial M} \cdot \mathbf{N}_{\partial M} = 1.$$

This is a linear algebra problem for each fixed value of  $\mathbf{q}$ . The procedure is to first find  $v_i$  such that the orthogonality condition above is satisfied (with  $v_i$  taking the place of  $n_i$ ), then normalize the result to obtain  $n_i$ . Let  $a_{ji} = \mathbf{T}_i \cdot \mathbf{T}'_j$ . Then the null space of the matrix  $A = [a_{ji}]$  (which is one-dimensional in this example) defines all possible values of  $v_i$  that satisfy the orthogonality constraints. A vector in this null space can be obtained

by multiplying an arbitrary vector with the null-space projector matrix in (A.40). For example, in the case of the upper half of the unit sphere in  $\mathbb{R}^4$  parameterized using  $\phi$ ,

$$A = \begin{pmatrix} \mathbf{T}'_1 \cdot \mathbf{T}_1 & \mathbf{T}'_1 \cdot \mathbf{T}_2 & \mathbf{T}'_1 \cdot \mathbf{T}_3 \\ \mathbf{T}'_2 \cdot \mathbf{T}_1 & \mathbf{T}'_2 \cdot \mathbf{T}_2 & \mathbf{T}'_2 \cdot \mathbf{T}_3 \end{pmatrix} = r_h \cdot \begin{pmatrix} -\sin s_1 \sin s_2 & \cos s_1 \sin s_2 & 0 \\ \cos s_1 \cos s_2 & \sin s_1 \cos s_2 & -\sin s_2 \end{pmatrix}.$$

Choosing the arbitrary vector  $\mathbf{e}_3$  gives

$$\mathbf{v} = [\mathbb{I} - A^T(AA^T)^{-1}A]\mathbf{e}_3 = \cos s_2 \cdot [\cos s_1 \sin s_2, \sin s_1 \sin s_2, \cos s_2]^T.$$

Normalizing the result then gives  $\mathbf{n} = \mathbf{v}/(\mathbf{v}, \mathbf{v})^{\frac{1}{2}}$ , or equivalently  $\mathbf{N}_{\partial M} = \mathbf{V}/(\mathbf{V} \cdot \mathbf{V})^{\frac{1}{2}}$ . These are respectively

$$\mathbf{n}(\phi(\mathbf{s})) = z(s_1, s_2) \begin{pmatrix} \cos s_1 \sin s_2 \\ \sin s_1 \sin s_2 \\ \cos s_2 \end{pmatrix} \quad \text{and} \quad \mathbf{N}_{\partial M} = \begin{pmatrix} \cos s_1 \sin s_2 \\ \sin s_1 \sin s_2 \\ \cos s_2 \\ 0 \end{pmatrix} \quad (7.32)$$

where  $z(s_1, s_2) = [\sin^2 s_2(\cos^2 s_1 \sin^2 s_2 + \sin^2 s_1) + \cos^2 s_2]^{-\frac{1}{2}}$  is the normalization factor required for  $(\mathbf{n}, \mathbf{n}) = \mathbf{n}^T G(\phi)\mathbf{n} = 1$ . This laborious calculation was not really required for this particular example, because  $\partial M$  is obtained in this case by slicing  $S^3$  with a hyper-plane parallel to  $\mathbb{R}^3$ , and it could be guessed that the result will be a sphere embedded in a copy of  $\mathbb{R}^3$  with Cartesian coordinates  $\{q_1, q_2, q_3\}$ , and therefore simply writing the outward-pointing normal to the unit sphere in  $\mathbb{R}^3$  up front would suffice. Or put differently, it is clear by inspection of (7.31) that the normal space of  $\partial M$  consists of linear combinations of the orthogonal unit vectors

$$\mathbf{N}_{\partial M}^{(1)} = \begin{pmatrix} \cos s_1 \sin s_2 \\ \sin s_1 \sin s_2 \\ \cos s_2 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{N}_{\partial M}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

And since  $\mathbf{N}_{\partial M}^{(2)}$  is not in the span of  $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3\}$ , it does not contribute to the normal of  $\partial M$  that is contained in the tangent space of  $M$ . That leaves only  $\mathbf{N}_{\partial M}^{(1)}$ , which is exactly  $\mathbf{N}_{\partial M}$ . While it is always nice when intuition can be used to obtain a solution, it is important to have a general mathematical procedure that can be used when intuition fails. And from the above example it can be useful to see how the general framework reduces to the intuitive result that is expected.

If instead of starting with  $\mathbf{n}$  and substituting  $n_i$  for  $v_i$  in (7.22) to obtain  $\mathbf{N}_{\partial M}$ , in a case such as this where the embedding provides information that can make it easy to calculate  $\mathbf{N}_{\partial M}$  directly, then (7.16) can be used to compute  $\mathbf{n}$  from  $\mathbf{N}_{\partial M}$ . Explicitly, in the case when the coordinates  $\{q_1, q_2, q_3\}$  are used,

$$\mathbf{n}(\mathbf{q}(\mathbf{s})) = [G(\mathbf{q}(\mathbf{s}))]^{-1} \frac{\partial \mathbf{x}^T}{\partial \mathbf{q}} \mathbf{N}_{\partial M} = h \begin{pmatrix} \cos s_1 \sin s_2 \\ \sin s_1 \sin s_2 \\ \cos s_2 \end{pmatrix}. \quad (7.33)$$

Note that this is not the same as  $\mathbf{n}(\phi(\mathbf{s}))$  in (7.32). The reason is that  $G(\mathbf{q})|_{\mathbf{q}=\mathbf{q}(\mathbf{s})} \neq G(\phi)|_{\phi=\phi(\mathbf{s})}$ . While  $\mathbf{n}(\mathbf{q}(\mathbf{s})) \neq \mathbf{n}(\phi(\mathbf{s}))$ , there is nevertheless a relationship between them. This relationship can be established by relating both to the extrinsically defined normal  $\mathbf{N}_{\partial M}$ , which is independent of coordinates. Referring back to (7.28) and restricting both sides to the submanifold  $\partial M$  defined by parameters  $\{s_1, s_2\}$  gives

$$G(\phi(\mathbf{s})) = \left. \frac{\partial \mathbf{q}^T}{\partial \phi} \right|_{\phi=\phi(\mathbf{s})} G(\mathbf{q}(\mathbf{s})) \left. \frac{\partial \mathbf{q}}{\partial \phi^T} \right|_{\phi=\phi(\mathbf{s})} \quad \text{where } G(\mathbf{q}(\mathbf{s})) = G(\mathbf{q}(\phi(\mathbf{s}))).$$

But from vector calculus,  $\mathbf{n}(\mathbf{q}(\mathbf{s}))$  in (7.33) can be rewritten as

$$\mathbf{n}(\mathbf{q}(\mathbf{s})) = [G(\mathbf{q}(\mathbf{s}))]^{-1} \frac{\partial \phi^T}{\partial \mathbf{q}} \frac{\partial \mathbf{x}^T}{\partial \phi} \mathbf{N}_{\partial M}.$$

Using (7.29), and comparing the above expression to

$$\mathbf{n}(\phi(\mathbf{s})) = [G(\phi(\mathbf{s}))]^{-1} \frac{\partial \mathbf{x}^T}{\partial \phi} \mathbf{N}_{\partial M}$$

leads to the conclusion that

$$\boxed{\mathbf{n}(\mathbf{q}(\mathbf{s})) = \left. \frac{\partial \mathbf{q}}{\partial \phi^T} \right|_{\phi=\phi(\mathbf{s})} \mathbf{n}(\phi(\mathbf{s}))}. \quad (7.34)$$

In other words, the conversion between  $\mathbf{n}(\mathbf{q}(\mathbf{s}))$  and  $\mathbf{n}(\phi(\mathbf{s}))$  can be performed with knowledge of only the metric tensor and the Jacobian matrix  $\partial \mathbf{q} / \partial \phi^T$ . And the conversion can be implemented as a push-forward from one coordinate system to the other. This operation is independent of the embedding, which only plays the indirect role of defining the metric.

### 7.3.4 The Divergence Theorem in Coordinates

In the case of a manifold  $M$  that is defined by the smooth embedding  $\mathbf{x}(\mathbf{q}) \in \mathbb{R}^n$  for  $\mathbf{q} \in D \subset \mathbb{R}^m$ , and the smooth boundary  $\partial M$  parameterized by  $\mathbf{q}(\mathbf{s}) \in \partial D \subset \mathbb{R}^m$ , the divergence theorem for  $\mathbb{R}^m$  can be used to define a coordinate-dependent version of the divergence theorem. From the definition in (5.49) (where now  $i \in \{1, \dots, m\}$ ), the integral of the divergence of a vector field  $\mathbf{F} = \sum_i f_i(\mathbf{q}) \frac{\partial \mathbf{x}}{\partial q_i} \in \mathbb{R}^n$  (or equivalently,  $\mathbf{f} = [f_1, \dots, f_m]^T$ ) can be converted as follows:

$$\begin{aligned} \int_M \operatorname{div}(\mathbf{f}) dV &= \int_D \operatorname{div}(\mathbf{f}) |G|^{\frac{1}{2}} dq_1 \dots dq_m \\ &= \int_D \sum_{i=1}^m \frac{\partial}{\partial q_i} (|G|^{\frac{1}{2}} f_i) dq_1 \dots dq_m \\ &= \int_{\partial D} \sum_{i=1}^m f_i(\mathbf{q}(\mathbf{s})) \nu_i(\mathbf{q}(\mathbf{s})) |J^T(\mathbf{s}) G(\mathbf{q}(\mathbf{s})) J(\mathbf{s})|^{\frac{1}{2}} ds_1 \dots ds_{m-1} \\ &= \int_{\partial D} \sum_{i,j=1}^m f_i(\mathbf{q}(\mathbf{s})) g_{ij}(\mathbf{q}(\mathbf{s})) g^{jk}(\mathbf{q}(\mathbf{s})) n_k(\mathbf{q}(\mathbf{s})) \\ &\quad |J^T(\mathbf{s}) G(\mathbf{q}(\mathbf{s})) J(\mathbf{s})|^{\frac{1}{2}} ds_1 \dots ds_{m-1} \\ &= \int_{\partial M} (\mathbf{f}, \mathbf{n}) dS. \end{aligned}$$

Here  $J_{ij} = \partial q_i / \partial s_j$ ,  $\boldsymbol{\nu}(\mathbf{q}(\mathbf{s})) \in \mathbb{R}^m$  is the normal to  $\partial D$ , and

$$dS = |J^T(\mathbf{s}) G(\mathbf{q}(\mathbf{s})) J(\mathbf{s})|^{\frac{1}{2}} ds_1 \dots ds_{m-1} \quad (7.35)$$

is the volume element for  $\partial M$  written in parametric form. The above step that converts the integral over  $D$  to an integral over  $\partial D$  is the divergence theorem for  $\mathbb{R}^m$ . The vector  $\mathbf{n} \in \mathbb{R}^n$  is the unique unit vector on the boundary point  $\mathbf{x} \in \partial M$  that points away from  $M$ , and results from pushing  $\nu(\mathbf{q}(\mathbf{s}))$  forward.

The divergence theorem is now illustrated for the hyper-spherical cap,  $M$ , and the associated bounding sphere,  $\partial M$ , discussed in Section 7.3.2. The bounding submanifold  $\partial M$  (which is a 2-sphere of radius  $r_h$ ) is described in terms of the coordinates  $\mathbf{q} = \mathbf{q}(\mathbf{s})$ , together with the normal expressed in these coordinates, as

$$\mathbf{q}(\mathbf{s}) = r_h \cdot \begin{pmatrix} \cos s_1 \sin s_2 \\ \sin s_1 \sin s_2 \\ \cos s_2 \end{pmatrix} \quad \text{and} \quad \mathbf{n}(\mathbf{q}(\mathbf{s})) = h \cdot \begin{pmatrix} \cos s_1 \sin s_2 \\ \sin s_1 \sin s_2 \\ \cos s_2 \end{pmatrix}. \quad (7.36)$$

The Jacobian transformation for  $\mathbf{q}(\mathbf{s})$  is

$$J(\mathbf{s}) = \frac{\partial \mathbf{q}}{\partial \mathbf{s}^T} = r_h \cdot \begin{pmatrix} -\sin s_1 \sin s_2 & \cos s_1 \cos s_2 \\ \cos s_1 \sin s_2 & \sin s_1 \cos s_2 \\ 0 & -\sin s_2 \end{pmatrix}.$$

Furthermore, it can be shown that after substituting (7.36) into (7.23),

$$dS = |\det J^T(\mathbf{s})G(\mathbf{q}(\mathbf{s}))J(\mathbf{s})|^2 ds_1 ds_2 = (r_h)^2 \sin s_2 ds_1 ds_2.$$

Note that in this particular example it was not necessary to go through the complicated calculation  $\det J^T(\mathbf{s})G(\mathbf{q}(\mathbf{s}))J(\mathbf{s})$  because  $\partial M$  can be parameterized as an embedded 2-manifold in  $\mathbb{R}^4$  as  $\mathbf{x}(\mathbf{q}(\mathbf{s})) = [r_h \cos s_1 \sin s_2, r_h \sin s_1 \sin s_2, r_h \cos s_2, h]^T$ , and its metric tensor can be computed directly. In other words, for this particular boundary manifold,  $J^T(\mathbf{s})G(\mathbf{q}(\mathbf{s}))J(\mathbf{s}) = J^T(\mathbf{s})J(\mathbf{s})$ .

As an example of a vector field, let

$$\mathbf{F}(\mathbf{q}) = (1 - q_1^2 - q_2^2 - q_3^2)^{\frac{1}{2}} q_1 \frac{\partial \mathbf{x}}{\partial q_1} \iff \mathbf{f}(\mathbf{q}) = [(1 - q_1^2 - q_2^2 - q_3^2)^{\frac{1}{2}} q_1, 0, 0]^T.$$

Then

$$\int_M \operatorname{div}(\mathbf{f}) dV = \int_{\|\mathbf{q}\| < r_h} \frac{\partial}{\partial q_1}(q_1) d\mathbf{q} = \frac{4}{3}\pi(r_h)^3,$$

since the integral of the number 1 over the interior of a sphere of radius  $r_h$  is just the volume of that sphere.

On the other hand,

$$\begin{aligned} \int_{\partial M} (\mathbf{f}, \mathbf{n}) dS &= \int_0^\pi \int_0^{2\pi} \left[ \frac{r_h}{h} \cos s_1 \sin s_2 \right] \cdot [h \cos s_1 \sin s_2] \cdot (r_h)^2 \sin s_2 ds_1 ds_2 \\ &= \frac{4}{3}\pi(r_h)^3. \end{aligned}$$

The equality of these two quantities therefore demonstrates the divergence theorem for embedded manifolds.

### 7.3.5 Integration by Parts on an Embedded Manifold

Let  $M$  be a smooth  $m$ -dimensional manifold embedded in  $\mathbb{R}^n$  and let  $\phi : M \rightarrow \mathbb{R}$  be a smooth function. In the discussion below,  $\phi(\mathbf{q})$  will be used as shorthand to denote

$\phi(\mathbf{x}(\mathbf{q}))$  where  $\mathbf{q} \in D \subset \mathbb{R}^m$  since there will be no ambiguity. Let  $\mathbf{v}(\mathbf{q})$  denote a vector field on  $M$  expressed in coordinates.

The extension of the integration-by-parts formulas (A.123) and (A.124) to the case of a differentiable  $m$ -dimensional embedded manifold in  $\mathbb{R}^n$  with  $B$  differentiable bounding sub-manifolds,<sup>7</sup> each of dimension  $m - 1$ , is

$$\int_D \sum_{i,j=1}^m v_i g^{ij} \frac{\partial \phi}{\partial q_j} |G|^{\frac{1}{2}} dq_1 \dots dq_m = \text{b.s.t.} - \int_D \phi \sum_{i,j=1}^m \frac{\partial}{\partial q_j} \left( v_i g^{ij} |G|^{\frac{1}{2}} \right) dq_1 \dots dq_m, \quad (7.37)$$

where the *bounding sub-manifold terms* (b.s.t.) are

$$\text{b.s.t.} =$$

$$\sum_{k=1}^B \int_{\partial D_k} \phi(\mathbf{q}(\mathbf{s}^{(k)})) \sum_{i,j=1}^m v_i(\mathbf{q}(\mathbf{s}^{(k)})) g^{ij}(\mathbf{q}(\mathbf{s}^{(k)})) \nu_j(\mathbf{s}^{(k)}) dS^{(k)}$$

where, in analogy with (7.35),

$$dS^{(k)} = |J^T(\mathbf{q}(\mathbf{s}^{(k)}))G(\mathbf{q}(\mathbf{s}^{(k)}))J(\mathbf{q}(\mathbf{s}^{(k)}))|^{\frac{1}{2}} ds_1^{(k)} \dots ds_{m-1}^{(k)}.$$

Here each  $\mathbf{s}^{(k)} = [s_1^{(k)}, \dots, s_{m-1}^{(k)}]^T$  is a parametrization of the  $k$ th bounding sub-manifold, and  $\nu_i(\mathbf{s}^{(k)})$  is the  $i$ th component of the image of the outward-pointing normal to the sub-manifold in the coordinate domain  $D$ . The Jacobian is defined as in the previous subsection.

This can be quite confusing, and requires some clarification. Recall that the normal directions to the  $m$ -dimensional manifold in  $\mathbb{R}^n$  are defined by  $n - m$  vectors,  $\mathbf{N}_1, \dots, \mathbf{N}_{n-m} \in \mathbb{R}^n$  such that

$$\mathbf{N}_i \cdot \frac{\partial \mathbf{x}}{\partial q_j} = 0$$

for all  $i = 1, \dots, n - m$  and  $j = 1, \dots, m$ . In contrast, a bounding sub-manifold has dimensions  $m - 1$ , with tangent vectors in  $\mathbb{R}^n$  defined by

$$\mathbf{T}'_k = \frac{\partial \mathbf{x}(\mathbf{q}(\mathbf{s}))}{\partial s_k},$$

for  $k = 1, \dots, m - 1$ . (Here the superscripts ( $j$ ) on  $\mathbf{s}^{(j)}$  have been suppressed while the focus is on a single bounding sub-manifold.) There are  $n - m + 1$  normal vectors to each point of this sub-manifold in  $\mathbb{R}^n$ .

The span of  $\{\mathbf{T}'_k\}$  is contained in the span of  $\{\mathbf{T}_i\}$ . The vectors in the normal space satisfy  $\mathbf{N}_l = \sum_i c_{li} \mathbf{T}_i$  under the constraint that  $\mathbf{N}_l \cdot \mathbf{T}'_k = 0$ .

Restricting the discussion to only those normals of the sub-manifold that are contained in the tangent to the original  $m$ -dimensional manifold imposes  $n - m$  constraints, yielding a single outward-pointing normal for each point on each bounding sub-manifold.

In the case of a manifold without boundary, the integration by parts formula (7.37) reduces to

$$\langle \mathbf{v}, \text{grad} \phi \rangle = -\langle \phi, \text{div} \mathbf{v} \rangle \quad (7.38)$$

<sup>7</sup>A sub-manifold of a manifold is itself a manifold of lower dimension.

because the bounding sub-manifold terms are zero. If the bounding sub-manifold terms are not zero, the divergence theorem with  $D$  viewed as a subset of  $\mathbb{R}^m$  together with a *localization argument*<sup>8</sup> yields

$$\operatorname{div}(\phi \mathbf{v}) = \phi \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \operatorname{grad} \phi. \quad (7.39)$$

This is also what would be obtained by directly applying the definition of divergence in (5.49) to the vector field  $\phi \mathbf{v}$ .

The integration-by-parts formula in (7.37) is now demonstrated on the same domain considered in the previous example. Let  $\mathbf{v}(\mathbf{q})$  and  $\phi(\mathbf{q})$  respectively be the vector and scalar fields defined in terms of the coordinates  $\mathbf{q}$  as

$$\sum_{i=1}^m v_i g^{ij} \doteq (1 - q_1^2 - q_2^2 - q_3^2)^{\frac{1}{2}} \delta_{1,j} \quad \text{and} \quad \phi(\mathbf{q}) = q_1. \quad (7.40)$$

Then the left-hand side of (7.37) becomes

$$\int_D \sum_{i,j=1}^3 v_i g^{ij} \frac{\partial \phi}{\partial q_j} |G|^{\frac{1}{2}} dq_1 dq_2 dq_3 = \int_{\|\mathbf{q}\| < r_h} dq_1 dq_2 dq_3 = \frac{4}{3} \pi (r_h)^3.$$

The second term on the right-hand side of (7.37) becomes

$$\int_D \phi \sum_{i,j=1}^3 \frac{\partial}{\partial q_j} \left( v_i g^{ij} |G|^{\frac{1}{2}} \right) dq_1 dq_2 dq_3 = \int_{\|\mathbf{q}\| < r_h} q_1 \frac{\partial}{\partial q_1} (1) dq_1 dq_2 dq_3 = 0.$$

In this case since the boundary consists of a single surface, b.s.t. reduces to

$$\text{b.s.t.} = \int_{\partial D} \phi(\mathbf{q}(\mathbf{s})) \sum_{i,j=1}^3 v_i(\mathbf{q}(\mathbf{s})) g^{ij}(\mathbf{q}(\mathbf{s})) \nu_j(\mathbf{s}) dS,$$

which, when substituting (7.36) and (7.40), reduces to

$$\text{b.s.t.} = \int_0^\pi \int_0^{2\pi} [r_h \cos s_1 \sin s_2] [\cos s_1 \sin s_2] (r_h)^2 \sin s_2 ds_1 ds_2 = \frac{4}{3} \pi (r_h)^3.$$

This therefore demonstrates the integration-by-parts formula (7.37).<sup>9</sup>

<sup>8</sup>There are two main kinds of localization arguments. The first is that if  $\int_D f dV = 0$  over a wide enough range of domains,  $D$ , then  $f = 0$ . The second is that for fixed  $D$  if  $\int_D f \phi dV = 0$  for a wide enough range of functions  $\phi$ , then  $f = 0$ . The first can be viewed as a subset of the second in which  $\phi$  is a window function that is equal to unity on various domains and zero otherwise.

<sup>9</sup>The reader should not be left with the impression that the volume of the sphere appears in every such calculation! These examples were chosen in a way so as to minimize the number of complicated integrals that need to be evaluated. While the formulas for the divergence theorem, integration by parts, etc., are general, computing complicated integrals adds little to the understanding of the underlying concepts.



### 7.3.6 Curvature

Once the metric tensor  $G = [g_{ij}]$  for an  $m$ -dimensional manifold in  $\mathbb{R}^n$  has been obtained as a matrix in a particular coordinate system, the Christoffel symbols  $\Gamma_{ij}^k$  and Riemannian curvature tensor  $R_{ijk}^l$  defined in (5.59) and (5.61) can be calculated. The only difference between the calculation for an  $m$ -manifold and a surface is that now indices range from 1 to  $m$  rather than from 1 to 2.

Given the Riemannian curvature tensor  $R_{ijk}^l$ ,

$$R_{ijkl} \doteq \sum_{l'} R_{ijk}^{l'} g^{l'l}. \quad (7.41)$$

It can be shown that this has the symmetries [20, 43, 48]

$$R_{ijkl} = R_{klij} \quad \text{and} \quad R_{ijkl} = -R_{jikl} = -R_{ijlk}$$

and obeys the (*first*) *Bianchi identity*:

$$R_{ijkl} + R_{kijl} + R_{jkil} = 0.$$

The *Ricci curvature tensor*,  $Ric(G) = [R_{ij}]$ , is obtained by contracting the Riemannian curvature tensor  $R_{ijk}^l$  as

$$R_{ij} \doteq \sum_k R_{ikj}^k = \sum_k R_{kij}^k = R_{ji}. \quad (7.42)$$

The notation  $Ric(G)$  indicates that it is completely determined by the metric tensor. The *scalar curvature* is computed from the Ricci curvature tensor as

$$k \doteq \sum_{i,j} g^{ij} R_{ij}. \quad (7.43)$$

Given four vector fields  $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}$  defined on an embedded manifold in the same way that  $\mathbf{f}$  is defined in (7.14), it is convenient to define

$$R(\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}) \doteq \sum_{ijkl} R_{ijkl} u_i v_j w_k z_l. \quad (7.44)$$

This, together with the pointwise inner product defined in (7.15) is used to define the *sectional curvature* associated with two vector fields,  $\mathbf{v}, \mathbf{w}$  on  $M$  [51]:

$$\kappa(\mathbf{v}, \mathbf{w}) \doteq \frac{R(\mathbf{v}, \mathbf{w}, \mathbf{v}, \mathbf{w})}{(\mathbf{v}, \mathbf{v})(\mathbf{w}, \mathbf{w}) - (\mathbf{v}, \mathbf{w})^2} \quad (7.45)$$

when  $\mathbf{v}$  and  $\mathbf{w}$  are orthogonal to each other this reduces to  $\kappa(\mathbf{v}, \mathbf{w}) = R(\mathbf{v}, \mathbf{w}, \mathbf{v}, \mathbf{w})$ . Given an orthonormal basis  $\{\mathbf{u}_i\}$  for the tangent space at  $\mathbf{x}(\mathbf{q})$ , the matrix with entries  $\kappa_{ij} = \kappa(\mathbf{u}_i, \mathbf{u}_j)$  contains the same information about the local geometry of the manifold as does the Riemannian and Ricci curvature tensors.

## 7.4 Covariant vs. Contravariant

### 7.4.1 Tensors

The volume element for a manifold can be expressed in two different coordinate systems,  $\{q_1, q_2, \dots, q_n\}$  and  $\{q'_1, q'_2, \dots, q'_n\}$ , and equated where the coordinates are compatible as

$$|G'(q')|^{\frac{1}{2}} dq'_1 dq'_2 \dots dq'_n = |G(q)|^{\frac{1}{2}} dq_1 dq_2 \dots dq_n.$$

Likewise, a differential length element,  $ds$ , for a curve can be computed in two different coordinate systems as

$$ds^2 = d\mathbf{q}^T G(\mathbf{q}) d\mathbf{q} = d\mathbf{q}'^T G(\mathbf{q}') d\mathbf{q}' = ds'^2. \quad (7.46)$$

In general a scalar function on a manifold can be defined relative to coordinates, and related to new coordinates as

$$f(\mathbf{q}) = f'(\mathbf{q}') \quad \text{where} \quad \mathbf{q}' = \mathbf{q}'(\mathbf{q}) \quad \text{and} \quad \mathbf{q} = \mathbf{q}(\mathbf{q}'). \quad (7.47)$$

From the chain rule,

$$d\mathbf{q}' = \frac{\partial \mathbf{q}'}{\partial \mathbf{q}^T} d\mathbf{q}. \quad (7.48)$$

This provides a rule for the conversion of the one-dimensional arrays of coordinate changes,  $d\mathbf{q}$  and  $d\mathbf{q}'$  (which can be thought of as column vectors). More generally, given a column vector that is a function on the manifold, expressed as  $\mathbf{v}(\mathbf{q})$ , then if the corresponding vector in the coordinates  $\mathbf{q}'$ , which is denoted as  $\mathbf{v}'(\mathbf{q}')$ , transforms in analogy with the chain rule, as

$$\mathbf{v}' = \frac{\partial \mathbf{q}'}{\partial \mathbf{q}^T} \mathbf{v}, \quad (7.49)$$

then  $\mathbf{v}$  (and  $\mathbf{v}'$ ) is called a *contravariant vector* with components  $v_i = v_i(\mathbf{q})$  (and likewise  $v'_i = v'_i(\mathbf{q}')$ ).

In contrast, given a function (also called a scalar field) on a manifold of the form in (7.47), then the chain rule gives

$$\frac{\partial f'}{\partial \mathbf{q}'^T} = \frac{\partial f}{\partial \mathbf{q}^T} \frac{\partial \mathbf{q}}{\partial \mathbf{q}'^T}. \quad (7.50)$$

This is a rule for transforming gradients, which are viewed here as row vectors. The generalization of (7.50) to row vectors other than gradients is

$$\mathbf{v}'^T = \mathbf{v}^T \frac{\partial \mathbf{q}}{\partial \mathbf{q}'^T} \quad \text{or} \quad \mathbf{v}' = \frac{\partial \mathbf{q}^T}{\partial \mathbf{q}'} \mathbf{v}. \quad (7.51)$$

In general a quantity such as  $\mathbf{v}$  (or  $\mathbf{v}'$ ) that follows the above transformation rule is called a *covariant vector* expressed in coordinates  $\mathbf{q}$  (or  $\mathbf{q}'$ ).

The concepts of co- and contra-variance are not limited to vector quantities. Referring back to (7.46), it is clear from (7.48) that

$$G'(\mathbf{q}') = \left( \frac{\partial \mathbf{q}}{\partial \mathbf{q}'^T} \right)^T G(\mathbf{q}) \frac{\partial \mathbf{q}}{\partial \mathbf{q}'^T} \quad \text{or} \quad g'_{ij} = \sum_{k,l=1}^n g_{kl} \frac{\partial q_k}{\partial q'_i} \frac{\partial q_l}{\partial q'_j}. \quad (7.52)$$

Taking the square root of the determinant of the above expression gives

$$|G'(\mathbf{q}')|^{\frac{1}{2}} = \left| \frac{\partial \mathbf{q}}{\partial \mathbf{q}'^T} \right| \cdot |G(\mathbf{q})|^{\frac{1}{2}}. \quad (7.53)$$

But since differential  $n$ -forms transform as

$$dq'_1 \wedge dq'_2 \wedge \dots \wedge dq'_n = \left| \frac{\partial \mathbf{q}'}{\partial \mathbf{q}^T} \right| dq_1 \wedge dq_2 \wedge \dots \wedge dq_n,$$

it follows that

$$\boxed{|G'(\mathbf{q}')|^{\frac{1}{2}} dq'_1 \wedge dq'_2 \wedge \dots \wedge dq'_n = |G(\mathbf{q})|^{\frac{1}{2}} dq_1 \wedge dq_2 \wedge \dots \wedge dq_n.} \quad (7.54)$$

In other words, this is invariant under coordinate changes.

The inverse of (7.52) defines a transformation rule

$$[G'(\mathbf{q}')]^{-1} = \left( \frac{\partial \mathbf{q}}{\partial \mathbf{q}'^T} \right)^{-1} [G(\mathbf{q})]^{-1} \left( \frac{\partial \mathbf{q}}{\partial \mathbf{q}'^T} \right)^{-T} = \frac{\partial \mathbf{q}'}{\partial \mathbf{q}^T} [G(\mathbf{q})]^{-1} \left( \frac{\partial \mathbf{q}'}{\partial \mathbf{q}^T} \right)^T \quad (7.55)$$

written in component form as

$$g'^{ij} = \sum_{k,l=1}^n g^{kl} \frac{\partial q'_i}{\partial q_k} \frac{\partial q'_j}{\partial q_l}. \quad (7.56)$$

More generally a mixed tensor of contravariant valency  $r$  and covariant valency  $s$ , denoted here as  $A$ , is a quantity expressed in terms of coordinates  $(q_1, \dots, q_n)$  as an array of  $n^{r+s}$  scalar functions  $a_{j_1, j_2, \dots, j_s}^{i_1, i_2, \dots, i_r}(q_1, \dots, q_n)$  such that the corresponding array of functions defined by a change of coordinates  $(q_1, \dots, q_n) \rightarrow (q'_1, \dots, q'_n)$  satisfies

$$a'_{j_1, j_2, \dots, j_s}^{i_1, i_2, \dots, i_r} = \sum_{\substack{k_1, k_2, \dots, k_r \\ l_1, l_2, \dots, l_s}} a_{l_1, l_2, \dots, l_s}^{k_1, k_2, \dots, k_r} \frac{\partial q'_{i_1}}{\partial q_{k_1}} \dots \frac{\partial q'_{i_r}}{\partial q_{k_r}} \cdot \frac{\partial q_{l_1}}{\partial q'_{j_1}} \dots \frac{\partial q_{l_s}}{\partial q'_{j_s}} \quad (7.57)$$

where each  $a'_{j_1, j_2, \dots, j_s}^{i_1, i_2, \dots, i_r}$  is a function of  $(q'_1, \dots, q'_n)$  and  $a_{l_1, l_2, \dots, l_s}^{k_1, k_2, \dots, k_r}$  is a function of  $(q_1, \dots, q_n)$ . When a tensor is (purely) covariant,  $r = 0$ , and when a tensor is (purely) contravariant,  $s = 0$ . The sum  $r + s$  is called the rank of the tensor. A scalar is a tensor of rank zero, and a vector is a tensor of rank 1.  $G = [g_{ij}]$  is a purely covariant tensor of rank 2, and  $G^{-1} = [g^{ij}]$  is a purely contravariant tensor of rank 2.  $G$  and  $G^{-1}$  are very special tensors because they can be used to change the valence of any tensor. For example, starting with a mixed rank three tensor defined by the functions  $\{a_{jk}^s\}$ , the metric tensor can be used to obtain a purely covariant tensor as

$$a_{ijk} = \sum_{s=1}^n a_{jk}^s g_{is}.$$

#### 7.4.2 Derivatives and Differentials

If  $G = [g_{ij}]$  and  $G^{-1} = [g^{ij}]$  are expressed in coordinates  $(q_1, \dots, q_n)$ , then the derivatives of these entries with respect to coordinates can be expressed in terms of themselves and the Christoffel symbols as [69]

$$\frac{\partial g_{jk}}{\partial q_l} = \sum_s (\Gamma_{jl}^s g_{sk} + g_{js} \Gamma_{kl}^s) \quad \text{and} \quad \frac{\partial g^{ij}}{\partial q_k} = - \sum_s (\Gamma_{sk}^i g^{sj} + g^{is} \Gamma_{sk}^j).$$

The *covariant differential of a contravariant vector* with  $i$ th entry  $v^i$  is a new contravariant vector with  $i$ th entry  $Dv^i$  defined as [69]

$$Dv^i \doteq dv^i + \sum_{jk} v^j \Gamma_{jk}^i dq_k. \tag{7.58}$$

The *covariant differential of a covariant vector* with  $i$ th entry  $v_j$  is a new covariant vector with  $j$ th entry  $Dv_j$  defined as [69]

$$Dv_j \doteq dv_j - \sum_{ik} v_i \Gamma_{jk}^i dq_k. \tag{7.59}$$

In contrast, the *covariant derivative of contravariant and covariant vectors* are respectively defined in component form as [69]

$$v^i_{;k} = \frac{\partial v^i}{\partial q_k} + \sum_j v^j \Gamma_{jk}^i \tag{7.60}$$

and

$$v_{j;k} = \frac{\partial v_j}{\partial q_k} - \sum_i v_i \Gamma_{jk}^i. \tag{7.61}$$

The first of these is a mixed second-order tensor, and the second is purely covariant.

Differentials and derivatives are related by the expressions

$$Dv^i = \sum_k v^i_{;k} dq_k \quad \text{and} \quad Dv_j = \sum_k v_{j;k} dq_k.$$

The covariant derivative of any covariant tensor can be defined in an analogous way. For example, the covariant derivative of  $a_{jk}$  is

$$a_{jk;l} \doteq \frac{\partial a_{jk}}{\partial q_l} - \sum_s (a_{sk} \Gamma_{jl}^s + a_{js} \Gamma_{kl}^s).$$

Since  $v_{j;k}$  is covariant, its second covariant derivative can be defined using the above definition. The Riemannian curvature tensor can then be viewed as the four-index array of scalar functions such that [69]

$$v_{j;k;l} - v_{j;l;k} = \sum_i v_i R^i_{jkl} \tag{7.62}$$

for any covariant vector  $v_i$ . The Riemannian curvature tensor  $R^i_{jkl}$  is a rank four tensor with covariant latency of three and contravariant latency of one.

The discussion of embedded manifolds presented in the previous section is reformulated in the language of differential forms in the next section.

## 7.5 Embedded Manifolds, Part II: Using Differential Forms

Consider the same situation as in the previous sections, now using the notation of differential forms. Let  $M \subset \mathbb{R}^m$  denote a connected and bounded domain in which a vector  $\mathbf{q} \in M$  is allowed to roam. Suppose that there is a mapping  $\mathbf{x} : M \rightarrow N \subset \mathbb{R}^n$  where  $n > m$ . The result is analogous to an  $m$ -dimensional parameterized “surface” in an  $n$ -dimensional space. The use of the word *surface* here is really not correct, since it implies that  $m = 2$ , much like the word *curve* corresponds to  $m = 1$ , regardless of the value of  $n$ . The word *embedding* will be used here to denote the generalization of the mapping that defines a curve or surface in  $\mathbb{R}^n$ . The geometrical object resulting from an embedding locally “looks like”  $\mathbb{R}^m$ , and is called an “ $m$ -dimensional manifold embedded in  $\mathbb{R}^n$ .” A precise mathematical definition of the word *manifold* will be provided later. For now, a manifold can be thought of as an embedded manifold<sup>10</sup> for which the following two properties hold:

$$\mathbf{x}(\mathbf{q}_1) = \mathbf{x}(\mathbf{q}_2) \implies \mathbf{q}_1 = \mathbf{q}_2 \quad (7.63)$$

and

$$\text{rank} \left( \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} \right) = m \quad \forall \mathbf{q} \in M. \quad (7.64)$$

These conditions guarantee that the embedded manifold observes constraints analogous to those imposed for simple curves in the plane and simply connected surfaces in  $\mathbb{R}^3$ . If (7.64) holds but (7.63) does not, then  $\mathbf{x} : M \rightarrow N$  is called an *immersion*.

### 7.5.1 Push-Forwards and Pull-Backs (Revisited)

Let  $\mathbf{q} \in M \subset \mathbb{R}^m$  and  $\mathbf{x} \in N \subset \mathbb{R}^n$ . Let  $\gamma : [0, 1] \rightarrow M$  be a differentiable mapping. In other words,  $\gamma(t)$  for  $t \in [0, 1]$  is a differentiable curve segment that exists in a part of  $\mathbb{R}^m$  denoted as  $M$ . Let  $f : N \rightarrow \mathbb{R}$  be a differentiable function. Let  $\psi : M \rightarrow N$  be a differentiable mapping. That is,  $\mathbf{x} = \psi(\mathbf{q})$ . This could also be written as  $\mathbf{x} = \mathbf{x}(\mathbf{q})$ , but in order to be consistent with the literature, the former notation in which the mapping and the result of the mapping are denoted with different symbols.

Define

$$\psi_* = \psi \circ \gamma \quad \text{such that} \quad \psi_* : [0, 1] \rightarrow N.$$

This is the image of the curve  $\gamma(t) \in M$  as it looks in  $N$ , i.e.,  $\psi_*(t) = \psi(\gamma(t))$ . It is called the “push-forward of  $\gamma$  by  $\psi$ .” Using the notation of the previous section, another way to denote the same thing is  $\psi_*(t) = \mathbf{x}(\mathbf{q}(t))$  where the curve is denoted as  $\mathbf{q} = \mathbf{q}(t)$  (rather than introducing the new name  $\gamma$  and writing  $\mathbf{q} = \gamma(t)$ ) and  $\mathbf{x} = \mathbf{x}(\mathbf{q})$  (rather than introducing the new name  $\psi$  for the mapping).

Now define

$$\psi^* = f \circ \psi \quad \text{such that} \quad \psi^* : M \rightarrow \mathbb{R}.$$

---

<sup>10</sup>From a pedagogical perspective, it might seem backwards to define “embedded manifold” first and “manifold” later, but there are benefits to this approach. For example, when defining the concept of “dog” to a child, a natural thing to do is to point to a “shaggy dog,” a “big dog,” a “sled dog,” etc., and then the intuition behind the concept of “dog” will emerge. In contrast, while the top down approach of first defining the concept of “animal” followed by the concept of “mammal” and then defining a dog as a mammalian animal that is a member of the canine genus and subspecies *canis lupus familiaris* may be more precise, that level of precision would not add much to the child’s understanding.

This is called the “pull-back of  $f$  by  $\psi$ .” For example, for each fixed value of  $t$  the mass density  $\rho^*(\mathbf{x}, t)$  defined in (1.41) is the pull-back of  $\rho(\mathbf{X}, t)$  under the map  $\mathbf{X}(\mathbf{x}, t)$ .

Push-forwards and pull-backs are dual operations in the sense that the former takes an object from a subset of the real line and produces an object in a higher-dimensional space (i.e., a curve segment), and the latter takes points in a high-dimensional space and returns a value on the real line (i.e., it is a function).

The tangent to the pushed-forward curve  $(\psi \circ \gamma)(t) = \psi(\gamma(t))$  is given by the chain rule as

$$\frac{d(\psi \circ \gamma)}{dt} = [D\psi]_{\gamma(t)} \gamma'(t)$$

where

$$[D\psi]_{\gamma(t)} = \left. \frac{\partial \psi}{\partial \mathbf{q}^T} \right|_{\mathbf{q}=\gamma(t)} \quad \text{and} \quad \gamma'(t) = \frac{d\gamma}{dt}.$$

Alternatively, this could be written as

$$[D\psi]_{\gamma(t)} = \left. \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} \right|_{\mathbf{q}(t)}.$$

In this notation, the differential of a pulled-back function is

$$d(f \circ \psi) = \frac{\partial}{\partial \mathbf{x}^T} (f(\psi(\mathbf{q}))) d\mathbf{q} = \frac{\partial f}{\partial \mathbf{x}^T} \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} d\mathbf{q} = d\mathbf{q}^T [D\mathbf{x}]^T \nabla_{\mathbf{x}} f.$$

### 7.5.2 Expressing Pull-Backs of Forms in Coordinates

Let  $k \leq \min(m, n)$ . Let  $\mathbf{x} = \mathbf{x}(\mathbf{q})$  (or equivalently  $\mathbf{x} = \psi(\mathbf{q})$ ) and  $\psi : M \rightarrow N$  where  $M \subset \mathbb{R}^m$  and  $N \subset \mathbb{R}^n$ . Let  $\omega_k$  be a  $k$ -form on  $N$ , written explicitly as

$$\omega_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} a_{i_1, \dots, i_k}(\mathbf{x}) dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k}. \tag{7.65}$$

From the chain rule,

$$dx_{i_j} = \mathbf{e}_{i_j}^T \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} d\mathbf{q}.$$

Therefore, pulling back this form to the coordinate patch  $M \ni \mathbf{q}$  yields

$$\alpha_k \doteq \psi^* \omega_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} a_{i_1, \dots, i_k}(\mathbf{x}(\mathbf{q})) \left( \mathbf{e}_{i_1}^T \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} d\mathbf{q} \right) \wedge \dots \wedge \left( \mathbf{e}_{i_k}^T \frac{\partial \mathbf{x}}{\partial \mathbf{q}^T} d\mathbf{q} \right). \tag{7.66}$$

This pulled-back  $k$ -form can be written as

$$\alpha_k = \sum_{1 \leq j_1 < \dots < j_k \leq m} \tilde{a}_{j_1, \dots, j_k}(\mathbf{q}) dq_{j_1} \wedge dq_{j_2} \wedge \dots \wedge dq_{j_k} \tag{7.67}$$

where  $\tilde{a}_{j_1, \dots, j_k}(\mathbf{q})$  results from collecting all of the Jacobian factors and combining with  $a_{i_1, \dots, i_k}(\mathbf{x}(\mathbf{q}))$ .

According to Schreiber [56], (7.66) can be written explicitly as

$$\psi^* \omega_k = \sum_{\substack{1 \leq i_1 < i_2 < \dots < i_k \leq n \\ 1 \leq j_1 < j_2 < \dots < j_k \leq m}} a_{i_1, \dots, i_k}(\mathbf{x}(\mathbf{q})) \frac{\partial(x_{i_1}, \dots, x_{i_k})}{\partial(q_{j_1}, \dots, q_{j_k})} dq_{j_1} \wedge dq_{j_2} \wedge \dots \wedge dq_{j_k} \tag{7.68}$$

where  $\partial(x_{i_1}, \dots, x_{i_k})/\partial(q_{j_1}, \dots, q_{j_k})$  is the determinant of the particular  $k \times k$  minor of the full Jacobian matrix with entries  $\partial x_{i_r}/\partial q_{j_s}$  where  $r$  and  $s$  run from 1 to  $k$ . Equation (7.68) results from

$$dx_{i_1} \wedge dx_{i_2} \wedge \dots \wedge dx_{i_k} = \sum_{1 \leq j_1 < j_2 < \dots < j_k \leq m} \frac{\partial(x_{i_1}, \dots, x_{i_k})}{\partial(q_{j_1}, \dots, q_{j_k})} dq_{j_1} \wedge dq_{j_2} \wedge \dots \wedge dq_{j_k}. \quad (7.69)$$

Therefore, comparing (7.66), (7.67), and (7.68), it becomes clear that

$$\tilde{a}_{j_1, \dots, j_k}(\mathbf{q}) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} a_{i_1, \dots, i_k}(\mathbf{x}(\mathbf{q})) \frac{\partial(x_{i_1}, \dots, x_{i_k})}{\partial(q_{j_1}, \dots, q_{j_k})}. \quad (7.70)$$

When there is only one mapping  $\psi : M \rightarrow N$ , in a particular problem it is convenient to simply use the  $\mathbf{x}(\mathbf{q})$  notation rather than  $\psi(\mathbf{q})$ , as was done above. However, when there are multiple mappings from  $M$  to  $N$ , this shorthand can lead to confusion.

### 7.5.3 Volume Element of an Embedded Manifold

In the special case when  $k = m \leq n$ , then all choices  $dq_{j_1} \wedge dq_{j_2} \wedge \dots \wedge dq_{j_m}$  for  $j_1 < j_2 < \dots < j_m$  reduce to  $dq_1 \wedge dq_2 \wedge \dots \wedge dq_m$ . Let

$$v_{i_1, \dots, i_m}(\mathbf{x}(\mathbf{q})) \doteq \frac{\partial(x_{i_1}, \dots, x_{i_m})}{\partial(q_1, \dots, q_m)}$$

for  $i_1 < i_2 < \dots < i_m$ . Let these  $\binom{n}{m}$  functions be viewed as the entries of a long vector,

$$\mathbf{v}(\mathbf{x}) \in \mathbb{R}^{\binom{n}{m}}. \quad (7.71)$$

The order in which these entries are arranged is unimportant in the current discussion.

Let  $\nu_m$  (rather than  $\omega_k$ ) be the form that results from letting  $v/\|\mathbf{v}\|$  be substituted for  $a$  in (7.65) when  $k = m$ . Expressed in terms of the coordinates  $\mathbf{q}$ , this becomes

$$a_{i_1, \dots, i_m}(\mathbf{x}(\mathbf{q})) = \frac{v_{i_1, \dots, i_m}(\mathbf{x}(\mathbf{q}))}{\|\mathbf{v}\|} = \frac{\frac{\partial(x_{i_1}, \dots, x_{i_m})}{\partial(q_1, \dots, q_m)}}{\left( \sum_{1 \leq l_1 < l_2 < \dots < l_m \leq m} \left| \frac{\partial(x_{l_1}, \dots, x_{l_m})}{\partial(q_1, \dots, q_m)} \right|^2 \right)^{\frac{1}{2}}}. \quad (7.72)$$

Then the pull-back of the form  $\nu_m$  can be calculated using (7.68). This is

$$\begin{aligned} \psi^* \nu_m &= \left[ \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq n} \left| \frac{\partial(x_{i_1}, \dots, x_{i_m})}{\partial(q_1, \dots, q_m)} \right|^2 \right]^{\frac{1}{2}} dq_1 \wedge dq_2 \wedge \dots \wedge dq_m \\ &= \|\mathbf{v}\| dq_1 \wedge dq_2 \wedge \dots \wedge dq_m, \end{aligned} \quad (7.73)$$

and this defines the volume integral for the embedded manifold:

$$\int_{\psi(M)} \nu_m = \int_M \psi^* \nu_m = \int_M dV. \quad (7.74)$$

The  $dV$  in the last equality is the volume element for the manifold, which can be computed in coordinates as discussed earlier.

A natural issue to address at this point is why the volume element can be written as in (7.73) on the one hand, and as  $|G|^{\frac{1}{2}}dq_1dq_2 \dots dq_m$  on the other. The reason for this is that given an  $n \times m$  matrix  $J$  with  $m < n$ , if  $J_i$  denotes the  $i$ th of the  $m \times m$  minors of this matrix, then

$$\det(J^T J) = \sum_{i=1}^{\binom{n}{m}} |\det J_i|^2. \tag{7.75}$$

This fact from linear algebra is independent of the way the minors are labeled as long as every minor is represented exactly once in the summation. For example, if

$$J = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \quad \text{then} \quad J^T J = \begin{pmatrix} a_{11}^2 + a_{21}^2 + a_{31}^2 & a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} \\ a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} & a_{12}^2 + a_{22}^2 + a_{32}^2 \end{pmatrix}$$

and one way to order the minors is

$$J_1 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}; \quad J_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{pmatrix}; \quad J_3 = \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

A straightforward calculation then shows that

$$\det(J^T J) = |\det J_1|^2 + |\det J_2|^2 + |\det J_3|^2,$$

which is a special case of (7.75).

This linear algebraic fact is useful in the setting of differential geometry because when  $J = \partial \mathbf{x} / \partial \mathbf{q}^T$  is the Jacobian for an  $m$ -dimensional smooth embedded manifold in  $\mathbb{R}^n$ , the metric tensor is written as  $G = J^T J$  and (7.75) becomes

$$\det G = \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq n} \left| \frac{\partial(x_{i_1}, \dots, x_{i_m})}{\partial(q_1, \dots, q_m)} \right|^2. \tag{7.76}$$

### 7.5.4 Conversion to Vector Notation

The  $\binom{n}{k}$  coefficients  $a_{i_1, \dots, i_k}$  for  $1 \leq i_1 < i_2 < \dots < i_k \leq n$  that define a multi-vector in  $A^k(\mathbb{R}^n)$  can be thought of as a column vector in  $\binom{n}{k}$ -dimensional space. In other words, in analogy with the way  $\vee$  operations were defined elsewhere in the text to convert an  $m \times n$  matrix into an  $m \cdot n$ -dimensional vector, a different  $\vee$  operator can be defined in the present context such that

$$\vee : A^k(\mathbb{R}^n) \rightarrow \mathbb{R}^{\binom{n}{k}}. \tag{7.77}$$

This is reminiscent of (7.71), though in that context  $\vee : \Omega^m(\mathbb{R}^n) \rightarrow \mathbb{R}^{\binom{n}{k}}$ . In other words, the object on which the  $\vee$  operation is acting is different in these two cases. If



$\mathbf{a} \in \mathbb{R} \binom{n}{k}$  is the vector resulting from a mapping such as (7.77) (which amounts to an arrangement of the coefficients  $a_{i_1, \dots, i_k}$  in a single column), then doing the same to the coefficients

$$v_{i_1, \dots, i_m}(\mathbf{q}) \doteq \frac{\partial(x_{i_1}, \dots, x_{i_m})}{\partial(q_1, \dots, q_m)}$$

will produce a vector  $\mathbf{v} \in \mathbb{R} \binom{n}{k}$ .

Let  $D \subset \mathbb{R}^m$  denote the coordinate domain that parameterizes the manifold, i.e.,  $\mathbf{x} : D \rightarrow M$ . Following Schreiber [56],

$$\int_{\psi(M)} \omega = \int_{\mathbf{q} \in M} a_{i_1, \dots, i_m}(\mathbf{x}(\mathbf{q})) v_{i_1, \dots, i_m}(\mathbf{q}) dq_1 \wedge dq_2 \wedge \dots \wedge dq_m \quad (7.78)$$

$$= \int_{\mathbf{q} \in D} \mathbf{a} \cdot \mathbf{v} dq_1 \wedge dq_2 \wedge \dots \wedge dq_m \quad (7.79)$$

$$= \int_{\mathbf{q} \in D} \mathbf{a} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} \|\mathbf{v}\| dq_1 \wedge dq_2 \wedge \dots \wedge dq_m \quad (7.80)$$

$$= \int_{\mathbf{q} \in D} \mathbf{a} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} |G|^{\frac{1}{2}} dq_1 dq_2 \dots dq_m \quad (7.81)$$

$$= \int_M \mathbf{a} \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|} dV. \quad (7.82)$$

### 7.5.5 General Properties of Differential Forms on Embedded Manifolds

If  $\omega_i$  and  $\alpha_i$  are  $r$  and  $s$  forms, respectively, and  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , then it can be shown that

$$\psi^*(\omega_1 + \omega_2) = \psi^*\omega_1 + \psi^*\omega_2 \quad (7.83)$$

$$\psi^*(f\omega) = \psi^*(f)\psi^*(\omega) \quad (7.84)$$

$$\omega \wedge \alpha = (-1)^{rs} \alpha \wedge \omega \quad (7.85)$$

$$\psi^*(\omega \wedge \alpha) = \psi^*(\omega) \wedge \psi^*(\alpha). \quad (7.86)$$

In addition, if  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^p$ , then  $\phi \circ \psi : \mathbb{R}^m \rightarrow \mathbb{R}^p$  and

$$(\phi \circ \psi)^*(\omega) = \psi^*(\phi^*(\omega)). \quad (7.87)$$

The exterior derivative of a differential form has the following properties:

$$d(\omega \wedge \alpha) = d\omega \wedge \alpha + (-1)^r \omega \wedge d\alpha \quad (7.88)$$

$$d(\psi^*\omega) = \psi^*(d\omega). \quad (7.89)$$

The proofs of some of the properties in (7.83)–(7.89) are left as exercises. They can be found in books on forms and calculus on manifolds such as [1, 6, 29, 38, 45, 56, 58]. The following section uses these properties in the context of manifolds that are not necessarily embedded.

## 7.6 Intrinsic Description of Riemannian Manifolds

Manifolds are one of the central objects studied in modern geometry. A manifold can be thought of as the generalization of simple curves and surfaces.<sup>11</sup> An  $n$ -dimensional manifold locally “looks like”  $\mathbb{R}^n$  in the sense that there is an invertible mapping between open subsets containing each point in a manifold, and open subsets in  $\mathbb{R}^n$ . There is no unique way to measure distance between points in an arbitrary abstract manifold. This requires the introduction of a Riemannian metric, after which point it is possible to measure the distance between points in a manifold. It is known that any  $n$ -dimensional manifold can be viewed as an “ $n$ -dimensional simple surface” that sits in  $\mathbb{R}^{2n+1}$  “in some way” [67].<sup>12</sup> In the special case when an  $n$ -dimensional manifold is embedded in  $\mathbb{R}^{n+1}$ , then the manifold is called a *hyper-surface*. For the case of manifolds embedded in a higher-dimensional Euclidean space, the way that it is embedded defines a Riemannian metric.

One way to measure distance between distant points in a manifold embedded in a Euclidean space would be the straight-line distance between the points using the norm of the difference of their vector positions in that space. This is one of the less elegant ways of measuring distance between points in a manifold. But for points that are close in this metric it is not a bad way to measure distance, and can be used to define a Riemannian metric tensor.

In principle, the coordinate-dependent extrinsic formulation of geometry for curves and surfaces in  $\mathbb{R}^n$  used in Chapter 5 and earlier in this chapter can be used for manifolds also. However, this is not the approach that is favored in modern mathematics. And so, to relate the methods developed in later chapters to the more popular intrinsic coordinate-free approach to modern geometry, some review is provided here.

Many excellent texts exist on modern differential geometry and differential topology of manifolds. These include Kobayashi and Nomizu [38], Guillemin and Pollack [29], Warner [64], and most recently Tu [63]. The definitions that are reviewed below can be found in any of these texts.

Let  $M$  be an  $n$ -dimensional manifold, as understood by the intuitive description provided earlier.<sup>13</sup> In order to make a precise mathematical definition, some additional ideas must first be introduced. First, an open neighborhood about any point in an embedded manifold can always be constructed by intersecting the manifold with an open ball in  $\mathbb{R}^{2n+1}$  centered on the point of interest. This fact is independent of the details of the embedding. Or, using a distance function between points  $x, u \in M$ ,  $U$  can be defined as the set of all points  $u$  such that  $0 < d(x, u) < \epsilon \in \mathbb{R}_{>0}$ . An  *$n$ -dimensional*

---

<sup>11</sup>The word *simple* denotes that a curve or surface does not intersect itself or form branches. If it did, then there could not be an invertible mapping between the subset containing the intersection or branch point with an open subset of the real line or plane.

<sup>12</sup>However, the way in which such a manifold is set into the higher-dimensional Euclidean space is not unique. As is illustrated by the example of a knot in  $\mathbb{R}^3$  (which is topologically the same manifold as a circle), the essence of some problems is to get not just “an” embedding, but rather the “right one.”

<sup>13</sup>Previously the dimension of  $M$  was denoted as  $m$  and it was defined to be embedded in  $\mathbb{R}^n$ . In the present context, there is no explicit embedding and  $n$  is used to denote the dimension of  $M$ . From the famous theorems of Whitney and Nash [50, 67] it is always possible to embed an  $n$ -dimensional manifold in  $\mathbb{R}^{2n+1}$ .

*proper coordinate chart about*  $x \in M$  is the pair  $(U, \phi)$  where  $U$  is an open neighborhood of  $x$  and  $\phi$  is an invertible mapping of the form  $\phi : U \rightarrow V \subset \mathbb{R}^n$  where  $V$  is open.<sup>14</sup>

A collection of coordinate charts  $\{(U_i, \phi_i)\}$  for  $i \in I$  ( $I$  is a set that indexes the charts) is called an *atlas*. The following conditions are imposed on the coordinate charts:

- $\{U_i, \phi_i\}$  exists so that for each  $x \in M$ ,  $x \in U_i$  for some  $i \in I$ .
- If  $(U_i, \phi_i)$  and  $(U_j, \phi_j)$  are any two coordinate charts in the atlas for which  $(U_i, \phi_i) \cap (U_j, \phi_j) \neq \emptyset$ , then the composed map

$$\phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \rightarrow \phi_j(U_i \cap U_j) \quad (7.90)$$

is continuous.

- All possible charts with the above two properties are contained in the atlas.

Since it is well-known what it means for a mapping between open sets in Euclidean space to be continuous, if the condition that the composite maps of the form  $\phi_j \circ \phi_i^{-1}$  are all continuous mappings between the open subsets  $\phi_i(U_i \cap U_j)$  and  $\phi_j(U_i \cap U_j)$  (which are both in  $\mathbb{R}^n$ ) for all  $i, j \in I$ , then we say that each  $\phi_i$  is a continuous mapping from  $U$  to  $\phi_i(U)$ .

In practice the manifolds most often encountered are even more well-behaved. Namely, if all of the functions  $\phi_j \circ \phi_i^{-1}$  are differentiable (with respect to any set of Cartesian coordinates imposed on the Euclidean space that contains the open sets  $\phi_i(U_i \cap U_j)$  and  $\phi_j(U_i \cap U_j)$ ), then  $M$  is called a *differentiable manifold*. If each  $\phi_j \circ \phi_i^{-1}$  can be differentiated an infinite number of times, then  $M$  is called a *smooth manifold*. And if each  $\phi_j \circ \phi_i^{-1}$  is an analytic function (i.e., a function for which a convergent Taylor series exists), then  $M$  is called an *analytic manifold*. An analytic manifold is always smooth, but it is possible to be smooth yet not analytic [63].

A differentiable manifold is called *orientable* if an atlas can be defined such that the sign of the Jacobian determinant of  $\phi_j \circ \phi_i^{-1}$  is positive for all  $i, j$  for which  $U_i \cap U_j \neq \emptyset$ . Since  $\phi_j \circ \phi_i^{-1}$  is a mapping between two open sets in  $\mathbb{R}^n$ , its Jacobian is computed using methods of multivariable calculus. Unless stated otherwise, all manifolds discussed throughout this book will be both orientable and analytic.<sup>15</sup>

A differential  $k$ -form,  $\omega$ , on a patch  $U$  in an  $n$ -dimensional manifold,  $M$ , can be defined with respect to a particular set of coordinates  $\mathbf{q} = [q_1, \dots, q_n]^T \in \phi_i(U_i)$ , and a set of smooth functions  $\{a_{i_1, i_2, \dots, i_k}(\mathbf{q})\}$  as

$$\omega = \sum_{i_1, i_2, \dots, i_k} a_{i_1, i_2, \dots, i_k}(\mathbf{q}) dq_{i_1} \wedge dq_{i_2} \wedge \dots \wedge dq_{i_k} \quad \text{where } 1 \leq i_1 < i_2 < \dots < i_k \leq n.$$

Here the set of coordinates  $\{q_i\}$  are treated in the same way as Cartesian coordinates in  $\mathbb{R}^n$ , and,  $\{dq_1, \dots, dq_n\}$  are interpreted according to the same rules as the differentials in  $\mathbb{R}^n$ . The set of all such  $k$ -forms on  $U$  is denoted as  $\Omega^k(U)$ .

Due to the properties of the wedge product from Chapter 6, the only non-zero contributions to the sum appear when there are no repeated indices. The above equation can be written more concisely as

<sup>14</sup>Note that the perspective here is reversed from that in classical surface theory. Instead of mapping from open sets in the coordinate domain to the manifold,  $\phi$  maps from the manifold to the coordinate domain. Stated another way, if  $\mathbf{q} \in \phi(U) \subset \mathbb{R}^n$  is a vector of local coordinates, then  $\phi^{-1}(\mathbf{q})$  is a local *parametrization* of the manifold.

<sup>15</sup>The Klein bottle and  $\mathbb{R}P^2$  are examples of nonorientable manifolds.

$$\omega = \sum_{I_k} a_{I_k} dq_{I_k} \tag{7.91}$$

where  $I_k = \{i_1, i_2, \dots, i_k\}$  is any subset of  $\{1, \dots, n\}$  consisting of  $k$  distinct numbers written in strictly increasing order,  $a_{I_k} = a_{i_1, i_2, \dots, i_k}$  and  $dq_{I_k} = dq_{i_1} \wedge dq_{i_2} \wedge \dots \wedge dq_{i_k}$ . In this notation, the exterior derivative of a  $k$ -form can be uniquely defined by the properties [63, 59]

$$\begin{aligned} d\omega &= \sum_{I_k} (da_{I_k}) \wedge dq_{I_k} \\ &= \sum_{I_k} \sum_j \frac{\partial a_{I_k}}{\partial q_j} dq_j \wedge dq_{I_k}. \end{aligned} \tag{7.92}$$

The first equality above results because of the defining property  $d(dq_{I_k}) = 0$ . Note that the introduction of the additional wedge product makes  $d\omega$  a  $(k + 1)$ -form whereas an arbitrary form is denoted here as  $\omega$ , an  $n$ -form on an  $n$ -dimensional manifold will be denoted here as  $\alpha$ . That is, for each patch  $U_i \in M$ ,  $\alpha \in \Omega^n(U_i)$  where  $\phi_i(U_i) \subset \mathbb{R}^n$ .

A beautiful theory of integration for  $n$ -forms on orientable  $n$ -dimensional manifolds has been developed. Let  $U_i \subset M$  and  $\alpha \in \Omega^n(U_i)$ . Then this  $n$ -form can be expressed in local coordinates  $\{q_1, \dots, q_n\}$  on  $\phi_i(U_i)$  as

$$(\phi_i^{-1})^* \alpha = a(q_1, q_2, \dots, q_n) dq_1 \wedge dq_2 \wedge \dots \wedge dq_n$$

and using (6.95) the integral of  $\alpha$  is defined as [13, 23, 19, 63]

$$\int_{U_i} \alpha = \int_{\phi_i(U_i)} (\phi_i^{-1})^* \alpha = \int_{\phi_i(U_i)} a(q_1, q_2, \dots, q_n) dq_1 dq_2 \dots dq_n. \tag{7.93}$$

This defines integration of an  $n$ -form on one patch of the manifold. If  $U_i$  and  $U_j$  are overlapping patches, then (7.90) holds, and due to the properties of the pull-back map [63],

$$\int_{\phi_i(U_i \cap U_j)} (\phi_i^{-1})^* \alpha = \int_{\phi_j(U_i \cap U_j)} (\phi_i \circ \phi_j^{-1})^* (\phi_i^{-1})^* \alpha = \int_{\phi_j(U_i \cap U_j)} (\phi_j^{-1})^* \alpha.$$

The  $n$ -form  $\alpha$  then can be integrated over the whole manifold by defining it in coordinates in each patch. The trick is to make sure that there is no double counting or missed spots. The two ways to do this are: (1) to break the manifold up into polytopes (such as hyper-cubes), that are conjoined by shared  $(n - 1)$ -dimensional faces but are otherwise disjoint, and integrate over each; or (2) introduce a partition of unity and blend the local descriptions of  $\alpha$  (which by definition must be the same on overlapping patches).

In Chapter 1, the concept of compactness was introduced to describe a body in  $\mathbb{R}^n$  that was closed and bounded, and therefore had finite volume. This working definition was used throughout the book. A more precise definition in the present context is to say that a *compact manifold* is a manifold that can be reconstructed from (or covered by) taking the union of a finite number of patches, each of which is bounded in its size.

It is sometimes useful to consider  $n$ -dimensional orientable *manifolds with boundary*. The boundary (which is also taken to be orientable) is then  $(n - 1)$ -dimensional. One way to view this is by starting with an orientable manifold,  $\tilde{M}$ , without boundary, and embedding an orientable submanifold,  $\partial M$ , of dimension  $n - 1$  in such a way that it partitions  $\tilde{M}$  into two disjoint components, each of which is an open set. This is analogous to the way that the famous Jordan curve theorem describes how a simple

closed curve partitions the plane into two disjoint parts, one that describes points on the interior of the curve and one that describes the exterior. If we call one of these parts  $M \subset \tilde{M}$ , then  $M \cup \partial M$  is a manifold with boundary. A manifold with boundary,  $M \subset \tilde{M}$ , is usually defined in books on differential geometry by piecing together patches for  $M$  and patches that locally look like the closed half space in  $\mathbb{R}^n$  defined by the constraint  $x_n \geq 0$ .

For example, if  $\tilde{M} = S^2$ , the unit sphere in  $\mathbb{R}^3$ , then by inscribing a simple closed curve on the sphere defines a boundary between two regions. Each of these open regions is a manifold. Taking the union of either one with the closed curve defines a manifold with boundary.

In the discussions above there has been very little geometric content because there was no mention of distance. The distinction between differential topology and differential geometry is that in geometric discussions a metric is required. Earlier, a Riemannian metric was induced by the way a manifold was embedded in Euclidean space. However, it is possible to define this in an intrinsic way. If  $T_x M$  denotes the tangent space to the smooth manifold  $M$  at the point  $x \in M$ , then the Riemannian metric is a family of functions .

$$g_x : T_x M \times T_x M \rightarrow \mathbb{R} \quad \forall x \in M \quad (7.94)$$

such that the function  $f(x) \doteq g_x(A(x), B(x))$  is differentiable for all  $x \in M$  and  $A(x), B(x) \in T_x M$ . Furthermore, if  $\{X_i(x)\}$  is a basis for  $T_x M$ , the matrix with entries  $g_{ij}(x) \doteq g_x(X_i(x), X_j(x))$  is symmetric in the arguments. A corresponding tensor is denoted as

$$G = \sum_{i,j=1}^n g_{ij}(x) dx_i \otimes dx_j, \quad (7.95)$$

where the tensor product  $\otimes$  in the above expression is between elements of the basis  $\{dx_i\}$  for the dual space  $(T_x M)^*$ , which is called the *cotangent space* of  $M$  at  $x$ . The tensor  $G$  is the *Riemannian metric tensor*. A smooth manifold equipped with a Riemannian metric tensor is called a *Riemannian manifold*. indexRiemannian!manifold

### 7.6.1 Computing Tangent Vectors and Boundary Normals in Local Coordinates

As an alternative to describing vector fields on manifolds in an ambient Euclidean space in which the manifold is taken to be embedded, it is possible to describe vector fields in terms of coordinate charts and mappings among the charts. For example, given a vector field  $\mathcal{V} = \sum_i v_i(\mathbf{q}) \partial / \partial q_i$  in coordinate system  $\{q_1, \dots, q_m\}$  and given  $\phi = \phi(\mathbf{q})$ , it is possible to use the mapping  $\phi$  to push forward the vector field and express it as  $\phi_* \mathcal{V} = \sum_i \mathbf{e}_i^T [\partial \phi / \partial \mathbf{q}^T] |_{\mathbf{q}(\phi)} \mathbf{v}(\mathbf{q}(\phi)) \partial / \partial \phi_i$ . This is nothing more than (6.80). The local geometry of the manifold is encoded in the transition between the maps and the metric.

As a concrete example, consider the vector field  $\mathcal{W}$  defined in terms of coordinate  $\{q_1, q_2, q_3\}$  for the hyper-spherical cap example presented in Section 7.3.2 as

$$\mathcal{W} \doteq w_1 \frac{\partial}{\partial q_1} + w_2 \frac{\partial}{\partial q_2} + w_3 \frac{\partial}{\partial q_3}, \quad (7.96)$$

or equivalently,  $\mathbf{w} = [w_1, w_2, w_3]^T$  where

$$w_i = \frac{q_i}{\sqrt{q_1^2 + q_2^2 + q_3^2}}.$$

Since the Jacobian matrix has already been computed, and since from (7.29)  $[\partial\phi/\partial\mathbf{q}^T]|_{\mathbf{q}(\phi)} = [\partial\mathbf{q}/\partial\phi^T]^{-1}$ , which was already computed explicitly, the vector field  $\psi_*\mathcal{W}$  can be written by inspection (together with some trigonometry) as

$$\psi_*\mathcal{W} = \tan\psi \frac{\partial}{\partial\psi}.$$

Now suppose that an abstract manifold  $M$  is defined locally in terms of the coordinates  $\mathbf{q}$  and metric  $G(\mathbf{q})$ . From the example above it should be clear how to transform vector fields between coordinates. In principle, as the whole manifold is traversed a series of such changes in coordinates can be made. Now suppose that a submanifold is defined locally in terms of coordinates as  $\mathbf{q}(\mathbf{s})$  where  $m - 1 = \dim(\mathbf{s}) = \dim(\mathbf{q}) - 1$ . The question then becomes, how can the normal to  $\partial M$  be defined without reference to any knowledge of how  $M$  might be embedded?

The answer is that analogs of the calculations performed for the embedded case follow when the abstract inner product  $(\partial/\partial q_i, \partial/\partial q_j) \doteq g_{ij}$  is defined for tangent vectors. If the coordinates  $\mathbf{s}$  are chosen in such a way  $\phi = [\mathbf{s}^T, s_m]^T$  is a full set of local coordinates for a neighborhood in  $M$  with  $s_m = 0$  locally defining the submanifold, then the normal direction for the submanifold in this coordinate system will be  $\partial/\partial s_m$ . If  $\mathbf{q} = \mathbf{q}(\phi)$ , then pushing  $\partial/\partial s_m$  forward will provide the description of the normal to  $\partial M$  in the coordinates  $\mathbf{q}$ .

### 7.6.2 Stokes' Theorem for Manifolds

Let  $M$  be a compact orientable manifold of dimension  $m$  with boundary  $\partial M$  of dimension  $m - 1$ . Let  $\omega$  denote an  $(m - 1)$ -form on  $M$ , and  $d\omega$  denote the  $m$ -form resulting from exterior differentiation of  $\omega$ . Then *Stokes' theorem* for manifolds is stated as [20, 23, 1]

$$\boxed{\int_M d\omega = \int_{\partial M} \omega.} \tag{7.97}$$

Suppose that  $M$  and  $\partial M$  are contained in a manifold  $\tilde{M}$  of dimension  $m$  and that their volume elements are respectively  $dV$  and  $dS$ . (Both of these are defined by the Riemannian metric on  $\tilde{M}$ .) At any point  $p \in \partial M$  let  $\mathbf{n}(p)$  denote the unit normal vector to  $\partial M$  that points away from  $M$ , and that is contained in the tangent space of  $\tilde{M}$  at  $p$ . When written as  $\mathbf{n}(p)$ , this vector can be thought of as an  $m$ -dimensional array of functions  $\{n_i\}$ . Alternatively, calligraphic “ $\mathcal{N}$ ” will denote the same vector field written as in (6.60).<sup>16</sup>

If the metric tensor for  $M$  in coordinates  $\mathbf{q}$  is  $G(\mathbf{q})$ , then the inner product of two vector fields,  $\mathcal{V} = \sum_i v_i \partial/\partial q_i$  and  $\mathcal{W} = \sum_i w_i \partial/\partial q_i$ , at the point in  $M$  defined by a specific value of  $\mathbf{q}$  is

$$(\mathcal{V}, \mathcal{W}) \doteq (\mathbf{v}, \mathbf{w}) = [\mathbf{v}(\mathbf{q})]^T [G(\mathbf{q})] \mathbf{w}(\mathbf{q}).$$

---

<sup>16</sup>The difference is that  $\mathbf{n}\phi = \phi\mathbf{n}$  is just scalar multiplication of the entries in the array  $\mathbf{n}$  by  $\phi$ , whereas  $\mathcal{N}\phi = \sum_i n_i \partial\phi/\partial q_i = \mathbf{n} \cdot \text{grad } \phi$  is not the same as  $\phi\mathcal{N} = \phi \sum_i n_i \partial/\partial q_i$ .

Given a vector field  $\mathbf{w}$  defined on the *tangent bundle*<sup>17</sup> of  $M$ , and scalar function  $\phi$  defined on  $M$ , the following (more explicit, and more specialized) forms of Stokes' theorem for orientable manifolds can be written as<sup>18</sup> (see, e.g., [43])

**Theorem 7.1.** *The Divergence Theorem for Manifolds with Boundary:*

$$\int_M \operatorname{div}(\mathbf{w}) dV = \int_{\partial M} (\mathbf{w}, \mathbf{n}) dS. \quad (7.98)$$

**Theorem 7.2.** *First Green's Theorem for Manifolds with Boundary:*

$$\int_M [\phi_1 \operatorname{div}(\operatorname{grad} \phi_2) + (\operatorname{grad} \phi_1, \operatorname{grad} \phi_2)] dV = \int_{\partial M} \phi_1 \mathcal{N} \phi_2 dS. \quad (7.99)$$

**Theorem 7.3.** *Second Green's Theorem for Manifolds with Boundary:*

$$\int_M [\phi_1 \operatorname{div}(\operatorname{grad} \phi_2) - \phi_2 \operatorname{div}(\operatorname{grad} \phi_1)] dV = \int_{\partial M} (\phi_1 \mathcal{N} \phi_2 - \phi_2 \mathcal{N} \phi_1) dS. \quad (7.100)$$

**Theorem 7.4.** *Integration-by-Parts for Manifolds with Boundary:*

$$\int_M (\operatorname{grad} \phi, \mathbf{w}) dV = \int_{\partial M} (\mathbf{w}, \mathbf{n}) \phi dS - \int_M \phi \operatorname{div}(\mathbf{w}) dV. \quad (7.101)$$

These can either be proved as special cases of Stokes' theorem using intrinsic and coordinate-free geometric techniques, or using the extrinsic and coordinate-dependent approach described earlier in this chapter for the case of an embedded manifold.

These theorems are now demonstrated with the example of a hyper-spherical cap and its boundary submanifold from Section 7.3.2.

### Example 1: Inner Product of Vector Fields

The calculations involved in computing the inner product of vector fields on a manifold are now illustrated with an example using both intrinsic and extrinsic approaches. Previous examples demonstrated the inner product of vector fields and Stokes' theorem for domains and surfaces in  $\mathbb{R}^3$ . However, it also applies to embedded manifolds as well as to manifolds defined by coordinate charts that need not be embedded in  $\mathbb{R}^n$ . We first consider the case of an embedded manifold with boundary, and then consider intrinsic calculations. Again, the example of an open hyper-spherical cap,  $M$ , with spherical boundary,  $\partial M$ , will be used as the example. In this example, one coordinate chart is sufficient to cover the whole manifold. When performing integrations over  $M$  and  $\partial M$ ,

<sup>17</sup>The tangent space at one point on a manifold is not the same object as the tangent spaces at another point on the same manifold. However, they are equivalent in that they have the same dimension, and for manifolds embedded in Euclidean space, one tangent space can be rigidly moved so as to coincide with another. The collection of all of these tangent spaces indexed by points on the manifold, together with a projection map, is called the tangent bundle. A single vector that is tangent to a manifold at a particular point is contained in a single tangent space. In contrast, a vector field "on the manifold" can be viewed as a mapping from the manifold to the tangent bundle.

<sup>18</sup> $M$  has  $m$ -dimensional volume element  $dV$  and  $\partial M$  has  $(m-1)$ -dimensional volume element  $dS$ .

they will therefore be performed in the coordinates  $\mathbf{q}$ . The range of parameters in the integrations will be

$$D = \{\mathbf{q} \mid \|\mathbf{q}\| < r_h\} \quad \text{and} \quad \partial D = \{\mathbf{q} \mid \|\mathbf{q}\| = r_h\}. \quad (7.102)$$

As an initial example, let

$$\mathbf{K}(\mathbf{q}) \doteq (1 - q_2^2 - q_3^2)^{-1} q_1 \frac{\partial \mathbf{x}}{\partial q_1} \quad \text{and} \quad \mathbf{H}(\mathbf{q}) \doteq (1 - q_1^2 - q_2^2 - q_3^2)^{\frac{3}{2}} \frac{\partial \mathbf{x}}{\partial q_1},$$

or equivalently,

$$\mathbf{k}(\mathbf{q}) = [(1 - q_2^2 - q_3^2)^{-1} q_1, 0, 0]^T \quad \text{and} \quad \mathbf{h}(\mathbf{q}) = [(1 - q_1^2 - q_2^2 - q_3^2)^{\frac{3}{2}}, 0, 0]^T.$$

Note that  $\mathbf{K} \in \mathbb{R}^4$ , but  $\mathbf{k} \in \mathbb{R}^3$ . These contain equivalent information as the modern notation

$$\mathcal{K} = k_1 \frac{\partial}{\partial q_1} \quad \text{and} \quad \mathcal{H} = h_1 \frac{\partial}{\partial q_1}.$$

Then (7.19) reduces to

$$\langle \mathbf{k}, \mathbf{h} \rangle = \int_D \|\mathbf{T}_1\|^2 q_1 |G(\mathbf{q})|^{\frac{1}{2}} d\mathbf{q} = \int_{\|\mathbf{q}\| < r_h} q_1 d\mathbf{q}.$$

This integral is most easily evaluated by converting  $\mathbf{q}$  to the spherical coordinates

$$q_1 = r \cos \phi \sin \theta; \quad q_2 = r \sin \phi \sin \theta; \quad q_3 = r \cos \theta.$$

Then

$$\langle \mathbf{k}, \mathbf{h} \rangle = \int_0^{r_h} \int_0^\pi \int_0^{2\pi} (r \cos \phi \sin \theta) r^2 \sin \theta d\phi d\theta dr = 0$$

since  $\int_0^{2\pi} \cos \phi d\phi = 0$ .

**Example 2: Divergence Theorem for Vector Fields on Manifolds Without Embedding**

Let us assume that  $G(\mathbf{q})$  is given as in (7.23), and that one coordinate chart is enough in this example. Since  $G$  is specified, the way that the sphere is embedded in  $\mathbb{R}^4$  can be completely forgotten, and all calculations can be performed in this chart.

For the vector field in (7.96),

$$\begin{aligned} \operatorname{div}(\mathbf{w}) &= |G(\mathbf{q})|^{-\frac{1}{2}} \sum_{i=1}^3 \frac{\partial}{\partial q_i} (|G(\mathbf{q})|^{\frac{1}{2}} w_i) \\ &= (1 - \|\mathbf{q}\|^2)^{\frac{1}{2}} \sum_{i=1}^3 \left[ (1 - \|\mathbf{q}\|^2)^{-\frac{3}{2}} q_i w_i + (1 - \|\mathbf{q}\|^2)^{-\frac{1}{2}} \frac{\partial w_i}{\partial q_i} \right] \\ &= \sum_{i=1}^3 \left[ (1 - \|\mathbf{q}\|^2)^{-1} q_i w_i + \frac{\partial w_i}{\partial q_i} \right] \\ &= (1 - \|\mathbf{q}\|^2)^{-1} \sum_{i=1}^3 q_i w_i + \sum_{i=1}^3 \frac{\partial w_i}{\partial q_i} \\ &= (1 - \|\mathbf{q}\|^2)^{-1} \|\mathbf{q}\| + 2\|\mathbf{q}\|^{-1}. \end{aligned}$$



Then

$$\begin{aligned}
 \int_M \operatorname{div}(\mathbf{w}) dV &= \int_D \left\{ (1 - \|\mathbf{q}\|^2)^{-1} \|\mathbf{q}\| + 2\|\mathbf{q}\|^{-1} \right\} |G(\mathbf{q})|^{\frac{1}{2}} dq_1 dq_2 dq_3 \\
 &= \int_0^{r_h} \int_0^\pi \int_0^{2\pi} \left\{ \frac{r}{[1-r^2]^{\frac{3}{2}}} + \frac{2}{r[1-r^2]^{\frac{1}{2}}} \right\} r^2 \sin \theta d\phi d\theta dr \\
 &= 4\pi \int_0^{r_h} \frac{r^3}{[1-r^2]^{\frac{3}{2}}} dr + 8\pi \int_0^{r_h} \frac{r}{[1-r^2]^{\frac{1}{2}}} dr \\
 &= 4\pi \left[ (1-r^2)^{\frac{1}{2}} + (1-r^2)^{-\frac{1}{2}} \right]_0^{r_h} - 8\pi \left[ (1-r^2)^{\frac{1}{2}} \right]_0^{r_h} \\
 &= \frac{4\pi}{h} (1-h^2).
 \end{aligned}$$

On the other hand, for this particular vector field and bounding surface,  $(\mathbf{w}, \mathbf{n}) = \mathbf{w}^T G \mathbf{n} = 1/h$  and so

$$\begin{aligned}
 \int_{\partial M} (\mathbf{w}, \mathbf{n}) dS &= \int_{\partial D} (\mathbf{w}^T G \mathbf{n}) |J^T(\mathbf{s}) G(\mathbf{q}(\mathbf{s})) J(\mathbf{s})| ds \\
 &= \int_0^\pi \int_0^{2\pi} (1/h) \cdot (r_h)^2 \sin s_2 ds_1 ds_2 \\
 &= \frac{4\pi}{h} (1-h^2).
 \end{aligned}$$

This illustrates Theorem 7.1 in the context of this particular example.

### Example 3: Integration by Parts on Manifolds Without Embedding

In addition to the vector field defined in (7.96), define the scalar function

$$\phi \doteq q_3^2.$$

With these and the metric tensor defined in (7.23), the integration-by-parts formula in Theorem 7.4 can be demonstrated. First, observe that

$$\begin{aligned}
 \int_M (\operatorname{grad} \phi, \mathbf{w}) dV &= \int_D (G^{-1} \nabla_{\mathbf{q}} \phi)^T G \mathbf{w} |G|^{\frac{1}{2}} d\mathbf{q} \\
 &= \int_D (\nabla_{\mathbf{q}} \phi)^T \mathbf{w} |G|^{\frac{1}{2}} d\mathbf{q} \\
 &= 2 \int_D q_3^2 \cdot \|\mathbf{q}\|^{-\frac{1}{2}} \cdot (1 - \|\mathbf{q}\|^2)^{-\frac{1}{2}} d\mathbf{q} \\
 &= 2 \int_0^{r_h} \int_0^\pi \int_0^{2\pi} (r \cos \theta)^2 (r)^{-1} (1-r^2)^{-\frac{1}{2}} r^2 \sin \theta d\phi d\theta dr \\
 &= 2 \cdot \left( \int_0^{r_h} \frac{r^3}{(1-r^2)^{\frac{1}{2}}} dr \right) \cdot \left( \int_0^\pi \int_0^{2\pi} \cos^2 \theta \sin \theta d\phi d\theta \right) \\
 &= \frac{8\pi}{3} \int_0^{r_h} \frac{r^3}{(1-r^2)^{\frac{1}{2}}} dr = \frac{8\pi}{3} \left[ \frac{1}{3} (1-r^2)^{\frac{3}{2}} - (1-r^2)^{\frac{1}{2}} \right]_0^{r_h} \\
 &= \frac{8\pi}{3} [h^3/3 - h + 2/3].
 \end{aligned}$$

For the particular vector field  $\mathbf{w} = \mathbf{n}/h$  and bounding surface  $\partial M$ ,  $\mathbf{w}^T G \mathbf{n} = 1/h$ . Therefore

$$\begin{aligned}
\int_{\partial M} (\mathbf{w}, \mathbf{n}) \phi \, dS &= \int_{\partial D} \phi(\mathbf{q}(\mathbf{s})) [\mathbf{w}(\mathbf{q}(\mathbf{s}))]^T G(\mathbf{q}(\mathbf{s})) \mathbf{n}(\mathbf{q}(\mathbf{s})) |J^T(\mathbf{s}) G(\mathbf{q}(\mathbf{s})) J(\mathbf{s})|^{1/2} \, d\mathbf{s} \\
&= \int_0^\pi \int_0^{2\pi} (r_h \cos s_2)^2 \cdot \frac{1}{h} \cdot (r_h)^2 \sin s_2 \, ds_1 \, ds_2 \\
&= \frac{(1-h^2)^2}{h} \int_0^\pi \int_0^{2\pi} \cos^2 s_2 \sin s_2 \, ds_1 \, ds_2 \\
&= \frac{4\pi}{3} \frac{(1-h^2)^2}{h}.
\end{aligned}$$

And

$$\begin{aligned}
\int_M \phi \operatorname{div}(\mathbf{w}) \, dV &= \int_D \phi(\mathbf{q}) \{ (1 - \|\mathbf{q}\|^2)^{-1} \|\mathbf{q}\| + 2\|\mathbf{q}\|^{-1} \} |G(\mathbf{q})|^{1/2} \, d\mathbf{q} \\
&= \int_D 2q_3^2 \|\mathbf{q}\|^{-1} (1 - \|\mathbf{q}\|^2)^{-1/2} \, d\mathbf{q} + \int_D q_3^2 (1 - \|\mathbf{q}\|^2)^{-3/2} \|\mathbf{q}\| \, d\mathbf{q} \\
&= \int_0^{r_h} \int_0^\pi \int_0^{2\pi} \left\{ \frac{2(r \cos \theta)^2}{(1-r^2)^{1/2} r} + \frac{(r \cos \theta)^2 r}{(1-r^2)^{3/2}} \right\} r^2 \sin \theta \, d\phi \, d\theta \, dr \\
&= \frac{4\pi}{3} \left[ \int_0^{r_h} \frac{2r^3 \, dr}{(1-r^2)^{1/2}} + \int_0^{r_h} \frac{r^5 \, dr}{(1-r^2)^{3/2}} \right] \\
&= \frac{4\pi}{3} [(2h^3/3 - 2h + 4/3) + (-h^3/3 + 2h + h^{-1} - 8/3)].
\end{aligned}$$

Substituting these into the formula in Theorem 7.4 verifies integration by parts for this example.

#### Example 4: Green's First Theorem for Vector Fields on Manifolds Without Embedding

In addition to the vector field defined in (7.96), define the scalar functions

$$\phi_1(\mathbf{q}) \doteq q_1^2 + q_2^2 \quad \text{and} \quad \phi_2(\mathbf{q}) \doteq q_3^2.$$

Then Green's theorems (7.2 and 7.3) can be demonstrated. Only Theorem 7.2 is demonstrated here.

To begin, observe that for this example

$$G^{-1} \nabla_{\mathbf{q}} \phi_2 = [-2q_1 q_3^2, -2q_2 q_3^2, 2(1 - q_3^2) q_3]^T$$

and so

$$\begin{aligned}
\sum_{i=1}^3 \frac{\partial}{\partial q_i} \left( |G(\mathbf{q})|^{1/2} \sum_{j=1}^3 g^{ij}(\mathbf{q}) \frac{\partial \phi_2}{\partial q_j} \right) &= -2 \frac{\partial}{\partial q_1} \left[ (1 - \|\mathbf{q}\|^2)^{-1/2} q_1 q_3^2 \right] \\
&\quad - 2 \frac{\partial}{\partial q_2} \left[ (1 - \|\mathbf{q}\|^2)^{-1/2} q_2 q_3^2 \right] \\
&\quad + 2 \frac{\partial}{\partial q_3} \left[ (1 - \|\mathbf{q}\|^2)^{-1/2} (1 - q_3^2) q_3 \right] \\
&= -2(1 - \|\mathbf{q}\|^2)^{-1/2} \left[ (1 - \|\mathbf{q}\|^2)^{-1} (q_1^2 + q_2^2) q_3^2 + 2q_3^2 \right] \\
&\quad + 2(1 - \|\mathbf{q}\|^2)^{-1/2} \left[ (1 - \|\mathbf{q}\|^2)^{-1} \right. \\
&\quad \left. (1 - q_3^2) q_3^2 + (1 - 3q_3^2) \right] \\
&= 2(1 - \|\mathbf{q}\|^2)^{-1/2} [1 - 4q_3^2]
\end{aligned}$$

and

$$(\nabla_{\mathbf{q}}\phi_1)^T G^{-1} \nabla_{\mathbf{q}}\phi_2 = -4(q_1^2 q_3^2 + q_2^2 q_3^2).$$

Also,

$$\sum_{i=1}^3 n_i \frac{\partial \phi_2}{\partial q_i} \Big|_{\mathbf{q}=\mathbf{q}(\mathbf{s})} = (h \cos s_2)(2 \cos s_2).$$

Now consider each of the three integrals in this theorem:

$$\begin{aligned} \int_M \phi_1 \operatorname{div}(\operatorname{grad} \phi_2) dV &= \int_D \phi_1 \sum_{i=1}^3 \frac{\partial}{\partial q_i} \left( |G(\mathbf{q})|^{\frac{1}{2}} \sum_{j=1}^3 g^{ij}(\mathbf{q}) \frac{\partial \phi_2}{\partial q_j} \right) d\mathbf{q} \\ &= 2 \int_D \frac{(q_1^2 + q_2^2)[1 - 4q_3^2]}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} d\mathbf{q} \\ &= 2 \int_0^{r_h} \int_0^\pi \int_0^{2\pi} \frac{[1 - 4r^2 \cos^2 \theta] r^2 \sin^2 \theta}{(1 - r^2)^{\frac{1}{2}}} r^2 \sin \theta dr d\phi d\theta \\ &= 2 \left( \int_0^\pi \int_0^{2\pi} \sin^3 \theta d\phi d\theta \right) \cdot \int_0^{r_h} \frac{r^4}{(1 - r^2)^{\frac{1}{2}}} dr \\ &\quad - 8 \left( \int_0^\pi \int_0^{2\pi} \cos^2 \theta \sin^3 \theta d\phi d\theta \right) \cdot \int_0^{r_h} \frac{r^6}{(1 - r^2)^{\frac{1}{2}}} dr \\ &= \frac{16\pi}{3} \int_0^{r_h} \frac{r^4}{(1 - r^2)^{\frac{1}{2}}} dr - \frac{64\pi}{15} \int_0^{r_h} \frac{r^6}{(1 - r^2)^{\frac{1}{2}}} dr \quad (7.103) \end{aligned}$$

$$\begin{aligned} \int_M (\operatorname{grad} \phi_1, \operatorname{grad} \phi_2) dV &= \int_D (\nabla_{\mathbf{q}}\phi_1)^T [G(\mathbf{q})]^{-1} (\nabla_{\mathbf{q}}\phi_2) |G(\mathbf{q})|^{\frac{1}{2}} d\mathbf{q} \\ &= \int_D \frac{-4(q_1^2 q_3^2 + q_2^2 q_3^2)}{(1 - \|\mathbf{q}\|^2)^{\frac{1}{2}}} d\mathbf{q} \\ &= \int_0^{r_h} \int_0^\pi \int_0^{2\pi} \frac{-4r^4 \sin^2 \theta \cos^2 \theta}{(1 - r^2)^{\frac{1}{2}}} r^2 \sin \theta dr d\phi d\theta \\ &= -4 \left( \int_0^\pi \int_0^{2\pi} \cos^2 \theta \sin^3 \theta d\phi d\theta \right) \cdot \int_0^{r_h} \frac{r^6}{(1 - r^2)^{\frac{1}{2}}} dr \\ &= -\frac{32\pi}{15} \int_0^{r_h} \frac{r^6}{(1 - r^2)^{\frac{1}{2}}} dr \quad (7.104) \end{aligned}$$

and

$$\begin{aligned} \int_{\partial M} \phi_1 \mathcal{N} \phi_2 dS &= \int_{\partial D} \phi_1 \cdot \left[ \sum_{i=1}^3 n_i \frac{\partial \phi_2}{\partial q_i} \right] \cdot |J^T(\mathbf{s})G(\mathbf{q}(\mathbf{s}))J(\mathbf{s})|^{\frac{1}{2}} d\mathbf{s} \\ &= \int_0^\pi \int_0^{2\pi} [(r_h \cos s_1 \sin s_2)^2 + (r_h \sin s_1 \sin s_2)^2] \quad (7.105) \\ &\quad \cdot (2hr_h \cos^2 s_2) \cdot (r_h)^2 \sin s_2 ds_1 ds_2 \\ &= 2h(r_h)^5 \int_0^\pi \int_0^{2\pi} \sin^3 s_2 \cos^2 s_2 ds_1 ds_2 \\ &= \frac{16\pi}{15} h(1 - h^2)^{\frac{5}{2}}. \quad (7.106) \end{aligned}$$

Green's First Theorem will then hold in this example if

$$\frac{1}{3} \int_0^{r_h} \frac{r^4}{(1-r^2)^{\frac{1}{2}}} dr - \frac{2}{5} \int_0^{r_h} \frac{r^6}{(1-r^2)^{\frac{1}{2}}} dr = \frac{1}{15} h(1-h^2)^{\frac{5}{2}}.$$

Indeed, this can be verified by consulting tables of integrals.

Equation (7.103) is a straightforward implementation of (5.50) and the factor of  $|G(\mathbf{q})|^{-\frac{1}{2}}$  from  $\text{div}(\text{grad}(\phi_2))$  cancels with the factor of  $|G(\mathbf{q})|^{\frac{1}{2}}$  in the definition of  $dV = |G(\mathbf{q})|^{\frac{1}{2}} d\mathbf{q}$ . The inner product in (7.104) is interpreted as in (7.14), where  $\text{grad} \phi = [G(\mathbf{q})]^{-1}(\nabla_{\mathbf{q}}\phi)$ , and so  $(\text{grad} \phi_1, \text{grad} \phi_2) = ([G(\mathbf{q})]^{-1}\nabla_{\mathbf{q}}\phi_1)^T [G(\mathbf{q})]([G(\mathbf{q})]^{-1}\nabla_{\mathbf{q}}\phi_2)$ , leading to the simplification in (7.104). And the vector field  $\mathcal{N}$  in the equation preceding (7.105) is interpreted as in Section 7.6.1.

**Example 5: Integration of Forms on Manifolds Without Embedding**

Let  $M$  be the manifold defined by (7.20) with  $\|\mathbf{q}\| < r_h \doteq (1-h^2)^{\frac{1}{2}}$ , and let  $\partial M \cong S_{r_h}^2$  be the boundary of  $M$ . If

$$\alpha = a_3(\mathbf{q}) dq_1 \wedge dq_2 - a_2(\mathbf{q}) dq_1 \wedge dq_3 + a_1(\mathbf{q}) dq_2 \wedge dq_3,$$

then

$$d\alpha = \left( \frac{\partial a_1}{\partial q_1} + \frac{\partial a_2}{\partial q_2} + \frac{\partial a_3}{\partial q_3} \right) dq_1 \wedge dq_2 \wedge dq_3$$

and

$$\int_M d\alpha = \int_{\|\mathbf{q}\| \leq r_h} \nabla \cdot \mathbf{a} d\mathbf{q}$$

where  $\nabla \cdot$  is the usual divergence operator in  $\mathbb{R}^3$ .

More specifically, define a 2-form

$$\alpha \doteq \frac{q_1}{\sqrt{q_1^2 + q_2^2 + q_3^2}} dq_1 \wedge dq_2 + \frac{q_2}{\sqrt{q_1^2 + q_2^2 + q_3^2}} dq_1 \wedge dq_3 + \frac{q_3}{\sqrt{q_1^2 + q_2^2 + q_3^2}} dq_2 \wedge dq_3.$$

The exterior derivative of this form is computed according to the rule (7.92) as

$$d\alpha = -\frac{q_2^2 + 2q_1q_3 + q_3^2}{(q_1^2 + q_2^2 + q_3^2)^{\frac{3}{2}}} dq_1 \wedge dq_2 \wedge dq_3.$$

Therefore,

$$\int_M d\alpha = - \int_{\|\mathbf{q}\| < r_h} \frac{q_2^2 + 2q_1q_3 + q_3^2}{(q_1^2 + q_2^2 + q_3^2)^{\frac{3}{2}}} dq_1 \wedge dq_2 \wedge dq_3.$$

This is most easily computed by converting to a spherical coordinate system such as

$$q_1 = r \cos \phi \sin \theta; \quad q_2 = r \sin \phi \sin \theta; \quad q_3 = r \cos \theta.$$

It follows from (6.77) that

$$dq_1 \wedge dq_2 \wedge dq_3 = -r^2 \sin \theta d\phi \wedge d\theta \wedge dr.$$

And since  $q_1^2 + q_2^2 + q_3^2 = r^2$ , and the negative signs in the above expressions cancel,

$$\begin{aligned} \int_M d\alpha &= \int_0^{r_h} \int_0^\pi \int_0^{2\pi} \frac{r^2 f(\phi, \theta)}{(q_3')^3} r^2 \sin \theta d\phi d\theta dr \\ &= \frac{(r_h)^2}{2} \int_0^\pi \int_0^{2\pi} f(\phi, \theta) \sin \theta d\phi d\theta \end{aligned}$$

where

$$f(\phi, \theta) = \cos^2 \phi \sin^2 \theta + 2 \cos \phi \sin \theta \cos \theta + \cos^2 \theta.$$

Evaluating the integrals using trigonometric identities gives

$$\int_M d\alpha = \frac{4\pi}{3}(r_h)^2. \quad (7.107)$$

Spherical coordinates can be used to parameterize this boundary as

$$q_1 = r_h \cos s_1 \sin s_2; \quad q_2 = r_h \sin s_1 \sin s_2; \quad q_3 = r_h \cos s_2.$$

In these coordinates,

$$\alpha = (r_h)^2 (\sin^2 s_1 \sin^3 s_2 - 2 \cos s_1 \sin^2 s_2 \cos s_2) ds_1 \wedge ds_2$$

and

$$\int_{\partial M} \alpha = (r_h)^2 \int_0^\pi \int_0^{2\pi} (\sin^2 s_1 \sin^3 s_2 - 2 \cos s_1 \sin^2 s_2 \cos s_2) ds_1 ds_2 = \frac{4\pi}{3}(r_h)^2,$$

which is the same as the result in (7.107), thus illustrating Stokes' theorem.

### 7.6.3 The Gauss–Bonnet–Chern Theorem

There are two ways to extend the Gauss–Bonnet theorem to higher dimensions. The first approach is extrinsic, viewing the manifold as being embedded in some Euclidean space. For an  $n$ -dimensional hyper-surface in  $\mathbb{R}^{n+1}$ , it is possible to define a normal line corresponding to each point, exactly as in the case of two-dimensional surfaces in  $\mathbb{R}^3$ . If this approach is taken for an  $n$ -dimensional manifold in  $\mathbb{R}^{n+p}$ , the problem arises that there is not a single normal vector for each point on the manifold, but rather a whole normal space of dimension  $p$ . Every vector in this normal space will be orthogonal to every vector in the tangent to the manifold. An  $(n+p+1)$ -dimensional plane spanned by the normal space and one of the tangent vectors can be used to “slice” the manifold. The result will be a one-dimensional curve, the signed curvature of which can be computed. This can be done for each of  $n$  independent tangent directions. The resulting curvatures can be used in a similar way to construct an analog of the Gaussian curvature for manifolds. This can then be integrated over a compact oriented Riemannian manifold to obtain a generalization of the Gauss–Bonnet theorem.<sup>19</sup> The case when  $p = 1$  was addressed by Hopf [33]. The case for arbitrary  $p$  was addressed by Allendoerfer and Weil [2, 3] and Fenchel [22].

The second approach, which is purely intrinsic, is due to Chern [16], and uses differential forms. It is considered to be more general than the approaches in which the manifold is embedded in Euclidean space, and thus the renaming of this result as the *Gauss–Bonnet–Chern theorem*. From a purely computational point of view, the end result can be written in coordinates as

$$\int_M k dV = \frac{1}{2} \mathcal{O}_{n+1} \chi(M) \quad (7.108)$$

<sup>19</sup>A compact manifold with Riemannian metric defined will have a total volume with respect to that metric that is finite.

where  $k$  is the appropriate generalization of the Gaussian curvature and  $\chi(M)$  is the Euler characteristic of  $M$ .

In local coordinates  $dV(q) = |G(q)|^{\frac{1}{2}} dq_1 \dots dq_n$  is the volume element for the manifold (defined with respect to an appropriate metric tensor,  $G(q)$ ). Recall from Section 2.3 that  $\mathcal{O}_{n+1}$  denotes the volume of the unit sphere in  $\mathbb{R}^{n+1}$ . The function that takes the place of curvature in the classical Gauss–Bonnet theorem is  $k(q) = \mathcal{P}(R(q))$  where [48]

$$\mathcal{P}(R) = \sum_{\sigma, \pi \in \Pi_n} \frac{\text{sgn}(\sigma)\text{sgn}(\pi)}{2^{n/2}n! \det G} R_{\sigma(1), \sigma(2), \pi(1), \pi(2)} R_{\sigma(3), \sigma(4), \pi(3), \pi(4)} \dots R_{\sigma(n-1), \sigma(n), \pi(n-1), \pi(n)}. \tag{7.109}$$

Here  $\text{sgn}(\pi)$  is the sign of a permutation  $\pi$ . The function  $\mathcal{P}(\cdot)$ , called the *Pfaffian*, converts the Riemannian metric tensor,  $R$ , into a scalar. Recall that the Riemannian metric tensor can be written in component form as  $R^l_{ijk}$  or  $R_{ijkl} = \sum_m g_{lm} R^m_{ijk}$ . If the manifold is embedded in Euclidean space and parameterized as  $\mathbf{x} = \mathbf{x}(q_1, \dots, q_n) \in \mathbb{R}^{n+p}$ , the elements of the metric tensor are then  $g_{ij} = \partial \mathbf{x} / \partial q_i \cdot \partial \mathbf{x} / \partial q_j$  and all of the other formulas from Section 5.4.2 that are based on the metric tensor (such as those for the Christoffel symbols,  $\Gamma^k_{ij}$ , and the elements of the Riemannian metric tensor,  $R^m_{ijk}$ ) still apply. The only difference is that the indices now all range from 1 to  $n$  rather than 1 to 2.

The definition in (7.109) can be thought of as the natural higher-dimensional generalization of (5.65), and can be restated in terms of the language of differential forms as follows. Define the *curvature 2-form* as [45]

$$\Omega^j_l = -\frac{1}{2} \sum_{hk} R^j_{l hk} dq_h \wedge dq_k. \tag{7.110}$$

If these are viewed as the entries in an  $n \times n$  skew-symmetric matrix  $\Omega = [\Omega^j_l]$ , then

$$\int_M k(q) dV(q) = \int_M \mathcal{P}(\Omega)$$

where the Pfaffian of an even-dimensional skew-symmetric matrix  $S$  can be reinterpreted as [58]

$$\mathcal{P}(S) \doteq \frac{1}{2^{n/2}(n/2)!} \sum_{\pi \in \Pi_n} \epsilon(\pi) S_{\pi(1), \pi(2)} \dots S_{\pi(n-1), \pi(n)}$$

and is related to the determinant of  $S$  as

$$[\mathcal{P}(S)]^2 = \det S.$$

In the case when  $n$  is odd, this will always be zero, and hence so too will be (7.108).

The *torsion 2-form* for  $M$  is defined as [45]

$$\Omega^j = \sum_{hk} \Gamma^j_{hk} dq_h \wedge dq_k. \tag{7.111}$$

Given an embedding, this form describes how the  $n$ -dimensional manifold “twists around” in  $\mathbb{R}^{n+p}$ .

### 7.7 Fiber Bundles and Connections

In this section two concepts that play important roles in modern differential geometry are reviewed. Section 7.7.1 focuses on the concept of *fiber bundles*, which includes

vector bundles, tangent bundles, normal bundles, and frame bundles, among others. Section 7.7.2 discusses the concept of affine and Riemannian connections. In the world of coordinate-free intrinsic differential geometry these concepts play a vital role. In the stochastic modeling problems that arise in engineering and biology where the manifolds of interest are usually embedded in Euclidean space in a way that is dictated by the problem, these concepts are less critical. However, they are provided here for completeness.

### 7.7.1 Fiber Bundles

In Chapter 5 we encountered the concept of tubular surfaces. A tubular surface can be decomposed in two natural ways. The planes normal to the backbone curve (i.e., those planes that have as their normal the tangent to the backbone curve) can be used to “chop up” a tubular surface into an infinite number of circles. Alternatively, a tubular surface can be viewed as an infinite number of parallel curves that are offset from the backbone by a fixed radius. It is natural to think of these offset curves as *fibers*. Each fiber can be associated with one point on one of the circles resulting from the intersection of a plane normal to the backbone curve and the tubular surface. Picking one of these circles, the tube can be thought of as a collection of fibers sprouting from the circle. The circle in this context is called a *base space*, and the whole tube can be thought of as a “bundle” of offset curves. Given the same example, we could pick one of the offset curves, and use it as the base space, in which case associated with each point of an offset curve would be one and only one circle. Then the circles become fibers, and the tube could be called a bundle of circles.

The above two scenarios define two different *fiber bundles* from the same tubular surface. The tubular surface, which is an example of a *total space* (also called the *entire* or *bundle space*), is the same in both cases and can be viewed locally as the direct product of a base space and a fiber space. In other words, the tube locally looks like a cylinder. It is not until the larger picture is viewed that the distinction between, for example, a knotted torus and a cylinder becomes clear. The way that the total space is decomposed into fibers and base space defines different fiber bundles.

Equipped with this intuitive picture, a more precise and general definition can now be understood more easily. A fiber bundle is a mathematical structure consisting of the four objects  $(E, B, \pi, F)$  where  $B$  is the base space,  $F$  is the fiber space,  $E$  is the entire (or total) space (which “locally looks like” a direct product  $B \times F$ ), and  $\pi$  is a continuous projection map  $\pi : E \rightarrow B$ . This map has the property that for any open neighborhood  $U_x \subset B$  of a point  $x \in B$ , the inverse image  $\pi^{-1}(U_x)$  can be mapped bijectively and continuously to an open subset of  $U_x \times F$ . A set  $\pi^{-1}(x) \subset \pi^{-1}(U_x)$  is called the *fiber over the point*  $x \in B$ . For a tubular surface in  $\mathbb{R}^3$  with a circle as the base space,  $\pi$  can be thought of as the operation of collapsing each offset curve to a single point on the circle. Then  $\pi^{-1}(x \in S^1)$  is a particular offset curve, or fiber.

A *vector bundle* is a special kind of fiber bundle in which the fibers are each vector spaces and the projection map satisfies some additional properties. Since every point on an  $m$ -dimensional manifold,  $x \in M$ , has an associated tangent space,  $TM_x$ , the total space consisting of all pairs  $(x, TM_x)$  where  $x \in M$  together with  $B = M$  and  $F \cong \mathbb{R}^m$ , and an appropriate projection map<sup>20</sup> is a special kind of vector bundle called a *tangent*

<sup>20</sup>The projection map,  $\pi$ , is defined as  $\pi : TM \rightarrow M$  where  $v_x \in TM_x$  gets mapped to  $x$ , i.e.,  $\pi(v_x) = x$ . For any tangent vector field  $X : M \rightarrow TM$ , the projection map is defined to satisfy the additional condition that  $\pi \circ X : M \rightarrow M$  is the identity map.

*bundle*. If  $M$  is  $m$ -dimensional, then the total space in this case will be  $2m$ -dimensional. Alternatively, for an  $m$ -dimensional manifold embedded in  $\mathbb{R}^n$ , each point  $x \in M$  has an associated normal space,  $NM_x$ , which is the  $(n-m)$ -dimensional orthogonal complement of  $TM_x$  in  $\mathbb{R}^n$ . The total space consisting of all pairs  $(x, NM_x)$  is  $n$ -dimensional. This total space together with  $B = M$ ,  $F \cong \mathbb{R}^{n-m}$ , and an appropriate projection map defines a *normal bundle*.

Other sorts of fiber bundles exist. (For example, the tubular surface example is not a vector bundle.) A framed simple curve or surface, or more generally a manifold with a frame attached at each point, defines a *frame bundle*. If associated with each point  $x \in M$  a sphere is attached, then the result is a sphere bundle. The previously described decomposition of the tubular surface into a bundle of circles is an example of this. A famous use of the concept of a fiber bundle is the *Hopf fibration* of the three sphere [33, 34]. Whereas  $S^3 \neq S^1 \times S^2$ , it is possible to view  $S^3$  as a fiber bundle in which  $S^2$  is the base space and each fiber is isomorphic to a copy of  $S^1$ .

The following example arose as part of the author's studies in the statistical mechanics of polymer chains. Consider a "semi-flexible polymer" (i.e., one whose shape can be described by a differentiable backbone curve). An example of such a polymer is the DNA molecule. Let the end positions be fixed at a distance that is shorter than the length of the backbone curve segment connecting the end points. As the molecule is subjected to forcing due to Brownian motion, the minimal energy shape of this polymer (highlighted as the dark path in Figure 7.7.1) will be perturbed, and will produce an ensemble of different shapes. Each of these shapes is called a *conformation*.

Each of the backbone curves corresponding to an individual conformation can be fully described at a certain level of detail as a framed curve. Each of these conformations may have a different length as they stretch or compress. However, each framed curve representing an individual conformation can be parameterized with the arc length  $s \in (0, 1)$  of the baseline (minimal energy) conformation, where the values  $s = 0$  and  $s = 1$  corresponding to the end constraints are excluded. Suppose that a continuous index set,  $\mathcal{Y}$ , is assigned to track all possible conformations. For example, at  $s = 1/2$  we can imagine recording the cloud of reference frames that are visited by the infinite number of possible conformations, and tag each one with an element of  $\mathcal{Y}$ . Each tag  $v \in \mathcal{Y}$  might be constructed with a lot of detailed facts about the conformation, including the curvature and torsion functions that define it. If enough of the attributes of each conformation are captured by each tag, then it is acceptable to assume as part of this model that for each  $v \in \mathcal{Y}$  there is only one conformation.

Then when considering all possible conformations, the full set of reference frames will be

$$E = \{g(s, v) \mid s \in (0, 1), v \in \mathcal{Y}\}.$$

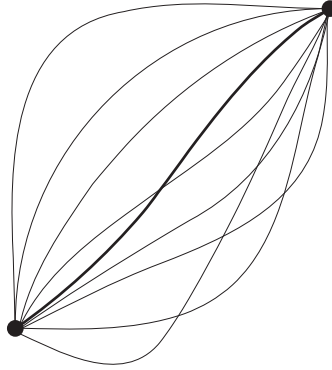
This set can be fibered in two ways. Perhaps the most intuitive way is to treat each conformation as a fiber that sprouts in both directions from the base space  $B = \{g(1/2, v) \mid v \in \mathcal{Y}\}$  and terminates at the fixed end points. The projection map in this case shrinks each fiber (i.e., each framed curve) to a specific tag,  $v \in \mathcal{Y}$ . And each fiber can be viewed as the inverse image of this projection map. Locally, any slice of this bundle of conformations from  $s = s_0$  to  $s = s_0 + \epsilon$  looks like the manifold  $U \times (s_0, s_0 + \epsilon)$  where  $U \subset SE(3)$  (the Lie group of rigid-body motions in three-dimensional space). And, if each tag is constructed with enough detail, it is possible to map each  $U \times (s_0, s_0 + \epsilon)$  to a specific  $v \in \mathcal{Y}$  for sufficiently small  $\epsilon \in \mathbb{R}_{>0}$ .

But, as with the torus example described earlier, there can be more than one way to describe a given space as a fiber bundle. Rather than treating conformations as fibers,



we can treat the unperturbed minimal energy conformation as the base space. Then, for each fixed value of  $s \in (0, 1)$  corresponding to a specific point on this baseline conformation, fibers can be defined as the subset of rigid-body motions attainable by the specific frame at arc-length  $s$  under forcing by Brownian motion:

$$F(s) = \{g(s, v) \mid v \in \mathcal{Y}\} \subset SE(3).$$



**Fig. 7.6.** A Conformational Bundle

A classic work on fiber bundles in which topological and group-theoretic aspects are addressed is Steenrod [60]. Kobayashi and Nomizu [38] also provide a comprehensive geometric treatment. However, they do not consider the sort of infinite dimensional conformational bundles discussed here and in [17].

### 7.7.2 Connections

For an  $m$ -dimensional manifold embedded in  $\mathbb{R}^n$ , many geometric calculations are straightforward, and are quite similar to those performed for a curve or surface in  $\mathbb{R}^3$ . This is because the position of every point on a curve in an embedded manifold is a parameterized curve in  $\mathbb{R}^n$ . The tangent vector to this curve is tangent to the manifold, and is also a vector in  $\mathbb{R}^n$ . Explicitly this tangent vector is obtained by simply taking the derivative of the parameterized vector of Cartesian coordinates of position along the curve in  $\mathbb{R}^n$ . The second derivative of position with respect to curve parameter will have a component in the tangent space and a component in the normal space. This is not a problem at all for curves on embedded manifolds. However, for an abstract manifold that is not embedded in  $\mathbb{R}^n$ , there is no ambient space that “ties together” each of the local coordinate charts, and there is no normal space defined in which the second derivative can have a component. However, it is still possible to consider the component of the derivative of the tangent to a curve in the tangent space. In the context of an embedded manifold, this component is the projection of the second derivative of position along a curve onto the tangent space, which defines the *covariant derivative* of a tangent vector.

In the case when a manifold is not embedded in  $\mathbb{R}^n$ , things are not as straightforward, since the tangent space at  $x(t)$  and  $x(t + dt)$  are two different spaces. Whereas this is not a big deal for embedded manifolds because in that context  $TM_{x(t)}$  and  $TM_{x(t+dt)}$  can be related by an affine motion in  $\mathbb{R}^n$ , in the case when the way the manifold is

embedded in  $\mathbb{R}^n$  is not known, an additional quantity must be defined to “connect” the properties of the tangent space at one point in a manifold with those of a tangent space of a nearby point in the manifold. This *connection* can be thought of as a rule for computing the part of the rate of change of a tangent vector that remains in the tangent space as a curve meanders through a manifold.

In Section 7.3.1, vector fields were defined in a coordinate-dependent way for manifolds embedded in  $\mathbb{R}^n$ . In that context, the basis vectors for a vector field were given as vectors  $\partial\mathbf{x}/\partial q_i \in \mathbb{R}^n$ . In the coordinate-free setting, the basis vectors are denoted as  $X_i$  for  $i = 1, \dots, m$  and there is no dependence of the definition on a particular set of coordinates or a particular embedding of the manifold in  $\mathbb{R}^n$ . In this context the set of all smooth vector fields on a manifold  $M$  is denoted as  $\mathfrak{X}(M)$ . For any  $V \in \mathfrak{X}(M)$  evaluated at any  $x \in M$  as  $V_x = \sum_i v_i(x)X_i$  this means that each  $v_i(x) \in C^\infty(M)$ . In this notation given any  $\phi \in C^\infty(M)$  the product  $\phi V_x$  is simply scalar multiplication, but when written in the reverse order,  $V_x(\phi)$  involves differentiation of  $\phi$ .

Given a Riemannian manifold,  $M$ , with metric tensor  $G = [g_{ij}]$  and Christoffel symbols  $\Gamma_{ij}^k = \Gamma_{ji}^k$ , a *symmetric Riemannian connection* can be defined to be the unique mapping that takes in any two smooth vector fields  $X, Y \in \mathfrak{X}(M)$  and produces a new one,  $\nabla_X Y \in \mathfrak{X}(M)$  with certain properties. In particular, if  $V, W, X, Y \in \mathfrak{X}(M)$  and  $\phi, \psi \in C^\infty(M)$ , the following properties are defined to hold for a symmetric Riemannian connection  $\nabla$  as stated in [9, 20]:

$$\nabla_{\phi V + \psi W} Y = \phi \nabla_V Y + \psi \nabla_W Y \tag{7.112}$$

$$\nabla_X (V + W) = \nabla_X V + \nabla_X W \tag{7.113}$$

$$\nabla_X (\phi Y) = \phi \nabla_X Y + X(\phi) Y \tag{7.114}$$

$$\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k. \tag{7.115}$$

Furthermore, the covariant derivative of a vector field  $V_{x(q(t))} = \sum_i v_k(t) X_k$  defined in terms of coordinates  $\mathbf{q}$  as

$$\frac{DV}{dt} \doteq \sum_k \left( \frac{dv_k}{dt} + \sum_{i,j} v_j \frac{dq_i}{dt} \Gamma_{ij}^k \right) X_k \tag{7.116}$$

satisfies the condition [20]

$$\frac{d}{dt}(V, W) = \left( \frac{DV}{dt}, W \right) + \left( V, \frac{DW}{dt} \right) \tag{7.117}$$

where  $(\cdot, \cdot)$  is the coordinate-free version of the inner product defined in coordinates in (7.15). This bare-bones description of connections can be augmented by further reading. See, for example, [11] and references therein.

Such tools are obviously important for physicists studying the large-scale structure of the physical universe since it is not at all obvious how the universe might be embedded in some larger Euclidean space. However, for engineering and biology problems, where the manifolds of interest are those that describe allowable motions of a physical system, the embedding in  $\mathbb{R}^n$  is often dictated by the problem itself. And it is therefore not necessary to pretend that the way in which the configuration manifold is embedded is not known.

## 7.8 The Heat Equation on a Riemannian Manifold

The heat equation on a Riemannian manifold can be defined as

$$\boxed{\frac{\partial f}{\partial t} = \operatorname{div}(\operatorname{grad} f)}. \quad (7.118)$$

An eigenfunction of the Laplacian<sup>21</sup> is a function that satisfies

$$\operatorname{div}(\operatorname{grad}\psi) = \lambda\psi$$

for some scalar number  $\lambda$ , called an eigenvalue. For a compact Riemannian manifold,  $M$ , the set of all such numbers is discrete (i.e., infinite but countable). Furthermore, given two eigenvalue–eigenfunction pairs,  $(\psi_i, \lambda_i)$  and  $(\psi_j, \lambda_j)$ , the rules for integration discussed in prior sections of this chapter can be used to show that for a compact Riemannian manifold without boundary

$$\lambda_i \neq \lambda_j \implies (\psi_i, \psi_j) = 0,$$

where  $(\cdot, \cdot)$  denotes the inner product of scalar functions on  $M$ , i.e., the integral over  $M$  of the product of the functions. The set of all eigenfunctions forms a complete orthogonal system of functions on  $M$  that can be taken to be orthonormal without loss of generality. Furthermore,  $\lambda_i \in \mathbb{R}$  and it is a non-positive number. For a compact Riemannian manifold without boundary, the eigenvalue with smallest absolute value will be  $\lambda_0 = 0$ .

Letting  $\mathbf{q}$  denote local coordinates, expanding  $f(\mathbf{q}, t) = \sum_i c_i(t)\psi_i(\mathbf{q})$ , and substituting into (7.118) then gives

$$f(\mathbf{q}, t) = \sum_{i=0}^{\infty} c_i(0)e^{-t|\lambda_i|}\psi_i(\mathbf{q}).$$

If  $f(\mathbf{q}, 0)$  is a pdf, then so too will be  $f(\mathbf{q}, t)$ . In this case  $c_0(0) = 1/\operatorname{Vol}(M)$ . And so

$$\lim_{t \rightarrow \infty} f(\mathbf{q}, t) = 1/\operatorname{Vol}(M).$$

The values of all of the coefficients  $\{c_i(0)\}$  are dictated by initial conditions. The rate at which the above limit is reached is dictated by the value of the eigenvalue with next smallest absolute value. The geometric meanings of eigenvalues of the Laplacian have been studied extensively. See [15, 54] for details and numerous references.

## 7.9 Chapter Summary

This chapter introduced the concept of a manifold together with the mathematical tools required to compute curvature and to integrate on manifolds. The concept and properties of convex polytopes were explored. The relationship between the Minkowski sum of polytopes and the convolution product of functions was examined.

<sup>21</sup>In differential geometry the Laplacian is usually defined as the negative of this, so that the eigenvalues are all non-negative. The notation used here differs from that convention in order to be more consistent with the way Laplacians are defined in engineering applications.

In this chapter polytopes and manifolds were treated. While it was not proven here that the formula for the Euler characteristic extends to higher-dimensional Euclidean spaces, this fact can be found in the literature. See, for example, [42]. In recent years some work has been done to merge these topics [8]. The differential topology of manifolds via the Gauss–Bonnet–Chern theorem and the use of the Euler characteristic in higher dimensions were briefly touched on. Works that focus on the connection between geometry and topology of manifolds include [10, 68, 69]. Other readable introductions to differential geometry and topology of manifolds include [5, 27, 40, 49]. Applications of differential geometry to mechanics are addressed in [1, 7, 11, 12].

In the next chapter stochastic differential equations on manifolds are discussed, and the corresponding Fokker–Planck equations are derived. This general theory is illustrated in the context of Brownian motion on the sphere. The specialized case of Brownian motion on Lie groups will be discussed in detail in Volume 2. It is important to keep in mind that the problems discussed in the next chapter involve the flow of probability density on a manifold. The corresponding partial differential (Fokker–Planck) equations are linear. This is very different than the topic of *Ricci flow*, which has received considerable attention in recent years. In Ricci flow, a non-linear partial differential equation of the form

$$\frac{\partial G}{\partial t} = -2Ric(G)$$

is propagated for given initial metric tensor  $G_0$ , where  $Ric(G)$  is the Ricci curvature tensor defined in (7.42). The long-time behavior of  $G(t)$  governed by this equation is then used to determine topological properties of the manifold described by  $G(t)$  as  $t \rightarrow \infty$ . While this subject is not addressed in this book other than in this paragraph, it is a hot area of research worth knowing about. For additional reading, see [31, 52, 62] and references therein.

## 7.10 Exercises

7.1. Verify that the Euler characteristic of the surfaces of the regular polyhedra in  $\mathbb{R}^3$  (Tetrahedron, Cube, Octahedron, Dodecahedron, Icosahedron) are all  $\chi(\partial B) = 2$ . Then divide up the polyhedral bodies defined by these surfaces into pyramidal cells, and verify that  $\chi(B) = 1$ .

7.2. The *Klein bottle* is like a 2-torus that is twisted in  $\mathbb{R}^4$ . It can be parameterized as [6]

$$\begin{aligned} x_1 &= (a + b \sin \theta) \cos \phi \\ x_2 &= (a + b \sin \theta) \sin \phi \\ x_3 &= b \cos \theta \cos \frac{\phi}{2} \\ x_4 &= b \cos \theta \sin \frac{\phi}{2} \end{aligned} \tag{7.119}$$

where  $0 \leq \theta, \phi \leq 2\pi$  and  $a > b$ . Demonstrate that this is not orientable.

7.3. Using integration by parts, prove that the gradient and divergence are the “dual” (or adjoint) of each other in the sense that (7.38) holds.

7.4. Which of the following are manifolds? (a) the standard unit circle in the plane; (b) the open disk (region enclosed by, but not including, the unit circle); (c) a figure-eight curve; (d) a two-sided cone; (e) a one-sided cone; (f) a sphere; (g) the set of all  $n \times n$

real matrices; (h) the intersection of a sphere and infinite cylinder in three-dimensional space.

7.5. The 2-torus examined in Section 5.4.5 is embedded in  $\mathbb{R}^3$ . It is possible to embed a 2-torus in  $\mathbb{R}^4$  as

$$\mathbf{x}(\theta, \phi) = [r \cos \theta, r \sin \theta, R \cos \phi, R \sin \phi]^T.$$

Compute the metric tensor and Gaussian curvature for this 2-torus, and use the Gauss–Bonnet theorem to verify that it has the same genus as the 2-torus in Section 5.4.5.

7.6. In analogy with (5.140), for a four-dimensional array of hyper-cubes compute  $f_0(\partial B)$ ,  $f_1(\partial B)$ ,  $f_2(\partial B)$ , and  $f_3(\partial B)$ .

7.7. Using the result of the previous problem, show that

$$\chi(\partial B) = f_0(\partial B) - f_1(\partial B) + f_2(\partial B) - f_3(\partial B) = 0,$$

and this remains true regardless of any sculpting, void formation, drilling, or cleaving operations.

7.8. Using Stokes' theorem, show that if  $\omega = \theta \wedge \tau$  is an  $(n-1)$ -form with  $\theta$  being a  $p$ -form and  $\tau$  being a  $q$ -form (so that  $p + q = n - 1$ ), then the following integration-by-parts formula holds [6]:

$$\int_M d\theta \wedge \tau = \int_{\partial M} \theta \wedge \tau - (-1)^p \int_M \theta \wedge d\tau, \quad (7.120)$$

and in particular, if  $\phi$  is a scalar function (i.e., a 0-form) and  $\alpha$  is an  $(n-1)$ -form, then from Stokes' theorem  $\omega = \phi \alpha$  satisfies

$$\int_M d\phi \wedge \alpha = \int_{\partial M} \phi \alpha - \int_M \phi d\alpha. \quad (7.121)$$

7.9. Consider two surfaces  $\mathbf{x}(u_1, u_2), \mathbf{y}(v_1, v_2) \in \mathbb{R}^3$  that respectively have Gaussian curvatures  $k_x(u_1, u_2)$  and  $k_y(v_1, v_2)$ . Embed the Cartesian product of these surfaces in  $\mathbb{R}^6$  using the rule

$$\mathbf{z}(u_1, u_2, v_1, v_2) = \begin{pmatrix} \mathbf{x}(u_1, u_2) \\ \mathbf{y}(v_1, v_2) \end{pmatrix}.$$

Using the Gauss–Bonnet–Chern theorem, show that (5.104) holds in this special case.

7.10. Using (6.96) as the starting point, the *Hodge star operator*,  $*$ , applied to a  $k$ -form,  $\beta_k$ , on an  $n$ -dimensional manifold,  $M$ , can be defined as  $*\beta_k$  such that for any  $k$ -form  $\alpha_k$  the following equality holds:

$$\int_M \alpha_k \wedge *\beta_k = \int_M \boldsymbol{\alpha} \cdot \boldsymbol{\beta} dV.$$

Here  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  can be thought of as column arrays of length  $\binom{n}{k}$  that define the forms  $\alpha_k$  and  $\beta_k$ , where each entry in these arrays is a real-valued function on  $M$ . From this definition: (a) choose your favorite three-dimensional manifold, and compute the Hodge star operator for generic 1-, 2-, and 3-forms on that manifold; (b) use the generalized Levi–Civita symbol to write  $*\beta_k$  for any  $\beta_k$  on any  $n$ -manifold.

## References

1. Abraham, R., Marsden, J.E., Ratiu, T., *Manifolds, Tensor Analysis, and Applications*, 2nd ed., Applied Mathematical Sciences 75, Springer, New York, 1988.
2. Allendoerfer, C.B., “The Euler number of a Riemannian manifold,” *Amer. J. Math.*, 62, pp. 243–248, 1940.
3. Allendoerfer, C.B., Weil, A., “The Gauss–Bonnet theorem for Riemannian polyhedra,” *Trans. Amer. Math. Soc.*, 53, pp. 101–129, 1943.
4. Apéry, F., *Models of the Real Projective Plane: Computer Graphics of Steiner and Boy Surfaces*, Vieweg, Braunschweig, 1987.
5. Berger, M., *A Panoramic View of Riemannian Geometry*, Springer, New York, 2003.
6. Bishop, R.L., Goldberg, S.I., *Tensor Analysis on Manifolds*, Dover, New York, 1980. (originally MacMillan, 1968).
7. Bloch, A.M., Baillieul, J., Crouch, P., Marsden, J., *Nonholonomic Mechanics and Control*, Springer, New York, 2007.
8. Bobenko, A.I., Schröder, P., Sullivan, J.M., Ziegler, G.M., eds., *Discrete Differential Geometry*, Oberwolfach Seminars, Vol. 38, Birkhäuser, Basel, 2008.
9. Boothby, W.M., *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press, New York, 1975.
10. Bott, R., Tu, L.W., *Differential Forms in Algebraic Topology*, Springer, New York, 1982.
11. Bullo, F., Lewis, A.D., *Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Simple Mechanical Control Systems*, Springer, New York, 2004.
12. Burke, W.L., *Applied Differential Geometry*, Cambridge University Press, London, 1985.
13. Cartan, H., *Differential Forms*, Hermann, Paris; Houghton Mifflin, Boston, 1970.
14. Charlap, L.S., *Bieberbach Groups and Flat Manifolds*, Springer-Verlag, New York, 1986.
15. Chavel, I., *Eigenvalues in Riemannian Geometry*, Academic Press, Orlando, FL, 1984.
16. Chern, S.-S., “A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds,” *Ann. Math.*, 45, pp. 747–752, 1944.
17. Chirikjian, G.S., “The stochastic elastica and excluded-volume perturbations of DNA conformational ensembles,” *Int. J. Non-Linear Mech.*, 43, pp. 1108–1120, 2008.
18. Chirikjian, G.S., Kyatkin, A.B., *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, Boca Raton, FL, 2001.
19. Darling, R.W.R., *Differential Forms and Connections*, Cambridge University Press, London, 1994.
20. do Carmo, M.P., *Riemannian Geometry*, Birkhäuser, Boston, 1992.
21. Farmer, D.W., *Groups and Symmetry*, American Mathematical Society, Providence, RI, 1996.
22. Fenchel, W., “On total curvatures of Riemannian manifolds: I,” *J. London Math. Soc.*, 15, pp. 15–22, 1940.
23. Flanders, H., *Differential Forms with Applications to the Physical Sciences*, Dover, New York, 1989.
24. Fukuda, K., Prodon, A., “Double description method revisited,” in *Combinatorics and Computer Science: 8th Franco-Japanese and 4th Franco-Chinese Conference, Brest, France, July 3–5, 1995: Selected Papers*, Vol. 112, M. Deza, R. Euler, I. Manoussakis, R. Euler, I. Manoussakis, eds., Lecture Notes in Computer Science, Springer, New York, 1996. For software see: [http://www.ifor.math.ethz.ch/~fukuda/cdd\\_home/index.html](http://www.ifor.math.ethz.ch/~fukuda/cdd_home/index.html)
25. Gruber, P.M., *Convex and Discrete Geometry*, Springer-Verlag, Berlin, 2007.
26. Grünbaum, B., *Convex Polytopes*, 2nd ed., Springer, New York, 2003.
27. Guggenheimer, H.W., *Differential Geometry*, Dover, New York, 1977.
28. Guibas, L., Ramshaw, L., Stolfi, J., “A kinetic framework for computational geometry,” *24th Annual Symposium on Foundations of Computer Science*, 7–9 November 1983, Tucson, AZ, pp. 100–111, 1983.
29. Guillemin, V., Pollack, A., *Differential Topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
30. Hadwiger, H., *Altes und Neues über Konvexe Körper*, Birkhäuser Verlag, Basel, 1955.

31. Hamilton, R.S., “Three-manifolds with positive Ricci curvature,” *J. Diff. Geom.*, 17, pp. 255–306, 1982.
32. Hammond, C., *The Basics of Crystallography and Diffraction*, Oxford University Press, Oxford, 1997.
33. Hopf, H., “Über die *curvatura integra* geschlossener Hyperflächen,” *Math. Ann.*, 95, pp. 340–376, 1925.
34. Hopf, H., “Über die Abbildungen der 3-Sphäre auf die Kugelfläche,” *Math. Ann.*, 104, pp. 637–665, 1931.
35. Johnson, C.K., Burnett, M.N., Dunbar, W.D., “Crystallographic topology and its applications,” in *Crystallographic Computing 7: Macromolecular Crystallographic Data*, edited by P.E. Bourne and K.D. Watenpaugh, Oxford University Press, Oxford, 1997.  
<http://www.ornl.gov/sci/ortep/topology/preprint.html>
36. Kavraki, L.E., “Computation of configuration-space obstacles using the fast Fourier transform,” *IEEE Trans. Robotics Automation*, 11, pp. 408–413, 1995.
37. Klein, F., *Vorlesungen über nicht-euklidische Geometrie*, Springer-Verlag, New York, 1968.
38. Kobayashi, S., Nomizu, K., *Foundations of Differential Geometry Vols. I and II*, John Wiley & Sons, New York, 1963 (Wiley Classics Library Edition 1996).
39. Ladd, M.F.C., *Symmetry in Molecules and Crystals*, Ellis Horwood/John Wiley & Sons, New York, 1989.
40. Lang, S., *Fundamentals of Differential Geometry*, Springer, New York, 1999.
41. Lattman, E.E., Loll, P.J., *Protein Crystallography: A Concise Guide*, The Johns Hopkins University Press, Baltimore, 2008.
42. Lawrence, J., “A short proof of Euler’s relation for convex polytopes,” *Can. Math. Bull.*, 40, pp. 471–474, 1997.
43. Lee, J.M., *Riemannian Manifolds: An Introduction to Curvature*, Springer, New York, 1997.
44. Lockwood, E.H., MacMillan, R.H., *Geometric Symmetry*, Cambridge University Press, London, 1978.
45. Lovelock, D., Rund, H., *Tensors, Differential Forms, and Variational Principles*, Dover, New York, 1989.
46. McPherson, A., *Introduction to Macromolecular Crystallography*, John Wiley & Sons, Hoboken, NJ, 2003.
47. Montesinos, J.M., *Classical Tessellations and Three-Manifolds*, Springer-Verlag, Berlin, 1987.
48. Morgan, F., *Riemannian Geometry : A Beginner’s Guide*, 2nd ed., A.K. Peters, Wellesley, MA, 1998.
49. Mukherjee, A., *Topics in Differential Topology*, Hindustan Book Agency, New Delhi, 2005.
50. Nash, J., “The embedding theorem for Riemannian manifolds,” *Ann. Math.*, 63, pp. 20–63, 1956.
51. Oprea, J., *Differential Geometry and Its Applications*, 2nd ed., The Mathematical Association of America, Washington, DC, 2007.
52. Perelman, G., “The entropy formula for the Ricci flow and its geometric applications,” <http://arXiv.org/math.DG/0211159v1> (2002). Updated Feb 1, 2008.
53. Rhodes, G., *Crystallography Made Crystal Clear*, 2nd ed., Academic Press, San Diego, CA, 2000.
54. Rosenberg, S., *The Laplacian on a Riemannian Manifold: An Introduction to Analysis on Manifolds* (London Mathematical Society Student Texts, No. 31), Cambridge University Press, London, 1997.
55. Satake, I., “On a generalization of the notion of a manifold,” *Proc. Nat. Acad. Sci. USA*, 42, pp. 359–363, 1956.
56. Schreiber, M., *Differential Forms: A Heuristic Introduction*, Universitext, Springer-Verlag, New York, 1977.
57. Scott, P., “The geometries of 3-manifolds,” *Bull. London Math. Soc.*, 15, pp. 401–487, 1983.
58. Spivak, M., *A Comprehensive Introduction to Differential Geometry*, Vols. 1, 2, Publish or Perish, Houston, TX, 1970.

59. Spivak, M., *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*, HarperCollins, New York, 1965.
60. Steenrod, N., *The Topology of Fibre Bundles*, Princeton University Press, Princeton, NJ, 1951 (reprinted 1999).
61. Thurston, W.P., *Three-Dimensional Geometry and Topology*, Vol. 1, (edited by S. Levy), Princeton University Press, Princeton, NJ, 1997.
62. Topping, P., *Lectures on the Ricci Flow*, London Mathematical Society Lecture Notes 325, Cambridge University Press, London, 2006.
63. Tu, L.W., *An Introduction to Manifolds*, Springer, New York, 2008.
64. Warner, F.W., *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, New York, 1983.
65. Weeks, J.R., *The Shape of Space*, Marcel Dekker, New York, 1985.
66. Weinstein, A., "Groupoids: Unifying internal and external symmetry," *Not. Amer. Math. Soc.*, 43, pp. 744–752, 1996.
67. Whitney, H., "Differentiable manifolds," *Ann. Math.*, 37, pp. 645–680, 1936.
68. Willmore, T.J., *Total Curvature in Riemannian Geometry*, Ellis Horwood/John Wiley & Sons, New York, 1982.
69. Yano, K., Bochner, S., *Curvature and Betti Numbers*, Annals of Mathematics Studies 32, Princeton University Press, Princeton, NJ, 1953.
70. Ziegler, G.M., *Lectures on Polytopes*, Springer, New York, 1995.



---

## Stochastic Processes on Manifolds

This chapter extends the discussion of stochastic differential equations and Fokker–Planck equations on Euclidean space initiated in Chapter 4 to the case of processes that evolve on a Riemannian manifold. The manifold either can be embedded in  $\mathbb{R}^n$  or can be an abstract manifold with Riemannian metric defined in coordinates. Section 8.1 formulates SDEs and Fokker–Planck equations in a coordinate patch. Section 8.2 formulates SDEs for an implicitly defined embedded manifold using Cartesian coordinates in the ambient space. Section 8.3 focuses on Stratonovich SDEs on manifolds. The subtleties involved in the conversion between Itô and Stratonovich formulations are explained. Section 8.4 explores entropy inequalities on manifolds. In Section 8.5 the following examples are used to illustrate the general methodology: (1) Brownian motion on the sphere and (2) the stochastic kinematic cart described in Chapter 1. Section 8.6 discusses methods for solving Fokker–Planck equations on manifolds. Exercises involving numerical implementations are provided at the end of the chapter.

The main points to take away from this chapter are:

- SDEs and Fokker–Planck equations can be formulated for stochastic processes in any coordinate patch of a manifold in a way that is very similar to the case of  $\mathbb{R}^n$ ;
- Stochastic processes on embedded manifolds can also be formulated extrinsically, i.e., using an implicit description of the manifold as a system of constraint equations;
- In some cases Fokker–Planck equations can be solved using separation of variables;
- Practical examples of this theory include Brownian motion on the sphere and the kinematic cart with noise.

### 8.1 The Fokker–Planck Equation for an Itô SDE on a Manifold: A Parametric Approach

The derivation of the Fokker–Planck equation governing the time evolution of pdfs on a Riemannian manifold proceeds in an analogous way to the derivation in  $\mathbb{R}^n$  that was provided in Section 4.5.6. This subject has been studied extensively in the mathematics literature. See, for example, the classic works of Yosida [30, 31, 32], Itô [11, 12], and McKean [17]. Aspects of diffusion processes on manifolds remain of interest today (see, e.g., [1, 5, 6, 8, 10, 13, 14]).

Unlike many derivations in the modern mathematics literature, the derivation of the Fokker–Planck equation for the case of a Riemannian manifold presented in this section is strictly coordinate-dependent.

The coordinates for a patch in a  $d$ -dimensional manifold are written as a column vector  $\mathbf{q} = [q_1, \dots, q_d]^T$ . In this context  $q_i(t)$  denotes a stochastic process corresponding to the coordinate  $q_i$ .

Consider the Itô SDE

$$d\mathbf{q} = \mathbf{h}(\mathbf{q}, t) + H(\mathbf{q}, t)d\mathbf{w} \tag{8.1}$$

where  $\mathbf{h}, \mathbf{q} \in \mathbb{R}^d$  and  $\mathbf{w} \in \mathbb{R}^m$ .

The notation  $d\mathbf{q}$  has been “double packed” in the sense that it has two meanings that are distinguished by their context. In the first meaning in (8.1),  $d\mathbf{q} = \mathbf{q}(t + dt) - \mathbf{q}(t)$ . However, it is also convenient to write  $d\mathbf{q} = dq_1dq_2 \dots dq_d$  as the volume element in parameter space. These two very different quantities typically do not appear in the same equation. If ever they do, then the volume element is denoted as  $d(\mathbf{q})$ .

The definition of the Wiener process and associated derivations proceed as in Chapter 4. The metric tensor for the manifold is  $G = [g_{ij}]$ , the inverse of which is denoted  $G^{-1} = [g^{ij}]$ . The only difference now is that the integration by parts required to isolate the function  $\epsilon(\mathbf{q})$  from the rest of the integrand in the final steps in the derivation in Section 4.5.6 will be weighted by  $|G|^{\frac{1}{2}}$  due to the fact that the volume element in the manifold  $M$  is  $dV(\mathbf{q}) = |G(\mathbf{q})|^{\frac{1}{2}}d\mathbf{q}$  where  $d\mathbf{q} = dq_1 \dots dq_d$ . In particular, if  $p(\mathbf{q}|\mathbf{s}, t)$  denotes the transition probability for the stochastic process corresponding to the SDE from the state  $\mathbf{s} = \mathbf{q}(t - dt)$  to  $\mathbf{q} = \mathbf{q}(t)$ , then

$$\begin{aligned} \int_M \frac{\partial \epsilon}{\partial s_i} h_i(\mathbf{s}, t) p(\mathbf{q}|\mathbf{s}, t) dV(\mathbf{s}) &= \int_{\mathbb{R}^d} \frac{\partial \epsilon}{\partial s_i} h_i(\mathbf{s}, t) p(\mathbf{q}|\mathbf{s}, t) |G(\mathbf{s})|^{\frac{1}{2}} d\mathbf{s} = \\ &- \int_{\mathbb{R}^d} \epsilon(\mathbf{s}) \frac{\partial}{\partial s_i} \left( h_i(\mathbf{s}, t) p(\mathbf{q}|\mathbf{s}, t) |G(\mathbf{s})|^{\frac{1}{2}} \right) d\mathbf{s}. \end{aligned}$$

The integration over all  $\mathbb{R}^d$  is valid since  $\epsilon(\mathbf{s})$  and its derivatives vanish outside of a compact subset of  $\mathbb{R}^d$  which can be assumed to be contained in the range of a single coordinate chart of  $M$ . Then integrating by parts twice yields

$$\begin{aligned} \int_M \frac{\partial^2 \epsilon}{\partial s_i \partial s_j} h_i(\mathbf{s}, t) p(\mathbf{q}|\mathbf{s}, t) dV(\mathbf{s}) &= \\ \int_{\mathbb{R}^d} \epsilon(\mathbf{s}) \frac{\partial^2}{\partial s_i \partial s_j} \left( h_i(\mathbf{s}, t) p(\mathbf{q}|\mathbf{s}, t) |G(\mathbf{s})|^{\frac{1}{2}} \right) d\mathbf{s}. \end{aligned}$$

Using the standard localization argument as in Section 4.5.6, extracting the functional which multiplies  $\epsilon(\mathbf{s})$ , and setting it equal to zero yields, after division by  $|G(\mathbf{q})|^{\frac{1}{2}}$ , the following Fokker–Planck equation:

$$\boxed{\frac{\partial f}{\partial t} + |G|^{-\frac{1}{2}} \sum_{i=1}^d \frac{\partial}{\partial q_i} \left( |G|^{\frac{1}{2}} h_i f \right) = \frac{1}{2} |G|^{-\frac{1}{2}} \sum_{i,j=1}^d \frac{\partial^2}{\partial q_i \partial q_j} \left( |G|^{\frac{1}{2}} \sum_{k=1}^m H_{ik} H_{kj}^T f \right)} \tag{8.2}$$

where the simplifying notation  $f(\mathbf{q}, t) \doteq p(\mathbf{q}|\mathbf{s}, t)$  is used. The second term on the left side of the equation above can be written as  $\text{div}(f\mathbf{h})$  (where the divergence operator is defined in (5.49)), and this raises questions about the differential-geometric interpretation of the right-hand side. In many cases of interest, the matrices  $H_{ik}(\mathbf{q}, t)$  will be the inverse of the Jacobian matrix, and hence in these cases  $\sum_k H_{ik}(\mathbf{q}, t) H_{kj}^T(\mathbf{q}, t) = \sum_k ((J_{ik})^{-1} ((J_{kj})^{-1})^T) = (g_{ij}(\mathbf{q}))^{-1} = (g^{ij}(\mathbf{q}))$ . In those cases, the Fokker–Planck equation on  $M$  becomes

$$\begin{aligned} \frac{\partial f(\mathbf{q}, t)}{\partial t} + |G(\mathbf{q})|^{-\frac{1}{2}} \sum_{i=1}^d \frac{\partial}{\partial q_i} \left( |G(\mathbf{q})|^{\frac{1}{2}} h_i(\mathbf{q}, t) f(\mathbf{q}, t) \right) = \\ \frac{1}{2} |G(\mathbf{q})|^{-\frac{1}{2}} \sum_{i,j=1}^d \frac{\partial^2}{\partial q_i \partial q_j} \left( |G(\mathbf{q})|^{\frac{1}{2}} (g^{ij}(\mathbf{q})) f(\mathbf{q}, t) \right). \end{aligned} \quad (8.3)$$

This equation is similar to, though not exactly the same as, the heat equation on  $M$  written in coordinate-dependent form. In fact, a straightforward calculation explained by Brockett [4] equates the Fokker–Planck and heat equation

$$\frac{\partial f(\mathbf{q}, t)}{\partial t} = \frac{1}{2} |G(\mathbf{q})|^{-\frac{1}{2}} \sum_{i,j=1}^d \frac{\partial}{\partial q_i} \left( |G(\mathbf{q})|^{\frac{1}{2}} (g^{ij}(\mathbf{q})) \frac{\partial}{\partial q_j} f(\mathbf{q}, t) \right) = \frac{1}{2} \nabla^2 f \quad (8.4)$$

in the special case when

$$h_i(\mathbf{q}) = \frac{1}{2} \sum_{j=1}^d \left( |G(\mathbf{q})|^{-\frac{1}{2}} (g^{ij}(\mathbf{q})) \frac{\partial |G(\mathbf{q})|^{\frac{1}{2}}}{\partial q_j} + \frac{\partial g^{ij}(\mathbf{q})}{\partial q_j} \right).$$

This is clear by expanding the term on the right-hand side of (8.3) as

$$\frac{1}{2} |G(\mathbf{q})|^{-\frac{1}{2}} \sum_{i,j=1}^d \frac{\partial}{\partial q_i} \left[ \frac{\partial}{\partial q_j} \left( |G(\mathbf{q})|^{\frac{1}{2}} (g^{ij}(\mathbf{q})) f(\mathbf{q}, t) \right) \right],$$

and observing the chain rule:

$$\begin{aligned} \frac{\partial}{\partial q_j} \left( |G(\mathbf{q})|^{\frac{1}{2}} (g^{ij}(\mathbf{q})) f(\mathbf{q}, t) \right) = \left( (g^{ij}(\mathbf{q})) \frac{\partial |G(\mathbf{q})|^{\frac{1}{2}}}{\partial q_j} + |G(\mathbf{q})|^{\frac{1}{2}} \frac{\partial g^{ij}}{\partial q_j} \right) f(\mathbf{q}, t) \\ + (g^{ij}(\mathbf{q})) |G(\mathbf{q})|^{\frac{1}{2}} \frac{\partial f(\mathbf{q}, t)}{\partial q_j}. \end{aligned}$$

## 8.2 Itô Stochastic Differential Equations on an Embedded Manifold: An Implicit Approach

### 8.2.1 The General Itô Case

Let  $\mathbf{x}(t) \in \mathbb{R}^n$  denote a vector-valued stochastic process. Define the vector-valued function  $\mathbf{a}(\mathbf{x}, t) \in \mathbb{R}^n$  and the matrix-valued function  $B(\mathbf{x}, t) \in \mathbb{R}^{n \times n}$ . Then  $\mathbf{x}(t)$  is defined by the stochastic differential equation (SDE)

$$d\mathbf{x} = \mathbf{a}(\mathbf{x}, t)dt + B(\mathbf{x}, t)d\mathbf{w} \quad (8.5)$$

where  $d\mathbf{w} \in \mathbb{R}^n$  is a vector of uncorrelated, unit-strength white noise processes. This is not the same as the equation defined in coordinates in the previous section. It does not have the same dimensions or variables.

Now suppose that a system of constraints  $\mathbf{y} = \mathbf{g}(\mathbf{x}, t) = \mathbf{0} \in \mathbb{R}^m$  is used to define a manifold,  $M$ , of dimension  $d = n - m$  embedded in  $\mathbb{R}^n$ . What must be true for the Itô SDE in (8.5), which is defined in the ambient Euclidean space, to evolve on the

manifold  $M$ ? The answer is provided by *Itô's rule* (4.55), which in the current context means that starting from a value  $\mathbf{x}_0$  such that  $\mathbf{g}(\mathbf{x}_0, t) = \mathbf{0}$  and satisfying the condition that  $d\mathbf{y} = \mathbf{0}$ , then in component form the following condition must hold:

$$\frac{\partial g_k}{\partial t} dt + \sum_{i=1}^n \frac{\partial g_k}{\partial x_i} dx_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 g_k}{\partial x_i \partial x_j} dx_i dx_j = 0 \tag{8.6}$$

when (8.5) is substituted in. This imposes conditions on the allowable  $\mathbf{a}$  and  $B$  such that the sample paths of the SDE will evolve on the time-evolving manifold  $M$ . In practice, the case when  $M$  does not change with time is more common. In that case,  $g(x, t) = g(x)$ .

### 8.2.2 Bilinear Itô Equations that Evolve on a Quadratic Hyper-Surface in $\mathbb{R}^n$

In [3], Brockett examined the properties of Itô SDEs of the form

$$d\mathbf{x} = A\mathbf{x} dt + \sum_{i=1}^m dw_i B_i \mathbf{x} \tag{8.7}$$

(where  $A$  and  $B_i$  are independent of  $\mathbf{x}$ ) and explained simple conditions under which this equation would evolve on a quadratic hyper-surface in  $\mathbb{R}^n$  of the form

$$\mathbf{x}^T Q \mathbf{x} = 1 \tag{8.8}$$

where  $Q = Q^T \in \mathbb{R}^n$ . These include ellipsoidal and hyperbolic hyper-surfaces.

Applying the derivative to (8.8) results in the condition  $d(\mathbf{x}^T Q \mathbf{x}) = 0$ , which is evaluated using Itô's rule (8.6), in the case of the equation  $y = \mathbf{x}^T Q \mathbf{x}$  (which has no subscript  $k$  since it is a single scalar equation), to yield

$$2(Q\mathbf{x})^T d\mathbf{x} + d\mathbf{x}^T Q d\mathbf{x} = 0. \tag{8.9}$$

Substitution of (8.7) into this equation gives

$$2\mathbf{x}^T \left[ Q A dt + \sum_i dw_i Q B_i \right] \mathbf{x} + \mathbf{x}^T \left[ (A^T dt + \sum_i dw_i B_i^T) Q (A dt + \sum_j dw_j B_j) \right] \mathbf{x} = 0. \tag{8.10}$$

From the rules of stochastic calculus,  $dt dw_i = 0$  and  $dw_i dw_j = \delta_{ij} dt$  (where equality is under the expectation operator), which reduce the second term to

$$\mathbf{x}^T \left[ (A^T dt + \sum_i dw_i B_i^T) Q (A dt + \sum_j dw_j B_j) \right] \mathbf{x} = \mathbf{x}^T \left[ \sum_i B_i^T Q B_i dt \right] \mathbf{x}.$$

Furthermore, since in general  $\mathbf{x}^T P \mathbf{x} = \frac{1}{2} \mathbf{x}^T (P + P^T) \mathbf{x}$ , the first term in (8.10) can be symmetrized as

$$2\mathbf{x}^T \left[ Q A dt + \sum_i dw_i Q B_i \right] \mathbf{x} = \mathbf{x}^T \left[ (Q A + A^T Q) dt + \sum_i dw_i (Q B_i + B_i^T Q) \right] \mathbf{x}.$$

Therefore, (8.10) is written by combining these two terms as

$$\mathbf{x}^T \left[ (QA + A^T Q + \sum_i B_i^T AB_i) dt + \sum_i dw_i (QB_i + B_i^T Q) \right] \mathbf{x} = 0.$$

It follows that sufficient conditions for (8.7) to “stay on” the manifold defined by (8.8) are

$$QA + A^T Q + \sum_{i=1}^m B_i^T AB_i = 0 \quad \text{and} \quad QB_j + B_j^T Q = 0 \quad \text{for } j = 1, \dots, m. \quad (8.11)$$

Each of the above symmetric  $n \times n$  matrix equations represents  $n(n+1)/2$  scalar constraints, and there are  $m+1$  such equations leading to  $(m+1)n(n+1)/2$  scalar constraints.

Of course, the SDEs in (8.7) are not the only ones that will “stay on” the manifold defined by (8.8). Returning to (8.5) and substituting this SDE into (8.9) yields the conditions

$$B^T Q \mathbf{x} = \mathbf{0} \quad \text{and} \quad \mathbf{x}^T Q \mathbf{a} + \frac{1}{2} \text{tr}(B^T QB) = 0. \quad (8.12)$$

This is a far less restrictive condition than (8.11) because it only imposes  $n+1$  scalar constraints.

### 8.3 Stratonovich SDEs and Fokker–Planck Equations on Manifolds

In analogy with the way Itô equations can be defined either parametrically or implicitly, the same can be done for Stratonovich SDEs. The general theory for the parametric case is presented in Section 8.3.1. The implicit formulation for Stratonovich SDEs on manifolds is discussed in Section 8.3.2.

#### 8.3.1 Stratonovich SDEs on Manifolds: Parametric Approach

The Stratonovich SDE corresponding to (8.1) is

$$d\mathbf{q} = \mathbf{h}^s(\mathbf{q}, t) + H^s(\mathbf{q}, t) \circledast dw \quad (8.13)$$

where  $\circledast$  is used to denote the Stratonovich interpretation of an SDE, and

$$h_i^s = h_i - \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^d H_{kj} \frac{\partial H_{ij}}{\partial q_k} \quad \text{and} \quad H_{ij}^s = H_{ij}. \quad (8.14)$$

If instead the SDE (8.13) is given and the corresponding Itô equation (8.1) is sought, then (8.14) is used in reverse to yield

$$h_i = h_i^s + \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^d H_{kj}^s \frac{\partial H_{ij}^s}{\partial q_k} \quad \text{and} \quad H_{ij} = H_{ij}^s. \quad (8.15)$$

Therefore, it follows from substitution of (8.15) into (8.2) that the Stratonovich version of a Fokker–Planck equation describing a process on a manifold is

$$\boxed{\begin{aligned} \frac{\partial f}{\partial t} + |G|^{-\frac{1}{2}} \sum_{i=1}^d \frac{\partial}{\partial q_i} \left[ \left( h_i^s + \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^d H_{kj}^s \frac{\partial H_{ij}^s}{\partial q_k} \right) f |G|^{\frac{1}{2}} \right] = \\ \frac{1}{2} |G|^{-\frac{1}{2}} \sum_{i,j=1}^d \frac{\partial^2}{\partial q_i \partial q_j} \left( \sum_{k=1}^m H_{ik}^s H_{jk}^s f |G|^{\frac{1}{2}} \right). \end{aligned}} \tag{8.16}$$

This is important because in many physical modeling problems, the following sort of Stratonovich SDE is presented:

$$J(\mathbf{q})d\mathbf{q} = \mathbf{b}(t) + B_0 \circledast d\mathbf{w} \tag{8.17}$$

where  $B_0$  is a constant coupling matrix. For example, if  $g(t)$  represents a rotational or full rigid-body motion, then infinitesimal motions are described in terms of a Jacobian matrix as

$$(g^{-1}\dot{g})^\vee dt = J(\mathbf{q})d\mathbf{q}, \tag{8.18}$$

where  $\vee$  is an operation that extracts the non-redundant information in  $g^{-1}\dot{g}$  and collects it in the form of a column vector. The Jacobian matrix is related to the metric tensor as  $G = J^T J$ .

And (8.17) is written as

$$d\mathbf{q} = [J(\mathbf{q})]^{-1}\mathbf{b}(t) + [J(\mathbf{q})]^{-1}B_0 \circledast d\mathbf{w}. \tag{8.19}$$

The interpretation of (8.17) is what allows for the simple expression in (8.18), rather than the extra terms that would be required when using Itô's rule. Clearly the final result in (8.19) now has a coupling matrix that is not constant, and so even if (8.17) could be interpreted as either Itô or Stratonovich, the result after the Stratonovich interpretation in (8.18) must thereafter be interpreted as a Stratonovich equation.

### 8.3.2 Stratonovich SDEs on Manifolds: Implicit Approach

Given a Stratonovich SDE in  $\mathbb{R}^n$  of the form

$$d\mathbf{x} = \mathbf{a}^s(\mathbf{x}, t)dt + B^s(\mathbf{x}, t) \circledast d\mathbf{w}, \tag{8.20}$$

conditions under which it will evolve on an implicitly defined manifold of dimension  $d = n - m$  follow from the rules of usual calculus. Namely, given the system of constraints  $\mathbf{g}(\mathbf{x}, t) = \mathbf{0} \in \mathbb{R}^m$  that define a manifold,  $M$ , then as long as the initial value  $\mathbf{x}(0) = \mathbf{x}_0 \in \mathbb{R}^n$  satisfies this constraint, the necessary condition for (8.20) to evolve on  $M$  at all future times is simply

$$\sum_{j=1}^n \frac{\partial g_i}{\partial x_j} dx_j = 0 \quad \text{for } i = 1, \dots, m. \tag{8.21}$$

For example, the condition that the system

$$d\mathbf{x} = A^s \mathbf{x} dt + \sum_{i=1}^m dw_i B_i^s \circledast \mathbf{x} \tag{8.22}$$

evolve on a quadratic hyper-surface in  $\mathbb{R}^n$  of the form in (8.8) is simply  $\mathbf{x}^T Q d\mathbf{x} = 0$ . However, the difficulty comes when trying to simplify the result since in the Stratonovich

calculus  $\langle x_i dw_j dw_k \rangle \neq x_i \langle dw_j dw_k \rangle$ . Therefore, at the stage where (8.7) is substituted into  $\mathbf{x}^T Q d\mathbf{x} = 0$  it is convenient to convert everything to Itô form to obtain the constraints on  $A^s$  and  $B_i^s$  for the process to evolve on the manifold defined by  $\mathbf{x}^T Q d\mathbf{x} = 0$ .

Having said this, a word of caution is in order about implicit equations and SDEs on manifolds. Both the formulation in this subsection and that in Section 8.2 are *mathematical* statements of the problem of SDEs on manifolds. This does not mean that they are good *numerical* ways to model SDEs on manifolds. In fact, when using the simplest codes for integrating SDEs, such as the Euler–Maruyama method referenced in Chapter 4, these implicit descriptions can give rise to “solutions” that rapidly diverge from the manifold of interest. In contrast, solutions to SDEs defined parametrically will always stay within the manifold if they remain in a coordinate patch and do not get close to singularities where the parametric description breaks down.

### 8.4 Entropy and Fokker–Planck Equations on Manifolds

The entropy of a probability density function on a manifold can be defined as

$$S(f) \doteq - \int_M f(x) \log f(x) dV = - \int_{\mathbf{q} \in D} f(\mathbf{q}) \log f(\mathbf{q}) |G(\mathbf{q})|^{\frac{1}{2}} d(\mathbf{q}) \tag{8.23}$$

where in the second equality  $f(\mathbf{q})$  is shorthand for  $f(x(\mathbf{q}))$  and  $D \subset \mathbb{R}^n$  is the coordinate domain (assuming that the whole manifold minus a set of measure zero can be parameterized by one such domain).

A natural issue to address is how the entropy  $S(f)$  behaves as a function of time when  $f(x; t)$  satisfies a Fokker–Planck equation. Differentiating (8.23) with respect to time gives

$$\frac{dS}{dt} = - \int_M \left\{ \frac{\partial f}{\partial t} \log f + \frac{\partial f}{\partial t} \right\} dV.$$

The Fokker–Planck equation (8.2) itself can be used to replace the partial derivatives with respect to time with derivatives in local coordinates in the manifold. Then integration by parts can be used. Taking the coordinate-free view, the formulas in (7.98)–(7.101) can be used to convert the integrals of differential operators over the manifold to more convenient integrals. In the case of a manifold without boundary, the same formulas apply with the boundary integrals set equal to zero.

It is easy to show that

$$\int_M \frac{\partial f}{\partial t} dV = \frac{d}{dt} \int_M f dV = 0$$

because probability density is preserved by the Fokker–Planck equation.

Taking a coordinate-dependent view, the remaining term is written as

$$\begin{aligned} \frac{dS}{dt} &= - \int_{\mathbf{q} \in D} \frac{\partial f}{\partial t} \log f |G|^{\frac{1}{2}} d(\mathbf{q}) \\ &= \int_{\mathbf{q} \in D} \left\{ \sum_{i=1}^d \frac{\partial}{\partial q_i} \left( |G|^{\frac{1}{2}} h_i f \right) - \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial q_i \partial q_j} \left( |G|^{\frac{1}{2}} \sum_{k=1}^m H_{ik} H_{kj}^T f \right) \right\} \log f d(\mathbf{q}). \end{aligned}$$

Integrating by parts, and ignoring the boundary terms, gives  $dS/dt$  equal to

$$- \int_{\mathbf{q} \in D} \left\{ \sum_{i=1}^d \frac{\partial f}{\partial q_i} h_i + \frac{1}{2} \sum_{i,j=1}^d \left[ -\frac{1}{f} \sum_{k=1}^m H_{ik} H_{kj}^T \frac{\partial f}{\partial q_i} \frac{\partial f}{\partial q_j} + \frac{\partial^2 f}{\partial q_i \partial q_j} \sum_{k=1}^m H_{ik} H_{kj}^T \right] \right\} |G|^{\frac{1}{2}} d(\mathbf{q}). \tag{8.24}$$

In general it is not guaranteed that this will be a non-negative quantity. For example, the Ornstein–Uhlenbeck process in  $\mathbb{R}^n$  forces an initial distribution to converge to the equilibrium one, regardless of whether the initial covariance is smaller or larger than the equilibrium covariance. However, if some constraints on the coefficient functions  $\{h_i(\mathbf{q}, t)\}$  and  $\{H_{ij}(\mathbf{q}, t)\}$  are preserved, then entropy can be shown to be non-decreasing. In particular, in cases when the first and third term vanish, the entropy will be non-decreasing because

$$\frac{1}{f} \sum_{i,j,k} \frac{\partial f}{\partial q_i} H_{ik} H_{kj}^T \frac{\partial f}{\partial q_j} \geq 0.$$

### 8.5 Examples

This section begins by seeking the answer to a simply stated question: What SDEs will describe processes that evolve on the unit sphere, and of these, which have a Fokker–Planck equation that is simply the heat equation? In principle since operations with Stratonovich integrals parallel those of standard calculus, it would seem that this should be straightforward. However, there are some subtle points that need to be kept in mind. This is first illustrated in the context of processes on the unit sphere in  $\mathbb{R}^3$ , and then for the stochastic kinematic cart that moves by translation and rotation on the plane  $\mathbb{R}^2$ .

#### 8.5.1 Stochastic Motion on the Unit Circle

Consider the SDE

$$\begin{aligned} dx_1 &= -\frac{1}{2}x_1 dt - x_2 dw \\ dx_2 &= -\frac{1}{2}x_2 dt + x_1 dw. \end{aligned} \tag{8.25}$$

Interpret  $(x_1, x_2)$  as Cartesian coordinates in the plane. Convert to polar coordinates,  $x_1 = x_1(r, \theta) = r \cos \theta$ , and  $x_2 = x_2(r, \theta) = r \sin \theta$ . In this problem the coordinates  $\mathbf{q} = [q_1, q_2]^T$  are  $q_1 = r$  and  $q_2 = \theta$ . And so,

$$\begin{pmatrix} \frac{\partial x_1}{\partial q_1} \\ \frac{\partial x_1}{\partial q_2} \end{pmatrix} = \begin{pmatrix} \cos \theta \\ -r \sin \theta \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{\partial x_2}{\partial q_1} \\ \frac{\partial x_2}{\partial q_2} \end{pmatrix} = \begin{pmatrix} \sin \theta \\ r \cos \theta \end{pmatrix}.$$

Likewise,

$$\begin{pmatrix} \frac{\partial^2 x_1}{\partial q_1 \partial q_1} & \frac{\partial^2 x_1}{\partial q_1 \partial q_2} \\ \frac{\partial^2 x_1}{\partial q_2 \partial q_1} & \frac{\partial^2 x_1}{\partial q_2 \partial q_2} \end{pmatrix} = \begin{pmatrix} 0 & -\sin \theta \\ -\sin \theta & -r \cos \theta \end{pmatrix}$$

and



$$\begin{pmatrix} \frac{\partial^2 x_2}{\partial q_1 \partial q_1} & \frac{\partial^2 x_2}{\partial q_1 \partial q_2} \\ \frac{\partial^2 x_2}{\partial q_2 \partial q_1} & \frac{\partial^2 x_2}{\partial q_2 \partial q_2} \end{pmatrix} = \begin{pmatrix} 0 & \cos \theta \\ \cos \theta & -r \sin \theta \end{pmatrix}.$$

Substitution into Itô's rule, which holds regardless of the SDE in (8.25), gives

$$dx_1 = \cos \theta dr - r \sin \theta d\theta - \sin \theta dr d\theta - \frac{1}{2} r \cos \theta (d\theta)^2 \quad (8.26)$$

$$dx_2 = \sin \theta dr + r \cos \theta d\theta + \cos \theta dr d\theta - \frac{1}{2} r \sin \theta (d\theta)^2. \quad (8.27)$$

Now, assume that an SDE in these new variables exists and can be written as

$$\begin{aligned} dr &= a_1 dt + b_1 dw \\ d\theta &= a_2 dt + b_2 dw \end{aligned}$$

where  $a_i = a_i(r, \theta)$  and  $b_i = b_i(r, \theta)$ .

Substitution of the above expressions into (8.26) and (8.27), and using the stochastic calculus rules  $dw^2 = dt$  and  $dt^2 = dt dw = 0$  gives

$$dx_1 = \left[ a_1 \cos \theta - a_2 r \sin \theta - b_1 b_2 \sin \theta - \frac{1}{2} b_2^2 r \cos \theta \right] dt + (b_1 \cos \theta - b_2 r \sin \theta) dw$$

and

$$dx_2 = \left[ a_1 \sin \theta + a_2 r \cos \theta + b_1 b_2 \cos \theta - \frac{1}{2} b_2^2 r \sin \theta \right] dt + (b_1 \sin \theta + b_2 r \cos \theta) dw.$$

Then substituting these into (8.25) forces  $a_1 = a_2 = b_1 = 0$  and  $b_2 = 1$ , resulting in the SDE

$$d\theta = dw.$$

This shows that (8.25) are Itô stochastic differential equations for a process that evolves only in  $\theta$ , with  $r$  remaining constant. In other words, this is a kind of stochastic motion on the circle.

### 8.5.2 The Unit Sphere in $\mathbb{R}^3$ : Parametric Formulation

Let the position of any point on the unit sphere,  $S^2$ , be parameterized as

$$\mathbf{u}(\phi, \theta) \doteq \begin{pmatrix} \cos \phi \sin \theta \\ \sin \phi \sin \theta \\ \cos \theta \end{pmatrix}. \quad (8.28)$$

It follows from the fact that  $\mathbf{u} \cdot \mathbf{u} = 1$  that taking the derivative of both sides yields  $\mathbf{u} \cdot d\mathbf{u} = 0$  where

$$d\mathbf{u} = \frac{\partial \mathbf{u}}{\partial \theta} d\theta + \frac{\partial \mathbf{u}}{\partial \phi} d\phi. \quad (8.29)$$

And since  $d\theta$  and  $d\phi$  are independent,

$$\mathbf{u} \cdot \frac{\partial \mathbf{u}}{\partial \theta} = \mathbf{u} \cdot \frac{\partial \mathbf{u}}{\partial \phi} = 0. \quad (8.30)$$

Of course, this can be verified by direct calculation. Furthermore, since

$$\frac{\partial \mathbf{u}}{\partial \theta} \cdot \frac{\partial \mathbf{u}}{\partial \theta} = 1 \quad \text{and} \quad \frac{\partial \mathbf{u}}{\partial \phi} \cdot \frac{\partial \mathbf{u}}{\partial \phi} = \sin^2 \theta,$$

the vectors

$$\mathbf{v}_1 \doteq \frac{\partial \mathbf{u}}{\partial \theta} \quad \text{and} \quad \mathbf{v}_2 \doteq \frac{1}{\sin \theta} \frac{\partial \mathbf{u}}{\partial \phi}$$

form an orthonormal basis for the tangent plane to the sphere at the point  $\mathbf{u}(\phi, \theta)$ , with  $\mathbf{v}_1 \times \mathbf{v}_2 = \mathbf{u}$ .

Indeed, any version of this coordinate system rotated around the vector  $\mathbf{u}$  of the form

$$\begin{aligned} \mathbf{v}'_1 &= \mathbf{v}_1 \cos \alpha - \mathbf{v}_2 \sin \alpha \\ \mathbf{v}'_2 &= \mathbf{v}_1 \sin \alpha + \mathbf{v}_2 \cos \alpha \end{aligned} \tag{8.31}$$

will also form an orthonormal basis for this tangent plane, where  $\alpha = \alpha(\phi, \theta)$  is an arbitrary smooth function. This will be relevant later, but for now the focus will be the basis  $\{\mathbf{v}_1, \mathbf{v}_2\}$ .

Consider the Stratonovich equation

$$d\mathbf{u} = \mathbf{v}_1 \circledast dw_1 + \mathbf{v}_2 \circledast dw_2,$$

which would seem like a reasonable definition of Brownian motion on the sphere. Taking the dot product of both sides with respect to  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , and observing (8.29), the resulting two scalar equations can be written as

$$\begin{pmatrix} d\theta \\ d\phi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1/\sin \theta \end{pmatrix} \circledast \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}. \tag{8.32}$$

The corresponding Fokker–Planck equation is

$$\frac{\partial f}{\partial t} = \frac{1}{2} \left[ \frac{\partial^2 f}{\partial \theta^2} + 2 \cot \theta \frac{\partial f}{\partial \theta} - f + \frac{1}{\sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} \right],$$

which is clearly not the heat equation.

Using the result of Exercise 8.2, a Stratonovich SDE that does correspond to the heat equation,

$$\frac{\partial f}{\partial t} = \frac{1}{2} \left[ \frac{\partial^2 f}{\partial \theta^2} + \frac{1}{\sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} \right],$$

is

$$\begin{pmatrix} d\theta \\ d\phi \end{pmatrix} = \frac{1}{2} \cot \theta \mathbf{e}_i + \begin{pmatrix} 1 & 0 \\ 0 & 1/\sin \theta \end{pmatrix} \circledast \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}. \tag{8.33}$$

Using the result of Exercise 8.3, the Itô SDE corresponding to (8.33) is of exactly the same form.

### 8.5.3 SDEs on Spheres and Rotations: Extrinsic Formulation

Let  $\mathbf{z} \in \mathbb{R}^n$  and consider the Itô SDE given in Øksendal [19]:

$$d\mathbf{z} = \mathbf{a}(\mathbf{z})dt + B(\mathbf{z})d\mathbf{w} \quad \text{where} \quad \mathbf{a}(\mathbf{z}) = -\frac{(n-1)}{2}\mathbf{z} \quad \text{and} \quad B(\mathbf{z}) = \mathbb{I} - \mathbf{z}\mathbf{z}^T. \quad (8.34)$$

It is easy to verify that (8.34) satisfies the conditions (8.12) for  $Q = cI$  for any  $c \in \mathbb{R}_{>0}$ , and hence if  $\|\mathbf{z}(0)\| = 1$ , the sample paths will stay on the sphere in  $n$ -dimensional space,  $S^{n-1}$ , for all values of time. This is left as an exercise.

The Itô SDE

$$\begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} = -\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} dt + \begin{pmatrix} x_2 & x_3 & 0 \\ -x_1 & 0 & x_3 \\ 0 & -x_1 & -x_2 \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \\ dw_3 \end{pmatrix}, \quad (8.35)$$

which can be thought of as a kind of spatial generalization of (8.25), or as a special case of (8.7), defines sample paths that evolve on the sphere  $S^2$ , as verified in Exercise 8.4.

Consider the matrix Itô SDE where  $R \in \mathbb{R}^{n \times n}$ :

$$dR = -\frac{(n-1)}{2}dt + \sum_{i,j=1}^n (E_{ij} - E_{ji})Rdw_{ij} \quad (8.36)$$

where  $dw_{ij}$  are  $n^2$  uncorrelated unit-strength white noises. If  $E_{ij}$  is the matrix with the number 1 in the  $ij$ th entry and zero everywhere else, Brockett [3, 4] showed that if  $R(0) \in SO(n)$ , then  $R(t) \in SO(n)$  for all  $t \geq 0$ .<sup>1</sup>

### 8.5.4 The SDE and Fokker–Planck Equation for the Kinematic Cart

Each matrix  $g(x, y, \theta)$  of the form in (1.1) for  $\theta \in [0, 2\pi)$  and  $x, y \in \mathbb{R}$  can be identified with a point on the manifold  $M = \mathbb{R}^2 \times S^1$ . In addition, the product of such matrices produces a matrix of the same kind. Explicitly, if

$$g_i = \begin{pmatrix} \cos \theta_i & -\sin \theta_i & x_i \\ \sin \theta_i & \cos \theta_i & y_i \\ 0 & 0 & 1 \end{pmatrix}$$

for  $i = 1, 2$ , then

$$g_1 g_2 = \begin{pmatrix} \cos(\theta_1 + \theta_2) & -\sin(\theta_1 + \theta_2) & x_1 + x_2 \cos \theta_1 - y_2 \sin \theta_1 \\ \sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) & y_1 + x_2 \sin \theta_1 + y_2 \cos \theta_1 \\ 0 & 0 & 1 \end{pmatrix}.$$

This product is an analytic function from  $M \times M \rightarrow M$ , which makes  $M$  (together with the operation of matrix multiplication) a Lie group (called the Special Euclidean group, or motion group, of the plane, and denoted as  $SE(2)$ ). Lie groups are not addressed formally in this volume, and  $M$  is treated simply as a manifold. The added structure provided by Lie groups makes the formulation of problems *easier* rather than harder.

<sup>1</sup> $SO(n)$  denotes the set of “special orthogonal”  $n \times n$  matrices defined by the condition  $RR^T = \mathbb{I}$  and  $\det R = +1$ . The set of all such matrices forms a group under matrix multiplication. This set is also an  $n(n-1)/2$ -dimensional manifold. In fact  $SO(n)$  is an example of a Lie group.

Lie groups are addressed in detail in Volume 2. For now, the manifold structure of  $M$  is sufficient to formulate the problem of the stochastic cart.

Consider the following variant on the SDE stated in (1.4) that describes the scenario in Figure 1.1:

$$\begin{pmatrix} dx \\ dy \\ d\theta \end{pmatrix} = \begin{pmatrix} r\omega \cos \theta \\ r\omega \sin \theta \\ 0 \end{pmatrix} dt + \sqrt{D} \begin{pmatrix} \frac{r}{2} \cos \theta & \frac{r}{2} \cos \theta \\ \frac{r}{2} \sin \theta & \frac{r}{2} \sin \theta \\ \frac{r}{L} & -\frac{r}{L} \end{pmatrix} \begin{pmatrix} dw_1 \\ dw_2 \end{pmatrix}. \quad (8.37)$$

Using the general formulation in (8.2), the Fokker–Planck equation becomes

$$\begin{aligned} \frac{\partial f'}{\partial t} = & -r\omega \cos \theta \frac{\partial f'}{\partial x} - r\omega \sin \theta \frac{\partial f'}{\partial y} \\ & + \frac{D}{2} \left( \frac{r^2}{2} \cos^2 \theta \frac{\partial^2 f'}{\partial x^2} + \frac{r^2}{2} \sin 2\theta \frac{\partial^2 f'}{\partial x \partial y} + \frac{r^2}{2} \sin^2 \theta \frac{\partial^2 f'}{\partial y^2} + \frac{2r^2}{L^2} \frac{\partial^2 f'}{\partial \theta^2} \right). \end{aligned} \quad (8.38)$$

The notation  $f'$  is used here to distinguish the solution to the above Fokker–Planck equation from the following one which arises in a variety of applications, as will be seen in Volume 2:

$$\frac{\partial f}{\partial t} + \alpha \left[ \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} \right] - \beta \frac{\partial^2 f}{\partial \theta^2} - \epsilon \left[ \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \right] = 0. \quad (8.39)$$

This diffusion equation with drift is highly degenerate when  $\epsilon = 0$ , which happens frequently in applications. See, for example, [28].

## 8.6 Solution Techniques

Once Fokker–Planck equations are derived, the goal becomes either solving them, or at least obtaining as many properties of the solutions as possible. The emphasis here will be solution methods. These fall into two categories (1) analytical solutions and (2) numerical solutions. Both kinds of solution methods are discussed below.

### 8.6.1 Finite Difference and Finite Elements

Finite-difference methods are standard in the solution of partial differential equations. In this method, differential operators applied to a function  $f(\mathbf{x})$  are approximated as

$$\frac{\partial f}{\partial x_i} \approx \frac{1}{\epsilon} [f(\mathbf{x} + \epsilon \mathbf{e}_i) - f(\mathbf{x})] \quad (8.40)$$

where  $\epsilon$  is a small positive number. Exactly how small is small is sometimes the subject of debate. One strategy for choosing  $\epsilon$  is to try a value, then try half of that value, and repeat. Each time compare how the approximated value of the partial derivative compared with the prior one. If the value is chosen to be too small, the limitations of machine precision will come into play to ruin the approximation. If the value of epsilon is too large relative to the size of the smallest fluctuations of the function, then the approximation will also fail. The value of  $\epsilon$  for which doubling and halving will cause the least effect on the estimate of the partial derivative is then a robust choice.

The approximation in (8.40) is often called a “forward difference,” in contrast to

$$\frac{\partial f}{\partial x_i} \approx \frac{1}{\epsilon} [f(\mathbf{x}) - f(\mathbf{x} - \epsilon \mathbf{e}_i)],$$

which is a “backward difference” and

$$\frac{\partial f}{\partial x_i} \approx \frac{1}{2\epsilon} [f(\mathbf{x} + \epsilon \mathbf{e}_i) - f(\mathbf{x} - \epsilon \mathbf{e}_i)],$$

which is called a “centered difference.”

In finite-difference schemes applied to linear PDEs such as the Fokker–Planck equation, the parametric domain is sampled on a regular grid. The value of the function  $f(\mathbf{q}, t)$  at each grid point  $\mathbf{q}$  then becomes a component in a vector of dimension  $N^d$  where  $d$  is the dimension of the manifold and  $N$  is the number of discretizations in each parameter. The original PDE then becomes a system of ODEs in this vector. This approach can be used for relatively low-dimensional problems (i.e., when  $d = 1, 2, 3$ ), but can be prohibitive for high-dimensional problems, even for moderate values of  $N$ .

Numerical losses in the finite-difference method can be substantial as the time parameter becomes large. The finite-element method (FEM) goes one step further to attempt to conserve quantities that should not vary. In the finite-element method the functions are not simply sampled on a grid, but expanded in a basis of local shape functions. These ensure that piecewise smooth functions are defined on polytopes in parameter space, and meet each other with differentiability conditions that are specified by the programmer. This means that a continuous solution results. Finite-element methods are used in engineering practice to model mechanical systems (i.e., solid mechanics, fluid mechanics, and heat transfer problems) because they do well at conserving mass, momentum, heat, etc. By extension, it would make sense that they would do well to conserve probability density when applied to a Fokker–Planck equation. However, they suffer from the same “curse of dimensionality” as finite differences.

### 8.6.2 Non-Parametric Density Estimation

In contrast to finite-difference methods and FEM, which are generic numerical tools for solving PDEs, numerical methods exist specifically for approximating the solutions to Fokker–Planck equations. This is because Fokker–Planck equations were derived from SDEs. And therefore, if a very large number of sample paths are generated from the SDE and stored, the histograms that result will approximate the desired pdfs.

In a sense, taking this approach would circumvent the need to derive a Fokker–Planck in the first place, since when it comes time to solving the Fokker–Planck equation the approach returns to the SDE. While this approach is valid, it has several limitations. First, the number of samples needed to recover the salient features of a pdf at each fixed value of time can be quite large. And this method also can suffer from the curse of dimensionality if a grid is established in the parametric domain to evaluate the pdf. The actual estimation of values of the pdf on the grid can be performed in a variety of ways. The simplest of these is the histogram method. More sophisticated schemes use kernel-based density estimation in which each sample point is replaced by a small probability density function (such as a Gaussian distribution). Each of these is called a kernel. Then the contributions of each kernel are added at each grid point to estimate the overall pdf.

### 8.6.3 Separation of Variables: Diffusion on $SE(2)$ as a Case Study

For particular kinds of linear partial differential equations, the standard solution method is separation of variables. When this method works, it is very convenient because it re-

duces the original multi-dimensional problem to many coupled single-dimensional ones. This is a powerful tool to circumvent the curse of dimensionality because the full multi-dimensional solution can be reconstructed at relatively low resolution in an efficient way.

The drawback of this method is that not every linear PDE can be separated. Conditions for separability were discussed in the context of the heat equation in Chapter 2. These conditions need to be applied on a case-by-case basis. This section therefore addresses the separation-of-variables solution of (8.39) which is a Fokker–Planck equation on the three-dimensional manifold of  $SE(2)$ . Note that when  $\alpha = 0$  this is nothing more than a special case of the driftless time-varying diffusion equation examined in Section 2.6.

The following subsections address various issues related to when the above equations can be solved using separation of variables.

### Transformation of Coordinates in the $SE(2)$ Diffusion Equation

Let  $\alpha = \alpha_0$  and  $\beta = \beta_0$  be positive constants,  $\epsilon = 0$ , and consider the following special case of (8.39):

$$\frac{\partial f}{\partial t} + \alpha_0 \left[ \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} \right] - \beta_0 \frac{\partial^2 f}{\partial \theta^2} = 0. \quad (8.41)$$

Can this equation be solved by separation of variables? In 1999, a graduate student in the author’s research group proclaimed “of course, just make a change of coordinates of the form

$$\begin{aligned} x' &= x \cos \theta + y \sin \theta \\ y' &= -x \sin \theta + y \cos \theta \\ \theta' &= \theta \end{aligned}$$

and all of the trigonometric coefficients will disappear.”

Indeed, if  $f(x, y, \theta; t) = f'(x', y', \theta'; t)$ , then by the chain rule

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial f'}{\partial x'} \frac{\partial x'}{\partial x} + \frac{\partial f'}{\partial y'} \frac{\partial y'}{\partial x} + \frac{\partial f'}{\partial \theta'} \frac{\partial \theta'}{\partial x} = \frac{\partial f'}{\partial x'} \cos \theta - \frac{\partial f'}{\partial y'} \sin \theta \\ \frac{\partial f}{\partial y} &= \frac{\partial f'}{\partial x'} \frac{\partial x'}{\partial y} + \frac{\partial f'}{\partial y'} \frac{\partial y'}{\partial y} + \frac{\partial f'}{\partial \theta'} \frac{\partial \theta'}{\partial y} = \frac{\partial f'}{\partial x'} \sin \theta + \frac{\partial f'}{\partial y'} \cos \theta \end{aligned}$$

and so

$$\cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} = \frac{\partial f'}{\partial x'}.$$

However, this is not the end of the story since

$$\frac{\partial f}{\partial \theta} = \frac{\partial f'}{\partial x'} \frac{\partial x'}{\partial \theta} + \frac{\partial f'}{\partial y'} \frac{\partial y'}{\partial \theta} + \frac{\partial f'}{\partial \theta'} \frac{\partial \theta'}{\partial \theta} \neq \frac{\partial f'}{\partial \theta'}.$$

Noting that

$$\begin{aligned} \frac{\partial x'}{\partial \theta} &= -x \sin \theta + y \cos \theta = y' \\ \frac{\partial y'}{\partial \theta} &= -x \cos \theta - y \sin \theta = -x' \\ \frac{\partial \theta'}{\partial \theta} &= 1, \end{aligned}$$

$$\frac{\partial f}{\partial \theta} = y' \frac{\partial f'}{\partial x'} - x' \frac{\partial f'}{\partial y'} + \frac{\partial f'}{\partial \theta'}.$$

This means that (8.41) is transformed to

$$\frac{\partial f'}{\partial t} + \alpha_0 \frac{\partial f'}{\partial x'} + \beta_0 \left( y' \frac{\partial f'}{\partial x'} - x' \frac{\partial f'}{\partial y'} + \frac{\partial f'}{\partial \theta'} \right)^2 f' = 0. \quad (8.42)$$

While it is true that the trigonometric terms have been removed, new terms have been introduced.

This begs the question: "Is it possible to find any coordinate system in which separation of variables will work for (8.41) or (8.42)?" To address this question, the method of symmetry operators will be attempted.

### Symmetry Operators for the $SE(2)$ Diffusion Equation

Here the methodology discussed in Section 2.8 is applied to (8.41). When written in terms of the original parameters, any symmetry operators will be of the form

$$L = X(x, y, \theta, t) \frac{\partial}{\partial x} + Y(x, y, \theta, t) \frac{\partial}{\partial y} + \Theta(x, y, \theta, t) \frac{\partial}{\partial \theta} + T(x, y, \theta, t) \frac{\partial}{\partial t} + Z(x, y, \theta, t).$$

It follows that

$$\begin{aligned} LQf &= X \frac{\partial^2 f}{\partial t \partial x} + \alpha_0 \cos \theta X \frac{\partial^2 f}{\partial x^2} + \alpha_0 \sin \theta X \frac{\partial^2 f}{\partial x \partial y} - \beta_0 X \frac{\partial^3 f}{\partial x \partial \theta^2} \\ &+ Y \frac{\partial^2 f}{\partial t \partial y} + \alpha_0 \cos \theta Y \frac{\partial^2 f}{\partial x \partial y} + \alpha_0 \sin \theta Y \frac{\partial^2 f}{\partial y^2} - \beta_0 Y \frac{\partial^3 f}{\partial y \partial \theta^2} \\ &+ \Theta \frac{\partial^2 f}{\partial t \partial \theta} + \alpha_0 \Theta \frac{\partial}{\partial \theta} \left( \cos \theta \frac{\partial f}{\partial x} \right) + \alpha_0 \Theta \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial f}{\partial y} \right) - \beta_0 \Theta \frac{\partial^3 f}{\partial \theta^3} \\ &+ T \frac{\partial^2 f}{\partial t^2} + \alpha_0 T \cos \theta \frac{\partial^2 f}{\partial x \partial t} + \alpha_0 T \sin \theta \frac{\partial^2 f}{\partial y \partial t} - \beta_0 T \frac{\partial^3 f}{\partial \theta^2 \partial t} \\ &+ Z \frac{\partial f}{\partial t} + \alpha_0 Z \cos \theta \frac{\partial f}{\partial x} + \alpha_0 Z \sin \theta \frac{\partial f}{\partial y} - \beta_0 Z \frac{\partial^2 f}{\partial \theta^2} \end{aligned}$$

and

$$\begin{aligned} QLf &= \frac{\partial}{\partial t} \left( X \frac{\partial f}{\partial x} \right) + \alpha_0 \cos \theta \frac{\partial}{\partial x} \left( X \frac{\partial f}{\partial x} \right) + \alpha_0 \sin \theta \frac{\partial}{\partial y} \left( X \frac{\partial f}{\partial x} \right) - \beta_0 \frac{\partial^2}{\partial \theta^2} \left( X \frac{\partial f}{\partial x} \right) \\ &+ \frac{\partial}{\partial t} \left( Y \frac{\partial f}{\partial y} \right) + \alpha_0 \cos \theta \frac{\partial}{\partial x} \left( Y \frac{\partial f}{\partial y} \right) + \alpha_0 \sin \theta \frac{\partial}{\partial y} \left( Y \frac{\partial f}{\partial y} \right) - \beta_0 \frac{\partial^2}{\partial \theta^2} \left( Y \frac{\partial f}{\partial y} \right) \\ &+ \frac{\partial}{\partial t} \left( \Theta \frac{\partial f}{\partial \theta} \right) + \alpha_0 \cos \theta \frac{\partial}{\partial x} \left( \Theta \frac{\partial f}{\partial \theta} \right) + \alpha_0 \sin \theta \frac{\partial}{\partial y} \left( \Theta \frac{\partial f}{\partial \theta} \right) - \beta_0 \frac{\partial^2}{\partial \theta^2} \left( \Theta \frac{\partial f}{\partial \theta} \right) \\ &+ \frac{\partial}{\partial t} \left( T \frac{\partial f}{\partial t} \right) + \alpha_0 \cos \theta \frac{\partial}{\partial x} \left( T \frac{\partial f}{\partial t} \right) + \alpha_0 \sin \theta \frac{\partial}{\partial y} \left( T \frac{\partial f}{\partial t} \right) - \beta_0 \frac{\partial^2}{\partial \theta^2} \left( T \frac{\partial f}{\partial t} \right) \\ &+ \frac{\partial}{\partial t} (Zf) + \alpha_0 \cos \theta \frac{\partial}{\partial x} (Zf) + \alpha_0 \sin \theta \frac{\partial}{\partial y} (Zf) - \beta_0 \frac{\partial^2}{\partial \theta^2} (Zf). \end{aligned}$$

Expanding these out further using the chain rule,

$$\begin{aligned}
LQf &= X \frac{\partial^2 f}{\partial t \partial x} + \alpha_0 \cos \theta X \frac{\partial^2 f}{\partial x^2} + \alpha_0 \sin \theta X \frac{\partial^2 f}{\partial x \partial y} - \beta_0 X \frac{\partial^3 f}{\partial x \partial \theta^2} \\
&+ Y \frac{\partial^2 f}{\partial t \partial y} + \alpha_0 \cos \theta Y \frac{\partial^2 f}{\partial x \partial y} + \alpha_0 \sin \theta Y \frac{\partial^2 f}{\partial y^2} - \beta_0 Y \frac{\partial^3 f}{\partial y \partial \theta^2} \\
&+ \Theta \frac{\partial^2 f}{\partial t \partial \theta} - \alpha_0 \Theta \sin \theta \frac{\partial f}{\partial x} + \alpha_0 \Theta \cos \theta \frac{\partial f^2}{\partial x \partial \theta} \\
&+ \alpha_0 \Theta \cos \theta \frac{\partial f}{\partial y} + \alpha_0 \Theta \sin \theta \frac{\partial f^2}{\partial y \partial \theta} - \beta_0 \Theta \frac{\partial^3 f}{\partial \theta^3} \\
&+ T \frac{\partial^2 f}{\partial t^2} + \alpha_0 T \cos \theta \frac{\partial^2 f}{\partial x \partial t} + \alpha_0 T \sin \theta \frac{\partial^2 f}{\partial y \partial t} - \beta_0 T \frac{\partial^3 f}{\partial \theta^2 \partial t} \\
&+ Z \frac{\partial f}{\partial t} + \alpha_0 Z \cos \theta \frac{\partial f}{\partial x} + \alpha_0 Z \sin \theta \frac{\partial f}{\partial y} - \beta_0 Z \frac{\partial^2 f}{\partial \theta^2}
\end{aligned}$$

and

$$\begin{aligned}
QLf &= \frac{\partial X}{\partial t} \frac{\partial f}{\partial x} + X \frac{\partial^2 f}{\partial x \partial t} + \alpha_0 \cos \theta \frac{\partial X}{\partial x} \frac{\partial f}{\partial x} + \alpha_0 \cos \theta X \frac{\partial^2 f}{\partial x^2} + \alpha_0 \sin \theta \frac{\partial X}{\partial y} \frac{\partial f}{\partial x} \\
&+ \alpha_0 \sin \theta X \frac{\partial^2 f}{\partial x \partial y} - \beta_0 \frac{\partial^2 X}{\partial \theta^2} \frac{\partial f}{\partial x} - 2\beta_0 \frac{\partial X}{\partial \theta} \frac{\partial^2 f}{\partial x \partial \theta} - \beta_0 X \frac{\partial^3 f}{\partial x \partial \theta^2} \\
&+ \frac{\partial Y}{\partial t} \frac{\partial f}{\partial y} + Y \frac{\partial^2 f}{\partial y \partial t} + \alpha_0 \cos \theta \frac{\partial Y}{\partial x} \frac{\partial f}{\partial y} + \alpha_0 \cos \theta Y \frac{\partial^2 f}{\partial y \partial x} \\
&+ \alpha_0 \sin \theta \frac{\partial Y}{\partial y} \frac{\partial f}{\partial y} + \alpha_0 \sin \theta Y \frac{\partial^2 f}{\partial y^2} - \beta_0 \frac{\partial^2 Y}{\partial \theta^2} \frac{\partial f}{\partial y} - 2\beta_0 \frac{\partial Y}{\partial \theta} \frac{\partial^2 f}{\partial y \partial \theta} - \beta_0 Y \frac{\partial^3 f}{\partial y \partial \theta^2} \\
&+ \frac{\partial \Theta}{\partial t} \frac{\partial f}{\partial \theta} + \Theta \frac{\partial^2 f}{\partial \theta \partial t} + \alpha_0 \cos \theta \frac{\partial \Theta}{\partial x} \frac{\partial f}{\partial \theta} + \alpha_0 \cos \theta \Theta \frac{\partial^2 f}{\partial \theta \partial x} \\
&+ \alpha_0 \sin \theta \frac{\partial \Theta}{\partial y} \frac{\partial f}{\partial \theta} + \alpha_0 \sin \theta \Theta \frac{\partial^2 f}{\partial \theta \partial y} - \beta_0 \frac{\partial^2 \Theta}{\partial \theta^2} \frac{\partial f}{\partial \theta} - 2\beta_0 \frac{\partial \Theta}{\partial \theta} \frac{\partial^2 f}{\partial \theta^2} - \beta_0 \Theta \frac{\partial^3 f}{\partial \theta^3} \\
&+ \frac{\partial T}{\partial t} \frac{\partial f}{\partial t} + T \frac{\partial^2 f}{\partial t^2} + \alpha_0 \cos \theta \frac{\partial T}{\partial x} \frac{\partial f}{\partial t} + \alpha_0 \cos \theta T \frac{\partial f^2}{\partial t \partial x} + \alpha_0 \sin \theta \frac{\partial T}{\partial y} \frac{\partial f}{\partial t} \\
&+ \alpha_0 \sin \theta T \frac{\partial^2 f}{\partial t \partial y} - \beta_0 \frac{\partial^2 T}{\partial \theta^2} \frac{\partial f}{\partial t} - 2\beta_0 \frac{\partial T}{\partial \theta} \frac{\partial^2 f}{\partial t \partial \theta} - \beta_0 T \frac{\partial^3 f}{\partial \theta^2 \partial t} \\
&+ \frac{\partial Z}{\partial t} f + Z \frac{\partial f}{\partial t} + \alpha_0 \cos \theta \frac{\partial Z}{\partial x} f + \alpha_0 \cos \theta Z \frac{\partial f}{\partial x} + \alpha_0 \sin \theta \frac{\partial Z}{\partial y} f + \alpha_0 \sin \theta Z \frac{\partial f}{\partial y} \\
&- \beta_0 \frac{\partial^2 Z}{\partial \theta^2} f - 2\beta_0 \frac{\partial Z}{\partial \theta} \frac{\partial f}{\partial \theta} - \beta_0 Z \frac{\partial^2 f}{\partial \theta^2}
\end{aligned}$$



$$\begin{aligned}
[Q, L]f &= \left( \frac{\partial X}{\partial t} + \alpha_0 \cos \theta \frac{\partial X}{\partial x} + \alpha_0 \sin \theta \frac{\partial X}{\partial y} - \beta_0 \frac{\partial^2 X}{\partial \theta^2} + \alpha_0 \Theta \sin \theta \right) \frac{\partial f}{\partial x} \\
&\quad - 2\beta_0 \frac{\partial X}{\partial \theta} \frac{\partial^2 f}{\partial x \partial \theta} \\
&\quad + \left( \frac{\partial Y}{\partial t} + \alpha_0 \cos \theta \frac{\partial Y}{\partial x} + \alpha_0 \sin \theta \frac{\partial Y}{\partial y} - \beta_0 \frac{\partial^2 Y}{\partial \theta^2} - \alpha_0 \Theta \cos \theta \right) \frac{\partial f}{\partial y} \\
&\quad - 2\beta_0 \frac{\partial Y}{\partial \theta} \frac{\partial^2 f}{\partial y \partial \theta} \\
&\quad + \left( \frac{\partial \Theta}{\partial t} + \alpha_0 \cos \theta \frac{\partial \Theta}{\partial x} + \alpha_0 \sin \theta \frac{\partial \Theta}{\partial y} \right) \frac{\partial f}{\partial \theta} - \beta_0 \frac{\partial^2 \Theta}{\partial \theta^2} \frac{\partial f}{\partial \theta} - 2\beta_0 \frac{\partial \Theta}{\partial \theta} \frac{\partial^2 f}{\partial \theta^2} \\
&\quad + \left( \frac{\partial T}{\partial t} + \alpha_0 \cos \theta \frac{\partial T}{\partial x} + \alpha_0 \sin \theta \frac{\partial T}{\partial y} - \beta_0 \frac{\partial^2 T}{\partial \theta^2} \right) \frac{\partial f}{\partial t} - 2\beta_0 \frac{\partial T}{\partial \theta} \frac{\partial^2 f}{\partial t \partial \theta} \\
&\quad + \left( \frac{\partial Z}{\partial t} + \alpha_0 \cos \theta \frac{\partial Z}{\partial x} + \alpha_0 \sin \theta \frac{\partial Z}{\partial y} - \beta_0 \frac{\partial^2 Z}{\partial \theta^2} \right) f - 2\beta_0 \frac{\partial Z}{\partial \theta} \frac{\partial f}{\partial \theta}
\end{aligned}$$

By the definition of a symmetry operator, multipliers of each partial derivative of  $f$  in this expression must equal the multipliers in

$$-RQf = -R \frac{\partial f}{\partial t} - R\alpha_0 \cos \theta \frac{\partial f}{\partial x} - R\alpha_0 \sin \theta \frac{\partial f}{\partial y} + R\beta_0 \frac{\partial^2 f}{\partial \theta^2}.$$

This results in the following equations:

$$\frac{\partial X}{\partial t} + \alpha_0 \cos \theta \frac{\partial X}{\partial x} + \alpha_0 \sin \theta \frac{\partial X}{\partial y} - \beta_0 \frac{\partial^2 X}{\partial \theta^2} + \alpha_0 \Theta \sin \theta = -R\alpha_0 \cos \theta$$

$$\frac{\partial Y}{\partial t} + \alpha_0 \cos \theta \frac{\partial Y}{\partial x} + \alpha_0 \sin \theta \frac{\partial Y}{\partial y} - \beta_0 \frac{\partial^2 Y}{\partial \theta^2} - \alpha_0 \Theta \cos \theta = -R\alpha_0 \sin \theta$$

$$\frac{\partial \Theta}{\partial t} + \alpha_0 \cos \theta \frac{\partial \Theta}{\partial x} + \alpha_0 \sin \theta \frac{\partial \Theta}{\partial y} = 0$$

$$\frac{\partial^2 \Theta}{\partial \theta^2} = 0$$

$$2 \frac{\partial \Theta}{\partial \theta} = -R$$

$$\frac{\partial T}{\partial t} + \alpha_0 \cos \theta \frac{\partial T}{\partial x} + \alpha_0 \sin \theta \frac{\partial T}{\partial y} - \beta_0 \frac{\partial^2 T}{\partial \theta^2} = -R$$

$$\frac{\partial Z}{\partial t} + \alpha_0 \cos \theta \frac{\partial Z}{\partial x} + \alpha_0 \sin \theta \frac{\partial Z}{\partial y} - \beta_0 \frac{\partial^2 Z}{\partial \theta^2} = 0$$

and

$$\frac{\partial X}{\partial \theta} = \frac{\partial Y}{\partial \theta} = \frac{\partial T}{\partial \theta} = \frac{\partial Z}{\partial \theta} = 0.$$

This leads to the reduction

$$\frac{\partial X}{\partial t} + \alpha_0 \cos \theta \frac{\partial X}{\partial x} + \alpha_0 \sin \theta \frac{\partial X}{\partial y} + \alpha_0 \Theta \sin \theta = 2\alpha_0 \cos \theta \frac{\partial \Theta}{\partial \theta} \quad (8.43)$$

$$\frac{\partial Y}{\partial t} + \alpha_0 \cos \theta \frac{\partial Y}{\partial x} + \alpha_0 \sin \theta \frac{\partial Y}{\partial y} - \alpha_0 \Theta \cos \theta = 2\alpha_0 \sin \theta \frac{\partial \Theta}{\partial \theta} \quad (8.44)$$

$$\frac{\partial \Theta}{\partial t} + \alpha_0 \cos \theta \frac{\partial \Theta}{\partial x} + \alpha_0 \sin \theta \frac{\partial \Theta}{\partial y} = 0 \quad (8.45)$$

$$\frac{\partial T}{\partial t} + \alpha_0 \cos \theta \frac{\partial T}{\partial x} + \alpha_0 \sin \theta \frac{\partial T}{\partial y} = 2 \frac{\partial \Theta}{\partial \theta} \quad (8.46)$$

$$\frac{\partial Z}{\partial t} + \alpha_0 \cos \theta \frac{\partial Z}{\partial x} + \alpha_0 \sin \theta \frac{\partial Z}{\partial y} = 0 \quad (8.47)$$

together with the conditions

$$X = X(x, y, t); \quad Y = Y(x, y, t); \quad T = T(x, y, t); \quad Z = Z(x, y, t)$$

and

$$\Theta(x, y, \theta, t) = c_1(x, y, t)\theta + c_2(x, y, t).$$

Since the trigonometric sequence  $\{1, \cos \theta, \sin \theta, \dots\}$  forms a basis for the set of square-integrable functions on the unit circle, and since  $Z$  does not depend on  $\theta$ , each coefficient of the terms  $1$ ,  $\cos \theta$ , and  $\sin \theta$  in (8.47) must be zero. This forces  $Z$  to be a constant, which is denoted here as  $Z = Z_0$ . A similar argument applied to (8.46) forces  $T = T(t)$  and the coefficient function  $c_1 = T'(t)$  in the definition of  $\Theta(x, y, \theta, t)$ . Equation (8.45) becomes

$$T''(t)\theta + \alpha_0 \cos \theta \frac{\partial c_2}{\partial x} + \alpha_0 \sin \theta \frac{\partial c_2}{\partial y} = 0.$$

Now the function  $h(\theta) = \theta$  can be expanded in a Taylor series, and since  $h(-\theta) = -h(\theta)$  the result will be a sine series of the form

$$h(\theta) = \sum_{k=1}^{\infty} a_k \sin k\theta.$$

The exact values of  $\{a_k\}$  can be obtained in a similar way as in Exercise 1.1. It suffices to say that since  $a_k \neq 0$  for at least one value of  $k > 1$ , it must be the case that

$$T''(t) = 0 \quad \rightarrow \quad \frac{\partial c_2}{\partial x} = 0; \quad \frac{\partial c_2}{\partial y} = 0.$$

This forces

$$T(t) = b_0 t + T_0 \quad \text{and} \quad c_2 = c_2(t) \quad (8.48)$$

where  $b_0$  and  $T_0$  are undetermined constants. Now turning to (8.43) and (8.44), the completeness of the Fourier basis forces  $\frac{\partial X}{\partial t} = \frac{\partial Y}{\partial t} = 0$ . Substituting (8.48) into these equations results in the equalities

$$\frac{\partial Y}{\partial x} = \Theta = T'(t)\theta + c_2(t)$$

$$\frac{\partial X}{\partial y} = -\Theta = -T'(t)\theta - c_2(t).$$

But since  $X$  and  $Y$  must both be independent of  $\theta$  and  $t$ , it must be that  $c_2(t) = b_0 = 0$ .

This means that  $\Theta = 0$  and  $X, Y, Z$ , and  $T$  are all constants, and the resulting Lie algebra of differential operators is spanned by the basis

$$L_1 = \frac{\partial}{\partial x}; \quad L_2 = \frac{\partial}{\partial y}; \quad L_3 = \frac{\partial}{\partial t}; \quad L_4 = 1.$$

This is a commutative Lie algebra as is observed from the condition

$$[L_i, L_j] = 0.$$

Therefore the Fokker–Planck equation in (8.41) is separable.

### Separation of the $SE(2)$ Diffusion Equation

Knowing that a separable solution exists is the first step to finding the solution. In classical separation of variables, a separable solution is assumed and substituted into the partial differential equation of interest. In the current context, this becomes

$$f(x, y, \theta; t) = f_x(x)f_y(y)f_\theta(\theta)f_t(t).$$

Examining the special case of (8.39) when  $\epsilon_0 = 0$  and  $\alpha_0 = 1$  that results in (8.41),

$$-\lambda_0 = \cos\theta \frac{f'_x(x)}{f_x(x)} - \sin\theta \frac{f'_y(y)}{f_y(y)} + \beta_0 \frac{f''_\theta(\theta)}{f_\theta(\theta)},$$

which can be separated by dividing by  $\cos\theta$  and isolating  $f'_\theta(x)/f_x(x)$  as

$$-\frac{f'_x(x)}{f_x(x)} = \frac{\lambda_0}{\cos\theta} - \tan\theta \frac{f'_y(y)}{f_y(y)} + \frac{\beta_0}{\cos\theta} \frac{f''_\theta(\theta)}{f_\theta(\theta)}.$$

Setting  $f'_x(x)/f_x(x) = -\mu_0$  means that  $f_x(x) = C_2 e^{-\mu_0 x}$ . Separating what remains by isolating  $f'_y(y)/f_y(y)$  (which requires a division by  $\tan\theta$ ) gives  $f'_y(y)/f_y(y) = -\nu_0$  or  $f_y(y) = C_3 e^{-\nu_0 y}$ . Finally,  $f_\theta$  is solved as

$$\beta_0 f''_\theta(\theta) + (\lambda_0 - \mu_0 \cos\theta + \nu_0 \sin\theta) f_\theta(\theta) = 0 \quad (8.49)$$

which should be solved subject to the periodic boundary conditions

$$f_\theta(0) = f_\theta(2\pi) \quad \text{and} \quad f'_\theta(0) = f'_\theta(2\pi).$$

Equation (8.49) could also have been obtained from the original equation by applying the 2D Fourier transform in  $x$  and  $y$  to  $f(x, y, \theta; t)$ , in which case  $\mu_0 = i\omega_1$  and  $\nu_0 = i\omega_2$ . Let the solution to (8.49) subject to these boundary conditions such that  $f_\theta(0) = 1$  be denoted as  $\hat{\Phi}_{\mu_0, \nu_0}^{\lambda_0}(\theta)$ . These solutions will contain freedom in scaling. The periodic boundary conditions are automatically satisfied by taking a solution of the form of a Fourier series:

$$\hat{\Phi}_{\mu_0, \nu_0}^{\lambda_0}(\theta) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \hat{\Phi}_{\mu_0, \nu_0}^{\lambda_0}(n) e^{in\theta}.$$

Substitution into (8.49) leads to recurrence relations of the form

$$(\lambda_0 - n^2) \hat{\Phi}_{\mu_0, \nu_0}^{\lambda_0}(n) - \frac{1}{2}(\mu_0 + \nu_0 i) \hat{\Phi}_{\mu_0, \nu_0}^{\lambda_0}(n-1) + \frac{1}{2}(\mu_0 + \nu_0 i) \hat{\Phi}_{\mu_0, \nu_0}^{\lambda_0}(n+1) = 0.$$

Then putting everything together,

$$f(x, y, \theta; t) = e^{-\lambda_0 t} e^{-\mu_0 x} e^{-\nu_0 y} \Phi_{\mu_0, \nu_0}^{\lambda_0}(\theta). \tag{8.50}$$

We know that  $\int_G f(g, t) dg = 1$ , and so

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\pi}^{\pi} f(x, y, \theta; t) d\theta dx dy = 1.$$

Therefore  $\mu_0$  and  $\nu_0$  cannot have real parts. It was suggested to the author by Profs. E. Kalnins and W. Miller that the solution based on (8.50) might be written in the form

$$f(x, y, \theta; t) = \sum_{\lambda} \int_{\omega \in \mathbb{R}^2} \hat{f}(\omega_1, \omega_2, \lambda; 0) e^{-\lambda t} e^{-i\omega_1 x} e^{-i\omega_2 y} \Phi_{-i\omega_1, -i\omega_2}^{\lambda}(\theta) d\omega \tag{8.51}$$

where rather than expanding  $\Phi_{-i\omega_1, -i\omega_2}^{\lambda}(\theta)$  in a Fourier series (which is divergent), that it be expressed either in terms of Mathieu functions as defined in [18], or using solutions to quantum mechanical analogies as described in [7]. Numerically, this would involve discretizing the Fourier integral and truncating the sum over allowable values of  $\lambda$ .

Returning to the general case in (8.39), substitution of an assumed separable solution and division by  $f$  results in

$$\frac{f'_t(t)}{f_t(t)} = \alpha(t) \cos \theta \frac{f'_x(x)}{f_x(x)} - \alpha(t) \sin \theta \frac{f'_y(y)}{f_y(y)} + \beta(t) \frac{f''_{\theta}(\theta)}{f_{\theta}(\theta)} + \epsilon(t) \left( \frac{f''_x(x)}{f_x(x)} + \frac{f''_y(y)}{f_y(y)} \right)$$

where a ' denotes differentiation in the context of functions of only one variable.

When  $\alpha$ ,  $\beta$ , and  $\epsilon$  are all independent functions of time that are not constant multiples of each other, the author has not found separable solutions of (8.39). In the special case when they are constants  $\alpha_0$ ,  $\beta_0$ , and  $\epsilon_0$ , then the term  $f'_t(t)/f_t(t)$  written on the left side of the equation, separated from the rest, depends only on  $t$  whereas the terms on the right side do not depend on  $t$ . Therefore both sides must be equal to a constant,  $-\lambda_0$ , and so  $f_t(t) = C_1 e^{-\lambda_0 t}$ . It is no longer obvious to the author how  $x$ ,  $y$ , and  $\theta$  would separate when  $\epsilon_0 \neq 0$ . This is one of the motivations for pursuing the non-commutative harmonic analysis tools that are developed in Volume 2.

## 8.7 Chapter Summary

In this chapter the theory of SDEs and their corresponding Fokker–Planck equations was extended from Euclidean space to the case of random processes on manifolds. This was done in two ways: (1) using a parametric approach in which the SDE is expressed in a local coordinate system and (2) defining an embedded manifold implicitly with a system of constraints and expressing the SDE in the Cartesian coordinates of the ambient Euclidean space. Examples illustrated the general methods. When it comes to simple numerical implementations, the parametric approach is generally more reliable because SDEs based on the implicit approach have the potential to generate sample paths that diverge from the manifold in which they are supposed to be contained.

Models of stochastic motion on spheres were illustrated in this chapter. More advanced treatments of stochastic motion on spheres and other manifolds can be found in [9, 15, 20, 24, 25, 26, 27]. A particularly important stochastic motion on a manifold that arises in molecular applications is that of Brownian motion on the rotation group

$SO(3)$  [21, 22, 23]. Since Fokker–Planck equations are related to the heat equation with a drift term, literature that connects the geometry of manifolds with the heat equation, such as references provided in the previous chapter and [2, 16, 21, 29], are relevant to studying the behavior of stochastic flows.

## 8.8 Exercises

8.1. As an alternative to the coordinate conversion in Section 8.5.1, show that (8.25) corresponds to Brownian motion on the circle by proving that this SDE satisfies the constraint  $g(\mathbf{x}) = x_1^2 + x_2^2 - 1 = 0$  when  $g(\mathbf{x}_0) = 0$  is observed for  $\mathbf{x}(0) = \mathbf{x}_0$ .

8.2. Modify (8.32) by including a drift term of the form  $\mathbf{a}(\theta, \phi)dt$ . Substitute this modified SDE into the Stratonovich form of the Fokker–Planck equation and verify that the choice of  $\mathbf{a}(\theta, \phi)$  that results in the heat equation is that defined by (8.33).

8.3. Show that the Itô equation corresponding to (8.33) has exactly the same form, and the corresponding Fokker–Planck equation is also the heat equation.

8.4. Verify that (8.34) for  $n = 3$  and (8.35) both define stochastic processes that evolve on the sphere  $S^2$ . Convert each of these into their equivalent spherical coordinate representation, and derive the corresponding the Fokker–Planck equations. Do (8.34) for  $n = 3$  and (8.35) define equivalent processes?

8.5. Verify that  $R(t)$  in (8.36) satisfies the constraint  $R(t)R^T(t) = \mathbb{I}$  for all values of time as long as the same constraint is satisfied at  $t = 0$ .

8.6. Prove that if  $\mathbf{q}$  parameterizes a whole manifold (up to a set of measure zero defined by singularities) and  $|G(\mathbf{q}_0)| \neq 0$ , then the solution to a Fokker–Planck equation on a manifold,  $f(\mathbf{q}, t)$ , satisfies the constraint

$$\int_{\mathbf{q}} f(\mathbf{q}, t) |G(\mathbf{q})|^{\frac{1}{2}} d\mathbf{q} = 1 \quad (8.52)$$

when  $f(\mathbf{q}, 0) = \delta(\mathbf{q} - \mathbf{q}_0)$ .

8.7. Using the fact that the volume element for  $SE(2)$  is of the form  $dx dy d\theta$  (i.e., the determinant of the metric tensor is equal to unity), derive (8.38) from (8.37).

8.8. Show that the Stratonovich SDE corresponding to (8.37) is of exactly the same form, and has the Fokker–Planck equation (8.38).

8.9. Starting with the Fokker–Planck equation in (8.39), work backwards and obtain one or more Itô SDEs that would give rise to it.

## References

1. Applebaum, D., Kunita, H., “Lévy flows on manifolds and Lévy processes on Lie groups,” *J. Math. Kyoto Univ.*, 33/34, pp. 1103–1123, 1993.
2. Atiyah, M., Bott, R., Patodi, V.K., “On the heat equation and the Index theorem,” *Invent. Math.*, 19, pp. 279–330, 1973.

3. Brockett, R.W., "Lie algebras and Lie groups in control theory," in *Geometric Methods in System Theory* (D.Q. Mayne and R.W. Brockett, eds.), Reidel, Dordrecht, 1973.
4. Brockett, R.W., "Notes on stochastic processes on manifolds," in *Systems and Control in the Twenty-First Century* (C.I. Byrnes et al. eds.), Birkhäuser, Boston, 1997.
5. Elworthy, K.D., *Stochastic Differential Equations on Manifolds*, Cambridge University Press, London, 1982.
6. Emery, M., *Stochastic Calculus in Manifolds*, Springer-Verlag, Berlin, 1989.
7. Flügge, S., *Practical Quantum Mechanics*, Vol. 1 and 2, Springer-Verlag, Berlin, 1971.
8. Hida, T., *Brownian Motion*, Applications of Math. No. 11, Springer, Berlin, 1980.
9. Hsu, E.P., *Stochastic Analysis on Manifolds*, Graduate Studies in Mathematics, Vol. 38, American Mathematical Society, Providence, RI, 2002.
10. Ikeda, N., Watanabe, S., *Stochastic Differential Equations and Diffusion Processes*, 2nd ed., North-Holland, Amsterdam, 1989.
11. Itô, K., "Stochastic differential equations in a differentiable manifold," *Nagoya Math. J.*, 1, pp. 35–47, 1950.
12. Itô, K., "Stochastic differential equations in a differentiable manifold (2)," *Sci. Univ. Kyoto Math. Ser. A*, 28, pp. 81–85, 1953.
13. Itô, K., McKean, H.P., Jr., *Diffusion Processes and their Sample Paths*, Springer, Berlin, 1996.
14. Kunita, H., *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, London, 1997.
15. Lewis, J., "Brownian motion on a submanifold of Euclidean space," *Bull. London Math. Soc.*, 18, pp. 616–620, 1986.
16. McKean, H., Jr., Singer, I.M., "Curvature and the eigenvalues of the Laplacian," *J. Diff. Geom.*, 1, pp. 43–69, 1967.
17. McKean, H.P., Jr., "Brownian motions on the 3-dimensional rotation group," *Mem. College Sci. Univ. Kyoto Ser. A*, 33, 1, pp. 25–38, 1960.
18. McLachlan, N.W., *Theory and Application of Mathieu Functions*, Oxford, Clarendon Press, 1951.
19. Øksendal, B., *Stochastic Differential Equations, An Introduction with Applications*, 5th ed., Springer, Berlin, 1998.
20. Orsingher, E., "Stochastic motions on the 3-sphere governed by wave and heat equations," *J. Appl. Prob.*, 24, pp. 315–327, 1987.
21. Perrin, P.F., "Mouvement Brownien D'un Ellipsoïde (I). Dispersion Diélectrique Pour des Molécules Ellipsoïdales," *J. Phys. Radium*, 7, pp. 497–511, 1934.
22. Perrin, P.F., "Mouvement Brownien D'un Ellipsoïde (II). Rotation Libre et Dépolarisation des Fluorescences. Translation et Diffusion de Molécules Ellipsoïdales," *J. Phys. Radium*, 7, pp. 1–11, 1936.
23. Perrin, P.F., "Étude Mathématique du Mouvement Brownien de Rotation," *Ann. Sci. Éc. Norm. Supér.*, 45, pp. 1–51, 1928.
24. Pinsky, M., "Isotropic transport process on a Riemannian manifold," *TAMS*, 218, pp. 353–360, 1976.
25. Pinsky, M., "Can you feel the shape of a manifold with Brownian motion?" in *Topics in Contemporary Probability and its Applications*, pp. 89–102, (J.L. Snell, ed.), CRC Press, Boca Raton, FL, 1995.
26. Roberts, P.H., Ursell, H.D., "Random walk on a sphere and on a Riemannian manifold," *Philos. Trans. R. Soc. London*, A252, pp. 317–356, 1960.
27. Stroock, D.W., *An Introduction to the Analysis of Paths on a Riemannian Manifold*, Mathematical Surveys and Monographs, Vol. 74, American Mathematical Society, Providence, RI, 2000.
28. Wang, Y., Zhou, Y., Maslen, D.K., Chirikjian, G.S., "Solving the phase-noise Fokker–Planck equation using the motion-group Fourier transform," *IEEE Trans. Commun.*, 54, pp. 868–877, 2006.
29. Yau, S.-T., "On the heat kernel of a complete Riemannian manifold," *J. Math. Pures Appl.*, Ser. 9, 57, pp. 191–201, 1978.

30. Yosida, K., "Integration of Fokker–Planck's equation in a compact Riemannian space," *Ark. Mat.*, 1, Nr. 9, pp. 71–75, 1949.
31. Yosida, K., "Brownian motion on the surface of the 3-sphere," *Ann. Math. Stat.*, 20, pp. 292–296, 1949.
32. Yosida, K., "Brownian motion in a homogeneous Riemannian space," *Pacific J. Math.*, 2, pp. 263–296, 1952.

---

## Summary

This volume presented the fundamentals of probability, parts of information theory, differential geometry, and stochastic processes at a level that is connected with physical modeling. The emphasis has been on reporting results that can be readily implemented as simple computer programs, though detailed numerical analysis has not been addressed. In this way it is hoped that a potentially useful language for describing physical problems from various engineering and scientific fields has been made accessible to a wider audience. Not only the terminology and concepts, but also the results of the theorems presented serve the goal of efficient physical description. In this context, efficiency means that the essence of any stochastic phenomenon drawn from a broad set of such phenomena can be captured with relatively simple equations in few variables. And these equations can be solved either analytically or numerically in a way that requires minimal calculations (either by human or computer). This goal is somewhat different than that of most books on stochastic processes. A common goal in other books is to train students of mathematics to learn how to prove theorems. While the ability to prove a theorem is at the center of a pure mathematician's skill set, the results that are spun off during that process sometimes need reinterpretation and restatement in less precise (but more accessible) language in order to be used by practitioners. In other words, rather than stating results in the classical definition–theorem–proof style aimed at pure mathematicians, this book is intended for mathematical modelers including engineers, computational biologists, physical scientists, numerical analysts, and applied and computational mathematicians.

A primary goal of mathematical modeling is to obtain the equations that govern a physical phenomenon. After that point, the rest becomes an issue of numerical implementation. In this volume many equations have been provided that can serve as potent descriptive tools. The combination of geometry, information theory, and stochastic calculus that is provided here can be applied directly to model engineering and biological problems. The numerous explicit examples and exercises make the presentation of what would otherwise be an abstract subject much more concrete. Additional examples can be found in the author's technical articles.

The emphasis here has been on continuous-time processes. This emphasis will continue in Volume 2, in which the topic of stochastic processes on Lie groups is addressed. The first three chapters in Volume 2 define, in a concrete way, the properties of Lie groups. These are special mathematical objects that have the benefits of both group theory and differential geometry behind them. Since a Lie group is both a group and a manifold, more detailed results about the theoretical performance of stochastic processes



on Lie groups can be made than for abstract manifolds. In addition, numerical methods based on harmonic analysis (Fourier expansions) on Lie groups become possible.

Topics that received considerable attention in the current volume, but which were not directly applied to stochastic processes here, are used to a large degree in Volume 2. These include differential forms, Weyl's tube theorem, Steiner's formula, and curvature integrals over bodies and their bounding surfaces. It will be shown that such things play important roles in the area of mathematics known as *integral geometry* or *geometric probability*.

Many other topics are covered in Volume 2 including: variational calculus, Shannon's information theory, ergodic theory, multivariate statistical analysis, statistical mechanics, and Fourier methods on Lie groups.

The topics covered in Volume 2 are not the only ones that follow naturally from the background that has been established in this volume. For those readers with more theoretical interests, but who are not inclined to go on to Volume 2, plenty of pointers to the literature have been provided throughout this volume. The following references cover material not addressed in Volume 2 that will also be of interest: [1, 2, 3, 4, 5].

## References

1. Bell, D.R., *Degenerate Stochastic Differential Equations and Hypoellipticity*, Pitman Monographs and Surveys in Pure and Applied Mathematics 79, Longman Group, Essex, England, 1995.
2. Belopolskaya, Ya. I., Dalecky, Yu. L., *Stochastic Equations and Differential Geometry*, Kluwer Academic, Dordrecht, 1990.
3. Hörmander, L., "Hypoelliptic second order differential equations," *Acta Math.*, 119, pp. 147–171, 1967.
4. Malliavin, P., *Stochastic Analysis*, Grundlehren der mathematischen Wissenschaften 313, Springer, Berlin, 1997.
5. Malliavin, P., *Géométrie Différentielle Stochastique*, University of Montréal Press, Montréal, 1978.

# A

---

## Review of Linear Algebra, Vector Calculus, and Systems Theory

Throughout this book, methods and terminology from the area of mathematics known as *linear algebra* are used to facilitate analytical and numerical calculations. Linear algebra is concerned with objects that can be scaled and added together (i.e., vectors), the properties of sets of such objects (i.e., vector spaces), and special relationships between such sets (i.e., linear mappings expressed as matrices). This appendix begins by reviewing the most relevant results from linear algebra that are used elsewhere in the book. Section A.1 reviews the definition and properties of vectors, vector spaces, inner products, and norms. Section A.2 reviews matrices, matrix norms, traces, determinants, etc. Section A.3 reviews the eigenvalue–eigenvector problem. Section A.4 reviews matrix decompositions. The theory of matrix perturbations is reviewed in Section A.5, the matrix exponential is reviewed in Section A.6, and Kronecker product of matrices is reviewed in Section A.7. Whereas the emphasis in this appendix (and throughout the book) is on real vector spaces, Section A.8 discusses the complex case. Basic linear systems theory is reviewed in Section A.9. The concept of a product integral, which is important for defining Brownian motions in Lie groups, is covered in Section A.10. Building on linear-algebraic foundations, concepts from vector and matrix calculus are reviewed in Section A.11. Section A.12 presents exercises.

### A.1 Vectors

The  $n$ -dimensional Euclidean space,  $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$  ( $n$  times), can be viewed as the set of all “vectors” (i.e., column arrays consisting of  $n$  real numbers,  $x_i \in \mathbb{R}$ ) of the form

$$\mathbf{x} \doteq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

A very special vector is the zero vector,  $\mathbf{0}$ , which has entries that are all equal to the number zero.

#### A.1.1 Vector Spaces

When equipped with the operation of *vector addition* for any two vectors,  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,

$$\mathbf{x} + \mathbf{y} \doteq \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix},$$

and *scalar multiplication* by any  $c \in \mathbb{R}$ ,

$$c \cdot \mathbf{x} \doteq \begin{bmatrix} c \cdot x_1 \\ c \cdot x_2 \\ \vdots \\ c \cdot x_n \end{bmatrix},$$

it can be shown that eight properties hold. Namely:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad (\text{A.1})$$

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}) \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n \quad (\text{A.2})$$

$$\mathbf{x} + \mathbf{0} = \mathbf{x} \in \mathbb{R}^n \quad (\text{A.3})$$

$$\exists -\mathbf{x} \in \mathbb{R}^n \quad \text{for each } \mathbf{x} \in \mathbb{R}^n \quad \text{s.t. } \mathbf{x} + (-\mathbf{x}) = \mathbf{0} \quad (\text{A.4})$$

$$\alpha \cdot (\mathbf{x} + \mathbf{y}) = \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y} \quad \forall \alpha \in \mathbb{R} \text{ and } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad (\text{A.5})$$

$$(\alpha + \beta) \cdot \mathbf{x} = \alpha \cdot \mathbf{x} + \beta \cdot \mathbf{x} \quad \forall \alpha, \beta \in \mathbb{R} \text{ and } \mathbf{x} \in \mathbb{R}^n \quad (\text{A.6})$$

$$(\alpha \cdot \beta) \cdot \mathbf{x} = \alpha \cdot (\beta \cdot \mathbf{x}) \quad \forall \alpha, \beta \in \mathbb{R} \text{ and } \mathbf{x} \in \mathbb{R}^n \quad (\text{A.7})$$

$$1 \cdot \mathbf{x} = \mathbf{x} \quad \forall \mathbf{x} \in \mathbb{R}^n. \quad (\text{A.8})$$

Here the symbol  $\exists$  means “there exists” and  $\forall$  means “for all.” The “ $\cdot$ ” in the above equations denotes scalar–scalar and scalar–vector multiplication.

The above properties each have names: (A.1) and (A.2) are respectively called commutativity and associativity of vector addition; (A.3) and (A.4) are respectively called the existence of an additive identity element and an additive inverse element for each element; (A.5), (A.6), and (A.7) are three different kinds of distributive laws; and (A.8) refers to the existence of a scalar that serves as a multiplicative identity.

These properties make  $(\mathbb{R}^n, +, \cdot)$  a *real vector space*. Moreover, any abstract set,  $X$ , that is closed under the operations of vector addition and scalar multiplication and satisfies the above eight properties is a real vector space  $(X, +, \cdot)$ . If the *field*<sup>1</sup> over which properties (A.5)–(A.7) hold is extended to include complex numbers, then the result is a *complex vector space*.

It is often convenient to decompose an arbitrary vector  $\mathbf{x} \in \mathbb{R}^n$  into a weighted sum of the form

$$\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n = \sum_{i=1}^n x_i \mathbf{e}_i.$$

Here the scalar–vector multiplication,  $\cdot$ , is implied. That is,  $x_i \mathbf{e}_i = x_i \cdot \mathbf{e}_i$ . It is often convenient to drop the dot, because the scalar product of two vectors (which will be defined shortly) is also denoted with a dot. In order to avoid confusion, the dot in the scalar–vector multiplication is henceforth suppressed.

Here the *natural basis vectors*  $\mathbf{e}_i$  are

<sup>1</sup>This can be thought of as the real or complex numbers. More generally a field is an algebraic structure that is closed under addition, subtraction, multiplication, and division. For example, the rational numbers form a field.

$$\mathbf{e}_1 \doteq \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \quad \mathbf{e}_2 \doteq \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}; \quad \dots \quad \mathbf{e}_n \doteq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

This (or any) basis is said to *span* the whole vector space. And “the span” of this basis is  $\mathbb{R}^n$ .

### A.1.2 Linear Mappings and Isomorphisms

In some situations, it will be convenient to take a more general perspective. For example, when considering the tangent planes at two different points on a two-dimensional surface in three-dimensional Euclidean space, these two planes are not the same plane since they sit in space in two different ways, but they nonetheless have much in common. It is clear that scalar multiples of vectors in either one of these planes can be added together and the result will remain within that plane, and all of the other rules in (A.1)–(A.8) will also follow. By attaching an origin and coordinate system to each plane at the point where it meets the surface, all vectors tangent to the surface at that point form a vector space. If two of these planes are labeled as  $V_1$  and  $V_2$ , it is clear that each one is “like”  $\mathbb{R}^2$ . In addition, given vectors in coordinate systems in either plane, it is possible to describe those two-dimensional vectors as three-dimensional vectors in the ambient three-dimensional space,  $E = \mathbb{R}^3$ , in which the surfaces sit. Both transformations between planes and from a plane into three-dimensional space are examples of *linear transformations* between two vector spaces of the form  $\mathbf{f} : V \rightarrow U$ , which are defined by the property

$$\mathbf{f}(a\mathbf{v}_1 + b\mathbf{v}_2) = a\mathbf{f}(\mathbf{v}_1) + b\mathbf{f}(\mathbf{v}_2) \quad (\text{A.9})$$

for all  $a, b \in \mathbb{R}$  (or  $\mathbb{C}$  if  $V$  is complex), and  $\mathbf{v}_i \in V$ .

Most linear transformations  $\mathbf{f} : V_1 \rightarrow V_2$  that will be encountered in this book will be of the form

$$\mathbf{f}(\mathbf{x}) = A\mathbf{x}$$

where the dimensions of the matrix  $A$  are  $\dim(V_2) \times \dim(V_1)$ . (Matrices, as well as matrix–vector multiplication, are defined in Section A.2.)

The concept of two planes being equivalent is made more precise by defining the more general concept of a *vector-space isomorphism*. Specifically, two vector spaces,  $V_1$  and  $V_2$ , are *isomorphic* if there exists an *invertible* linear transformation between them. And this is reflected in the matrix  $A$  being invertible. (More about matrices will follow.) When two vector spaces are isomorphic, the notation  $V_1 \cong V_2$  is used.

### A.1.3 The Scalar Product and Vector Norm

The *scalar product* (also called the *inner product*) of two vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  is defined as

$$\mathbf{x} \cdot \mathbf{y} \doteq x_1y_1 + x_2y_2 + \dots + x_ny_n = \sum_{i=1}^n x_iy_i.$$

Sometimes it is more convenient to write  $\mathbf{x} \cdot \mathbf{y}$  as  $(\mathbf{x}, \mathbf{y})$ . The comma in this notation is critical. With this operation, it becomes clear that  $x_i = \mathbf{x} \cdot \mathbf{e}_i$ . Note that the inner product is linear in each argument. For example,

$$\mathbf{x} \cdot (\alpha_1 \mathbf{y}_1 + \alpha_2 \mathbf{y}_2) = \alpha_1 (\mathbf{x} \cdot \mathbf{y}_1) + \alpha_2 (\mathbf{x} \cdot \mathbf{y}_2).$$

Linearity in the first argument follows from the fact that the inner product is symmetric:  $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ .

The vector space  $\mathbb{R}^n$  together with the inner-product operation is called an *inner-product space*. The norm of a vector can be defined using the inner product as

$$\|\mathbf{x}\| \doteq \sqrt{(\mathbf{x}, \mathbf{x})}. \quad (\text{A.10})$$

If  $\mathbf{x} \neq \mathbf{0}$ , this will always be a positive quantity, and for any  $c \in \mathbb{R}$ ,

$$\|c\mathbf{x}\| = |c| \|\mathbf{x}\|. \quad (\text{A.11})$$

The *triangle inequality* states that

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|. \quad (\text{A.12})$$

This is exactly a statement (in vector form) of the ancient fact that the sum of the lengths of any two sides of a triangle can be no less than the length of the third side.

Furthermore, the well-known *Cauchy–Schwarz inequality* states that

$$(\mathbf{x}, \mathbf{y}) \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|. \quad (\text{A.13})$$

This is used extensively throughout the rest of the book, and it is important to know where it comes from.

The proof of the Cauchy–Schwarz inequality is actually quite straightforward. To start, define  $f(t)$  as

$$f(t) = (\mathbf{x} + t\mathbf{y}, \mathbf{x} + t\mathbf{y}) = \|\mathbf{x} + t\mathbf{y}\|^2 \geq 0.$$

Expanding out the inner product results in a quadratic equation in  $t$ :

$$f(t) = (\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y})t + (\mathbf{y}, \mathbf{y})t^2 \geq 0.$$

Since the minimum of  $f(t)$  occurs when  $f'(t) = 0$  (i.e., when  $t = -(\mathbf{x}, \mathbf{y})/(\mathbf{y}, \mathbf{y})$ ), the minimal value of  $f(t)$  is

$$f(-(\mathbf{x}, \mathbf{y})/(\mathbf{y}, \mathbf{y})) = (\mathbf{x}, \mathbf{x}) - (\mathbf{x}, \mathbf{y})^2/(\mathbf{y}, \mathbf{y})$$

when  $\mathbf{y} \neq \mathbf{0}$ . Since  $f(t) \geq 0$  for all values of  $t$ , the Cauchy–Schwarz inequality follows. In the case when  $\mathbf{y} = \mathbf{0}$ , the Cauchy–Schwarz inequality reduces to the equality  $0 = 0$ .

Alternatively, the Cauchy–Schwarz inequality is obtained for vectors in  $\mathbb{R}^n$  from *Lagrange’s equality* [2]:

$$\left( \sum_{k=1}^n a_k^2 \right) \left( \sum_{k=1}^n b_k^2 \right) - \left( \sum_{k=1}^n a_k b_k \right)^2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n (a_i b_j - a_j b_i)^2 \quad (\text{A.14})$$

by observing that the right-hand side of the equality is always non-negative. Lagrange’s equality can be proved by induction.

The norm in (A.10) is often called the “2-norm” to distinguish it from the more general vector “p-norm”

$$\|\mathbf{x}\|_p \doteq \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

for  $1 \leq p \leq \infty$ , which also satisfies (A.11) and (A.12). The vector space  $\mathbb{R}^n$  together with the norm operation is called a *normed vector space*. Furthermore, if instead of vectors with real-valued entries, we consider vectors with complex-valued entries, then the inner product

$$(\mathbf{x}, \mathbf{y}) \doteq \sum_{i=1}^n \overline{x_i} y_i$$

can be defined where for any complex number  $c = a + b\sqrt{-1}$ , the notation  $\bar{c} = a - b\sqrt{-1}$  defines the complex conjugate of  $c$ . In doing so (A.10)–(A.13) all still hold, with  $|c| = \sqrt{c\bar{c}} = \sqrt{a^2 + b^2}$  replacing the absolute value in (A.11).

#### A.1.4 The Gram–Schmidt Orthogonalization Process

Let  $V = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  be a *nonorthogonal basis* for  $\mathbb{R}^n$ . That is, any vector  $\mathbf{x} \in \mathbb{R}^n$  can be expressed as  $\mathbf{x} = \sum_{i=1}^n \chi_i \mathbf{v}_i$  for an appropriate choice of real numbers  $\chi_1, \chi_2, \dots, \chi_n$ . Any collection of  $m < n$  vectors  $\{\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_m}\}$  with  $1 \leq i_1 < \dots < i_m \leq n$  is said to *span* a *vector subspace* of  $\mathbb{R}^n$  (i.e., a vector space strictly contained in  $\mathbb{R}^n$ ). This vector subspace is denoted as  $\text{span}\{\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_m}\}$ .

An orthonormal basis for  $\mathbb{R}^n$  can be constructed from  $V$  as follows. First normalize  $\mathbf{v}_1$  and define

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}.$$

Then define  $\mathbf{u}_2$  by removing the part of  $\mathbf{v}_2$  that is parallel to  $\mathbf{u}_1$  and normalizing what remains:

$$\mathbf{u}_2 = \frac{\mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1}{\|\mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1\|}.$$

It is easy to verify that  $\mathbf{u}_1 \cdot \mathbf{u}_1 = \mathbf{u}_2 \cdot \mathbf{u}_2 = 1$  and  $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$ . The process can then be performed recursively by removing the parts of  $\mathbf{v}_i$  that are parallel to each of the  $\{\mathbf{u}_1, \dots, \mathbf{u}_{i-1}\}$ . Then  $\mathbf{u}_i$  is defined as the unit vector of what remains. In other words, the following formula is used recursively for  $i = 2, 3, \dots, n$ :

$$\mathbf{u}_i = \frac{\mathbf{v}_i - \sum_{k=1}^{i-1} (\mathbf{v}_i \cdot \mathbf{u}_k)\mathbf{u}_k}{\|\mathbf{v}_i - \sum_{k=1}^{i-1} (\mathbf{v}_i \cdot \mathbf{u}_k)\mathbf{u}_k\|}. \quad (\text{A.15})$$

This process is repeated until a full set of orthonormal basis vectors  $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  is constructed.

Gram–Schmidt orthogonalization works equally well on inner product spaces other than  $\mathbb{R}^n$ , such as the space of square-integrable functions  $L^2(\mathbb{R})$ , where the dot product is replaced with the inner product defined for that space.

#### A.1.5 Dual Spaces

Given the space of  $n$ -dimensional real column vectors, each of which is described by  $\mathbf{x} \in \mathbb{R}^n$ , it is possible to define the linear mapping

$$a : \mathbb{R}^n \rightarrow \mathbb{R} \quad \text{where} \quad a(\mathbf{x}) \doteq \mathbf{a} \cdot \mathbf{x}$$

for some fixed  $\mathbf{a} \in \mathbb{R}^n$ . This linear mapping can be written as  $a(\mathbf{x}) = \mathbf{a}^T \mathbf{x}$  where  $\mathbf{a}^T$  is an  $n$ -dimensional row vector, called the *transpose* of  $\mathbf{a}$ .

The fact that the function  $a(\mathbf{x})$  is linear is clear, since

$$a(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2) = \alpha a(\mathbf{x}_1) + \beta a(\mathbf{x}_2).$$

Furthermore, given two such functionals,  $a(\mathbf{x})$  and  $b(\mathbf{x})$ , together with scalars  $\alpha$  and  $\beta$ , it is possible to define the functional

$$(\alpha a + \beta b)(\mathbf{x}) = \alpha \mathbf{a}^T \mathbf{x} + \beta \mathbf{b}^T \mathbf{x}.$$

That is, linear functionals can be scaled and added like vectors and the space of linear functionals “acts like”  $\mathbb{R}^n$ . This is not surprising, because each linear functional  $a(\mathbf{x})$  is defined by an element  $\mathbf{a} \in \mathbb{R}^n$ . The space of all linear functionals is a vector space called the *dual space* of  $\mathbb{R}^n$ , and can be thought of intuitively as the collection of all  $n$ -dimensional row vectors.

If  $V = \mathbb{R}^n$  and the *dual space* is denoted as  $V^*$ , then the inner product of two vectors in  $V$  instead can be thought of as a product between one vector in  $V$  and one in  $V^*$ . And if  $V$  has the basis  $\{\mathbf{e}_i\}$ , then  $V^*$  has the basis  $\{\mathbf{e}_i^*\}$  such that  $\mathbf{e}_i^* \mathbf{e}_j = \delta_{ij}$ . In the present context when everything is real, the  $*$  has no meaning other than transpose, but when the discussion is broadened to include vectors with complex entries, or infinite-dimensional spaces of functions, the value of the dual space concept becomes more apparent. For more, see [12, 23].

### A.1.6 The Vector Product in $\mathbb{R}^3$

Given two vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ , the *vector product* (or *cross product*) is defined as

$$\mathbf{a} \times \mathbf{b} \doteq \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.$$

A real matrix is called *skew-symmetric* (or *anti-symmetric*) if its transpose is equal to its negative. Any  $3 \times 3$  skew-symmetric matrix,  $S = -S^T$ , can be written as

$$S = \begin{pmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{pmatrix}, \quad (\text{A.16})$$

where  $s_1$ ,  $s_2$ , and  $s_3$  can be viewed as the components of a vector  $\mathbf{s} \in \mathbb{R}^3$ , called the *dual vector* of  $S$ .

The relationship between skew-symmetric matrices and the cross product is

$$S\mathbf{x} = \mathbf{s} \times \mathbf{x}. \quad (\text{A.17})$$

The *triple product* of three vectors,  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ , is

$$\det[\mathbf{a}, \mathbf{b}, \mathbf{c}] \doteq \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

This has the geometric meaning of the volume of the region in  $\mathbb{R}^3$  defined by all vectors of the form  $\mathbf{x}(u, v, w) = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$  where  $(u, v, w) \in [0, 1] \times [0, 1] \times [0, 1]$ .

The above concepts only apply to vectors in  $\mathbb{R}^3$ .  $\mathbb{R}^3$  (viewed as a vector space) when augmented with the cross-product operation, becomes a new kind of space with richer structure. This is similar to the way in which an inner-product space or normed vector space is richer than a vector space.

A *Lie algebra* is a special kind of vector space,  $V$ , with an additional operation  $[\cdot, \cdot]$ , such that for any  $x, y, z \in V$ ,  $[x, y] \in V$  and for any  $\alpha \in \mathbb{C}$  the following properties are satisfied:

$$[x + y, z] = [x, z] + [y, z] \quad (\text{A.18})$$

$$[z, x + y] = [z, x] + [z, y] \quad (\text{A.19})$$

$$[\alpha x, y] = [x, \alpha y] = \alpha[x, y] \quad (\text{A.20})$$

$$[x, y] = -[y, x] \quad (\text{A.21})$$

$$0 = [x, [y, z]] + [y, [z, x]] + [z, [x, y]]. \quad (\text{A.22})$$

The first three of these properties, (A.18)–(A.20), are true for any “algebra,” whereas property (A.21) (which is called anti-symmetry) and (A.22) (which is called the *Jacobi identity*) turn the algebra into a Lie algebra. To distinguish a Lie algebra from a generic vector space,  $V$ , it is sometimes written as  $(V, [\cdot, \cdot])$ . The operation  $[x, y]$  is called the *Lie bracket* of the vectors  $x$  and  $y$ . The property  $[x, x] = 0$  follows automatically. Note that Lie algebras are not generally associative with respect to the Lie bracket operation.

For example,  $\mathbb{R}^3$  together with the cross product  $[\mathbf{a}, \mathbf{b}] = \mathbf{a} \times \mathbf{b}$  makes  $\mathbb{R}^3$  a Lie algebra. (See Exercise A.9(a).)

## A.2 Matrices

An  $m \times n$  matrix  $A$  is an array of real or complex numbers:

$$A = [a_{ij}] \doteq \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & a_{m-1,n} \\ a_{m1} & \cdots & a_{m,n-1} & a_{mn} \end{pmatrix}.$$

The numbers  $m$  and  $n$  are called the *dimensions* of  $A$ . The element (or entry) in the  $i$ th row and  $j$ th column of the  $m \times n$  matrix  $A$  is denoted as  $a_{ij}$ . Likewise, the elements of any matrix denoted with an upper case letter are generally written as subscripted lower case letters.

Sometimes it is convenient to write this as an array of  $n$  column vectors:

$$A = [\mathbf{a}_1, \dots, \mathbf{a}_n] \quad \text{where} \quad \mathbf{a}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix} \in \mathbb{R}^m \quad \text{for } i = 1, 2, \dots, n.$$

Addition of two matrices with the same dimensions is defined by the scalar addition of elements with the same indices:

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}].$$

Multiplication of a scalar and a matrix is defined as the matrix

$$c \cdot A = [c \cdot a_{ij}] \doteq \begin{pmatrix} c \cdot a_{11} & c \cdot a_{12} & \cdots & c \cdot a_{1n} \\ c \cdot a_{21} & c \cdot a_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & c \cdot a_{m-1,n} \\ c \cdot a_{m1} & \cdots & c \cdot a_{m,n-1} & c \cdot a_{mn} \end{pmatrix}.$$



The complex conjugate of a matrix  $A$  is the matrix consisting of the complex conjugate of all of its entries:  $\overline{A} = [\overline{a_{ij}}]$ .

The *transpose* of a matrix  $A$ , denoted as  $A^T$ , is the matrix resulting by interchanging the role of the rows and the columns:

$$A^T \doteq \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & \vdots \\ \vdots & \vdots & \ddots & a_{m,n-1} \\ a_{1n} & \cdots & a_{m-1,n} & a_{mn} \end{pmatrix}. \quad (\text{A.23})$$

In other words,  $[a_{ij}]^T = [a_{ji}]$ .

The *Hermitian conjugate* of a matrix is the complex conjugate and transpose:

$$A^* = [a_{ij}]^* \doteq \overline{A}^T = [\overline{a_{ji}}].$$

### A.2.1 Matrix Multiplication and the Trace

Given an  $m \times n$  matrix  $A$ , and an  $n \times p$  matrix  $B$  the  $(i, j)$ th element of the product  $AB$  is defined as

$$(AB)_{ij} \doteq \sum_{k=1}^n a_{ik} b_{kj}.$$

The particular label for  $k$  is unimportant because it is summed over all values, i.e.,  $k$  in the above equation can be replaced with  $l$  (or any other letter not already being used), and the meaning will be the same.

When three square  $n \times n$  matrices are multiplied, the  $(i, j)$ th element of the result is

$$(ABC)_{ij} = \sum_{k=1}^n \sum_{l=1}^n a_{ik} b_{kl} c_{lj}.$$

This can be broken up in two ways as

$$(ABC)_{ij} = \sum_{k=1}^n a_{ik} \left( \sum_{l=1}^n b_{kl} c_{lj} \right) = \sum_{l=1}^n \left( \sum_{k=1}^n a_{ik} b_{kl} \right) c_{lj}.$$

In terms of the original matrices, this is the *associative law*

$$ABC = A(BC) = (AB)C.$$

Note that the order of the matrices when written from left to right stays the same and so there is no need to compute a product between  $A$  and  $C$ . But there is nonetheless some choice in the way that the matrices can be grouped together when performing the constituent pairwise matrix products.

When  $A$  is  $n \times n$ , the *trace* of  $A$  is defined as

$$\boxed{\text{trace}(A) \doteq \sum_{i=1}^n a_{ii}.} \quad (\text{A.24})$$

A convenient shorthand for this is  $\text{tr}(A)$ . Note that for square matrices  $A$  and  $B$  with the same dimensions and a scalar,  $c$ ,

$$\begin{aligned} \text{tr}(c \cdot A) &= c \cdot \text{tr}(A) & \text{tr}(A + B) &= \text{tr}(A) + \text{tr}(B) \\ \text{tr}(AB) &= \text{tr}(BA) & \text{tr}(A^T) &= \text{tr}(A). \end{aligned} \tag{A.25}$$

Given  $n \times n$  matrices  $A = [a_{ij}]$  and  $B = [b_{ij}]$ , computing the product  $C = AB$  by the definition

$$c_{ik} = \sum_{j=1}^n a_{ij} b_{jk}$$

uses  $n$  multiplications and  $n - 1$  additions for each fixed pair of  $(i, k)$ . Doing this for all  $i, k \in [1, n]$  then uses  $n^3$  scalar multiplications and  $n^2(n - 1)$  additions. However, if the matrices have special structure, then this computational cost can be reduced tremendously. For example, the product of  $n \times n$  diagonal matrices (i.e.,  $A$  with  $a_{ij} = 0$  when  $i \neq j$ ) can be computed using  $n$  multiplications and  $n - 1$  additions.

Other methods for reducing the complexity of matrix multiplication even when they have no special structure also exist [20, 21].

### A.2.2 The Determinant

A *determinant* of an  $n \times n$  matrix,  $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ , is a scalar-valued function,  $\det A$  (also denoted as  $\det(A)$  and  $|A|$ ) with the following properties:

1. Multilinearity

$$\begin{aligned} &\det[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{i-1}, \alpha \mathbf{v} + \beta \mathbf{w}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_n] = \\ &\alpha \det[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{i-1}, \mathbf{v}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_n] + \beta \det[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{i-1}, \mathbf{w}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_n]. \end{aligned}$$

2. Anti-symmetry

$$\det[\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n] = -\det[\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n].$$

3. Normalization

$$\det \mathbb{I} = \det[\mathbf{e}_1, \dots, \mathbf{e}_n] = 1.$$

Here  $\mathbb{I} = [\delta_{ij}]$  is the identity matrix consisting of diagonal entries with a value of unity and all other entries with a value of zero. Similarly,  $\mathbb{O}$  denotes the matrix with entries that are all zeros.

It can be shown (see, e.g., [14]) that these three properties are satisfied by a single unique function which exists for every square matrix. Therefore, we refer to “the” determinant rather than “a” determinant. Furthermore, the above conditions could have been stated by decomposing  $A$  into rows rather than columns. It then becomes clear that Gaussian elimination computations in exact arithmetic (which correspond to the row version of #1 above) leave the determinant unchanged.

The determinant function satisfying the above three properties can be defined by the Leibniz formula

$$\det A \doteq \sum_{\sigma \in \Pi_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} = \sum_{j=1}^{n!} \operatorname{sgn}(\sigma_j) \prod_{i=1}^n a_{i,\sigma_j(i)} \quad (\text{A.26})$$

where  $\sigma$  is a permutation<sup>2</sup> of the numbers  $(1, 2, \dots, n)$ , and the sign (or *signature*)  $\operatorname{sgn}(\sigma) = +1$  for *even permutations*, and  $\operatorname{sgn}(\sigma) = -1$  for *odd permutations*.<sup>3</sup> An even permutation is one defined by an even number of pairwise swaps, and an odd permutation is one defined by an odd number of pairwise swaps. Every permutation is either even or odd, but cannot be both. For example, the cyclic permutation

$$(1, 2, \dots, n) \rightarrow (n, 1, 2, \dots, n-1)$$

will be even when  $n$  is odd, and it will be odd when  $n$  is even since it can be realized as the product (i.e., composition) of  $n-1$  pairwise swaps.

As examples of permutations, the elements of  $S_3$  are denoted as

$$\begin{aligned} \sigma_0 &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}; & \sigma_1 &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}; & \sigma_2 &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}; \\ \sigma_3 &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}; & \sigma_4 &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}; & \sigma_5 &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}. \end{aligned}$$

These are *not* matrices. They represent assignments of the upper numbers to the lower ones. For example,  $\sigma_1(1) = 2$  and  $\sigma_4(3) = 1$ .

The signs of the permutations listed above are

$$\begin{aligned} \operatorname{sgn}(\sigma_0) &= \operatorname{sgn}(\sigma_1) = \operatorname{sgn}(\sigma_2) = +1 \\ \operatorname{sgn}(\sigma_3) &= \operatorname{sgn}(\sigma_4) = \operatorname{sgn}(\sigma_5) = -1. \end{aligned}$$

The formula (A.26) is independent of the way we label the  $n!$  elements of  $\Pi_n$ . Due to the factorial growth in  $|\Pi_n|$ , (A.26) is not a practical method for computing the determinant of large matrices. In practice, the properties #1 and #2 from the beginning of this subsection are used together with matrix decompositions and the fact that the determinant of a matrix is equal to the product of its eigenvalues (a fact that will be reviewed in Section A.3). Having said this, (A.26) can be useful. For example, if any column or row of  $A$  consists of all zeros, (A.26) indicates that  $\det A = 0$ .

The determinant has several very useful properties listed below:

If  $A$  and  $B$  are square matrices with the same dimensions,

$$\det(AB) = \det(A)\det(B). \quad (\text{A.27})$$

If  $A$  is square,

$$\det(A^T) = \det(A). \quad (\text{A.28})$$

If  $A^{-1}$  exists (see next subsection), then

$$\det(A^{-1}) = 1/\det(A). \quad (\text{A.29})$$

If  $P$  is invertible, then

<sup>2</sup>See Chapter 6 for a more detailed discussion of permutations.

<sup>3</sup>The name  $\Pi_n$  stands for “permutation group on  $n$  elements.” It is also called the “symmetric group” on  $n$  elements.

$$\det(P^{-1}AP) = \det(A). \quad (\text{A.30})$$

If  $A$  is  $m \times m$  and  $B$  is  $n \times n$ , then the  $(m+n) \times (m+n)$  matrix

$$A \oplus B \doteq \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

(called the *direct sum*) of  $A$  and  $B$  has the determinant

$$\det(A \oplus B) = \det(A)\det(B). \quad (\text{A.31})$$

Note also that

$$\text{trace}(A \oplus B) = \text{trace}(A) + \text{trace}(B). \quad (\text{A.32})$$

### A.2.3 The Inverse of a Matrix

Given a square matrix  $A$  with  $\det(A) \neq 0$ , there exists a matrix  $A^{-1}$  called the *inverse* of  $A$ , which is the unique matrix such that

$$AA^{-1} = A^{-1}A = \mathbb{I}$$

where  $\mathbb{I}$  is the identity matrix with the same dimensions as  $A$ . The set of all invertible  $n \times n$  matrices with complex (or real) entries is denoted as  $GL(n, \mathbb{C})$  (or  $GL(n, \mathbb{R})$ ).

It follows that when the inverse of the product  $AB$  exists, it must have the property that  $(AB)^{-1}(AB) = \mathbb{I}$ . But since  $\det(AB) = \det(A)\det(B)$ , the necessary and sufficient conditions for  $(AB)^{-1}$  to exist are that  $\det(A) \neq 0$  and  $\det(B) \neq 0$ . If these conditions hold, then from the associativity of matrix multiplication,

$$(B^{-1}A^{-1})(AB) = (B^{-1}(A^{-1}A)B) = B^{-1}B = \mathbb{I}.$$

From the uniqueness property of the inverse of a matrix, it follows that

$$(AB)^{-1} = B^{-1}A^{-1}. \quad (\text{A.33})$$

A similar-looking rule, called the *transpose rule*, states that

$$(AB)^T = B^T A^T \quad \text{and} \quad (AB)^* = B^* A^*. \quad (\text{A.34})$$

It can be shown that when the inverse of a matrix exists, the transpose and Hermitian conjugate operations commute with the inverse:

$$(A^{-1})^T = (A^T)^{-1} \quad \text{and} \quad (A^{-1})^* = (A^*)^{-1}.$$

Sometimes the abbreviation  $A^{-T}$  is used to denote the combination of transpose and inverse.

It is useful to compute the inverse of the sum of two matrices. The following identity is derived in [8]:

$$(A + B^T C)^{-1} = A^{-1} - A^{-1} B^T (\mathbb{I} + C^T A^{-1} B)^{-1} C A^{-1} \quad (\text{A.35})$$

under the assumption that  $A + B^T C$ ,  $A$ , and  $\mathbb{I} + C^T A^{-1} B$  are all invertible.

Following [8], (A.35) is proved by first observing that

$$\mathbb{I} = (A + B^T C)^{-1}(A + B^T C)$$

and expanding the right side so that

$$\mathbb{I} = (A + B^T C)^{-1}A + (A + B^T C)^{-1}B^T C.$$

Then multiplying both sides on the right by  $A^{-1}$ ,

$$A^{-1} = (A + B^T C)^{-1} + (A + B^T C)^{-1}B^T C A^{-1}, \quad (\text{A.36})$$

which can be rearranged as

$$(A + B^T C)^{-1}B^T C A^{-1} = A^{-1} - (A + B^T C)^{-1}.$$

Following [8], and returning to (A.36) and multiplying both sides on the right by  $B^T$  gives

$$A^{-1}B^T = (A + B^T C)^{-1}B^T + (A + B^T C)^{-1}B^T C A^{-1}B^T = (A + B^T C)^{-1}B^T(\mathbb{I} + C A^{-1}B^T).$$

Multiplying the first and last terms in the above double equality by  $(\mathbb{I} + C A^{-1}B^T)^{-1}C A^{-1}$  on their right sides gives

$$A^{-1}B^T(\mathbb{I} + C A^{-1}B^T)^{-1}C A^{-1} = (A + B^T C)^{-1}B^T C A^{-1}.$$

But the right side of this expression is the same as the rightmost term in (A.36). Therefore making this substitution results in (A.35).

A similar identity is stated in [15] as

$$(A + S B T^T)^{-1} = A^{-1} - A^{-1}S(B^{-1} + T^T A^{-1}S)^{-1}T^T A^{-1}, \quad (\text{A.37})$$

and is left as an exercise to prove.

#### A.2.4 Pseudo-Inverses and Null Spaces

In some applications it happens that a relationship of the form

$$J\mathbf{x} = \mathbf{b} \quad (\text{A.38})$$

is presented, where  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{b} \in \mathbb{R}^m$ , and  $J \in \mathbb{R}^{m \times n}$  where  $m \neq n$ , and the goal is “to find as good of an approximation to a solution as possible.” If  $m < n$  and  $J$  has  $m$  independent rows, then in general an infinite number of solutions will exist. If  $m > n$ , then in general no solution will exist. However, in both of these situations it can happen that if  $\mathbf{x}$  and  $\mathbf{b}$  are specially chosen vectors, then unique solutions can exist.

In the general case when  $m < n$  and  $J$  has  $m$  independent rows (i.e., it has full *row rank*), then a common way to find a “good” solution is to treat (A.38) as a constraint imposed on the minimization of a quadratic cost function of the form  $C_1 = \frac{1}{2}\mathbf{x}^T W \mathbf{x}$  where  $W = W^T \in \mathbb{R}^{n \times n}$  is a *positive definite* matrix. That is,  $W$  is chosen in such a way that  $C_1 > 0$  for all  $\mathbf{x} \in \mathbb{R}^n$  except  $\mathbf{x} = \mathbf{0}$ , in which case  $C_1 = 0$ .

The resulting solution (which can be obtained using the method of Lagrange multipliers described in Section A.11.1) is [3]

$$\mathbf{x} = J_W^+ \mathbf{b} \quad \text{where} \quad J_W^+ \doteq W^{-1} J^T (J W^{-1} J^T)^{-1}. \quad (\text{A.39})$$

The matrix  $J_W^+$  is called the weighted *pseudo-inverse* of  $J$  with weighting matrix  $W$ . In the event that  $\text{rank}(J) < m$ , and so an exact solution may not exist, then a *damped pseudo-inverse* can be defined as

$$J_{W,\epsilon}^+ \doteq W^{-1} J^T (JW^{-1} J^T + \epsilon \mathbb{I})^{-1}$$

where  $\mathbb{I}$  is the  $n \times n$  identity matrix and  $\epsilon \in \mathbb{R}_{>0}$ . Typically, the larger the value of  $\epsilon$ , the worse the approximate solution will be.

The *null space projector* matrix for  $J$  is defined as

$$N_{J,W} \doteq \mathbb{I} - J_W^+ J. \quad (\text{A.40})$$

When  $J$  is full rank, then the solution  $J_W^+ \mathbf{b}$  can be thought of as rejecting the contribution of any linear combination of columns of  $N_{J,W}$  because

$$\begin{aligned} N_{J,W} J_W^+ &= [\mathbb{I} - W^{-1} J^T (JW^{-1} J^T)^{-1} J] [W^{-1} J^T (JW^{-1} J^T)^{-1}] \\ &= [W^{-1} J^T (JW^{-1} J^T)^{-1} - W^{-1} J^T (JW^{-1} J^T)^{-1} (JW^{-1} J^T) (JW^{-1} J^T)^{-1}] \\ &= [W^{-1} J^T (JW^{-1} J^T)^{-1} - W^{-1} J^T (JW^{-1} J^T)^{-1}] \\ &= \mathbb{O}_m, \end{aligned}$$

and similarly

$$J N_{J,W} = \mathbb{O}_m.$$

In the case when  $m > n$ , the “best” approximate solution to (A.38) can be obtained as that which minimizes a cost function of the form  $C_2 = (J\mathbf{x} - \mathbf{b})^T M (J\mathbf{x} - \mathbf{b})$  where  $M = M^T \in \mathbb{R}^{m \times m}$  is chosen to be positive definite. This result is denoted as

$$\mathbf{x} = J_M^\dagger \mathbf{b} \quad \text{where} \quad J_M^\dagger = (J^T M J)^{-1} J^T M. \quad (\text{A.41})$$

This will not solve the equation, but will provide an approximation that minimizes the cost  $C_2$ , as long as  $J$  has rank  $n$ . In the event that the rank of  $J$  is less than  $n$ , then a damped version of this pseudo-inverse also exists, and can be written in the form

$$J_{M,\epsilon}^\dagger = (J^T M J + \epsilon \mathbb{I})^{-1} J^T M.$$

All of the versions of these pseudo-inverses have practical applications. For example, see [4] for applications to the design of robotic manipulator arms.

### A.2.5 Special Kinds of Matrices

Many special kinds of matrices are defined in terms of the transpose and Hermitian conjugate. For example, a *symmetric matrix* is one for which  $A = A^T$ . A *skew-symmetric matrix* is one for which  $A = -A^T$ . An *orthogonal matrix* is one for which  $AA^T = \mathbb{I}$ . A *Hermitian matrix* is one for which  $A = A^*$ . A *skew-Hermitian matrix* is one for which  $A = -A^*$ . A *unitary matrix* is one for which  $AA^* = \mathbb{I}$ . All of these are examples of *normal* matrices, which have the property that  $AA^* = A^*A$ .

The properties of the determinant immediately indicate that for all unitary matrices

$$\det(AA^*) = \det(A)\det(A^*) = \det(\mathbb{I}) = 1.$$

Furthermore, since the determinant is unchanged under transposition,  $\det(A^*) = \det(\bar{A})$ . But since the operation of complex conjugation distributes over scalar multiplication and addition, it must also do so for the determinant since (A.26) is a combination

of products and sums of the scalar elements. Therefore, if  $A$  is unitary  $\overline{\det(A)}\det(A) = 1$ , indicating that the determinant of a unitary matrix is of the form  $\det(A) = e^{i\theta}$  for some  $\theta \in [0, 2\pi)$ . For real orthogonal matrices,  $|\det(A)|^2 = 1$ , indicating that  $\det(A) = \pm 1$ .

Two “special” sets of matrices that are encountered frequently are

$$SO(n) = \{A \mid AA^T = \mathbb{I} \text{ and } \det(A) = +1\}$$

and

$$SU(n) = \{A \mid AA^* = \mathbb{I} \text{ and } \det(A) = +1\}.$$

These are respectively the “special orthogonal” and “special unitary” matrices.

A rotation in  $n$ -dimensional space is described with a special orthogonal matrix. In the case of three-dimensional rotation about a fixed axis by an angle  $\phi$ , the rotation only has one degree of freedom. In particular, for counterclockwise rotations about the  $\mathbf{e}_1$ ,  $\mathbf{e}_2$ , and  $\mathbf{e}_3$  axes:

$$R_1(\phi) \doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{pmatrix} \quad (\text{A.42})$$

$$R_2(\phi) \doteq \begin{pmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{pmatrix} \quad (\text{A.43})$$

$$R_3(\phi) \doteq \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (\text{A.44})$$

## A.2.6 Matrix Norms

Throughout this text, the *Frobenius norm* of a square  $n \times n$  matrix is used:

$$\|A\| \doteq \left( \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}} = (\text{tr}(AA^*))^{\frac{1}{2}}. \quad (\text{A.45})$$

$A$  can have either real or complex entries, and  $|a_{ij}|$  is interpreted as the absolute value of a real number, or modulus of this complex number. When  $A$  is infinite-dimensional,  $\|A\|$  is also called the *Hilbert-Schmidt* norm.

In general a *matrix norm* must satisfy the properties

$$\|A + B\| \leq \|A\| + \|B\| \quad (\text{A.46})$$

$$\|cA\| = |c| \|A\| \quad (\text{A.47})$$

$$\|A\| \geq 0 \quad (\text{A.48})$$

$$\|A\| = 0 \iff A = \mathbb{O}. \quad (\text{A.49})$$

The last quality above is referred to as positive definiteness.

In addition, the Frobenius norm has the *sub-multiplicative* property:

$$\|AB\| \leq \|A\| \|B\|; \quad (\text{A.50})$$

it is invariant under Hermitian conjugation:

$$\|A^*\| = \|A\|;$$

and is invariant under products with arbitrary unitary matrices of the same dimensions as  $A$ :

$$\|AU\| = \|UA\| = \|A\|.$$

Throughout the text,  $\|\cdot\|$  denotes the Frobenius norm. This norm is easy to compute, and has many nice properties such as those mentioned above. However, one desirable property of norms to use in the analysis of limiting behaviors of Fourier transforms of pdfs on Lie groups in Volume 2 is that the norm of the identity matrix should be equal to unity. Unfortunately, the Frobenius norm returns  $\|\mathbb{I}\| = \sqrt{n}$ . And if we define  $\|A\|' = \|A\|/\sqrt{n}$ , then this new norm does not have the sub-multiplicative property. For this reason, in some problems it is useful to use an alternative norm,  $\|\cdot\|$  that possesses both the sub-multiplicative property  $\|AB\| \leq \|A\| \cdot \|B\|$  and  $\|\mathbb{I}\| = 1$ . An infinite number of such norms exist. These are the *induced norms*:

$$\|A\|_p \doteq \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

where

$$\|\mathbf{x}\|_p \doteq \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}.$$

The matrix and vector  $p$ -norms are said to be *consistent* with each other in the sense that  $\|A\mathbf{x}\|_p \leq \|A\|_p \|\mathbf{x}\|_p$ . More generally, any norm that remains sub-multiplicative when applied to the product of matrices with any compatible dimensions (including matrix-vector products) is called consistent. Three examples are the induced 1-norm, 2-norm, and  $\infty$ -norms:

$$\|A\|_1 \doteq \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$$

$$\|A\|_2 \doteq \max_{\mathbf{x} \neq \mathbf{0}} \left( \frac{\mathbf{x}^* A^* A \mathbf{x}}{\mathbf{x}^* \mathbf{x}} \right)^{\frac{1}{2}}$$

and

$$\|A\|_\infty \doteq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

The trouble is, with the exception of  $\|\cdot\|_2$ , these norms are not invariant under products with unitary matrices and under Hermitian conjugation. Therefore, as a second “special” norm for use in the analysis of probability problems on groups, the following notation will be used:

$$\boxed{\|A\|} \doteq \|A\|_2 \tag{A.51}$$

which has the properties

$$\|A\| = \|AU\| = \|UA\| = \|A^*\| \quad \text{and} \quad \|U\| = 1 \quad \forall \quad U \in SU(n). \tag{A.52}$$

One final set of norms that have some of the desirable properties are



$$\|A\|_R \doteq \max_{1 \leq i \leq n} \sqrt{\sum_{j=1}^n |a_{ij}|^2} = \max_{1 \leq i \leq n} (\mathbf{e}_i^T A A^* \mathbf{e}_i)^{\frac{1}{2}} \quad (\text{A.53})$$

and

$$\|A\|_L \doteq \max_{1 \leq j \leq n} \sqrt{\sum_{i=1}^n |a_{ij}|^2} = \max_{1 \leq j \leq n} (\mathbf{e}_j^T A^* A \mathbf{e}_j)^{\frac{1}{2}}. \quad (\text{A.54})$$

To see that  $\|A\|_R$  is a norm, simply evaluate the following:

$$\begin{aligned} \|A + B\|_R &= \max_{1 \leq i \leq n} \sqrt{\sum_{j=1}^n |a_{ij} + b_{ij}|^2} \\ &\leq \max_{1 \leq i \leq n} \left[ \sqrt{\sum_{j=1}^n |a_{ij}|^2} + \sqrt{\sum_{j=1}^n |b_{ij}|^2} \right] \\ &\leq \max_{1 \leq i \leq n} \sqrt{\sum_{j=1}^n |a_{ij}|^2} + \max_{1 \leq i \leq n} \sqrt{\sum_{j=1}^n |b_{ij}|^2} \\ &= \|A\|_R + \|B\|_R. \end{aligned}$$

Clearly  $\|cA\|_R = |c| \cdot \|A\|_R$  and it is positive definite. And while it does not appear to be sub-multiplicative, it does have some other useful properties under multiplication:

$$\begin{aligned} \|AB\|_R &= \max_{1 \leq i \leq n} \sqrt{\left| \sum_{j=1}^n \sum_{k=1}^n a_{ik} b_{kj} \right|^2} \\ &\leq \max_{1 \leq i \leq n} \sqrt{\sum_{j=1}^n \left( \sum_{k=1}^n |a_{ik}|^2 \right) \left( \sum_{l=1}^n |b_{lj}|^2 \right)} \quad (\text{A.55}) \\ &= \max_{1 \leq i \leq n} \sqrt{\sum_{k=1}^n |a_{ik}|^2} \sqrt{\sum_{j,l=1}^n |b_{lj}|^2} \\ &= \|A\|_R \cdot \|B\|. \end{aligned}$$

And so

$$\|AB\|_R \leq \|A\|_R \cdot \|B\|. \quad (\text{A.56})$$

If in the above derivation at the point where the Cauchy–Schwarz inequality is used at (A.55), the weaker condition

$$\sum_{l=1}^n |b_{lj}|^2 \leq n \cdot \max_{1 \leq l \leq n} |b_{lj}|^2$$

is substituted, then the following inequality results:

$$\|AB\|_R \leq \sqrt{n} \|A\|_R \cdot \|B\|_R. \quad (\text{A.57})$$

Similar arguments yield analogous relationships for  $\|AB\|_L$ .

Note that  $\|AU\|_R = \|A\|_R$  is “right” invariant under multiplication by a unitary matrix and  $\|UA\|_L = \|A\|_L$  is “left” invariant. Unlike the Frobenius norm  $\|\cdot\|$  and the induced two-norm  $\|\cdot\|$ , these norms are not bi-invariant. However, they are not as costly to compute as  $\|\cdot\|$ , yet have the nice property that  $\|\mathbb{I}\|_L = \|\mathbb{I}\|_R = 1$ .

In fact, there are an infinite number of possible matrix norms (just as there are an infinite number of vector norms), but the additional useful properties (A.52) are usually not satisfied by norms, making  $\|\cdot\|$  and  $\|\cdot\|$  particularly useful in the context of the problems addressed in this book and its companion volume.

A set of norms that do not have the complete set of desired properties, but are nonetheless useful in some circumstances, is

$$\|A\|_p' \doteq \left( \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^p \right)^{\frac{1}{p}} \tag{A.58}$$

for  $1 < p \neq 2$ .

As has been illustrated, various norms exist for matrices and vectors. And while matrix and vector  $p$ -norms are consistent, these are not the only consistent norms that satisfy  $\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|$ . For example, the Frobenius matrix norm together with the vector 2-norm satisfies this condition. For any consistent norm,

$$\|\mathbf{x}\| = \|\mathbb{I}\mathbf{x}\| \leq \|\mathbb{I}\| \cdot \|\mathbf{x}\| \implies 1 \leq \|\mathbb{I}\|.$$

### A.2.7 Matrix Inequalities

The Cauchy–Schwarz inequality for vectors as presented in (A.13) can be extended to matrices in several ways.

For example, if  $A, B \in \mathbb{R}^{n \times n}$ , then [17]

$$[\text{tr}(A^T B)]^2 \leq \text{tr}(A^T A) \text{tr}(B^T B) \tag{A.59}$$

(with equality if and only if  $B = \alpha A$  for  $\alpha \in \mathbb{R}$ ), and

$$\text{tr}(A^T B)^2 \leq \text{tr}[(A^T A)(B^T B)] \tag{A.60}$$

(with equality if and only if  $AB^T = BA^T$ ). As a consequence,

$$\text{tr}A^2 \leq \|A\|^2$$

with equality holding if and only if  $A = A^T$ .

## A.3 Eigenvalues and Eigenvectors

Given a square matrix,  $A$ , with real or complex entries, an *eigenvector* is any unit vector  $\mathbf{v}$  such that

$$A\mathbf{v} = \lambda\mathbf{v} \tag{A.61}$$

where  $\lambda$  is a scalar called the *eigenvalue*. These can be computed as the roots of the *characteristic polynomial*

$$\det(A - \lambda\mathbb{I}) = 0. \quad (\text{A.62})$$

From the definition of a consistent norm, it follows that

$$\|A\| \cdot \|\mathbf{v}\| \geq \|A\mathbf{v}\| = \|\lambda\mathbf{v}\| = |\lambda| \cdot \|\mathbf{v}\| \quad \implies \quad \|A\| \geq |\lambda|.$$

If  $A$  is  $n \times n$ , then there will always be  $n$  eigenvalues, though they may not be unique. For example, the identity matrix has the eigenvalues  $\lambda_k = 1$  for  $k = 1, \dots, n$ . In contrast, it does not have to be the case that a matrix has  $n$  eigenvectors. (For example, see the discussion of Jordan blocks in Section A.4.1.)

It can be shown that the trace and determinant of a matrix can be expressed in terms of its eigenvalues as

$$\operatorname{tr}(A) = \sum_{i=1}^n \lambda_i \quad \text{and} \quad \det(A) = \prod_{i=1}^n \lambda_i. \quad (\text{A.63})$$

In general, the eigenvalues of a matrix will be complex numbers, even if the matrix has real entries. However, in some matrices with special structure, the eigenvalues will be real. Below the eigenvalues in the case when  $A = A^* \in \mathbb{C}^{n \times n}$  are examined, and special properties of the eigenvectors are derived. The results presented hold as a special case when  $A = A^T \in \mathbb{R}^{n \times n}$ .

**Theorem A.1.** *Eigenvalues of Hermitian matrices are real.*

Proof: To show that something is real, all that is required is to show that it is equal to its own complex conjugate. Recall that given  $a, b \in \mathbb{R}$ , a complex number  $c = a + b\sqrt{-1}$  has a conjugate  $\bar{c} = a - b\sqrt{-1}$ . If  $c = \bar{c}$ , then  $b = 0$  and therefore  $c = a$  is real. The complex conjugate of a vector or matrix is just the complex conjugate of its elements. Furthermore, the complex conjugate of a product is the product of the complex conjugates. Therefore, given  $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$ , applying the Hermitian conjugate to both sides yields

$$\mathbf{u}_i^* A^* = \overline{\lambda_i} \mathbf{u}_i^*.$$

In this derivation  $A^* = A$  because  $A$  is Hermitian, and so making this substitution and multiplying on the right by  $\mathbf{u}_i$  gives

$$\mathbf{u}_i^* A \mathbf{u}_i = \overline{\lambda_i} \mathbf{u}_i^* \mathbf{u}_i.$$

In contrast, starting with  $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$  and multiplying on the left by  $\mathbf{u}_i^*$  gives

$$\mathbf{u}_i^* A \mathbf{u}_i = \lambda_i \mathbf{u}_i^* \mathbf{u}_i.$$

Since the left sides of both of these equations are the same, the right sides can be equated to give

$$\overline{\lambda_i} \mathbf{u}_i^* \mathbf{u}_i = \lambda_i \mathbf{u}_i^* \mathbf{u}_i.$$

Dividing by  $\mathbf{u}_i^* \mathbf{u}_i$ , which is a positive real number, gives  $\lambda_i = \overline{\lambda_i}$  which means that the imaginary part of  $\lambda_i$  is zero, or equivalently  $\lambda_i \in \mathbb{R}$ .

**Theorem A.2.** *Eigenvectors of Hermitian matrices corresponding to distinct eigenvalues are orthogonal.*

Proof: Given a Hermitian matrix  $A$  and two of its eigenvalues  $\lambda_i \neq \lambda_j$  with corresponding eigenvectors  $\mathbf{u}_i$  and  $\mathbf{u}_j$ , by definition the following is true:  $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$ , and  $A\mathbf{u}_j = \lambda_j\mathbf{u}_j$ . Multiplying the first of these on the left by  $\mathbf{u}_j^*$ , and multiplying the second one on the left by  $\mathbf{u}_i^*$ , these become

$$\mathbf{u}_j^*A\mathbf{u}_i = \lambda_i\mathbf{u}_j^*\mathbf{u}_i \quad \text{and} \quad \mathbf{u}_i^*A\mathbf{u}_j = \lambda_j\mathbf{u}_i^*\mathbf{u}_j.$$

Taking the Hermitian conjugate of the second expression gives

$$\mathbf{u}_j^*A^*\mathbf{u}_i = \overline{\lambda_j}\mathbf{u}_j^*\mathbf{u}_i.$$

But since  $A$  is Hermitian,  $A^* = A$ , and from Theorem A.1,  $\overline{\lambda_j} = \lambda_j$ . Therefore, combining the above yields

$$\lambda_j\mathbf{u}_i^*\mathbf{u}_j = \lambda_i\mathbf{u}_i^*\mathbf{u}_i.$$

Subtracting one from the other and using the fact that  $\mathbf{u}_i^*\mathbf{u}_i = \mathbf{u}_i^*\mathbf{u}_j$  then gives

$$(\lambda_i - \lambda_j)\mathbf{u}_i^*\mathbf{u}_j = 0.$$

Since  $\lambda_i \neq \lambda_j$ , division by their difference yields  $\mathbf{u}_i^*\mathbf{u}_j = 0$ , which is a statement of orthogonality.

Any invertible  $n \times n$  matrix consisting of real or complex entries which has distinct eigenvalues (i.e., none of the eigenvalues are repeated) can be written as

$$A[\mathbf{u}_1, \dots, \mathbf{u}_n] = [\lambda_1\mathbf{u}_1, \dots, \lambda_n\mathbf{u}_n] \quad \text{or} \quad AU = UA,$$

where

$$U = [\mathbf{u}_1, \dots, \mathbf{u}_n] \quad \text{and} \quad A = \begin{pmatrix} \lambda_1 & 0 & \dots & \dots & 0 \\ 0 & \lambda_2 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \lambda_{n-1} & 0 \\ 0 & \dots & \dots & 0 & \lambda_n \end{pmatrix}.$$

In other words,  $A$  can be decomposed as

$$A = U\Lambda U^{-1}. \tag{A.64}$$

In fact, this would be true even if there are repeated eigenvalues (as long as there are  $n$  linearly independent eigenvectors), but this will not be proven here.

In the special case when  $A = A^*$ , Theorems A.1 and A.2 above indicate that  $\Lambda$  will have all real entries, and  $U$  will be unitary, and so  $U^{-1} = U^*$ . In the case when  $A$  has real entries and  $A = A^T$ , then  $U$  becomes a real orthogonal matrix.

An  $n \times n$  positive-definite Hermitian matrix,  $A$ , is one for which

$$\mathbf{x}^*A\mathbf{x} \geq 0$$

for all  $\mathbf{x} \in \mathbb{C}^n$  with equality holding only when  $\mathbf{x} = \mathbf{0}$ . This is equivalent to all of the eigenvalues being positive.

## A.4 Matrix Decompositions

### A.4.1 Jordan Blocks and the Jordan Decomposition

A *Jordan block* corresponding to a  $k$ -fold repeated eigenvalue is a  $k \times k$  matrix with the repeated eigenvalue on its diagonal and the number 1 in the superdiagonal. For example,

$$J_2(\lambda) = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix},$$

$$J_3(\lambda) = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix},$$

etc.

Jordan blocks are degenerate in the sense that (to within a sign change) the unit vector  $\mathbf{e}_1 \in \mathbb{R}^k$  is the only eigenvector of  $J_k(\lambda)$ , and the corresponding eigenvalue is  $\lambda$  repeated  $k$  times:

$$[J_k(\lambda)]\mathbf{e}_1 = \lambda\mathbf{e}_1.$$

The determinant and trace of a Jordan block are respectively

$$\det[J_k(\lambda)] = \lambda^k \quad \text{and} \quad \text{tr}[J_k(\lambda)] = k \cdot \lambda.$$

From this it is clear that  $J_k(\lambda)$  is invertible when  $\lambda \neq 0$ .

The notation  $n_i J_i$  stands for the  $n_i$ -fold direct sum of  $J_i$  with itself:

$$n_i J_i \doteq J_i \oplus J_i \oplus \dots \oplus J_i = \begin{pmatrix} J_i & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J_i \end{pmatrix}.$$

The notation

$$\sum_{i=1}^m \bigoplus A_i \doteq A_1 \oplus A_2 \oplus \dots \oplus A_m$$

can be useful shorthand. Note that  $J_k(\lambda)$  is a  $k \times k$  matrix, or equivalently  $\dim(J_k) = k$ . Therefore,

$$\dim \left( \sum_{i=1}^m \bigoplus A_i \right) = \sum_{i=1}^m \dim A_i.$$

Every matrix  $A \in \mathbb{C}^{n \times n}$  can be written in the form

$$A = T J T^{-1} \tag{A.65}$$

where  $T$  is an invertible matrix and

$$J = \sum_{j=1}^q \bigoplus \left( \sum_{i=1}^m \bigoplus n_i^j J_i(\lambda_j) \right)$$

is the direct sum of a direct sum of Jordan blocks with  $m$  being the dimension of the largest Jordan block,  $q$  being the number of different eigenvalues, and  $n_i^j$  being the number of times  $J_i(\lambda_j)$  is repeated in the decomposition of  $A$ . Note that

$$\sum_{j=1}^q \sum_{i=1}^m i \cdot n_i^j = \dim(A).$$

The matrix  $J$  in (A.65) is called the *Jordan normal form* of  $A$ .

For instance, if

$$J = J_1(\lambda_1) \oplus J_1(\lambda_2) \oplus J_2(\lambda_3) \oplus J_2(\lambda_3) \oplus J_3(\lambda_4) \oplus J_5(\lambda_5) \oplus J_6(\lambda_5),$$

then  $m = 6$ ,  $q = 5$ , and all values of  $n_i^j$  are zero except for

$$n_1^1 = n_1^2 = n_3^4 = n_5^5 = n_6^5 = 1; \quad n_2^3 = 2.$$

#### A.4.2 Decompositions into Products of Special Matrices

When linear algebraic operations are performed on a computer, a number of other matrix decompositions are useful. Several of these are reviewed here. For more details see [11, 13, 22].

(*QR Decomposition*) [13]: For any  $n \times n$  matrix,  $A$ , with complex entries, it is possible to find an  $n \times n$  unitary matrix  $Q$  such that

$$A = QR$$

where  $R$  is upper triangular. In the case when  $A$  is real, it is possible to take  $Q \in SO(n)$  and  $R$  real.

(*Cholesky Decomposition*) [13]: Let  $B$  be an  $n \times n$  complex matrix that is decomposable as  $B = A^*A$  for some  $n \times n$  complex matrix  $A$ . Then  $B$  can be decomposed as  $B = LL^*$  where  $L$  is lower triangular with non-negative diagonal entries.

(*Schur Decomposition*) [13]: For any  $A \in \mathbb{C}^{n \times n}$ , it is possible to find a matrix  $U \in SU(n)$  such that

$$U^*AU = T$$

where  $T$  is upper (or lower) triangular with the eigenvalues of  $A$  on its diagonal.

Note: This does *not* mean that for real  $A$  that  $U$  will necessarily be real orthogonal. For example, if  $A = -A^T$ , then  $Q^T A Q$  for  $Q \in O(n)$  will also be skew symmetric and hence cannot be upper triangular.

(*Unitary Diagonalizability of normal matrices*) [13]: Recall that a normal matrix is one for which  $AA^* = A^*A$ . Any such matrix can be decomposed as  $A = U\Lambda U^*$  where  $UU^* = \mathbb{I}$ . Examples of normal matrices include Hermitian, skew-Hermitian, unitary, real orthogonal, and real symmetric matrices.

(*Singular Value Decomposition* or *SVD*) [11]: For any real  $m \times n$  matrix  $A$ , there exist orthogonal matrices  $U \in O(m)$  and  $V \in O(n)$  such that

$$A = U\Lambda V^T$$

where  $\Lambda$  is an  $m \times n$  matrix with real entries  $\Lambda_{ij} = \sigma_i \delta_{ij}$ . The value  $\sigma_i$  is called the  $i$ th largest singular value of  $A$ . If  $A$  has complex entries, then the decomposition becomes

$$A = U\Lambda V^*$$

with  $U$  and  $V$  unitary rather than orthogonal.

(*Polar Decomposition*): Every real square matrix  $A$  can be written as the product of a symmetric matrix  $S_1 = UAU^T$  and the orthogonal matrix

$$R = UV^T, \quad (\text{A.66})$$

or as the product of  $R$  and  $S_2 = VAV^T$ . Hence it is possible to write

$$A = S_1 R = R S_2. \quad (\text{A.67})$$

In the case when  $\det A \neq 0$ ,  $R$  can be calculated as

$$R = A(A^T A)^{-\frac{1}{2}} \quad (\text{A.68})$$

(the negative fractional root makes sense for a symmetric positive definite matrix). Note that (A.66) is always a stable numerical technique for finding  $R$ , whereas (A.68) becomes unstable as  $\det(A)$  becomes small.

(*LU-Decomposition*) [11]: For  $n \times n$  real matrices such that  $\det(A_i) \neq 0$  for  $i = 1, \dots, n$ ,<sup>4</sup> it is possible to write

$$A = LU$$

where  $L$  is a unique lower triangular matrix and  $U$  is a unique upper triangular matrix.

See the classic references [10, 29] for other general properties of matrices.

### A.4.3 Decompositions into Blocks

It is often convenient to partition a matrix into blocks that are themselves matrices as

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

When  $M$  is square, the most common kind of partitioning would be one in which  $A$  and  $D$  are square. Even when  $A$  and  $D$  are square, their dimensions will generally be different. In the case when  $M = M^T$ , it necessarily means that  $A = A^T$ ,  $D = D^T$ , and  $C = B^T$ .

If  $M$  is square and invertible, its inverse can be written in terms of the blocks  $A, B, C, D$ . This is accomplished by decomposing  $M$  into a product of the form

$$M = \begin{pmatrix} \mathbb{I} & 0 \\ L & \mathbb{I} \end{pmatrix} \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} \mathbb{I} & U \\ 0 & \mathbb{I} \end{pmatrix} \quad (\text{A.69})$$

where  $L$  and  $U$  are general matrices,  $P$  and  $Q$  are general invertible matrices, and  $\mathbb{I}$  is the identity of appropriate dimension. Multiplying this out gives four matrix equations in the unknown matrices that can be solved in terms of the originally given blocks. Explicitly,

$$A = P; \quad B = PU; \quad LP = C; \quad D = LPU + Q.$$

Therefore,

$$P = A; \quad U = A^{-1}B; \quad L = CA^{-1}; \quad Q = D - CA^{-1}B.$$

---

<sup>4</sup>The notation  $A_i$  denotes the  $i \times i$  matrix formed by the first  $i$  rows and  $i$  columns of the matrix  $A$ .

This means that  $M^{-1}$  can be computed by applying the rule for the inverse of products of matrices to (A.69) as

$$\begin{aligned} M^{-1} &= \begin{pmatrix} \mathbb{I} & U \\ 0 & \mathbb{I} \end{pmatrix}^{-1} \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{I} & 0 \\ L & \mathbb{I} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \mathbb{I} & -U \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{pmatrix} \begin{pmatrix} \mathbb{I} & 0 \\ -L & \mathbb{I} \end{pmatrix} \\ &= \begin{pmatrix} P^{-1} + UQ^{-1}L & -UQ^{-1} \\ -Q^{-1}L & Q^{-1} \end{pmatrix}. \end{aligned}$$

Then substituting the above expressions for  $L, U, P, Q$  in terms of  $A, B, C, D$  gives an expression for the inverse of  $M$  in terms of its blocks:

$$\boxed{\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}}. \tag{A.70}$$

By recursively using the above procedure, a matrix can be decomposed into smaller blocks, and the inverse of the original matrix can be expressed in terms of the smaller blocks.

### A.5 Matrix Perturbation Theory

This section reviews an important theorem regarding the norm of the inverse of a matrix.

**Theorem A.3.** [27] *Let  $\|\cdot\|$  denote a consistent matrix norm. Given  $A \in GL(n, \mathbb{C})$ , and  $\tilde{A} = A + E$ , then if  $\tilde{A}$  is non-singular,*

$$\frac{\|\tilde{A}^{-1} - A^{-1}\|}{\|\tilde{A}^{-1}\|} \leq \|A^{-1}E\|. \tag{A.71}$$

Furthermore, if  $\|A^{-1}E\| < 1$ , then  $\tilde{A}$  must be non-singular and

$$\frac{\|\tilde{A}^{-1} - A^{-1}\|}{\|A^{-1}\|} \leq \frac{\|A^{-1}E\|}{1 - \|A^{-1}E\|}. \tag{A.72}$$

Proof: Since  $\tilde{A}^{-1}$  exists,  $\tilde{A}\tilde{A}^{-1} = (A + E)\tilde{A}^{-1} = \mathbb{I}$ . Multiplying on the left by  $A^{-1}$  results in  $(I + A^{-1}E)\tilde{A}^{-1} = A^{-1}$ . Therefore,

$$\tilde{A}^{-1} - A^{-1} = -A^{-1}E\tilde{A}^{-1}. \tag{A.73}$$

Taking the norm of both sides gives

$$\|\tilde{A}^{-1} - A^{-1}\| = \|A^{-1}E\tilde{A}^{-1}\| \leq \|A^{-1}E\| \|\tilde{A}^{-1}\|,$$

and hence (A.71).

Instead of assuming that  $\tilde{A}$  is non-singular, if we assume that  $\|A^{-1}E\| < 1$ , then  $\mathbb{I} + A^{-1}E$  must be non-singular. This is because  $\lambda(\mathbb{I} + A^{-1}E) = 1 + \lambda(A^{-1}E) > 0$ , which



follows when  $\|\cdot\|$  is consistent because consistency implies  $|\lambda(A^{-1}E)| \leq \|A^{-1}E\| < 1$ . Then  $\tilde{A}$  must be non-singular, since  $\tilde{A} = A(I + A^{-1}E)$ . Again taking the norm of (A.73), but this time grouping terms in a different order,

$$\|\tilde{A}^{-1}\| = \|A^{-1} - A^{-1}E\tilde{A}^{-1}\| \leq \|A^{-1}\| + \|A^{-1}E\| \|\tilde{A}^{-1}\|.$$

This can be rearranged as

$$\frac{\|\tilde{A}^{-1}\|}{\|A^{-1}\|} \leq \frac{1}{1 - \|A^{-1}E\|}.$$

Multiplying the left side of this inequality with the left side of (A.71), and likewise for the right sides, yields (A.72).

## A.6 The Matrix Exponential

Given the  $n \times n$  matrices  $X$  and  $A$  where  $X = X(t)$  is a function of time and  $A$  is constant, the solution to the differential equation

$$\frac{d}{dt}(X) = AX \tag{A.74}$$

subject to the initial conditions  $X(0) = \mathbb{I}$  is

$$X(t) = \exp(tA).$$

The matrix exponential of any square matrix  $B$  is defined by the Taylor series:

$$\boxed{\exp B = \mathbb{I} + B + B^2/2 + B^3/3! + B^4/4! + \dots} \tag{A.75}$$

The matrix exponential has some very interesting and useful properties. The exponential of the direct sum of two square matrices is the direct sum of their exponentials:

$$\exp(A \oplus B) = e^A \oplus e^B. \tag{A.76}$$

Below, a proof of the following equality is provided:

$$\boxed{\det(\exp A) = e^{\text{tr}(A)}}. \tag{A.77}$$

Using the notation

$$\det X = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{vmatrix},$$

it follows from the product rule for differentiation and the defining properties of the determinant that

$$\begin{aligned} \frac{d}{dt}(\det X) = & \begin{vmatrix} \frac{dx_{11}}{dt} & \frac{dx_{12}}{dt} & \dots & \frac{dx_{1n}}{dt} \\ x_{21} & x_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{vmatrix} + \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \frac{dx_{21}}{dt} & \frac{dx_{22}}{dt} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{vmatrix} + \dots + \\ & \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dx_{n-1,1}}{dt} & \frac{dx_{n-1,2}}{dt} & \dots & \frac{dx_{n-1,n}}{dt} \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{vmatrix} + \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dx_{n1}}{dt} & \frac{dx_{n2}}{dt} & \dots & \frac{dx_{nn}}{dt} \end{vmatrix}. \end{aligned} \tag{A.78}$$

Equation (A.74) is written in component form as

$$\frac{dx_{ik}}{dt} = \sum_{j=1}^n a_{ij}x_{jk}.$$

After making this substitution, the  $i$ th term in (A.78) becomes

$$\begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \sum_{j=1}^n a_{ij}x_{j1} & \sum_{j=1}^n a_{ij}x_{j2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{vmatrix} = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ a_{ii}x_{i1} & a_{ii}x_{i2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{vmatrix}.$$

This follows by subtracting  $a_{ij}$  times the  $j$ th row of  $X$  from the  $i$ th row of the left side for all  $j \neq i$ .

The result is then

$$\frac{d}{dt}(\det X) = \text{trace}(A)(\det X). \tag{A.79}$$

Since  $\det X(0) = 1$ , this implies

$$\det X = \exp(\text{trace}(A)t).$$

Evaluation of both sides at  $t = 1$  yields (A.77).

A sufficient condition for the equalities

$$\exp(A + B) = \exp A \exp B = \exp B \exp A \tag{A.80}$$

to hold is

$$AB = BA. \tag{A.81}$$

This can be verified by expanding both sides in a Taylor series and equating term by term. What is perhaps less obvious is that sometimes the first and/or the second equality in (A.80) can be true when  $A$  and  $B$  do not commute. For example, Fréchet [9] observed that when

$$A = \begin{pmatrix} 0 & 2\pi \\ -2\pi & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$AB \neq BA$  but it is nonetheless true that

$$e^A e^B = e^B e^A.$$

In [26] it is shown that for the non-commuting complex matrices

$$A = \begin{pmatrix} \pi i & 0 \\ 0 & -\pi i \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 \\ 0 & -2\pi i \end{pmatrix}$$

$$e^{A+B} = e^A e^B.$$

Other examples can be found in [9, 31].

Having said this,  $AB = BA$  is a necessary condition for  $\exp t(A+B) = \exp tA \exp tB = \exp tB \exp tA$  to hold for all values of  $t \in \mathbb{R}_{>0}$ . Furthermore, (A.81) becomes a necessary condition for (A.80) if  $A$  and  $B$  are Hermitian [26].

## A.7 Kronecker Products and Kronecker Sums

The *Kronecker product* can be defined for any two matrices

$$H = \begin{pmatrix} h_{11} & h_{12} & \dots & h_{1q} \\ h_{21} & h_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ h_{p1} & h_{p2} & \dots & h_{pq} \end{pmatrix} \in \mathbb{R}^{p \times q}$$

and

$$K = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{rs} \\ k_{21} & k_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ k_{r1} & k_{r2} & \dots & k_{rs} \end{pmatrix} \in \mathbb{R}^{r \times s}$$

as [7]

$$H \widehat{\otimes} K \triangleq \begin{pmatrix} h_{11}K & h_{12}K & \dots & h_{1q}K \\ h_{21}K & h_{22}K & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ h_{p1}K & h_{p2}K & \dots & h_{pq}K \end{pmatrix} \in \mathbb{R}^{pr \times qs}.$$

It follows immediately that

$$(H \widehat{\otimes} K)^T = H^T \widehat{\otimes} K^T.$$

Note that this is in the opposite order than the “transpose rule” for the transpose of the matrix product:  $(HK)^T = K^T H^T$ .

The Kronecker product has the interesting property that for matrices  $A, B, C, D$  of compatible dimensions,

$$(A \widehat{\otimes} B)(C \widehat{\otimes} D) = (AC) \widehat{\otimes} (BD). \quad (\text{A.82})$$

Note that in general,  $H \widehat{\otimes} K \neq K \widehat{\otimes} H$ . However, when  $H$  and  $K$  are both square, there exists a permutation matrix  $P$  such that

$$H \widehat{\otimes} K = P(K \widehat{\otimes} H)P^T,$$

which means that  $\lambda_i(K \widehat{\otimes} H) = \lambda_i(H \widehat{\otimes} K)$  for all values of  $i$ . Furthermore, if  $H$  and  $K$  are invertible,

$$(K \widehat{\otimes} H)^{-1} = K^{-1} \widehat{\otimes} H^{-1}.$$

In general, given  $X \in \mathbb{R}^{q \times s}$ , it is possible to write  $\mathbf{x} \in \mathbb{R}^{q \cdot s}$  by sequentially stacking columns of  $X$  on top of each other. This operation is denoted as  $\mathbf{x} = (X)^\vee$ . It is easy to verify that

$$(HXK^T)^\vee = (K \widehat{\otimes} H)\mathbf{x} \tag{A.83}$$

where  $\widehat{\otimes}$  is the Kronecker product.

Since an  $n$ -dimensional column vector,  $\mathbf{a}$ , can be viewed as an  $n \times 1$  matrix, and its transpose is an  $n$ -dimensional row vector, it is a well-defined operation to “take the vector of a vector” as

$$\mathbf{a}^\vee = (\mathbf{a}^T)^\vee = \mathbf{a}.$$

Furthermore, for vectors  $\mathbf{a}$  and  $\mathbf{b}$  (not necessarily of the same dimensions) the following equalities hold [17]:<sup>5</sup>

$$(\mathbf{a}\mathbf{b}^T)^\vee = \mathbf{b} \widehat{\otimes} \mathbf{a} \tag{A.84}$$

and

$$(A^\vee)^T B^\vee = \text{tr}(A^T B) \tag{A.85}$$

where  $A$  and  $B$  have dimensions such that  $A^T B$  makes sense. Note that  $[(A^T)^\vee]^T \neq A^\vee$ . For example, in the  $2 \times 2$  case

$$A^\vee = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{pmatrix} \neq [(A^T)^\vee]^T = [a_{11}, a_{12}, a_{21}, a_{22}].$$

The *Kronecker sum* of two square matrices  $A \in \mathbb{R}^{m \times m}$  and  $B \in \mathbb{R}^{n \times n}$  is the matrix

$$A \widehat{\oplus} B = A \widehat{\otimes} \mathbb{I}_n + \mathbb{I}_m \widehat{\otimes} B$$

where  $\mathbb{I}_m$  is the  $m$ -dimensional identity matrix. This Kronecker sum is *not* the same as the direct sum  $A \oplus B$ . It does not even have the same dimensions. This should be clear since

$$I_m \widehat{\otimes} B = B \oplus B \oplus \dots \oplus B$$

(the  $m$ -fold direct sum of  $B$  with itself).

An interesting property of the Kronecker sum is that

$$\exp A \widehat{\oplus} B = e^A \widehat{\otimes} e^B.$$

Note the difference with (A.76).

Another useful application of the Kronecker sum is in solving equations of the form

$$AX + XB^T = C$$

for given  $A, B, C$  and unknown  $X$ , all of which are square. Application of the  $\vee$  operator as in (A.83) converts this to

$$(A \widehat{\oplus} B)\mathbf{x} = \mathbf{c} \implies \mathbf{x} = (A \widehat{\oplus} B)^{-1}\mathbf{c},$$

from which  $X$  can be obtained.

---

<sup>5</sup>Note that  $\mathbf{a}\mathbf{b}^T = \mathbf{a} \otimes \mathbf{b}$ , where  $\otimes$  is the tensor product discussed in Section 6.4. This is a good way to remember that  $\otimes \neq \widehat{\otimes}$ .

## A.8 Complex Numbers and Fourier Analysis

Throughout these volumes, complex numbers and Fourier analysis are used extensively. However, the following restrictions in scope apply: (1) only real vector spaces are used; (2) only real-valued functions (in particular, probability densities) are of concern; (3) only real Lie groups are studied. The reason for this limitation in scope is to avoid the intricacies associated with taking Taylor series of functions on complex Lie groups and other more mundane problems that can lead to significant confusion when introducing complex numbers. For example, in mathematical notation, the inner product of two complex-valued functions on the real interval  $[a, b]$  is

$$(f_1, f_2) = \int_a^b f_1(x) \overline{f_2(x)} dx,$$

whereas in physics the conjugate is over  $f_1(x)$  rather than  $f_2(x)$ .

But how is it possible to address problems of interest without using complex numbers? It is actually quite simple to circumvent the use of complex numbers. The fundamental properties of complex arithmetic revolve around the way complex numbers are added, multiplied, and conjugated. If  $c_1 = a_1 + ib_1$  and  $c_2 = a_2 + ib_2$ , then the sum and product of these two numbers, and the conjugation of  $c = a + ib$  are

$$c_1 + c_2 = (a_1 + a_2) + (b_1 + b_2)i; \quad c_1 \cdot c_2 = (a_1 a_2 - b_1 b_2) + (b_1 a_2 + a_1 b_2)i; \quad \bar{c} = a - ib \quad (\text{A.86})$$

where  $a_k, b_k \in \mathbb{R}$  and  $i = \sqrt{-1}$  and the usual scalar arithmetic operations are followed to produce the equalities in (A.86). In a sense, the simplicity of scalar operations is retained at the expense of adding the abstraction of  $i = \sqrt{-1}$ . Of course, on one level the very concept of  $\sqrt{-1}$  is absurd. But, the elegance and compactness of expressions such as

$$(e^{i\theta})^m = e^{im\theta} = \cos m\theta + i \sin m\theta \quad (\text{A.87})$$

make it worth accepting the concept. But this does not mean that the concept is *necessary* when doing the calculations in this book. Rather, in all of the problems addressed in this book, complex numbers are nothing more than a *convenient shorthand* for things that can be expressed as real quantities in higher dimensions.

For example, referring back to the properties in (A.86), rather than using complex numbers, it is possible to introduce the vectors  $\mathbf{c}_k = [a_k, b_k]^T \in \mathbb{R}^2$  with the properties

$$\mathbf{c}_1 + \mathbf{c}_2 = \begin{pmatrix} a_1 + a_2 \\ b_1 + b_2 \end{pmatrix}; \quad \mathbf{c}_1 \times \mathbf{c}_2 = \begin{pmatrix} a_1 a_2 - b_1 b_2 \\ b_1 a_2 + a_1 b_2 \end{pmatrix}; \quad \bar{\mathbf{c}} = \begin{pmatrix} a \\ -b \end{pmatrix}. \quad (\text{A.88})$$

Or, without introducing the operator  $\times$ , real  $2 \times 2$  matrices of the form

$$M(c) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

can be defined, and using only real matrix operations,

$$M(c_1)M(c_2) = M(c_1 + c_2); \quad M(c_1)M(c_2) = M(c_1 \cdot c_2); \quad [M(c)]^T = M(\bar{c}).$$

Taking this point of view, (A.87) is equivalent to

$$[R(\theta)]^m = R(m\theta); \quad \text{where } R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

An obvious place where complex notation seems required is in the discussion of unitary groups. After all, in Volume 2, when it comes time to discuss the concepts of inner products on Lie algebras, computation of Jacobians, adjoint matrices, Killing forms, etc., definitions are only provided for real Lie groups. Fortunately, elements of  $SU(n)$  can be identified with a subgroup of  $SO(2n)$ .

Given the unitary matrix  $U$ ,

$$U = A + iB \in \mathbb{C}^{n \times n}; \quad \text{where } A, B \in \mathbb{R}^{n \times n}$$

the constraint

$$UU^* = \mathbb{I} \implies AA^T + BB^T = \mathbb{I}; \quad BA^T - AB^T = \mathbb{O}.$$

If

$$R(U) = \begin{pmatrix} A & B \\ -B & A \end{pmatrix},$$

then

$$R(U_1 U_2) = R(U_1) R(U_2); \quad R(\mathbb{I}) = \mathbb{I} \oplus \mathbb{I}; \quad R(U^*) = [R(U)]^T$$

and so it is easy to see that a unitary matrix can be represented as a higher-dimensional orthogonal matrix.

But what about the Lie algebras and the exponential map? For example, if  $S = -S^*$  is a skew-Hermitian matrix, then  $U = \exp S$  will be special unitary. Letting  $S = W + iV$  where  $W$  and  $V$  are both real matrices, it becomes clear that  $W = -W^T$  and  $V = V^T$ . Therefore,

$$\Omega(S) = \begin{pmatrix} W & V \\ -V^T & W \end{pmatrix}$$

is skew-symmetric, and

$$\exp \Omega(S) = R(\exp(S)). \tag{A.89}$$

In fact, most of the Lie groups of practical interest consist of elements that are  $n \times n$  complex matrices that can be viewed instead as a group with elements that are  $(2n) \times (2n)$  real matrices.

This means that, for example, if it is desired to compute Jacobian matrices or invariant integration measures for groups such as  $SU(2)$ , this can be done without introducing an inner product on a complex vector space. Rather, these groups can be viewed as being equivalent to higher-dimensional groups of real matrices, and the computations demonstrated for the evaluation of Jacobians for real matrix groups can be applied here as well.

In summary, while complex numbers are used extensively throughout the text in the context of both classical and non-commutative Fourier expansions, the functions being expanded as well as the arguments of those functions can always be viewed as real, rather than complex, quantities. And while complex number notation leads to elegant simplifications in the way quantities are written, there is nothing necessary about their use in the class of problems discussed in this book. Taking this point of view, relatively simple tools such as the classical Taylor series for real-valued functions and operations on spaces of real-valued functions on real matrix Lie groups can be understood and applied without the extra effort required to master the theory of complex functions. Of course, this statement does not generalize to other areas of mathematics and its applications.

## A.9 Important Inequalities from the Theory of Linear Systems

In this section, three fundamental results from the theory of linear systems of ordinary differential equations are presented. Extensive use of the property

$$\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\| \quad (\text{A.90})$$

is made where  $\|\cdot\|$  is any vector norm and the corresponding induced matrix norm. Equation (A.90) also holds when  $\|A\|$  is the Frobenius norm of  $A$  and  $\|\mathbf{x}\|$  is the 2-norm of  $\mathbf{x}$ .

1. Let  $A \in \mathbb{R}^{n \times n}$  be a constant matrix and  $\mathbf{x}(t), \mathbf{g}(t) \in \mathbb{R}^n$  be vector-valued functions of time. The solution to

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x} + \mathbf{g}(t) \quad \text{with} \quad \mathbf{x}(0) = \mathbf{x}_0$$

is

$$\mathbf{x}(t) = \exp(At)\mathbf{x}_0 + \int_0^t \exp(A(t-\tau))\mathbf{g}(\tau)d\tau. \quad (\text{A.91})$$

From the properties of matrix norms, it is easy to see if  $0 > -a > \text{Re}(\lambda_i(A))$  for all values of  $i$  that for some positive constant scalar  $c$  the following holds:

$$\begin{aligned} \|\mathbf{x}(t)\| &= \|\exp(At)\mathbf{x}_0 + \int_0^t \exp(A(t-\tau))\mathbf{g}(\tau)d\tau\| \\ &\leq \|\exp(At)\mathbf{x}_0\| + \|\int_0^t \exp(A(t-\tau))\mathbf{g}(\tau)d\tau\| \\ &\leq \|\exp(At)\| \|\mathbf{x}_0\| + \int_0^t \|\exp(A(t-\tau))\| \|\mathbf{g}(\tau)\| d\tau \\ &\leq ce^{-at} \|\mathbf{x}_0\| + c \int_0^t e^{-a(t-\tau)} \|\mathbf{g}(\tau)\| d\tau. \end{aligned}$$

Hence, if  $\|\mathbf{g}(\tau)\| \leq \gamma$  for some scalar constant  $\gamma$ , it follows that the integral in the above expression will be less than

$$\gamma \int_0^t e^{-a(t-\tau)} d\tau = \gamma e^{-at} \int_0^t e^{a\tau} d\tau = \gamma e^{-at} (e^{at} - 1)/a \leq \gamma/a,$$

and hence  $\|\mathbf{x}(t)\|$  will be bounded. Likewise, if

$$f(t) = \int_0^t e^{a\tau} \|\mathbf{g}(\tau)\| d\tau$$

is bounded by a constant, then  $\|\mathbf{x}(t)\|$  will decay to zero as  $t \rightarrow \infty$ .

2. The *Bellman–Gronwall lemma* states that for any two functions  $u(t)$  and  $v(t)$  that are continuous on  $0 \leq t \leq \infty$ , and satisfy the inequality

$$u(t) \leq \alpha + \int_0^t v(s)u(s)ds \quad (\text{A.92})$$

for some  $\alpha > 0$  and  $t > 0$ , it must be the case that

$$u(t) \leq \alpha \exp\left(\int_0^t v(s)ds\right). \tag{A.93}$$

To prove this, first multiply both sides of (A.92) by  $v(t)$  and divide by  $\alpha + \int_0^t v(s)u(s)ds$  resulting in

$$\frac{u(t)v(t)}{\alpha + \int_0^t v(s)u(s)ds} \leq v(t).$$

Integrating both sides from 0 to  $t$  gives

$$\log\left[\alpha + \int_0^t v(s)u(s)ds\right] - \log \alpha \leq \int_0^t v(s)ds.$$

Adding  $\log \alpha$  to both sides and exponentiating preserves the inequality and results in (A.93).

3. The “solution” for the system

$$\frac{d\mathbf{x}}{dt} = (A + B(t))\mathbf{x} \quad \text{with} \quad \mathbf{x}(0) = \mathbf{x}_0$$

(where  $B(t) \in \mathbb{R}^{n \times n}$  is a matrix-valued function of time) is

$$\mathbf{x}(t) = \exp(At)\mathbf{x}_0 + \int_0^t \exp(A(t - \tau))B(\tau)\mathbf{x}(\tau)d\tau. \tag{A.94}$$

(Of course, this is not truly a solution because  $\mathbf{x}$  appears on both sides of the equation, but it is nonetheless a useful expression.) This can be used to write the following inequalities:

$$\begin{aligned} \|\mathbf{x}(t)\| &= \|\exp(At)\mathbf{x}_0 + \int_0^t \exp(A(t - \tau))B(\tau)\mathbf{x}(\tau)d\tau\| \\ &\leq \|\exp(At)\| \|\mathbf{x}_0\| + \|\int_0^t \exp(A(t - \tau))B(\tau)\mathbf{x}(\tau)d\tau\| \\ &\leq \|\exp(At)\| \|\mathbf{x}_0\| + \int_0^t \|\exp(A(t - \tau))\| \|B(\tau)\| \|\mathbf{x}(\tau)\| d\tau. \end{aligned}$$

If  $a > 0$  is a number that bounds from below the absolute value of the real part of all the eigenvalues of  $A$  such that

$$\|\exp(At)\| \leq ce^{-at},$$

then

$$\|\mathbf{x}(t)\| \leq ce^{-at} \|\mathbf{x}_0\| + c \int_0^t e^{-a(t-\tau)} \|B(\tau)\| \|\mathbf{x}(\tau)\| d\tau.$$

The Bellman–Gronwall lemma can be used as follows. Multiply both sides by  $e^{at}$  and let  $\alpha = c$ ,  $u(t) = \|\mathbf{x}(t)\|e^{at}$ , and  $v(t) = \|B(\tau)\|$ .

The Bellman–Gronwall lemma then indicates that systems for which  $A$  has eigenvalues all with negative real parts and, for example,

$$\int_0^\infty \|B(t)\| dt < \infty$$

or

$$\|B(t)\| < \beta$$

for a sufficiently small real number  $\beta$ , will be stable in the sense that  $\lim_{t \rightarrow \infty} \|\mathbf{x}(t)\| = 0$ .



## A.10 The State-Transition Matrix and the Product Integral

Given a general (possibly non-linear) scalar differential equation with initial conditions of the form

$$\frac{dx}{dt} = f(x, t) \quad \text{and} \quad x(0) = x_0, \quad (\text{A.95})$$

the simplest numerical integration scheme approximates the solution  $x(t)$  at regularly spaced increments of time,  $\Delta t$ , by replacing the derivative in (A.95) with the *finite-difference approximation*

$$\dot{x}(t) \approx \frac{1}{\Delta t} [x(t + \Delta t) - x(t)]$$

for  $0 < \Delta t \ll 1$ . Substituting this approximation into (A.95) results in a difference equation of the form

$$x_{n+1} = x_n + \Delta t f(x_n, t_n) \quad \text{where} \quad t_n = n\Delta t \quad (\text{A.96})$$

and  $x_n$  is the approximation to  $x(t_n)$  for  $n = 0, 1, 2, \dots$ . This scheme (which is called *Euler integration*) is known to diverge from the actual solution as  $n$  gets large. However, if  $n$  is relatively small, and as  $\Delta t \rightarrow 0$ , the approximation is not bad when  $f(x, t)$  is well behaved. And (A.96) has some convenient properties that will be used in the sequel.

In the special case when  $f(x, t) = a(t)x$ , and hence (A.95) is a scalar time-varying linear *ordinary differential equation (ODE)*, the exact solution can be written in closed form as

$$x(t) = x_0 \exp \int_0^t a(s) ds. \quad (\text{A.97})$$

How does the approximate solution in (A.96) compare with this? The statement

$$x_{n+1} = [1 + a(n\Delta t) \Delta t] x_n$$

is equivalent to

$$x_n = x_0 \prod_{k=1}^n [1 + a(k\Delta t) \Delta t].$$

If  $n \rightarrow \infty$  and  $\Delta t \rightarrow 0$  in such a way that  $0 < t_n < \infty$ , then

$$x_n \rightarrow x_0 \exp \left( \Delta t \sum_{k=1}^n a(k\Delta t) \right) = x_0 \exp \int_0^{t_n} a(\tau) d\tau = x(t_n).$$

In other words, the numerical integration scheme in (A.96) can be thought of as an algorithm that produces the correct solution under the special conditions described above.

Now consider the time-invariant system of  $n \times n$  matrix differential equations

$$\frac{dX}{dt} = AX \quad \text{where} \quad X(0) = X_0. \quad (\text{A.98})$$

The solution to this system is known to be of the form

$$X(t) = [\exp(At)] X_0. \quad (\text{A.99})$$

It is easy to show that this is a solution using the definition of the matrix exponential as a Taylor series, direct substitution into (A.98), and matching each term in the resulting

Taylor series. The fact that it is the only solution follows from existence and uniqueness theorems from the classical theory of ordinary differential equations.

In contrast, a numerical approximation of the solution to the system (A.98) can be made in analogy with the scalar case as

$$X_{n+1} = X_n + \Delta t A X_n = [\mathbb{I} + \Delta t A] X_n \approx \exp(\Delta t A) X_n.$$

The approximation above becomes accurate as  $\Delta t \rightarrow 0$ , and so the actual solution is obtained at the discrete sample points in that limit:

$$X_n = \left( \prod_{k=1}^n \exp(\Delta t A) \right) X_0 = \exp \left( \Delta t \sum_{k=1}^n A \right) X_0 = \exp(t_n A) X_0 = X(t_n).$$

Now if  $A = A(t)$  in (A.98), the solution will no longer be (A.99). However, if

$$\left[ A(t), \int_0^t A(s) ds \right] = 0 \tag{A.100}$$

for all values of  $t$  (where in this context the brackets mean  $[A, B] = AB - BA$ ), then

$$X(t) = \exp \left( \int_0^t A(s) ds \right) X_0. \tag{A.101}$$

And the same numerical scheme gives

$$X_{n+1} = \left[ \mathbb{I} + \int_{t_n}^{t_{n+1}} A(s) ds \right] X_n \tag{A.102}$$

and

$$\begin{aligned} X_n &= \left( \prod_{k=1}^n \exp \int_{t_{k-1}}^{t_k} A(s) ds \right) X_0 = \exp \left( \sum_{k=1}^n \int_{t_{k-1}}^{t_k} A(s) ds \right) X_0 \\ &= \exp \left( \int_0^{t_n} A(s) ds \right) X_0 = X(t_n). \end{aligned}$$

In the more general case when (A.100) does not necessarily hold, a unique solution will still exist, and it will be of the form

$$X(t) = \Phi(t, 0) X_0$$

where the *state transition matrix* can be thought of as the limit

$$\Phi(t_2, t_1) = \lim_{k \rightarrow \infty} \Phi_k(t_2, t_1) \quad \text{where} \quad \Phi_k(t_2, t_1) = \mathbb{I} + \int_{t_1}^{t_2} A(\sigma) \Phi_{k-1}(\sigma, t_1) d\sigma. \tag{A.103}$$

Or, written in a different way [6, 24],

$$\begin{aligned} \Phi(t_2, t_1) &= \mathbb{I} + \int_{t_1}^{t_2} A(\sigma_1) d\sigma_1 + \int_{t_1}^{t_2} A(\sigma_1) \left[ \int_{t_1}^{\sigma_1} A(\sigma_2) d\sigma_2 \right] d\sigma_1 \\ &\quad + \int_{t_1}^{t_2} A(\sigma_1) \left[ \int_{t_1}^{\sigma_1} \left[ \int_{t_1}^{\sigma_2} A(\sigma_3) d\sigma_3 \right] A(\sigma_2) d\sigma_2 \right] d\sigma_1 + \dots \tag{A.104} \end{aligned}$$

In addition to being called the state transition matrix, this is sometimes called the *matrizant* or *matricant*. It is easy to see that if the solution from  $t = t_0$  to  $t = t_1$  is  $X(t_1) = \Phi(t_1, t_0)X(t_0)$  where  $t_1 > t_0$ , and then if  $X(t_1)$  is used as the initial conditions for a solution evaluated at  $t_2 > t_1$ , then  $X(t_2) = \Phi(t_2, t_1)X(t_1)$ . It then follows that

$$X(t_2) = \Phi(t_2, t_1)\Phi(t_1, t_0)X(t_0) \implies \Phi(t_2, t_0) = \Phi(t_2, t_1)\Phi(t_1, t_0). \quad (\text{A.105})$$

On the other hand, even when (A.100) does not hold, (A.102) will still hold, and for very small values of  $\Delta t$  it is possible to write

$$X_{n+1} \approx \exp \left[ \int_{t_n}^{t_{n+1}} A(s) ds \right] X_n \quad (\text{A.106})$$

since (A.102) can be thought of as the first two terms in the expansion of the Taylor series for the matrix exponential, with all other terms being insignificant since they involve higher powers of  $\Delta t$ .

It follows from back substituting (A.102) into itself that

$$X_{n+1} \approx \exp \left[ \int_{t_n}^{t_{n+1}} A(s) ds \right] \exp \left[ \int_{t_{n-1}}^{t_n} A(s) ds \right] \dots \exp \left[ \int_{t_1}^{t_0} A(s) ds \right] X_0,$$

or equivalently,

$$X(t_{n+1}) = \prod_{k=0}^n \exp \left( \int_{t_k}^{t_{k+1}} A(s) ds \right) X_0 \quad (\text{A.107})$$

where the order in the products is understood to be in decreasing values of  $k$  written from left to right.

In the limit at  $\Delta t \rightarrow 0$  and  $n \rightarrow \infty$  in such a way that their product is the finite value,  $t$ , the above approximate solution becomes exact, and the corresponding state-transition matrix can be written as the *product integral* [1, 18, 30]:

$$\Phi(t, t_0) = \lim_{n \rightarrow \infty} \prod_{k=0}^n \exp \left( \int_{t_k}^{t_{k+1}} A(s) ds \right). \quad (\text{A.108})$$

Shorthand that is sometimes used for this is [19]<sup>6</sup>

$$\boxed{\Phi(t, t_0) \doteq \int_{t_0 < s < t} \exp[A(s) ds]}. \quad (\text{A.109})$$

Interestingly, even though  $\Phi(t, t_0)$  cannot be written as a single matrix exponential unless (A.100) holds, (A.79) still holds for the case when  $A = A(t)$ , and so

$$\det \Phi(t, t_0) = \exp \int_{t_0}^t \text{tr} A(s) ds.$$

The interpretations in (A.103) and (A.107) both have certain advantages from the perspective of numerical approximation. For example, if the system has many dimensions, then for small values of time, (A.103) has the advantage that the numerical

<sup>6</sup>The most rigorous definition of the product integral summarized by this shorthand is somewhat more involved than described here.

evaluation of the matrix exponential is not required. However, in the case when the system of equations describes a process evolving subject to constraints, then (A.109) will usually observe those constraints better than will (A.103).

For example, the constraint  $X^T(t)X(t) = \mathbb{I}$  indicates that  $X(t)$  is an orthogonal matrix. Taking the time derivative of both sides means that  $\dot{X}X^T$  must be skew-symmetric. Therefore, the system  $\dot{X} = A(t)X$  where  $A(t)$  is a specified skew-symmetric matrix and  $X(0)$  is orthogonal, should produce an  $X(t)$  that is orthogonal. However, if (A.103) is truncated at a finite number of nested integrals and used as an approximate solution, or if (A.102) is iterated with a finite value of  $\Delta t$ , errors will add up and the solution will no longer obey the constraint  $X^T(t_n)X(t_n) = \mathbb{I}$  as the value of  $t_n$  increases. In contrast, if  $A(s)$  is skew-symmetric, then each integral inside of each exponential in (A.107) will also be skew-symmetric. And since the exponential of a skew-symmetric matrix is always orthogonal, the solution obtained by the product integral will, in principle, obey the orthogonality constraint that the true solution should have, since the product of orthogonal matrices is orthogonal.

Now just because a numerical approximation observes constraints is not a guarantee that it is accurate. That is, it is a necessary property of a solution, but it is not a sufficient property. And furthermore, the product-integral formulation is not the only way to enforce constraints. For example, it is possible to use one of the other numerical approximations discussed above and incorporate a correction step that first generates an approximate solution, and then modifies that approximation so as to be consistent with the constraints. Another alternative would be to parameterize all possible states that are consistent with the constraints, and replace the original linear system of ODEs with a smaller number of non-linear ODEs in the parameter space, the solution of which will necessarily observe the constraints. A more detailed discussion of this is given in Volume 2.

## A.11 Vector Calculus

Vector calculus addresses differentiation and integration of functions and vector fields in analogy with the way calculus in one dimension works. The subsections that follow serve as a review of the basic results of vector calculus. Section A.11.1 begins with a review of basic optimization. Section A.11.2 focuses on differential operations in Euclidean space using Cartesian coordinates, and Section A.11.3 addresses the spatial generalizations of the fundamental theorem of calculus (i.e., the theorems named after Gauss, Green, and Stokes). Section A.11.4 discusses integration by parts in  $\mathbb{R}^n$ . Section A.11.5 reviews the chain rule. And Section A.11.6 serves as an introduction to matrix calculus.

### A.11.1 Optimization in $\mathbb{R}^n$

In this section optimization problems in  $\mathbb{R}^n$  are reviewed. Given a smooth function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , classical calculus provides the necessary and sufficient conditions for a particular point  $\mathbf{x}_0$  to be a *critical point* or *extremum* (i.e., a local minimum, local maximum, or saddle point). The necessary condition for a critical point is that the gradient vector evaluated at  $\mathbf{x}_0$  is zero:

$$\left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} = \mathbf{0}. \quad (\text{A.110})$$

Computing the matrix of second derivatives,  $\partial^2 f / \partial x_i \partial x_j$ , and examining its properties determines what kind of critical point  $\mathbf{x}_0$  is. In particular, the  $n \times n$  symmetric matrix

$$H(\mathbf{x}_0) = \left. \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^T} \right|_{\mathbf{x}_0}, \quad (\text{A.111})$$

which is called the *Hessian*, describes the shape of the local landscape in the neighborhood of  $\mathbf{x}_0$ . And the eigenvalues of  $H$  characterize the type of critical point. If all of the eigenvalues of  $H(\mathbf{x}_0)$  are positive, then  $\mathbf{x}_0$  is a local minimum. If all of the eigenvalues of  $H(\mathbf{x}_0)$  are negative, then  $\mathbf{x}_0$  is a local maximum. If they are mixed, then the result is a saddle.

Often in optimization problems, it is desirable to minimize a function subject to *constraints*. This changes the problem stated above from one of finding the critical points of  $f(\mathbf{x})$  where  $\mathbf{x}$  can take any value in the  $n$ -dimensional space,  $\mathbb{R}^n$ , to one of finding constrained extrema that are contained within an extrinsically defined hyper-surface  $h(\mathbf{x}) = 0$ . Sometimes multiple constraints of the form  $h_i(\mathbf{x}) = 0$  for  $i = 1, \dots, m < n$  are provided. When constraints are present, the goal becomes one of finding the values  $\mathbf{x}_0$  that extremize  $f(\mathbf{x})$  while exactly satisfying all of these  $h_i(\mathbf{x}) = 0$ . The combination of  $m$  constraint surfaces, when intersected in an  $n$ -dimensional space generally gives an  $(n - m)$ -dimensional manifold. If a parametrization  $\mathbf{x} = \mathbf{x}(\mathbf{u})$  can be found where  $\mathbf{u} \in \mathbb{R}^{n-m}$  such that  $h_i(\mathbf{x}(\mathbf{u})) = 0$  for  $i = 1, \dots, m$ , then the original problem can be reduced to the minimization of the function  $\tilde{f}(\mathbf{u}) = f(\mathbf{x}(\mathbf{u}))$ . While in general it can be difficult to find such a parametrization, it is still useful to think of the problem in this way. The necessary conditions for finding a constrained extremum are then  $\partial \tilde{f} / \partial u_i = 0$  for  $i = 1, \dots, n - m$ . By the chain rule (see (1.33)), this necessary condition is written as

$$\frac{\partial f}{\partial \mathbf{x}} \frac{\partial \mathbf{x}^T}{\partial \mathbf{u}} = \mathbf{0}.$$

If  $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \dots, h_m(\mathbf{x})]^T$ , then since  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ , it follows that  $\mathbf{h}(\mathbf{x}(\mathbf{u})) = \mathbf{0}$  and

$$\frac{\partial \mathbf{h}^T}{\partial \mathbf{x}} \frac{\partial \mathbf{x}^T}{\partial \mathbf{u}} = \mathbf{0}.$$

This says that  $\partial f / \partial \mathbf{x}$  and  $\partial \mathbf{h}^T / \partial \mathbf{x}$  are both orthogonal (or normal) to  $\partial \mathbf{x}^T / \partial \mathbf{u}$ , which describes the tangent to the constraints. Since the tangent space is  $m$ -dimensional, and the original space was  $n$ -dimensional, the fact that the  $n - m$  vectors  $\partial h_i / \partial \mathbf{x}$  are normal to the constraints means that they form a basis for all normal vectors. Since  $\partial f / \partial \mathbf{x}$  is one such normal vector, then it must be possible to expand it as a linear combination of the basis for normal vectors. Therefore, it must be possible to find constants,  $\lambda_i$  for  $i = 1, \dots, m$ , such that

$$\frac{\partial f}{\partial \mathbf{x}} = \sum_{i=1}^m \lambda_i \frac{\partial h_i}{\partial \mathbf{x}} \quad \text{or} \quad \frac{\partial f}{\partial \mathbf{x}} = \frac{\partial \mathbf{h}^T}{\partial \mathbf{x}} \boldsymbol{\lambda}. \quad (\text{A.112})$$

The constants  $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]^T$  are called *Lagrange multipliers*. The *method of Lagrange multipliers*, as stated in (A.112), enforces constraints while working in the original coordinates,  $\mathbf{x}$ . And while (A.112) was derived by assuming that the constraints could be described parametrically, the end result does not require that such a parametrization be found. For other explanations, see [16, 28].

In practice, when seeking to minimize a function  $f(\mathbf{x})$  subject to constraints  $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ , Lagrange multipliers are integrated into a modified cost function,  $c(\mathbf{x}) = f(\mathbf{x}) +$

$\boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x})$ . Then simultaneously solving  $\partial c / \partial \mathbf{x} = \mathbf{0}$  (which is (A.112)), and  $\partial c / \partial \boldsymbol{\lambda} = \mathbf{h}(\mathbf{x}) = \mathbf{0}$ , provides  $n + m$  equations to be solved in the  $n + m$  variables  $[\mathbf{x}^T, \boldsymbol{\lambda}^T]^T \in \mathbb{R}^{n+m}$ . This may seem wasteful since doing calculations in an  $(n + m)$ -dimensional space is more costly than in the  $(n - m)$ -dimensional space that would have resulted if parametric constraints had been obtained, but it is easier said than done to find parametric descriptions of constraints. And so, the method of Lagrange multipliers is widely used in practice.

### A.11.2 Differential Operators in $\mathbb{R}^n$

The gradient of a differentiable scalar function,  $\phi(\mathbf{x})$  for  $\mathbf{x} \in \mathbb{R}^n$ , is defined as the vector  $\text{grad}(\phi) \in \mathbb{R}^n$  with  $i$ th entry

$$\mathbf{e}_i \cdot \text{grad}(\phi) = \frac{\partial \phi}{\partial x_i}.$$

Usually the notation  $\text{grad}(\phi) = \nabla \phi$  is used as shorthand, where the “del” operator,  $\nabla$ , itself is viewed as a vector of the form

$$\nabla = \sum_{i=1}^n \mathbf{e}_i \frac{\partial}{\partial x_i}.$$

The *divergence* of a vector field (i.e., vector-valued function of vector-valued argument),  $\mathbf{f}(\mathbf{x})$ , is defined as

$$\text{div}(\mathbf{f}) \doteq \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} = \nabla \cdot \mathbf{f}. \quad (\text{A.113})$$

The *Laplacian* is defined as

$$\text{div}(\text{grad}(\phi)) \doteq \sum_{i=1}^n \frac{\partial^2 \phi}{\partial x_i^2} = \nabla \cdot (\nabla \phi). \quad (\text{A.114})$$

This is often denoted as  $\nabla^2 \phi$  or  $\Delta \phi$ .

In three-dimensional space, we can also define a *curl operator*,

$$\text{curl}(\mathbf{f}) = \nabla \times \mathbf{f} = \left[ \frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \right]^T. \quad (\text{A.115})$$

These operators play central roles in mechanics and theory of electromagnetism.

### A.11.3 Integral Theorems in $\mathbb{R}^2$ and $\mathbb{R}^3$

The Fundamental Theorem of Calculus states that for a differentiable real-valued function on the interval  $[a, b]$ ,  $f : [a, b] \rightarrow \mathbb{R}$ ,

$$\int_a^b \frac{df}{dx} dx = f(b) - f(a). \quad (\text{A.116})$$

A natural question to ask is, “How does this extend to higher dimensions?” Consider a region in the plane bounded by differentiable curves. Each curve can be assigned an

orientation defined such that the region is “always on the left” as a curve is traversed, as shown in Figure 5.4.

Let  $\mathbf{x}$  be a vector-valued function ( $\mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ) defined on the closure of the region (i.e., the union of its interior and the boundary curves). The value of the function on the oriented boundary curves is, in a sense, like the values  $f(a)$  and  $f(b)$  in (A.116). It should not be surprising, then, that evaluating the integral of “some kind of derivative” of  $\mathbf{f}(\mathbf{x})$  on the interior of the region should relate values evaluated on the boundary. In fact, the detailed relationship is written as

$$\int_{B \subset \mathbb{R}^2} \operatorname{div}(\mathbf{f}) dx_1 dx_2 = \sum_i \int_{C_i} \mathbf{f}(\mathbf{x}^{(i)}(s)) \cdot \mathbf{n}^{(i)}(s) ds \tag{A.117}$$

where  $i = 1, 2, \dots, m$  enumerates the boundary curves,  $C_1, \dots, C_m$ , each of which has a parametric description  $\mathbf{x}^{(i)}(s)$  with tangent  $\mathbf{t}^{(i)}(s)$  pointing along the direction of increase of  $s$  (which is a dummy variable of integration that need not be arc length). The unit normal that points away from the region at each boundary point is  $\mathbf{n}^{(i)}(s)$ , which by definition satisfies  $\mathbf{n}^{(i)}(s) \cdot \mathbf{t}^{(i)}(s) = 0$ .

Note that the right-hand side of (A.116) can be re-written as  $(+1)f(b) + (-1)f(a)$  where  $+1$  denotes the “direction” on the real line pointing away from the interval  $[a, b]$  when  $x = b$ , and  $-1$  points away from the interval when  $x = a$  and  $a < b$ . Hence, the issue of “orientation” was present even in the one-dimensional case.

By slightly changing notation, if we let  $\mathbf{f} = [f_1, f_2]^T = [h_2, -h_1]$ , and recognize that a vector  $\mathbf{v} \in \mathbb{R}^2$  can always be constructed to be orthogonal to a given vector  $\mathbf{u} \in \mathbb{R}^2$  as

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \pm \begin{pmatrix} -u_2 \\ u_1 \end{pmatrix} = \pm \begin{pmatrix} -\mathbf{u} \cdot \mathbf{e}_2 \\ \mathbf{u} \cdot \mathbf{e}_1 \end{pmatrix}, \tag{A.118}$$

then (A.117) is re-written as *Green’s theorem*:

$$\int_{B \subset \mathbb{R}^2} \left( \frac{\partial h_2}{\partial x_1} - \frac{\partial h_1}{\partial x_2} \right) dx_1 dx_2 = \sum_i \int_{C_i} \mathbf{f}(\mathbf{x}^{(i)}) \cdot d\mathbf{x}^{(i)} \tag{A.119}$$

where  $d\mathbf{x}^{(i)} = \mathbf{t}^{(i)} ds$ , and the construction (A.118) is used to relate the normal and tangent, with the choice  $\pm$  such that the curves are oriented as described earlier.

The spatial generalization of (A.117) is the *divergence theorem* (or *Gauss’ theorem*):

$$\int_B \operatorname{div}(\mathbf{f}) dx_1 dx_2 dx_3 = \sum_i \int_{\partial B_i} \mathbf{f}(\mathbf{x}^{(i)}) \cdot \mathbf{n}^{(i)} dS \tag{A.120}$$

where now  $\mathbf{x}_i$  ranges over each of the bounding surfaces of the domain (or “body”),  $B$ , with  $\mathbf{n}^{(i)}$  denoting the outward-pointing normal to the surface, and  $dS$  is an element of surface area.

The spatial generalization of (A.119) is *Stokes’ theorem*,

$$\sum_i \int_{S_i} \operatorname{curl}(\mathbf{f}) \cdot \mathbf{n}^{(i)} dS = \sum_j \int_{C_j} \mathbf{f} \cdot d\mathbf{x}^{(j)} \tag{A.121}$$

where  $j$  can be used to enumerate oriented curves on the exterior surface of the body. Collectively, these curves “carve out” oriented surface patches and strips. In contrast,  $i$  runs over all interior surfaces and the part of the exterior surface defined by the curves.

The extension of these theorems to higher dimensions (and even to non-Euclidean spaces such as group manifolds and homogeneous spaces) is possible. This is facilitated

by the use of differential forms, which are used in place of the cross product. Recall that the cross product was used to define the normal to a surface, and the element of surface area is also defined using the cross product. And so as written, (A.120) and (A.121) are limited to three-dimensional space.

Differential forms and the generalization of Stokes' theorem, Green's theorem, and the divergence theorem are discussed in Chapter 6.

#### A.11.4 Integration by Parts in $\mathbb{R}^n$

The inner product of two real-valued functions on the interval  $[a, b] \subset \mathbb{R}$  is defined as

$$(f, h) = \int_a^b f(x)h(x)dx. \quad (\text{A.122})$$

Clearly  $(f, h) = (h, f)$ .

Using the notation  $f' = df/dx$ , the familiar integration-by-parts formula,

$$\int_a^b f(x)h'(s)dx = f(b)h(b) - f(a)h(a) - \int_a^b h(x)f'(s)dx, \quad (\text{A.123})$$

can be written compactly as

$$(f, h') = fh|_a^b - (h, f').$$

Let  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $\mathbf{v} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Let  $D \in \mathbb{R}^n$  be a *compact (i.e., closed and bounded) domain* with smooth boundary  $\partial D$ . The integration-by-parts formula (A.123) generalizes to  $n$ -dimensional domains  $D \subset \mathbb{R}^n$  as

$$\int_D \sum_{i=1}^n \frac{\partial \phi}{\partial x_i} v_i dV = \int_{\partial D} \phi \sum_{i=1}^n v_i n_i dS - \int_D \sum_{i=1}^n \phi \frac{\partial v_i}{\partial x_i} dV,$$

which is written more compactly as

$$\int_D (\text{grad } \phi) \cdot \mathbf{v} dV = \int_{\partial D} \phi \mathbf{v} \cdot \mathbf{n} dS - \int_D \phi \text{div}(\mathbf{v}) dV, \quad (\text{A.124})$$

where  $\mathbf{n} \in \mathbb{R}^n$  is the unit outward-pointing normal to the bounding surface  $\partial D$ . Using the divergence theorem, (A.124) can be restated as

$$\int_D (\text{grad } \phi) \cdot \mathbf{v} dV = \int_D \{\text{div}(\phi \mathbf{v}) - \phi \text{div}(\mathbf{v})\} dV.$$

#### A.11.5 The Chain Rule

Given a scalar-valued function of multiple arguments,  $f(x_1, \dots, x_n)$ , a vector-valued function of a single argument,  $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$ , or a vector-valued function of vector-valued argument, classical multi-variate calculus addresses how rates of change of these quantities relate to each other. For example, the chain rule states that

$$\frac{df}{dt} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt}. \quad (\text{A.125})$$



Given the system of equations  $y_i = f_i(x_1, \dots, x_n)$  for  $i = 1, \dots, m$ ,

$$\boxed{\frac{dy_i}{dt} = \sum_{j=1}^n \frac{\partial f_i}{\partial x_j} \frac{dx_j}{dt}}. \quad (\text{A.126})$$

Treating the variables  $x_j$  and  $y_i$  as entries in a vector, the above can be written as

$$\frac{d\mathbf{y}}{dt} = J(\mathbf{x}) \frac{d\mathbf{x}}{dt}$$

where the  $m \times n$  *Jacobian matrix*  $J(\mathbf{x})$  is defined as

$$J(\mathbf{x}) = \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \iff J_{ij} = \frac{\partial f_i}{\partial x_j}.$$

The chain rule can be iterated. If  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ , and  $\mathbf{z} = \mathbf{g}(\mathbf{y})$ , then

$$\frac{d\mathbf{z}}{dt} = \frac{\partial \mathbf{z}}{\partial \mathbf{y}^T} \frac{d\mathbf{y}}{dt} = \frac{\partial \mathbf{z}}{\partial \mathbf{y}^T} \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \frac{d\mathbf{x}}{dt}.$$

Here the order of multiplication of the Jacobian matrices matters.

Using this notation, the gradient of a scalar function  $f = f(x_1, \dots, x_n)$  (which was defined earlier as a column vector) is written as

$$\text{grad} f = \frac{\partial f}{\partial \mathbf{x}} = \left( \frac{\partial f}{\partial \mathbf{x}^T} \right)^T.$$

The chain rule is an important tool that allows the differential operations in Section A.11.2 and integral theorems in Section A.11.3 to be expressed in different coordinate systems. For example, the expressions for divergence, curl, etc. take on different forms when expressed in cylindrical or spherical coordinates than in Cartesian coordinates. And likewise, the expressions for volume elements and surface area differ in appearance when expressed in various coordinate systems. The chain rule links these expressions.

### A.11.6 Matrix Differential Calculus

In some contexts, matrix-valued functions of a scalar, vector, or even another matrix arise in applications. For example, the state-transition matrix in Section A.10 gives the solution  $X(t) = \Phi(t, t_0)$  for the linear system  $dX/dt = A(t)X$  subject to given initial conditions  $X(t_0)$ .

Given two such matrix-valued functions of scalar argument,  $X(t)$  and  $Y(t)$ , and constant matrices  $A$  and  $B$  such that the products  $X(t)Y(t)$  and  $AX(t)B$  make sense, matrix differential calculus in one variable gives

$$\frac{d}{dt}(XY) = \frac{dX}{dt}Y + X \frac{dY}{dt} \quad \text{and} \quad \frac{d}{dt}(AXB) = A \frac{dX}{dt}B.$$

From the first of these expressions, the derivative of the inverse of a matrix can be computed from the fact that  $XX^{-1} = \mathbb{I}$  and  $d\mathbb{I}/dt = \mathbb{O}$  as

$$\frac{d}{dt}(XX^{-1}) = \frac{dX}{dt}X^{-1} + X \frac{dX^{-1}}{dt} = \mathbb{O} \iff \frac{dX^{-1}}{dt} = -X^{-1} \frac{dX}{dt} X^{-1}. \quad (\text{A.127})$$

In contrast to matrix-valued functions of scalar argument, it is also possible to have scalar-valued functions of matrix argument. For example, the functions  $\text{tr}(X)$ ,  $\det X$ , and  $\|X\|$  all take in matrices as their arguments and return scalars. The derivative of a real-valued function,  $\phi(X)$ , with respect to the  $(m \times n)$ -dimensional real matrix argument can be defined in terms of components as

$$\left[ \frac{\partial \phi}{\partial X} \right]_{ij} = \frac{\partial \phi}{\partial X_{ij}} \quad \text{for } (i, j) \in [1, \dots, m] \times [1, \dots, n]. \quad (\text{A.128})$$

Within the context of this notation, it is straightforward to show for constant matrices  $A$  and  $B$  that

$$\frac{\partial \text{tr} X}{\partial X} = \mathbb{I}; \quad \frac{\partial \text{tr}(A^T X)}{\partial X} = A; \quad \frac{\partial \text{tr}(AX^{-1})}{\partial X} = -(X^{-1}AX^{-1})^T \quad (\text{A.129})$$

and

$$\frac{\partial \text{tr}(XAXB)}{\partial X} = (AXB + BXA)^T; \quad \frac{\partial \text{tr}(XAX^T B)}{\partial X} = B^T X A^T + B X A. \quad (\text{A.130})$$

In addition to matrix-valued functions of scalar argument, and scalar-valued functions of matrix argument, the problem of matrix-valued functions of matrix-valued arguments sometimes arises in applications. While the system of scalar equations in multiple scalar variables of the form  $y_i = f_i(x_1, \dots, x_n)$  for  $i = 1, \dots, m$  was treated as a vector-valued function of vector-valued argument in Section A.11.5, if  $n = r \cdot s$  and  $m = p \cdot q$  for positive integers  $p, q, r, s$ , then rather than viewing this as the vector expression  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ , it could be viewed as the matrix expression  $Y = F(X)$  where  $\mathbf{x} = X^\vee$ ,  $\mathbf{y} = Y^\vee$ , and  $\mathbf{f} = F^\vee$  are the long column vectors obtained by stacking columns of the matrices to which the  $\vee$  operator is applied. And if two such expressions exist such that  $Y = F(X)$  and  $Z = G(Y)$ , it is natural to ask what the chain rule looks like. Several possible notations exist. One could define the Jacobian corresponding to the derivative of a matrix with respect to a matrix as a three-dimensional array; or one could treat the operator  $\partial/\partial X$  as a matrix (in analogy with the way  $\nabla$  is treated as a vector) and define a Jacobian as a Kronecker product  $(\partial/\partial X) \hat{\otimes} F$ . While these and other possible concepts exist, one that is very convenient in many applications is to convert  $X \in \mathbb{R}^{r \times s}$  and  $Y = F(X) \in \mathbb{R}^{p \times q}$  back to vectors and then use the definition [17]

$$DF(X) = \frac{\partial F^\vee}{\partial [X^\vee]^T} \in \mathbb{R}^{p \cdot q \times r \cdot s}.$$

Then if  $X = X(t)$ ,

$$\frac{dY^\vee}{dt} = DF(X) \frac{dX^\vee}{dt}$$

and the chain rule for concatenated transformations of the form  $Z = G(F(X))$  can be expressed simply as the matrix product

$$\frac{dZ^\vee}{dt} = DG(F(X))DF(X) \frac{dX^\vee}{dt}$$

without having to worry about how matrices multiply multi-dimensional arrays or any of the other problems associated with other definitions of Jacobians associated with matrix-valued functions of matrix argument.

For a more detailed treatment of matrix differential calculus, see [17].

## A.12 Exercises

A.1. Show that: (a) the “triple product”  $[\mathbf{a}, \mathbf{b}, \mathbf{c}] \doteq \det(\mathbf{a}, \mathbf{b}, \mathbf{c})$  has the property

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}). \quad (\text{A.131})$$

and (b) any vector  $\mathbf{v} \in \mathbb{R}^3$  can be expressed in terms of three non-coplanar vectors  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$  as

$$\mathbf{v} = \frac{\det(\mathbf{v}, \mathbf{b}, \mathbf{c})}{\det(\mathbf{a}, \mathbf{b}, \mathbf{c})} \mathbf{a} + \frac{\det(\mathbf{a}, \mathbf{v}, \mathbf{c})}{\det(\mathbf{a}, \mathbf{b}, \mathbf{c})} \mathbf{b} + \frac{\det(\mathbf{a}, \mathbf{b}, \mathbf{v})}{\det(\mathbf{a}, \mathbf{b}, \mathbf{c})} \mathbf{c}. \quad (\text{A.132})$$

A.2. Show that for a  $2 \times 2$  matrix,  $A$ , the 2-norm is related to the Frobenius norm and determinant of  $A$  as [14]

$$\|A\|_2^2 = \frac{1}{2} \left( \|A\|^2 + \sqrt{\|A\|^4 - 4|\det(A)|^2} \right).$$

A.3. Solve the above expression for  $\|A\|^2$  as a function of  $\|A\|_2^2$  and  $\det A$ .

A.4. Show that the 2-norm is the same as max eigenvalue

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^*A)}.$$

A.5. Show that for an  $n \times n$  matrix,  $A$ , with *characteristic polynomial*  $p(\lambda) = \det(\lambda\mathbb{I} - A) = \lambda^n - I_1(A)\lambda^{n-1} + \dots + (-1)^{n-1}I_{n-1}(A)\lambda + (-1)^n I_n(A) = 0$ , that  $I_1(A) = \text{tr}(A)$  and  $I_n(A) = \det(A)$ .

A.6. Using the properties of the trace and determinant, prove that if  $\det(P) \neq 0$ , then: (a)  $\det(PAP^{-1}) = \det(A)$  and (b)  $\text{tr}(PAP^{-1}) = \text{tr}(A)$ . (c) Is this true for all of the *scalar invariants*<sup>7</sup>  $I_k(A)$  in the characteristic polynomial  $p_A(\lambda) = 0$ ?

A.7. For the matrices

$$A_1 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad A_2 = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \quad A_3 = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

compute the following by hand, without the assistance of a software package: (a)  $\text{tr}(A_k)$ ; (b)  $\|A_k\|$ ; (c)  $\|A_k\|_2$ ; (d)  $\det(A_k)$ ; (e) all of the eigenvalues,  $\lambda_i(A_k)$ , and eigenvectors,  $\mathbf{v}_i(A_k)$ ; (f) show that  $(\mathbf{v}_i(A_k), \mathbf{v}_j(A_k)) = 0$  if  $\lambda_i(A_k) \neq \lambda_j(A_k)$ .

A.8. Compute by hand analytically the eigenvalues and eigenvectors of the following matrices: (a) the arbitrary real skew-symmetric matrix in (A.16) and (b) the arbitrary rotation around the  $\mathbf{e}_3$  axis denoted as  $R_3(\theta)$  in (A.44).

A.9. (a) Show that the cross product makes  $\mathbb{R}^3$  a Lie algebra. (b) Show that the set of skew-symmetric matrices together with the *matrix commutator*

<sup>7</sup>The characteristic polynomial for any matrix  $A \in \mathbb{C}^{n \times n}$  is defined as  $p_A(\lambda) = \det(\lambda\mathbb{I} - A)$ . It can be written in the form  $p_A(\lambda) = \lambda^n - I_1(A)\lambda^{n-1} + \dots + (-1)^k I_k(A)\lambda^{n-k} + \dots + (-1)^n I_n(A)$  where  $I_k(A)$  is called the  $k$ th scalar invariant of  $A$ . For example,  $I_1(A) = \text{tr}(A)$  and  $I_n(A) = \det(A)$ .

$$[A, B] = AB - BA \quad (\text{A.133})$$

is a Lie algebra. (c) Show that there is a bijective (one-to-one and onto) mapping between these two Lie algebras.

A.10. Prove (A.27). Hint: Start with the case when  $\det A \neq 0$ , and let  $f(B) = \det(AB)/\det A$ , and show that  $f(B)$  satisfies the three properties that are unique to the determinant.

A.11. Prove the sub-multiplicative property (A.50) for the Frobenius and induced 2-norm.

A.12. Determine whether or not the Frobenius and induced 2-norms are invariant under: (a) transformations of the form  $A \rightarrow UAV$  where  $U$  and  $V$  are unitary; (b) arbitrary similarity transformations of the form  $A \rightarrow PAP^{-1}$  where  $\det P \neq 0$ .

A.13. The norm  $\|A\|_\infty$  is sub-multiplicative and  $\|\mathbb{I}\|_\infty = 1$ , but is  $\|U\|_\infty \leq 1$  for  $U \in SU(n)$ ?

A.14. Show that: (a) As  $p \rightarrow \infty$  the norm in (A.58) approaches

$$\|A\|'_\infty = \max_{1 \leq i, j \leq n} |a_{ij}|$$

(b) The norm  $\|A\|'_\infty$  has the property  $\|\mathbb{I}\|'_\infty = 1$  and  $\|U\|'_\infty \leq 1$  for all  $U \in SU(n)$ .

(c) This norm lacks the sub-multiplicative property. If we define  $\|A\|''_\infty = n \cdot \|A\|'_\infty$ , will it become sub-multiplicative? Will  $\|\mathbb{I}\|''_\infty = 1$  and  $\|U\|''_\infty \leq 1$ ?

A.15. Show that if  $\|A\|'$  and  $\|A\|''$  are matrix norms, then

$$\|A\|''' = \max\{\|A\|', \|A\|''\} \quad (\text{A.134})$$

is also a matrix norm.

A.16. Show that if  $\|A\|'$  is a matrix norm, and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a function with the properties

$$\begin{aligned} \phi(x) &\geq 0 \quad \text{with} \quad \phi(x) = 0 \iff x = 0 \\ \phi(xy) &\leq \phi(x)\phi(y) \quad \forall \quad x, y \in \mathbb{R}_{>0} \\ \phi(ax + by) &\leq a\phi(x) + b\phi(y) \quad \forall \quad a, b \in \mathbb{R}_{>0}, \end{aligned} \quad (\text{A.135})$$

then  $\phi(\|A\|')$  is a matrix norm.

A.17. Use (A.27) and the fact that  $AA^{-1} = \mathbb{I}$  to prove (A.29) and (A.30).

A.18. Use the Jordan decomposition together with the results of Exercises A.10 and A.17 above to prove (A.28).

A.19. Using the proof of (A.35) as a guide, prove (A.37).

A.20. Prove that [14]

$$\left[ \frac{d}{dt} \det(A + tB) \right]_{t=0} = \det A \operatorname{tr}(A^{-1}B). \quad (\text{A.136})$$

Hint: First prove it for the case  $A = \mathbb{I}$  by observing the part of the function  $f(t) = \det(\mathbb{I} + tB)$  that is linear in the parameter  $t$ .

A.21. Verify the following vector identities for arbitrary  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^3$ :

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c} \quad (\text{A.137})$$

$$(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d}) = \det(\mathbf{a}, \mathbf{c}, \mathbf{d})\mathbf{b} - \det(\mathbf{b}, \mathbf{c}, \mathbf{d})\mathbf{a} \quad (\text{A.138})$$

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}) \quad (\text{A.139})$$

A.22. Show that for all matrices  $A, B, C$  of compatible dimensions, the Kronecker product satisfies

$$\begin{aligned} A \widehat{\otimes} (B + C) &= A \widehat{\otimes} B + A \widehat{\otimes} C \\ (A + B) \widehat{\otimes} C &= A \widehat{\otimes} C + B \widehat{\otimes} C \\ (\alpha A) \widehat{\otimes} B &= A \widehat{\otimes} (\alpha B) = \alpha (A \widehat{\otimes} B) \\ (A \widehat{\otimes} B) \widehat{\otimes} C &= A \widehat{\otimes} (B \widehat{\otimes} C). \end{aligned}$$

A.23. Show that in general if  $A \in \mathbb{R}^{m \times m}$  and  $B \in \mathbb{R}^{n \times n}$ , then

$$\text{tr}(A \widehat{\otimes} B) = \text{tr} A \text{tr} B \quad \text{and} \quad \det(A \widehat{\otimes} B) = \det A^n \det B^m.$$

A.24. Verify the following for matrices of compatible dimensions [17]:

$$\begin{aligned} (ABC)^\vee &= (C^T \widehat{\otimes} A) B^\vee \\ \text{tr}(ABCD) &= [(D^T)^\vee]^T (C^T \widehat{\otimes} A) B^\vee = [D^\vee]^T (A \widehat{\otimes} C^T) (B^T)^\vee. \end{aligned}$$

A.25. Assuming  $A \in \mathbb{R}^{N \times N}$  and  $P \in GL(N, \mathbb{R})$  (i.e.,  $P \in \mathbb{R}^{N \times N}$  and it is invertible), show the following:

- (a)  $\exp(P^{-1}AP) = P^{-1} \exp(A)P$ ;
- (b)  $\exp(-At) = [\exp(At)]^{-1}$ .

A.26. The *Cayley–Hamilton theorem* states that any matrix satisfies its own characteristic polynomial (i.e., if  $p(\lambda) = 0$ , then  $p(\mathbf{A}) = \mathbf{0}$ ). Use the Cayley–Hamilton theorem to compute by hand a closed-form expression for  $\exp(tA)$  where

$$A = \begin{pmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{pmatrix}.$$

A.27. Determine the stability of the following system:

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{pmatrix} -2 & 1 & 2 \\ 0 & -1 & 6 \\ 0 & 0 & -3 \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} e^{-2t} \\ \cos 12t \\ 0 \end{bmatrix}.$$

A.28. Let  $A \in \mathbb{R}^{n \times n}$  be a constant matrix and  $\mathbf{x}(t), \mathbf{g}(t) \in \mathbb{R}^n$  be vector-valued functions of time. The solution to

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x} + \mathbf{g}(t) \quad \text{with } \mathbf{x}(0) = \mathbf{x}_0$$

is

$$\mathbf{x}(t) = \exp(At)\mathbf{x}_0 + \int_0^t \exp(A(t-\tau))\mathbf{g}(\tau)d\tau. \quad (\text{A.140})$$

Similarly, given

$$\frac{d\mathbf{x}}{dt} = (A + B(t))\mathbf{x} \quad \text{with } \mathbf{x}(0) = \mathbf{x}_0$$

(where  $B(t) \in \mathbb{R}^{n \times n}$  is a matrix-valued function of time) it is possible to write

$$\mathbf{x}(t) = \exp(At)\mathbf{x}_0 + \int_0^t \exp(A(t-\tau))B(\tau)\mathbf{x}(\tau)d\tau. \quad (\text{A.141})$$

(a) Prove Equation (A.91).

(b) Prove Equation (A.94).

A.29. Use Equation (A.91) and/or Equation (A.94) and/or the Bellman–Gronwall inequality to determine the behavior of  $x(t)$  governed by the following equations as  $t \rightarrow \infty$ :

(a)  $\ddot{x} + \dot{x} + (1 + e^{-t})x = 0$

(b)  $\ddot{x} + \dot{x} + (1 + 0.2 \cos t)x = 0$

(c)  $\ddot{x} + \dot{x} + x = \cos t$

(d)  $\ddot{x} + \dot{x} + x = e^{-t}$

(e)  $\ddot{x} + \dot{x} + x = e^{2t}$ .

Hint: Rewrite the above second-order differential equations as a system of first-order differential equations in terms of the vector  $\mathbf{x}(t) = [x_1(t), x_2(t)]^T$  where  $x_1 = x$  and  $x_2 = \dot{x}$ .

A.30. Verify the following well-known formulas, assuming  $\phi_i(\mathbf{x})$ ,  $\mathbf{f}(\mathbf{x})$ , and  $\mathbf{g}(\mathbf{x})$  are sufficiently differentiable [5, 25]:

$$\nabla(\phi_1\phi_2) = \phi_1\nabla\phi_2 + \phi_2\nabla\phi_1 \quad (\text{A.142})$$

$$\nabla \cdot (\phi \mathbf{f}) = \phi \nabla \cdot \mathbf{f} + \mathbf{f} \cdot \nabla \phi \quad (\text{A.143})$$

$$\nabla \times (\phi \mathbf{f}) = \phi \nabla \times \mathbf{f} + (\nabla \phi) \times \mathbf{f} \quad (\text{A.144})$$

$$\nabla \times (\nabla \phi) = \mathbf{0} \quad (\text{A.145})$$

$$\nabla \cdot (\nabla \times \mathbf{f}) = 0 \quad (\text{A.146})$$

$$\nabla \cdot (\nabla \phi_1 \times \nabla \phi_2) = 0 \quad (\text{A.147})$$

$$\nabla \cdot (\mathbf{f} \times \mathbf{g}) = \mathbf{g} \cdot (\nabla \times \mathbf{f}) - \mathbf{f} \cdot (\nabla \times \mathbf{g}) \quad (\text{A.148})$$

$$\nabla^2(\phi_1\phi_2) = \phi_1\nabla^2\phi_2 + 2(\nabla\phi_1) \cdot (\nabla\phi_2) + \phi_2\nabla^2\phi_1 \quad (\text{A.149})$$

$$\nabla \cdot (\phi_1 \nabla \phi_2) = \phi_1 \nabla^2 \phi_2 + \nabla \phi_1 \cdot \nabla \phi_2 \quad (\text{A.150})$$

$$\nabla \times (\nabla \times \mathbf{f}) = \nabla(\nabla \cdot \mathbf{f}) - \nabla^2 \mathbf{f} \quad (\text{A.151})$$

where  $\nabla^2 \mathbf{f} \doteq \sum_{i=1}^3 \mathbf{e}_i \nabla^2 f_i$ .

A.31. Given the Hermitian positive definite matrix  $A$ , prove that

$$|\mathbf{x}^* \mathbf{A} \mathbf{y}|^2 \leq (\mathbf{x}^* \mathbf{A} \mathbf{x})(\mathbf{y}^* \mathbf{A} \mathbf{y}) \quad (\text{A.152})$$

with equality if and only if  $\mathbf{x} = c\mathbf{y}$  for some  $c \in \mathbb{C}$ .

A.32. Prove the equalities in (A.129).

A.33. Prove the equalities in (A.130).

## References

1. Birkhoff, G., "On product integration," *J. Math. Phys.*, 16, pp. 104–132, 1937.
2. Boyce, W.E., DiPrima, R.C., *Elementary Differential Equations and Boundary Value Problems*, 7th ed., John Wiley & Sons, New York, 2001.
3. Campbell, S.L., Meyer, C.D., Jr., *Generalized Inverses of Linear Transformations*, Pitman, London, 1979.
4. Chirikjian, G.S., "Kinematic synthesis of mechanisms and robotic manipulators with binary actuators." *ASME J. Mech. Design*, 117, pp. 573–580, 1995.
5. Davis, H.F., Snider, A.D., *Introduction to Vector Analysis*, 7th ed., William C. Brown, Dubuque, IA, 1995.
6. Davis, J.H., *Foundations of Deterministic and Stochastic Control*, Birkhäuser, Boston, 2002.
7. Edwards, H.M., *Linear Algebra*, Birkhäuser, Boston, 1995.
8. Elbert, T.F., *Estimation and Control of Systems*, Van Nostrand Reinhold, New York, 1984.
9. Fréchet, M., "Les Solutions Non Commutables De L'Équation Matricielle  $e^X \cdot e^Y = e^{X+Y}$ ," *Rend. Circ. Mat. Palermo*, 1, pp. 11–21, 1952 (also Vol. 2, pp. 71–72, 1953).
10. Gantmacher, F.R., *The Theory of Matrices*, Chelsea, New York, 1959.
11. Golub, G.H., Van Loan, C.F., *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, 1989.
12. Halmos, P., *Finite Dimensional Vector Spaces*, Springer, New York, 1974.
13. Horn, R.A., Johnson, C.R., *Matrix Analysis*, Cambridge University Press, New York, 1985.
14. Hubbard, J.H., Burke-Hubbard, B., *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*, Prentice-Hall, Upper Saddle River, NJ, 1999.
15. Kanatani, K., *Statistical Optimization for Geometric Computation*, Dover, New York, 2005.
16. Larson, R., Hostetler, R.P., Edwards, B.H., *Calculus*, 8th ed., Houghton Mifflin, Boston, 2005.
17. Magnus, J.R., Neudecker, H., *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd ed., John Wiley & Sons, New York, 1999.
18. Magnus, W., "On the exponential solution of differential equations for a linear operator," *Comm. Pure Appl. Math.*, 7, pp. 649–673, 1954.
19. McKean, H.P., Jr. *Stochastic Integrals*, Academic Press, New York, 1969 (reissued in 2005 by the AMS).
20. Pan, V., "How can we speed up matrix multiplication," *SIAM Rev.*, 26, pp. 393–416, 1984.
21. Pan, V., *How to Multiply Matrices Fast*, Springer-Verlag, Berlin, 1984.
22. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., *Numerical Recipes in C*, Cambridge University Press, New York, 1988.

23. Rudin, W., *Functional Analysis*, McGraw-Hill, New York, 1991.
24. Rugh, W.J., *Linear Systems Theory*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1995.
25. Schey, H.M., *Div, Grad, Curl, and All That: An Informal Text on Vector Calculus*, 4th ed., W.W. Norton, New York, 2005.
26. So, W., "Equality cases in matrix exponential inequalities," *SIAM J. Matrix Anal. Appl.*, 13, pp. 1154–1158, 1992.
27. Stewart, G.W., Sun, J.-Q., *Matrix Perturbations*, Academic Press, San Diego, 1990.
28. Thomas, G.B., Jr., Finney, R.L., *Thomas' Calculus*, 10th ed., Addison Wesley, Reading, MA, 2000.
29. Wedderburn, J.H.M., *Lectures on Matrices*, Dover, New York, 1964.
30. Wei, J., Norman, E., "On global representations of the solutions of linear differential equations as a product of exponentials," *Proc. Amer. Math. Soc.*, 15, pp. 327–334, 1964.
31. Wermuth, E.M.E., Brenner, J., Leite, F.S., Queiro, J.F., "Computing matrix exponentials," *SIAM Rev.*, 31, pp. 125–126, 1989.



---

# Index

- $A = [a_{ij}]$  (See matrix), 321  
 $(U, \phi)$  (See coordinate chart), 267  
 $(f_1 * f_2)(x)$  (See convolution), 36  
 $(l \circ n)(\cdot)$  or  $l(n(\cdot))$  (See composed mapping), 13  
 $(\mathbf{x}, \mathbf{y})$  (See inner product), 319  
 $*\omega$  (See Hodge star operator (applied to differential form  $\omega$ )), 224  
 $+$  (See addition), 315  
 $=$  (See equality), 4  
 $A^*$  (See Hermitian conjugate (of  $A$ )), 322  
 $A^{-1}$  (See inverse of a matrix), 325  
 $B$  (See body), 178  
 $B_1 \times B_2$  (See Cartesian product of two bodies), 179  
 $C^0(S^1)$  (See continuous functions on the circle), 6  
 $C^n(a, b)$  (See  $n$ -times continuously differentiable function on the interval  $(a, b)$ ), 5  
 $D_{FI}(f_1 \parallel f_2)$  (See Fisher information divergence), 76  
 $D_{KL}(f_1 \parallel f_2)$  (See Kullback–Leibler divergence), 71, 76  
 $D_{ij}(t)$  (See time-varying diffusion constants), 50  
 $D\mathbf{f}, \frac{\partial \mathbf{f}}{\partial \mathbf{x}^T}, J$  (See Jacobian (of  $\mathbf{f}$ )), 18  
 $E[\mathbf{X}]$  (See expected value (of  $\mathbf{X}$ )), 64  
 $E^n$  (See  $n$ -dimensional Euclidean space), 8  
 $F(\boldsymbol{\theta}; f)$  (See Fisher information matrix), 78  
 $F(f)$  (See Fisher information, (3.40)), 78  
 $G$  (See metric tensor, or in other contexts, group), 161  
 $G \setminus \mathbb{R}^n$  (See unit cell), 244  
 $I(f_1, f_2; f)$  (See mutual information), 76  
 $J(\mathbf{x})$ , also denoted as  $D\mathbf{y}$  and  $\partial \mathbf{y} / \partial \mathbf{x}^T$  (See Jacobian matrix), 354  
 $J_k(\lambda)$  (See Jordan block), 334  
 $K$  (See total Gaussian curvature), 166  
 $L^2(S^1)$  (See square-integrable functions on the circle), 6  
 $L^p(S^1)$  (See functions on the circle, the integral of the  $p$ th power of which is finite), 6  
 $M$  (See total mean Gaussian curvature), 166  
 $N(p)$  (See entropy power), 85  
 $O(g(x))$  (See Big-O notation), 6  
 $R_{ijk}^l$  (See Riemannian curvature tensor), 165  
 $S(f)$  (See entropy), 34  
 $S(f_{X|Y}; f)$  (See conditional entropy), 74  
 $S(f_Y; f)$ , or  $S(f_Y)$  (See marginal entropy), 74  
 $SE(2)$  (See special Euclidean group of the plane), 239  
 $SO(3)$  (See special orthogonal group), 240  
 $SO(n)$  (See special orthogonal  $n \times n$  matrices), 328  
 $SU(n)$  (See special unitary  $n \times n$  matrices), 328  
 $S^1$  (See unit circle), 5, 6  
 $S^{n-1}$  (See hypersphere in  $\mathbb{R}^n$ ), 5  
 $S^{n-1}$  (See unit sphere in  $n$ -dimensional space), 6  
 $S_1 \times S_2$  (See Cartesian product (of sets  $S_1$  and  $S_2$ )), 15  
 $S_r^n$  (See sphere of radius  $r$  in  $(n+1)$ -dimensional space), 6  
 $T^3$  (See three-torus), 244  
 $T_x \mathbb{R}^n$  (See tangent space to  $\mathbb{R}^n$  at the point  $\mathbf{x}$ ), 213  
 $V$  (See volume, or in other contexts, vector space), 167  
 $V(C)$  (See volume of a convex body,  $C$ ), 235  
 $V^*$  (See dual (of the vector space  $V$ )), 206, 320  
 $Vol(B^n)$  (See volume of a unit ball in  $\mathbb{R}^n$ ), 45  
 $Vol(S^{n-1})$  (See volume of unit sphere in  $\mathbb{R}^n$ ), 44, 45

- $[g^{ij}]$  (See inverse of the metric tensor), 161  
 $[g_{ij}]$  (See metric tensor), 161  
 $[x, y]$  (See Lie bracket), 321  
 $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$  (See triple product), 320  
 $\Gamma(\alpha)$  (See Gamma function), 44  
 $\Gamma_{ij}^k$  (See Christoffel symbol), 164  
 $\mathbb{I}$  (See identity matrix), 26  
 $\mathbb{R}$  (See real numbers), 4  
 $\mathbb{R}^n$  (See  $n$ -dimensional real vector space), 8, 315  
 $\mathbb{Z}$  (See integers), 102  
 $\Lambda^p V$ , or  $\Lambda^p(V)$  (See  $p$ th exterior power of  $V$ ), 209  
 $\iff$  (See iff, or “if and only if”), 11  
 $\implies$  (See “implies”), 12  
 $\mathbb{O}$  (the zero matrix), 53  
 $\Omega^k(U)$  (See differential  $k$ -forms on  $U$ ), 267  
 $\Omega^k(\mathbb{R}^n)$  (See set of differential  $k$ -forms on  $\mathbb{R}^n$ ), 197  
 $\Omega^j$  (See torsion 2-form), 278  
 $\Omega_i^j$  (See curvature 2-form), 278  
 $\Phi(\mathbf{x})$  (See convex function), 71  
 $\Phi(t_2, t_1)$  (See state transition matrix), 347  
 $\Pi_n$  (See permutation, or symmetric, group), 201, 323  
 $\Sigma$  (See covariance matrix), 40, 66  
 $\binom{n}{k}$  (See binomial coefficients), 39  
 $\mu$  (See mean), 40  
 $\prod_{t_0 < s < t}$  (See product integral), 348  
 $\cap$  (See intersection), 12  
 $\cdot$  (See scalar product), 317  
 $\chi(B)$  (See Euler characteristic (of a body  $B$ )), 176  
 $\chi(S)$  (See Euler characteristic (of closed surface  $S$ )), 170  
 $\chi(\partial B)$  (See Euler characteristic (for the boundary of body  $B$ )), 159, 178  
 $\cup$  (See union), 12  
 $\delta_{ij}$  (See Kronecker delta function), 102, 203  
 $\doteq$  (See defining equality), 4  
 $\emptyset$  (See empty set), 12  
 $\epsilon_{ijk}$  (See Levi-Civita symbol), 203  
 $\equiv$  (See identically equal), 4  
 $\exists$  (See “there exists”), 316  
 $\exp B$  (See exponential (of matrix  $B$ )), 338  
 $\forall$  (See “for all”), 13, 316  
 $\frac{d^2 f}{dx^2}$  (See the second derivative of  $f(x)$ ), 5  
 $\frac{df}{dx}$  (See the derivative of  $f(x)$ ), 5  
 $\gamma(S)$  (See genus (of closed surface  $S$ )), 170  
 $\hat{f}(\omega)$  (See Fourier transform of  $f(x)$ ), 37  
 $\hat{f}(n)$  (See Fourier coefficient of  $f(\theta)$ ), 48  
 $\in$  (See “in”), 12  
 $\kappa(s)$  (See unsigned (or absolute) curvature (of an arc length-parameterized curve)), 155  
 $\kappa(t)$  (See unsigned curvature (of an arbitrarily parameterized curve)), 156  
 $\kappa_g(s)$  (See geodesic curvature), 163  
 $\kappa_n(s)$  (See normal curvature), 163  
 $\langle \alpha(\mathbf{x}) \rangle$  (See ensemble average (of  $\alpha(\mathbf{x})$ )), 65  
 $\langle \phi(x) | y \rangle$  (See conditional expectation), 71  
 $\longrightarrow$  (See “goes to”), 12  
 $\mu(S_i)$  (See valuation (on  $S_i$ )), 16  
 $\nabla^2$  (See Laplacian), 351  
 $\oint$  (See integral around a closed curve), 157  
 $\omega_1$  (See differential 1-form), 194  
 $\omega_2$  (See differential 2-form), 195  
 $\oplus$  (See direct sum), 325  
 $\otimes$  (See tensor product), 206  
 $\partial B$  (See boundary (of body  $B$ )), 158, 178  
 $\phi(\mathbf{x}) = 0$  (See implicit surface), 180  
 $\pi$  (See a permutation), 201  
 $\rho(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$  (See Gaussian distribution, multivariate), 40  
 $\rho(x; \mu, \sigma^2)$  (See Gaussian distribution, univariate), 34  
 $\rho_W(\theta; \mu, \sigma)$  (See Gaussian distribution wrapped around the circle), 47  
 $\mathbb{R}_{>0}$  (See positive real numbers), 14, 15  
 $\mathbb{R}_{\geq 0}$  (See non-negative real numbers), 14, 15  
 $\sim$  (See equivalence relation), 12  
 $\subset$  (See strictly contained), 12  
 $\subseteq$  (See contained, or subset), 12  
 $\tau(t)$  (See torsion (of an arbitrarily parameterized curve)), 156  
 $\times$  (See cross product), 320  
 $\|\cdot\|$  (See sub-multiplicative matrix norm), 329  
 $\mathbf{u}(t)$  (See unit tangent vector (to a curve)), 155  
 $\varphi(\mathbf{v})$  (See dual vector), 206  
 $\hat{\oplus}$  (See Kronecker sum), 341  
 $\hat{\otimes}$  (See Kronecker product), 340  
 $\mathbf{x}^T$  (See transpose (of  $\mathbf{x}$ )), 319  
 $\|A\|$  (See Frobenius norm (of matrix  $A$ )), 328  
 $\|A\|_p$  (See induced norm (of matrix  $A$ )), 329  
 $\|\mathbf{x}\|$  (See vector norm), 318  
 $\|\mathbf{x}\|_p$  (See  $p$ -norm), 319  
 $d(\mathbf{x}) = dx_1 dx_2 \cdots dx_n$ , or  $d\mathbf{x}$  for short (See differential volume element for  $\mathbb{R}^n$ ), 40  
 $d(d\omega) = 0$  (See (double) exterior derivative of a  $k$ -form), 199  
 $d(s_1, s_2)$  (See metric, or distance, function), 15  
 $dA$  (See differential area element), 20  
 $dS$  (See element of surface area), 161

- $dV$  (See differential volume element), 20  
 $d\omega$  (meaning 1: differential element in frequency space), 37  
 $d\omega$  (meaning 2: exterior derivative (of the differential form  $\omega$ )), 196  
 $d\mathbf{x}$  (meaning 1: vector with infinitesimal length), 18  
 $d\mathbf{x}$  (meaning 2: infinitesimal volume element  $dx_1 dx_2 \cdots dx_n$ ), 19  
 $dw_i(t) = w_i(t+dt) - w_i(t)$  (See increments of a Wiener process, or white noise), 109  
 $dx_1 \wedge dx_2$  (See wedge product (of  $dx_1$  and  $dx_2$ )), 195  
 $dx_i(t)$ ,  $i$ th entry of  $d\mathbf{x}(t) = \mathbf{x}(t+dt) - \mathbf{x}(t)$ ,  
 Not to be confused with the differential used in integration (See (4.38)), 114  
 $f(\mathbf{x}_1 | \mathbf{x}_2)$  (See conditional density), 42  
 $f(k; n, p)$  (See binomial distribution), 39  
 $f(x) = O(g(x))$  (See Big-O notation), 7  
 $f(x; a)$  (See family of functions (parameterized by  $a$ )), 8  
 $k(q_1, q_2)$  (See Gaussian curvature (of a surface)), 165  
 $k(s)$  (See signed curvature (of an arc-length-parameterized curve)), 155  
 $m(q_1, q_2)$  (See mean (sectional) curvature of a surface), 166  
 $p(\mathbf{x}, t; \mathbf{x}_2, t_2; \mathbf{x}_3, t_3; \dots; \mathbf{x}_n, t_n)$  (See random process, pdf for a), 105  
 $p(\mathbf{x} | \mathbf{y}, t)$  (See transition probability), 121  
 $w_i(t)$  (See Wiener process), 108  
 $R_i(\phi)$  (See a fundamental rotation), 328  
 $T^2$  (See 2-torus), 171  
 $\mathbf{e}_k$  (See  $k$ th natural basis vector for  $\mathbb{R}^n$ ), 316  
 $\mathbf{n}_1(s)$  (See principal normal (for a space curve)), 155  
 $\mathbf{n}_2(s)$  (See binormal (of a space curve)), 156  
 $\mathbf{o}(s; r)$  (See offset curve), 171  
 $\mathbf{o}(t_1, t_2; r)$  (See offset surfaces, or tubes), 175  
 $\mathbf{w}(t)$  (See Wiener process,  $m$ -dimensional), 111  
 $\mathcal{F}(f)$  (See Fourier transform of  $f$ ), 37  
 $\mathcal{F}^{(1)}(d\mathbf{q}, d\mathbf{q})$  (See first fundamental form), 161  
 $\mathcal{F}^{(2)}(d\mathbf{q}, d\mathbf{q})$  (See second fundamental form), 162  
 $\mathcal{F}^{-1}(\hat{f})$  (See inverse Fourier transform), 37  
 $\mathcal{N}(D)$  (See space of “nice functions” on domain  $D$ ), 9  
 $\mathcal{O}_n$  (See volume of a (unit) sphere in  $\mathbb{R}^n$ ), 45  
 $\mathcal{P}(R)$  (See Pfaffian), 278  
 $\mathfrak{X}(M)$  (See set of smooth vector fields on  $M$ ), 281  
 $\text{curl}(\mathbf{f})$  (See curl operator (applied to  $\mathbf{f}$ )), 351  
 $\det(A)$  (See determinant (of matrix  $A$ )), 323  
 $\text{div}(\mathbf{f})$  (See divergence (of  $\mathbf{f}$ )), 162, 351  
 $\text{div}(\text{grad}f)$  (See Laplacian (of  $f$ )), 162  
 $\text{erf}(x)$  (See error function), 46  
 $\text{grad}(\phi)$ ,  $\nabla_{\mathbf{x}}\phi$ , or  $\nabla\phi$ , or  $\frac{\partial\phi}{\partial\mathbf{x}}$  (See gradient (of  $\phi(\mathbf{x})$ )), 18, 25, 161, 180  
 $\text{sgn}(\pi)$  (See signature (of permutation  $\pi$ )), 201, 323  
 $\text{span}\{\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_m}\}$  (See span of a set of vectors), 319  
 $\text{trace}(A)$  or  $\text{tr}(A)$  (See trace (of matrix  $A$ )), 323  
 $|D\mathbf{f}|$ , or  $\frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}$ ,  $J$  (See Jacobian determinant), 18  
 $|G|$  (See metric tensor determinant, or in other contexts, number of elements in a group), 161  
 $|S|$  (See number of elements in a finite set), 14  
 Abbena, E., 191  
 Abhyankar, S.S., 190  
 Abraham, R., 232, 286  
 Ackerman, M., vii  
 action (of a group), 17  
 Adams, C.C., 190  
 addition  
     matrix, 321  
     vector, 315  
 additive measure, 16  
 admissible deformation, 21  
 affine transformation, 152  
 Aguilar, A., 61  
 Aleksandrov–Fenchel inequality, 236  
 Algebra, 10  
 algebraic equations, 150  
 Allendoerfer, C.B., 286  
 alternating tensor, 203  
 Amari, S., 29  
 analytic  
     function, 9  
     manifold, 267  
 angular momentum, 25  
 anti-commuting, 195  
 anti-symmetric (Same as skew-symmetric), 320  
 Apéry, F., 286  
 Applebaum, D., 97, 309  
 arc length, 155  
 Arfken, G.B., 29  
 Arnol’d, V.I., viii  
 associative law  
     for group operations, 16  
     for matrix multiplication, 322

- asymmetric unit, 241, 242
- Atiyah, M., 309
- atlas, 267
- axioms, 11
  
- backbone, 169
- Baillieul, J., 286
- balls
  - boundary of, 45
  - volume of, 43–45
- band limit, 6
- Barron, A.R., 97
- Barroso, V., 99
- base space, 279
- basis
  - non-orthogonal, 319
  - orthogonal, 319
- basis vectors for  $\mathbb{R}^n$ , 316
- Bates, P.W., 190
- Bayes' rule, 69
- bell curve, 32
- Bell, D.R., 314
- Bellman–Gronwall lemma, 344
- Belopolskaya, Ya. I., 314
- Ben-Israel, A., 190
- Benedetto, J.J., 60
- Berger, M., 286
- Bertsekas, D., 97
- Bhattacharya, R., 98
- bi-unit cube (and Stokes' theorem), 224
- bi-variate distribution, 70
- bi-variate pdf, 68
- Bianchi identity, 257
- Big-O notation, 6
- bijective mapping, 14
- binary operations, 15
- binomial
  - coefficients, 39
  - distribution, 39, 103
  - theorem, 39
- binormal of a space curve, 156
- Birkhoff, G., 29, 360
- Bishop, R., 190, 286
- Blachman, N.M., 98
- Blackmore, D., 190
- Bloch, A.M., 286
- Bloomenthal, J., 190
- Bluman, G., 60
- Bobenko, A.I., 286
- Bochner, S., 288
- body, 178
  - planar, 158
  - spatial, 159
- Boothby, W.M., 286
  
- Borel measurability, 63
- Bott, R., 286, 309
- Bottema, O., 190
- Bouleau, N., 138
- boundary, 158
- boundary of a body, 178
- bounding sub-manifold terms, 255
- Boy's surface, 241
- Boyce, W.E., 360
- Boyer, C., 60
- Bracewell, R.N., 60
- Brakke, K.A., 190
- Braunstein, S.L., 98
- Brenner, J., 361
- bridging
  - index, 157
  - number, 157
- Brockett, R.W., 310
- Brown, L.D., 98
- Brownian motion, 3, 85, 102, 103, 111
- Bruckstein, A.M., 191
- Brunn–Minkowski inequality, 235
- Bullo, F., 286
- bundle space, 279
- Burke, W.L., 286
- Burke–Hubbard, B., 360
- Burnett, M.N., 287
- Burrus, C.S., 60
- Buttazzo, G., 190
  
- calculus, 10
  - matrix, 353–355
  - multivariable, 18, 349–353
- Campbell, S.L., 360
- Canny, J., 192
- cart
  - kinematic, 3
  - stochastic, 3
- cart-like robot, 2
- Cartan, E., 193
- Cartan, H., 232, 286
- Cartesian product, 15
  - and Euler characteristic, 179
  - of two bodies, 179
- Casimir, H.B.G., 138
- Cauchy distribution, 68
- Cauchy–Bunyakovsky–Schwarz (CBS)
  - inequality, 79
- Cauchy–Schwarz inequality, 83, 318
  - for matrices, 331
- Caves, C.M., 98
- Cayley–Hamilton theorem, 358
- cells, 176
- central limit theorem

- for real line, 90–95
- Chan, T.F., 190
- change of coordinates
  - and entropy, 88
  - and integration, 21
- change of variables, 65
- Chapman–Kolmogorov equation, 108, 118
  - and derivation of the Fokker–Planck equation, 120
- characteristic polynomial, 356
- Charlap, L.S., 286
- Chavel, I., 286
- Chazvini, M., 190
- Chen, B.-Y., 190
- Chern, S.-S., 286
- Chern–Lashof inequality, 171
- Chirikjian, G.S., x, 61, 98, 190, 286, 310, 360
- Cholesky decomposition, 335
- Chopp, D.L., 191
- Christoffel symbol, 164
- Clausius–Duhem inequality, 27
- clipped Gaussian distributions
  - multi-dimensional, 46
  - one-dimensional, 45
  - variance of, 46
- closed form, 226
- closure
  - under conditioning, 68
  - under convolution, 68
  - under marginalization, 68
- co-dimension, 248
- Codazzi–Mainardi equations, 188
- Cole, J., 60
- commutative diagrams, 17, 19
- commutativity, 17
- compact, 10
  - domain, 353
  - manifold, 268
  - support, 89
- complex
  - numbers, 341–343
  - vector space, 316
- composed mapping, 13
- conditional
  - density, 41
  - entropy, 74
  - expectation, 68–73
    - convolution and, 72
    - Jensen’s inequality and, 71
  - mean, 69
- conditioning, 65
- conformation (of a polymer), 281
- conformational bundle, 281
- conformations (of a robot arm), 143
- connected, 10
- connection
  - affine, 281
  - Riemannian, 281
  - symmetric, 281
- conservation of mass, 24
- consistent matrix norm, 329, 337
- contained, 12
- continuity equation
  - for fluids, 24
  - for heat flow, 26
  - for solids, 23
- continuous entropy, 73
- continuous function, 5
- continuously differentiable, 5
- continuum mechanics, 20
- convex
  - function, 71
  - polytopes, 234
- convolution, 36, 65
  - and implicit surfaces, 180
  - information and, 78
  - on the integers, 102
- Cooley, J.W., 61
- coordinate changes
  - and Fokker–Planck equations, 130
  - and Itô SDEs, 135
  - and SDEs, 130
  - and Stratonovich SDEs, 136
- coordinate chart, 267
- coordinates
  - non-linear transformations of, 57
- corollaries, 11
- Costa, M.H., 98
- cotangent space, 269
- covariance matrix, 41, 66
- Cover, T.M., 98
- Cramér, H., 98
- Cramér–Rao bound, 81–84
- Crassidis, J.L., 98
- cross product (See vector product), 320
- Crouch, P., 286
- crystallographic space group, 244
- crystallography, 241
- cumulative distribution function, 34
- curl operator, 351
- curvature
  - 2-form, 278
  - flow, 179, 185
  - Gaussian, 165
  - geodesic, 163
  - mean, 165
  - normal, 163
  - of an implicit surface, 181

- of planar curve, 155
- of surfaces, 163
- principal, 165
- Ricci, 257
- Riemannian, 165
- signed, 155
- unsigned, 155
- curves
  - differential geometry of, 155
  - implicit, 179–185
  - simple, 157
- cyclic permutation, 324
  
- Dalecky, Yu. L., 314
- damping, 124
- Darboux frame, 163
- Darling, R.W.R., 232, 286
- Davis, H.F., 360
- Davis, J.H., 360
- de Bruijn identity, 84–85
- decomposition
  - of a matrix, 331–337
  - spectral, 126
- defining equality, 4
- Dembo, A., 98
- densities
  - conditional, 40
  - marginal, 40
- derivative, 5
- detailed balance, 127
- determinant of a matrix, 323
- diffeomorphism, 213, 241
- differential
  - 2-form, 194
  - area element, 20
  - entropy, 73
  - equations, 10
  - forms
    - in Euclidean space, 194–200
  - geometry
    - of curves, 155
    - of surfaces, 159
  - manifold, 267
  - volume element, 20
  - volume element for  $\mathbb{R}^n$ , 40
- diffusion constants
  - time-varying, 50
- diffusion equations, 38
- DiPrima, R.C., 360
- direct kinematics, 142
- direct sum, 325
- discrete entropy, 90
- dispersion
  - measures of, 95
- distance function, 15
- distribution
  - bi-variate, 70
  - binomial, 39
  - Cauchy, 68
  - Gaussian, 29
  - maximum entropy, 40
  - multi-Gaussian, 92
  - multi-variate, 70
  - tri-variate, 70
- distributions
  - parametric, 45
- divergence, 351
  - Fisher information, 76
  - information-theoretic, 76
  - of a vector field, 162
- divergence theorem, 9, 160, 352
  - and forms, 224
  - for implicitly defined surface, 184
  - for manifolds with boundary, 271
- DNA, 3, 281
- do Carmo, M., 191
- do Carmo, M.P., 232, 286
- dodecahedron, 237, 244
- Dombrowski, P., 191
- donut, 10, 144
- Doob, J.L., 98, 138
- dot product (See scalar product), 317
- double exterior derivative of a  $k$ -form, 199
- dual
  - of a vector space, 206
  - spaces, 319–320
  - vector, 206, 320
- dummy variable, 13
- Dunbar, W.D., 287
- Durrett, R., 138
  
- E. coli*, 3
- edges, 158, 170
- Edwards, B.H., 360
- Edwards, H.M., 360
- eigenvalues, 26, 127, 331
- eigenvectors, 127, 331
- Einstein, A., 138
- Elbert, T.F., 360
- elbow down, 144
- elbow up, 144
- element of surface area, 161
- elimination theory, 146
- ellipsoid
  - metric tensor for, 167
  - of revolution, 167
  - surface area of the, 168
  - total Gaussian curvature of the, 168

- total mean curvature of the, 168
- volume of the, 168
- Elworthy, K.D., 310
- embedding, 141, 266
- Emery, M., 310
- empty set, 12
- end effector, 143
- ensemble average, 65
- entire space, 279
- entropy, 34
  - and change of coordinates, 88
  - and discretization, 89
  - conditional, 74
  - continuous, 73
  - differential, 73
  - discrete, 90
  - marginal, 74
  - power inequality, 75
  - statistical mechanical, 90
  - thermodynamic, 27
- entropy power inequality, 85–87
- equality
  - defining, 4
  - different meanings of, 4
  - in the mean-squared sense, 5, 6
  - philosophical view, 7
- equations
  - algebraic, 150
  - polynomial, 150
  - rate-linearized, 150
- equivalence
  - classes, 12
  - relation, 12
- ergodic (ergodicity), 109
- error
  - function, 46
  - mean-squared, 5
- estimation of parameters, 81
- estimators
  - unbiased, 82
- Euclidean space, 8, 64, 315
- Euler
  - integration, 346
- Euler characteristic, 158, 170, 175–179
  - in  $N$  dimensions, 176
  - of a body, 178
  - of a boundary, 178
  - of a Cartesian product, 179
- Euler parameters, 240
- Euler–Maruyama integration, 115
- Eulerian description, 22
- Evans, L.C., 191
- even function, 32
- evolution equations, 53
- evolving surfaces, 185
- exact form, 226
- expectation, 63–73
  - conditional, 68–73
- expectation, or expected value, 64
- exponential (of a matrix), 338–340
- exterior
  - algebra, 204–221
  - calculus, 195
  - derivative, 195
  - power, 209
  - product, 207
- faces, 158, 170
- factorial, 39
- family
  - of functions, 8, 11
  - of parametric distributions, 45
  - of sets, 16
- Farmer, D.W., 286
- Farouki, R.T., 191
- Fary, I., 191
- Fary–Milnor theorem, 157
- fast marching methods, 184
- Faugeras, O., 191
- Feder, M., 99
- Fedkiw, R.P., 192
- Feller, W., 98
- Fenchel’s theorem, 157
- Fenchel, W., 191, 286
- FFT (Fast Fourier Transform), 237
- fiber bundles, 278
- fiber space, 279
- fibers, 172
- Fick’s law, 25
- field, 316
- filtrations, 63
- finite-difference approximation, 346
- Finney, R.L., 361
- first fundamental form, 161
- Fisher information, 77
  - divergence, 76
  - matrix, 77
- Fisher, R.A., 98
- Flanders, H., 232, 286
- Flannery, B.P., 360
- fluctuation–dissipation theorem, 130
- Flügge, S., 310
- Fokker, A.D., 138
- Fokker–Planck equation, 3, 38
  - and coordinate changes, 130
  - derivation of, 120–121
  - in Euclidean space, 123–127
- folded Gaussian distribution, 47–48

- for all, 13
- form
  - closed, 226
  - exact, 226
- forms
  - one-, 194
  - quadratic, 195
  - two-, 194
  - zero-, 194
- forward kinematics, 142
- Fourier
  - analysis, 5, 10
  - coefficient, 48
  - reconstruction formula, 37
  - series, 48
  - transform, 37
- Fourier's law of heat conduction, 25
- Fourier, J.B.J., 61
- Fowler, R.H., 138
- Fox, R.H., 191
- Fréchet, M., 360
- Frenet frames, 156
- Frieden, B.R., 98, 99
- Fris, I., 61
- Frobenius norm, 328
- Fukuda, K., 286
- function, 11, 14
  - analytic, 9
  - band-limited, 6
  - complex-valued, 5
  - continuous, 5
  - continuously differentiable, 5
  - convex, 71
  - error, 46
  - even, 32
  - family of, 8, 11
  - Fourier series of, 48
  - indicator, 14
  - integrable, 33
  - Kronecker delta, 102
  - matrix-valued, 26
  - natural logarithm, 75
  - nice, 9, 36
  - non-anticipating, 118
  - non-pathological, 9
  - odd, 32
  - on the unit circle, 48
  - periodic, 48
  - smooth, 29
  - well-behaved, 9
- fundamental form
  - first, 161
  - second, 161
- Fundamental Theorem of Calculus, 5, 351
- Gage, M., 191
- Gamma function, 44
- Gantmacher, F.R., 360
- Gard, T.C., 138
- Gardiner, C.W., 138
- Gardner, R.J., 98
- Gauss' theorem (See divergence theorem), 160, 352
- Gauss–Bonnet theorem, 170
- Gauss–Bonnet–Chern theorem, 277
- Gaussian
  - curvature, 165
    - of an implicit surface, 181
  - distribution, 29–48
    - clipped, 45–47
    - folded (or wrapped), 47–48
    - multi-dimensional, 39–43
    - multivariate, 40
    - on the real line, 31–39
    - wrapped around the circle, 47
  - integrals, 42
  - random processes, 106–108
- genus, 157, 170
- geodesic curvature, 163
- geodesics, 165
- Geometry, 10
- geometry
  - algebraic, 11
  - differential, 11
  - of curves and surfaces, 139–190
- $GL(n, \mathbb{C})$ , 325
- $GL(n, \mathbb{R})$ , 325
- gluing, 244
- Gnedenko, B.V., 98
- goes to, or is mapped to, 12
- Goldberg, S.I., 286
- Goldman, R., 191
- Golub, G.H., 360
- González, M., 61
- gradient, 18, 25, 162, 180
- Gram–Schmidt orthogonalization process, 251, 319
- Gray, A., 61, 191
- Grayson, M., 191
- Green's theorem, 352
  - and forms, 223
  - for manifolds with boundary, 271
- Grenander, U., 98
- Greub, W.H., 232
- Greville, T.N.E., 190
- Gromoll, D., 191
- group, 16
  - action, 17



- permutation, 323
- symmetric, 323
- transformation, 17
- group manifold, 239
- groupoid, 244
- Gruber, P.M., 286
- Grünbaum, B., 286
- Guggenheimer, H.W., 191, 232, 286
- Guibas, L., 286
- Guillemin, V., 232, 286
- Guo, D., 99
  
- Hadwiger, H., 191, 286
- Haker, S., 192
- Halmos, P., 360
- Hamilton, R.S., 191, 287
- Hammond, C., 287
- Hardy, G.I., 98
- heat, 25
  - conduction, 25
  - current density, 26
  - flow, 20
  - flux, 26
- heat equation, 26, 38, 48–58
  - multi-dimensional, 51–53
  - on a Riemannian manifold, 283
  - on the circle, 50–51
  - on the line, 48–50
  - symmetry analysis of, 53–58
- Heaviside step function, 34
- heavy tails, 67
- helix (right-handed circular), 187
- Hendricks, H., 98
- Hermitian conjugate, 322
- Hermitian matrix, 327
- Hermitian positive definite, 360
- Hida, T., 310
- Higham, D.J., 138
- Hilbert–Schmidt norm, 328
- hip implant, 147
- Hodge star operator, 204, 224, 285
- Hodge, W.V.D., 191
- homothetic, 236
- Hopf, H., 287
- Hörmander, L., 314
- Horn, R.A., 360
- Hostetler, R.P., 360
- Hsu, E.P., 310
- Hubbard, J.H., 360
- Huisken, G., 191
- hyper-sphere, 5
- hyper-surface, 233, 266
  
- icosahedron, 237
- identity matrix, 21, 26
- if, 11
- iff, 11
- Ikeda, N., 310
- image, 13
- immersion, 261
- implicit equation, 149
- implicit surfaces, 179–185
  - Gaussian curvature of, 181
  - mean curvature of, 181
- in, or “is an element of”, 12
- increment (of a Wiener process), 108
- independent (statistically), 67
- index set, 200
- indexing set, 16
- indicator function, 14, 237
- induced norms, 329
- inequality
  - Aleksandrov–Fenchel, 236
  - Brunn–Minkowski, 235
- information
  - and convolution, 78
  - and entropy, 76
- information theory, 73–80
- information-theoretic divergence, 76
- injective mapping, 14
- inner product (See scalar product), 317
- inner-product space, 317
- integral
  - around a closed curve, 157
  - Riemann, 114
- integral geometry, 166
- integrating by parts, 50
- integration by parts, 47
  - for manifolds with boundary, 271
  - in  $\mathbb{R}^n$ , 353
- intersection, 12
- interval
  - closed, 5
  - open, 5
- intrinsic
  - quantity, 164
  - volumes, 236
- intrinsic geometry
  - of space curves, 155
  - of surfaces, 159
- inverse
  - of a matrix, 325–326
  - of the metric tensor, 161
- inverse Fourier transform, 37
- inverse function theorem, 21
- inverse kinematics, 144
  - analytical solution, 145
  - incremental linearization, 144

- invertible mapping, 13, 14
- isomorphisms (between vector spaces), 317
- Itô, K., 310
- iterations; using inverse Jacobian, 151
- Itô
  - integral, 113
  - SDE, 114
    - and coordinate changes, 135
    - in Cartesian coordinates, 132
    - in polar coordinates, 133
    - relationship with Stratonovich SDE, 134
  - stochastic calculus, 112
  - stochastic differential equations
    - in  $\mathbb{R}^d$ , 112–114
  - stochastic integral, 116–119
- Itô, K., 138, 310
- Itoh, Y., 98
- Itô's rule, 119–120
  
- Jacobi identity, 321
- Jacobian matrix, 18, 119, 131, 354
- Jaynes, E.T., 98
- Jensen's inequality, 71
- Jensen, J.L.W.V., 98
- Johnson, C.K., 287
- Johnson, C.R., 360
- Johnson, O.T., 98
- Jordan block, 334
- Jordan curve theorem, 157
- Joshi, A.W., 29
- Juan, O., 191
- Junkins, J.L., 98
  
- k*-form, 208
- k*-vector, 207
- Kalnins, E., 61
- Kanatani, K., 360
- Karatzas, I., 138
- Karlin, S., 138
- Kass, M., 191
- Katsoulakis, M.A., 191
- Kavraki, L.E., 287
- Keriven, R., 191
- Kho, A.T., 191
- Kimmel, R., 191
- kinematic cart, 3
- Klein bottle, 244
- Klein, F., vii, 287
- Klingenberg, W., 191
- Kloedon, P.E., 138
- Knight, F.B., 138
- knotted curves, 157
- Kobayashi, S., 287
- Kohli, D., 191
  
- Kolmogorov's forward equation (See Fokker–Planck equation), 38
- Kolmogorov, A.N., 98, 138
- Körner, T.W., 61
- Krempf, E., 29
- Kreyszig, E., 29
- Kronecker
  - delta, 202
  - product, 340–341
  - sum, 340–341
- Kronecker delta, 125
- Kronecker delta function, 102
- Kronecker product, 340, 358
- Kronecker sum, 341
- Kuiper, N.H., 191
- Kullback, S., 98
- Kullback–Leibler divergence, 71, 76
- Kunita, H., 309, 310
- Kuo, H.-H., 138
- Kyatkin, A.B., 61, 98, 286
  
- Ladd, M.F.C., 287
- Lagrange multipliers, 35
- Lagrangian description, 22
- Lai, W.M., 29
- Lang, S., 232, 287
- Langevin, P., 104, 138
- Langevin, R., 192
- Laplace–Beltrami operator (See Laplacian), 162
- Laplacian, 162, 351
  - for a Riemannian manifold, 283
- Larson, R., 360
- lattice (partially ordered), 16
- Lattman, E.E., 287
- Lawler, G.F., 99
- Lawrence, J., 287
- Lee, J.M., 287
- Leibniz formula, 323
- Leite, F.S., 361
- lemmas, 11
- Lépingle, D., 138
- Leu, M.C., 190
- level set, 179
- level set methods, 184
- Levi–Civita symbol, 203
- Lévy, P., 138
- Lewis, A.D., 286
- Lewis, J., 310
- lexicographical ordering, 207, 231
- Lie
  - algebra, 320
  - bracket, 212, 320
  - group, 17, 284

- six-dimensional, 58
- Lie, S., 61
- Lindell, I.V., 232
- linear algebra, 10, 315–331
- linear mappings, 317
- linear systems theory, 343–345
- linear transformations, 317
- Linnik, Y.V., 99
- Lipschitz condition, 114
- Lipschutz, M.M., 192
- Littlewood, J.E., 98
- local volume change, 40
- Lockwood, E.H., 287
- Loll, P.J., 287
- Lovelock, D., 232, 287
- LU-decomposition, 336
  
- Mace, G.E., 29
- MacLane, S., 29
- MacMillan, R.H., 287
- Madiman, M., 99
- Magnus, J.R., 360
- Magnus, W., 360
- Malliavin, P., 314
- Malvern, L.E., 29
- manifold, 228, 233, 260–282
  - analytic, 267
  - compact, 277
  - heat equation on a, 282–283
  - smooth, 267
  - stochastic processes on a, 288–309
  - with boundary, 246, 268
- manifolds
  - examples of, 238
- Manocha, D., 192
- mapping, 11
  - bijective, 14
  - composed, 13
  - injective, 14
  - invertible, 13
  - linear, 317
  - one-to-one, 14
  - onto, 14
  - surjective, 14
- marginal
  - density, 41
  - entropy, 74
- marginalization, 65
- Markov
  - property, 105
  - random processes, 106–108, 118
- Markovian (See Markov), 108
- Marsden, J., 286
- Marsden, J.E., 232, 286
  
- martingales, 63
- Maruyama, G., 138
- Maslen, D.K., 310
- material description, 22
- mathematical notation, 10
- matrix, 321–331
  - commutator, 356
  - decompositions, 331–337
  - exponential of a, 338–340
  - Hermitian, 327
  - inequalities, 331
  - inverse of a, 325–326
  - Lie groups, 57
  - normal, 327
  - norms, 328–331
    - Frobenius, 328
    - induced, 329
    - sub-multiplicative, 328
  - orthogonal, 327
  - perturbations, 337–338
  - positive definite, 326
  - skew-Hermitian, 327
  - skew-symmetric, 327
  - special orthogonal, 328
  - special unitary, 328
  - unitary, 327
  - version of Cauchy–Schwarz, 331
- matrix calculus, 353–355
- matrix Lie group, 239
- matrix norm
  - consistent, 329
- matrix norm, consistent, 337
- matrix-valued function, 26
- matrizant, or matricant (See state transition matrix), 348
- maximum entropy
  - distribution, 40
  - property, 34
- McKean, H.P., Jr., 138, 310, 360
- McLachlan, N.W., 310
- McPherson, A., 287
- McShane, E.J., 138
- McWhorter, L., 99
- mean, 33, 66
  - conditional, 69
  - curvature, 165
    - of an implicit surface, 181
    - total, 170
  - squared error, 5
- measures of dispersion, 95
- mechanics
  - continuum, 20
- medial circle, 169
- median, 33

- medical imaging, 147
- Meeks, W.H., 191
- metric, 15
- metric tensor, 161
  - for the ellipsoid, 167
  - for the sphere, 166
  - for the torus, 169
- Meyer, C.D., Jr., 360
- Meyer, W., 191
- Miller, W., Jr., 61
- Millman, R.S., 192
- Millstein, G.N., 138
- Milnor, J., 192
- Minkowski sum, 235
- mixed volumes, 236
- mode, 33
- moment of momentum, 25
- momentum balance, 24
- Montesinos, J.M., 287
- Morgan, F., 287
- motion
  - Brownian, 3
  - rigid body, 152
  - rigid-body, 97, 294
  - rotational, 153, 309
- Mukherjee, A., 232, 287
- multi-Gaussian distributions, 92
- multi-set, 204
- multi-variate distribution, 70
- multi-vector, 207
- multilinearity, 323
- multinomial expansion, 92
- multivariable calculus, 18, 349–353
- multivariate Gaussian distribution, 40
- Mumford, D., 192
- Munkres, J.R., 29
- mutual information, 76
  
- $n$ -dimensional space, 8
- $n$ -times continuously differentiable, 5
- Nagaoka, H., 29
- Nash, J., 287
- natural basis vectors for  $\mathbb{R}^n$ , 316
- natural logarithm function, 75
- necessary conditions, 11
- Neff, C.A., 191
- Neudecker, H., 360
- Neyfeh, A., 29
- nice functions, 36
  - on a domain  $D$ , 9
  - on an interval, 9
- Nikolov, B., 99
- noise, 103
- Nomizu, K., 287
  
- non-anticipating functions, 118
- non-linear transformations of coordinates, 57
- non-negative real numbers, 14, 15
- non-orientable surface, 244
- non-orthogonal basis, 319
- non-pathological function, 9
- non-stationary random processes, 104–106
- Norbert Wiener, 111
- norm of a matrix, 328–331
- normal
  - curvature, 163
  - distribution, 34
  - distribution (See Gaussian distribution), 29
  - matrix, 327
- Norman, E., 361
- normed vector space, 319
- null space, 326
- number of
  - edges, 170
  - faces, 170
  - vertices, 170
- numbers
  - complex, 14
  - real, 5, 14
- numerical approximations
  - finite difference, 301
  - of SDE sample paths, 115
- numerical integration of SDEs, 115
  
- O-U (See Ornstein–Uhlenbeck), 123
- odd function, 32
- ODEs, 349
- offset curves, 171
- offset surfaces (and tubes), 171–175
- Øksendal, B., 138, 310
- Olver, P.J., 61, 192
- one-to-one, 14
- only if, 11
- Onsager relations, 127
- Onsager, L., 138
- onto, 14
- open interval, 5
- Oprea, J., 192, 287
- orbifold, 244
- ordinary differential equation (ODE), 346
- orientable differential manifold, 267
- Ornstein, L.S., 139
- Ornstein–Uhlenbeck process, 123
- Orsingher, E., 310
- orthogonal matrix, 327
- Osher, S.J., 192
- Osserman, R., 192
- Osvatic, M., 191

- outer product, 206
- Ovsiannikov, L.V., 61
- p-norm, 318
- Pan, V., 360
- parameter estimation, 81
- parametric distributions, 67
- parametrization of a manifold, 266, 267
- Parker, G.D., 192
- Parks, T.W., 60
- partially ordered lattice, 16
- partition, 13
- Patera, J., 61
- Patodi, V.K., 309
- Patrangenaru, V., 98
- pdf, 63
  - (See probability density function), 63
  - mean of, 33
  - median of, 33
  - mode of, 33
  - spread of, 33
  - standard deviation of, 33
  - variance of, 33
  - weighted sum of, 87
- Pedoe, D., 191
- Penec, X., 99
- Perelman, G., 287
- permutation
  - group, 201
  - sign of a, 202
- permutation group, 323, 324
- Perrin, P.F., 310
- Pfaffian, 278
- Pham, B., 192
- phase factor, 39
- Pinsky, M., 29, 310
- Pitts, D.R., 29
- planar
  - body
    - simple, 157
  - region
    - simply connected, 157
- Planck, M., 138
- Platen, E., 138
- polar
  - coordinates
    - and Gaussians, 32
    - and Itô SDEs, 132
    - and Stratonovich SDEs, 134
  - decomposition, 336
- Pollack, A., 232, 286
- Poluyanov, L.V., 61
- Pólya, G., 98
- polyhedra, 234–237
  - volume of, 235
- polynomial
  - elimination, 146
  - equations, 150
- polytopes, 234–237
- pose (or position and orientation), 3
- positive definite matrix, 326
- positive real numbers, 14, 15
- Postelnicu, G., 191
- postulates, 11
- pre-image, 13
- Press, W.H., 360
- principal curvatures, 165
- principal normal vector
  - to a space curve, 155
- probability density function, 63–68
- probability density function (See pdf), 33, 64
- process
  - random, 103
  - return-to-equilibrium, 127
- Prodon, A., 286
- product
  - exterior, 207
  - outer, 206
  - tensor, 206
  - wedge, 207
- product integral, 345–349
- projection
  - stereographic, 7
- projective geometry
  - and medical imaging, 147–154
- proper subset, 12
- Protter, P., 139
- pseudo-inverse, 326
- pull-back, 21, 214
- punctured sphere, 7
- push-forward, 21, 214
- QR decomposition, 335
- quadratic form, 195
- Queiro, J.F., 361
- quermassintegrals, 236
- Raghavan, M., 192
- Ramshaw, L., 286
- random
  - noise, 103
  - processes, 102, 104–108
    - Gaussian, 106–108
    - Markov, 106–108
    - non-stationary, 104–106
  - pdfs for, 105
    - stationary, 104–105
    - strong stationarity, 105–106
    - weak stationarity, 105–106

- walker, 1, 102
- walks
  - on the integers, 101–104
- rank (of a matrix), 326
- Rao, C.R., 99
- Rastamian, R., 188
- rate-linearized equations, 150
- Ratiu, T., 232, 286
- Ratnanather, J.T., 192
- real numbers, 4, 5
- real projective plane, 240, 244
- referential state, 21
- reflexive property, 12
- Rényi, A., 61, 139
- return-to-equilibrium process, 127
- reverse kinematics (See inverse kinematics), 144
- Reynolds transport theorem, 24, 29
- Rhodes, G., 287
- ribbons, 172
- Ricci
  - curvature tensor, 257
  - flow, 284
- Riemann integral, 112, 114
- Riemann–Stieltjes integral, 112
- Riemannian
  - curvature, 165, 257
  - manifold, 266
  - metric, 247, 269
- Riemannian metric tensor, 269
- right-handed circular helix, 187
- rigid body, 3
- Ripley, B.D., 139
- Risken, H., 139
- RMSD (See root-mean-square deviation), 93
- Roberts, P.H., 310
- robotic manipulators, 142–147
- Rockafellar, R.T., 99
- Rogers, L.C.G., 139
- Rolfen, D., 192
- Roman surface, 241
- root-mean-square deviation, 93
- Ros, A., 192
- Rosenberg, S., 287
- Rosenburg, H., 192
- rotations
  - in  $n$ -dimensional space, 133
  - in the plane, 17
  - spatial, 148
- Roth, B., 190, 192
- Rubin, D., 29
- Rudin, W., 361
- Rugh, W.J., 361
- ruled surface, 188
- Rund, H., 232, 287
- Salamon, S., 191
- Samorodnitsky, G., 99
- sample paths, 115
- San Jose Estepar, R., 192
- Satake, I., 287
- scalar (dot) product, 22
- scalar multiplication, 315
- scalar product, 317–319
- Scharf, L.L., 99
- Scheinerman, E.R., x
- Schervish, M.J., 99
- Schey, H.M., 361
- Schreiber, M., 232, 287
- Schröder, P., 286
- Schubert, H., 192
- Schur decomposition, 335
- Scott, P., 287
- sculpting operations, 158
- SDEs, 114
  - and coordinate changes, 130
  - Stratonovich, 122
- SDEs (See stochastic differential equations), 101
- second fundamental form, 161, 162
- Second Law of Thermodynamics, 27
- sectional curvature, 165, 257
- semi-flexible polymer, 281
- set
  - countably infinite, 16
  - finite, 16
  - indexing, 16
  - uncountably infinite, 16
- Sethian, J.A., 192
- sets, 11
- Shah, J., 192
- Shannon entropy, 64, 80
- Shannon, C. E., 61
- Shannon, C.E., 99
- shear, 152
- Shiohama, K., 192
- Shoemake, K., 190
- Shreve, S.E., 138
- sign (or signature) of a permutation, 202, 324
- signed curvature, 155
- simple
  - curve, 157
  - planar body, 157
- simply connected planar region, 157
- Singer, I.M., 310
- singular value decomposition, 336
- singularities (of a robot arm), 143
- Sissom, L.E., 29

- skew-Hermitian matrix, 327
- skew-symmetric, 195
  - (Same as anti-symmetric), 320
  - matrix, 327
- Smith, S.T., 99
- smooth, 19
  - function, 29
  - manifold, 267
  - vector fields, 281
- Snider, A.D., 360
- So, W., 361
- Sommese, A.J., 192
- Soner, H.M., 192
- space curve
  - binormal of, 156
  - curvature of, 155
  - torsion of, 156
- spaces, 15
- span, 317, 319
- spatial
  - body, 159
  - description, 22
- special
  - orthogonal matrices, 328
  - unitary matrices, 328
- special Euclidean group, 239
- special orthogonal group, 240
- spectral decomposition, 126
- sphere, 166
  - metric tensor for the, 166
  - punctured, 7
  - surface area of the, 167
  - total Gaussian curvature of the, 167
  - total mean curvature of the, 167
  - volume in  $\mathbb{R}^n$ , 43–45
  - volume of the, 167
- spherical coordinates, 148
- Spivak, M., 287, 288
- spread, 33
- Spruck, J., 191
- square-integrable function
  - on the circle, 6
- Stam, A.J., 99
- standard deviation, 33
- state
  - space, 124
  - transition matrix, 347
- stationarity
  - strong, 105–106
  - weak, 105–106
- stationary
  - strictly, 106
  - wide-sense, 106
- stationary random processes, 104–105
- Steenrod, N., 288
- Stein, D., x
- Steiner's formula
  - in  $\mathbb{R}^3$ , 175
  - multi-dimensional version, 236
- stereographic projection, 7
- Stewart, G.W., 361
- stiffness, 124
- Stirling series, 39
- Stirling's formula, 39
- stochastic
  - calculus
    - Itô, 112, 119
    - Stratonovich, 121
  - differential equations
    - and changes of coordinates, 130–136
    - in Euclidean space, 99–137
    - Itô, 112–114
    - numerical approximation of, 114–116
    - Stratonovich, 130
  - integral
    - Itô, 116–119
    - Stratonovich, 130
  - process
    - Ornstein–Uhlenbeck, 123
  - processes (See random processes), 102, 104
  - systems, 6
- Stokes' theorem, 352
  - and forms, 224
  - for bi-unit cube, 224
  - for implicitly defined surface, 184
  - for manifolds, 270
  - in cylindrical coordinates, 225
  - light form of, 9
- Stolfi, J., 286
- Stratonovich
  - integral, 113
  - SDE, 122, 130
    - and coordinate changes, 136
    - in polar coordinates, 134
    - relationship with Itô SDE, 134
  - stochastic calculus, 121
  - stochastic integral, 130
- Stratonovich, R.L., 139
- strictly
  - contained, 12
  - stationary, 106
- strongly stationary Markov process, 107
- Stroock, D., 139
- Stroock, D.W., 310
- sub-multiplicative matrix norms, 328
- sublinear growth, 75
- subsets, 12
- sufficient conditions, 11

- Sullivan, J.M., 192, 286
- Sun, J.-Q., 361
- superquadric surfaces, 225
- support (of a function), 89, 237
- surface
  - Boy's, 241
  - non-orientable, 244
  - Roman, 241
- surface area element, 161
- surfaces
  - differential geometry of, 159
  - evolving, 185
  - implicit, 179–185
  - ruled, 188
- surjective mapping, 14
- SVD, 336
- symmetric
  - group, 201, 323
  - property, 12
- symmetries
  - in parameters, 54
  - infinitesimal, 54
- symmetry operators, 56
- systems
  - linear, 338
  - stochastic, 6
  
- tails (heavy), 67
- Takagi, R., 192
- tangent
  - bundle, 271
  - space to  $\mathbb{R}^n$  at the point  $\mathbf{x}$ , 213
  - vector (of a curve), 155
- Taqqu, M.S., 99
- Taylor series, 7
  - in derivation of Fokker–Planck equation, 121
- Taylor, H.M., 138
- temperature, 25
- tensor product, 206
- Terzopoulos, D., 191
- tessellation, 241
- tetrahedron, 236
- Teukolsky, S.A., 360
- Theorema Egregium, 165
- theorems, 11
- thermal
  - conduction, 25
  - conductivity matrix, 26
- thermodynamic entropy, 27
- Thermodynamics, 27
- Thomas, G.B., Jr., 361
- Thomas, J.A., 98
- three-torus, 244
  
- Thurston, W.P., 288
- time-varying diffusion constants, 50
- topology, 10
  - algebraic, 11
  - differential, 11
- Topping, P., 288
- toroidal
  - body, 179
  - surface, 179
  - void, 177
- toroidal world, 3
- torsion
  - 2-form, 278
  - of a space curve, 156
- torus, 10, 169, 171, 243
  - making from a square, 144
  - metric tensor for the, 169
  - surface area of the, 169
  - total Gaussian curvature of the, 169
  - total mean curvature of the, 169
  - volume of the, 169
- total
  - Gaussian curvature, 166
  - mean curvature, 166
    - of knotted tori, 174
    - of surfaces, 170
  - space, 279
- Touzi, N., 192
- trace of a matrix, 323
- transformation groups, 17
- transformations
  - affine, 152
  - linear, 317
  - non-linear, 57
- transition probability, 121
- transitive
  - group action, 17
  - property, 12
- transport phenomena, 20
- transpose
  - of a matrix, 322
  - of a vector, 319
  - rule, 325
- transposition, 202
- Tretyakov, M.V., 138
- tri-variate distribution, 70
- triangle inequality, 15
- triple product, 320, 356
- Tu, L.W., 232, 286, 288
- tubes
  - and offset surfaces, 171–175
  - knotted, 174
  - of curves in  $\mathbb{R}^2$ , 171
  - of curves in  $\mathbb{R}^3$ , 172



- of surfaces, 175
- tubular
  - body, 173
  - surface, 173
- Tukey, J., 61
- Uhlenbeck, G.E., 139
- unbiased estimators, 82
- unimodal distribution, 32
- union, 12
- unit
  - circle, 5
  - strength Wiener process, 112
  - tangent vector, 155
- unit cell, 242, 244
- unitary
  - diagonalizability, 335
  - matrix, 327, 343
- unsigned curvature, 155
- Ursell, H.D., 310
- valuation, 16
- van Kampen, N.G., 139
- Van Loan, C., 61
- Van Loan, C.F., 360
- Varadhan, S.R.S., 139
- variables
  - change of, 65
- variance, 33
- vector, 315
  - addition, 315
  - dual, 206
  - field (divergence of), 162
  - multiplication by scalars, 315
  - norm, 317–319
  - normal, 155
  - product, 320–321
  - space, 315–317
    - complex, 316
    - dual, 206
    - isomorphisms, 317
    - normed, 319
  - subspace, 319
- vector calculus, 349–353
- Verdú, S., 99
- vertices, 158
  - number of, 170
- Vese, L.A., 190
- Vetterling, W.T., 360
- Villani, C., 99
- Visintin, A., 190
- volume
  - of a convex body, 235
  - of balls, 43–45
  - of polyhedra, 235
  - of spheres, 43–45
- Voss, K., 192
- wallpaper pattern, 241
- Wampler, C.W., 192
- Wang, L.P., 190
- Wang, M.C., 139
- Wang, Y., 310
- Warner, F.W., 232, 288
- Watanabe, S., 139, 310
- Weaver, W., 99
- Weber, H.J., 29
- Wedderburn, J.H.M., 361
- wedge product, 195, 207, 208
- Weeks, J.R., 288
- Wei, G.W., 190
- Wei, J., 361
- weighted sum of pdfs, 87
- Weil, A., 286
- Weinstein, A., 288
- Weisner, L., 61
- well-behaved function, 9
- Wermuth, E.M.E., 361
- Westin, C.F., 192
- Weyl's tube theorem, 175
- Weyl, H., 192
- white noise, 109
- Whitney, H., 288
- wide-sense stationary, 106
- Wiener process, 108–112
  - $m$ -dimensional, 111
  - abstract definitions of, 111–112
  - as driving force for SDEs, 112–114
  - unit strength, 112
- Wiener stochastic integral, 109
- Wiener, N., 139
- Williams, D., 99, 139
- Willmore, T.J., 192, 288
- Winternitz, P., 61
- Witkin, A., 191
- workspace of a robot arm, 144
- wrapped Gaussian distribution, 47–48
- x-ray source, 147
- Xavier, J., 99
- Yano, K., 288
- Yau, S.-T., 310
- Yip, N.K., 192
- Yosida, K., 311
- Younes, L., 192
- Zamir, R., 99
- Zhang, S., 192

Zhao, S., 190

Zhou, Y., 310

Ziegler, G.M., 286, 288

Zimmermann, G., 60

Zolotarev, V.M., 99

Zweck, J., 192

# Applied and Numerical Harmonic Analysis

---

- J.M. Cooper: *Introduction to Partial Differential Equations with MATLAB* (ISBN 978-0-8176-3967-9)
- C.E. D'Attellis and E.M. Fernández-Berdaguer: *Wavelet Theory and Harmonic Analysis in Applied Sciences* (ISBN 978-0-8176-3953-2)
- H.G. Feichtinger and T. Strohmer: *Gabor Analysis and Algorithms* (ISBN 978-0-8176-3959-4)
- T.M. Peters, J.H.T. Bates, G.B. Pike, P. Munger, and J.C. Williams: *The Fourier Transform in Biomedical Engineering* (ISBN 978-0-8176-3941-9)
- A.I. Saichev and W.A. Woyczyński: *Distributions in the Physical and Engineering Sciences* (ISBN 978-0-8176-3924-2)
- R. Tolimieri and M. An: *Time-Frequency Representations* (ISBN 978-0-8176-3918-1)
- G.T. Herman: *Geometry of Digital Spaces* (ISBN 978-0-8176-3897-9)
- A. Procházka, J. Uhlíř, P.J.W. Rayner, and N.G. Kingsbury: *Signal Analysis and Prediction* (ISBN 978-0-8176-4042-2)
- J. Ramanathan: *Methods of Applied Fourier Analysis* (ISBN 978-0-8176-3963-1)
- A. Teolis: *Computational Signal Processing with Wavelets* (ISBN 978-0-8176-3909-9)
- W.O. Bray and Č.V. Stanojević: *Analysis of Divergence* (ISBN 978-0-8176-4058-3)
- G.T. Herman and A. Kuba: *Discrete Tomography* (ISBN 978-0-8176-4101-6)
- J.J. Benedetto and P.J.S.G. Ferreira: *Modern Sampling Theory* (ISBN 978-0-8176-4023-1)
- A. Abbate, C.M. DeCusatis, and P.K. Das: *Wavelets and Subbands* (ISBN 978-0-8176-4136-8)
- L. Debnath: *Wavelet Transforms and Time-Frequency Signal Analysis* (ISBN 978-0-8176-4104-7)
- K. Gröchenig: *Foundations of Time-Frequency Analysis* (ISBN 978-0-8176-4022-4)
- D.F. Walnut: *An Introduction to Wavelet Analysis* (ISBN 978-0-8176-3962-4)
- O. Bratteli and P. Jorgensen: *Wavelets through a Looking Glass* (ISBN 978-0-8176-4280-8)
- H.G. Feichtinger and T. Strohmer: *Advances in Gabor Analysis* (ISBN 978-0-8176-4239-6)
- O. Christensen: *An Introduction to Frames and Riesz Bases* (ISBN 978-0-8176-4295-2)
- L. Debnath: *Wavelets and Signal Processing* (ISBN 978-0-8176-4235-8)
- J. Davis: *Methods of Applied Mathematics with a MATLAB Overview* (ISBN 978-0-8176-4331-7)
- G. Bi and Y. Zeng: *Transforms and Fast Algorithms for Signal Analysis and Representations* (ISBN 978-0-8176-4279-2)
- J.J. Benedetto and A. Zayed: *Sampling, Wavelets, and Tomography* (ISBN 978-0-8176-4304-1)
- E. Prestini: *The Evolution of Applied Harmonic Analysis* (ISBN 978-0-8176-4125-2)
- O. Christensen and K.L. Christensen: *Approximation Theory* (ISBN 978-0-8176-3600-5)
- L. Brandolini, L. Colzani, A. Iosevich, and G. Travaglini: *Fourier Analysis and Convexity* (ISBN 978-0-8176-3263-2)
- W. Freeden and V. Michel: *Multiscale Potential Theory* (ISBN 978-0-8176-4105-4)
- O. Calin and D.-C. Chang: *Geometric Mechanics on Riemannian Manifolds* (ISBN 978-0-8176-4354-6)

## Applied and Numerical Harmonic Analysis (Cont'd)

---

J.A. Hogan and J.D. Lakey: *Time-Frequency and Time-Scale Methods* (ISBN 978-0-8176-4276-1)

C. Heil: *Harmonic Analysis and Applications* (ISBN 978-0-8176-3778-1)

K. Borre, D.M. Akos, N. Bertelsen, P. Rinder, and S.H. Jensen: *A Software-Defined GPS and Galileo Receiver* (ISBN 978-0-8176-4390-4)

T. Qian, V. Mang I, and Y. Xu: *Wavelet Analysis and Applications* (ISBN 978-3-7643-7777-9)

G.T. Herman and A. Kuba: *Advances in Discrete Tomography and Its Applications* (ISBN 978-0-8176-3614-2)

M.C. Fu, R.A. Jarrow, J.-Y. J. Yen, and R.J. Elliott: *Advances in Mathematical Finance* (ISBN 978-0-8176-4544-1)

O. Christensen: *Frames and Bases* (ISBN 978-0-8176-4677-6)

P.E.T. Jorgensen, K.D. Merrill, and J.A. Packer: *Representations, Wavelets, and Frames* (ISBN 978-0-8176-4682-0)

M. An, A.K. Brodzik, and R. Tolimieri: *Ideal Sequence Design in Time-Frequency Space* (ISBN 978-0-8176-4737-7)

S.G. Krantz: *Explorations in Harmonic Analysis* (ISBN 978-0-8176-4668-4)

G.S. Chirikjian: *Stochastic Models, Information Theory, and Lie Groups, Volume I* (ISBN 978-0-8176-4802-2)

C. Cabrelli and J.L. Torrea: *Recent Developments in Real and Harmonic Analysis* (ISBN 978-0-8176-4531-1)