Mark H. Holmes

TEXTS IN APPLIED MATHEMATICS 56

Introduction to the Foundations of Applied Mathematics

Springer

Texts in Applied Mathematics 56

Editors
J.E. Marsden
L. Sirovich
S.S. Antman

Advisors
G. Iooss
P. Holmes
D. Barkley
M. Dellnitz
P. Newton

Texts in Applied Mathematics

1. *Sirovich*: Introduction to Applied Mathematics.
2. *Wiggins*: Introduction to Applied Nonlinear Dynamical Systems and Chaos.
3. *Hale/Koçak*: Dynamics and Bifurcations.
4. *Chorin/Marsden*: A Mathematical Introduction to Fluid Mechanics, 3rd ed.
5. *Hubbard/West*: Differential Equations: A Dynamical Systems Approach: Ordinary Differential Equations.
6. *Sontag*: Mathematical Control Theory: Deterministic Finite Dimensional Systems, 2nd ed.
7. *Perko*: Differential Equations and Dynamical Systems, 3rd ed.
8. *Seaborn*: Hypergeometric Functions and Their Applications.
9. *Pipkin*: A Course on Integral Equations.
10. *Hoppensteadt/Peskin*: Modeling and Simulation in Medicine and the Life Sciences, 2nd ed.
11. *Braun*: Differential Equations and Their Applications, 4th ed.
12. *Stoer/Bulirsch*: Introduction to Numerical Analysis, 3rd ed.
13. *Renardy/Rogers*: An Introduction to Partial Differential Equations.
14. *Banks*: Growth and Diffusion Phenomena: Mathematical Frameworks and Applications.
15. *Brenner/Scott*: The Mathematical Theory of Finite Element Methods, 2nd ed.
16. *Van de Velde*: Concurrent Scientific Computing.
17. *Marsden/Ratiu*: Introduction to Mechanics and Symmetry, 2nd ed.
18. *Hubbard/West*: Differential Equations: A Dynamical Systems Approach: Higher-Dimensional Systems.
19. *Kaplan/Glass*: Understanding Nonlinear Dynamics.
20. *Holmes*: Introduction to Perturbation Methods.
21. *Curtain/Zwart*: An Introduction to Infinite-Dimensional Linear Systems Theory.
22. *Thomas*: Numerical Partial Differential Equations: Finitc Difference Methods.
23. *Taylor*: Partial Differential Equations: Basic Theory.
24. *Merkin*: Introduction to the Theory of Stability of Motion.
25. *Naber*: Topology, Geometry, and Gauge Fields: Foundations.
26. *Polderman/Willems*: Introduction to Mathematical Systems Theory: A Behavioral Approach.
27. *Reddy*: Introductory Functional Analysis with Applications to Boundary-Value Problems and Finite Elements.
28. *Gustafson/Wilcox*: Analytical and Computational Methods of Advanced Engineering Mathematics.
29. *Tveito/Winther*: Introduction to Partial Differential Equations: A Computational Approach.
30. *Gasquet/Witomski*: Fourier Analysis and Applications: Filtering, Numerical Computation, Wavelets.

(continued after index)

Mark H. Holmes

Introduction to the Foundations of Applied Mathematics

Springer

Mark H. Holmes
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
110 8th Street
Troy NY 12180-3590
USA
holmes@rpi.edu

Series Editors

J.E. Marsden
Control and Dynamical Systems, 107–81
California Institute of Technology
Pasadena, CA 91125
USA
marsden@cds.caltech.edu

L. Sirovich
Division of Applied Mathematics
Brown University
Providence, RI 02912
USA
lawrence.sirovich@mssm.edu

S.S. Antman
Department of Mathematics
and
Institute for Physical Science
and Technology
University of Maryland
College Park, MD 20742-4015
USA
ssa@math.umd.edu

ISSN 0939-2475
ISBN 978-0-387-87749-5 e-ISBN 978-0-387-87765-5
DOI 10.1007/978-0-387-87765-5
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009929235

Mathematics Subject Classification (2000): 74-01; 76R50; 76A02; 76M55; 35Q30; 35Q80; 92C45; 74A05; 74A10

© 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 (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.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Colette, Matthew and Marianna

Preface

FOAM. This acronym has been used for over fifty years at Rensselaer to designate an upper-division course entitled, Foundations of Applied Mathematics. This course was started by George Handelman in 1956, when he came to Rensselaer from the Carnegie Institute of Technology. His objective was to closely integrate mathematical and physical reasoning, and in the process enable students to obtain a qualitative understanding of the world we live in. FOAM was soon taken over by a young faculty member, Lee Segel. About this time a similar course, Introduction to Applied Mathematics, was introduced by Chia-Ch'iao Lin at the Massachusetts Institute of Technology. Together Lin and Segel, with help from Handelman, produced one of the landmark textbooks in applied mathematics, *Mathematics Applied to Deterministic Problems in the Natural Sciences*. This was originally published in 1974, and republished in 1988 by the Society for Industrial and Applied Mathematics, in their Classics Series.

This textbook comes from the author teaching FOAM over the last few years. In this sense, it is an updated version of the Lin and Segel textbook. The objective is definitely the same, which is the construction, analysis, and interpretation of mathematical models to help us understand the world we live in. However, there are some significant differences. Lin and Segel, like many recent modeling books, is based on a case study format. This means that the mathematical ideas are introduced in the context of a particular application. There are certainly good reasons why this is done, and one is the immediate relevance of the mathematics. There are also disadvantages, and one pointed out by Lin and Segel is the fragmentary nature of the development. However, there is another, more important reason for not following a case studies approach. Science evolves, and this means that the problems of current interest continually change. What does not change as quickly is the approach used to derive the relevant mathematical models, and the methods used to analyze the models. Consequently, this book is written in such a way as to establish the mathematical ideas underlying model development independently of a specific application. This does not mean applications are not

considered, they are, and connections with experiment are a staple of this book.

The first two chapters establish some of the basic mathematical tools that are needed. The model development starts in Chapter 3, with the study of kinetics. The goal of this chapter is to understand how to model interacting populations. This does not account for the spatial motion of the populations, and this is the objective of Chapters 4 and 5. What remains is to account for the forces in the system, and this is done in Chapter 6. The last three chapters concern the application to specific problems and the generalization of the material to more geometrically realistic systems. The book, as well as the individual chapters, is written in such a way that the material becomes more sophisticated as you progress. This provides some flexibility in how the book is used, allowing consideration for the breadth and depth of the material covered.

The principal objective of this book is the derivation and analysis of mathematical models. Consequently, after deriving a model, it is necessary to have a way to solve the resulting mathematical problem. A few of the methods developed here are standard topics in upper-division applied math courses, and in this sense there is some overlap with the material covered in those courses. Examples are the Fourier and Laplace transforms, and the method of characteristics. On the other hand, other methods that are used here are not standard, and this includes perturbation approximations and similarity solutions. There are also unique methods, not found in traditional textbooks, that rely on both the mathematical and physical characteristics of the problem.

The prerequisite for this text is a lower-division course in differential equations. The implication is that you have also taken two or three semesters of calculus, which includes some component of matrix algebra. The one topic from calculus that is absolutely essential is Taylor's theorem, and for this reason a short summary is included in the appendix. Some of the more sophisticated results from calculus, related to multidimensional integral theorems, are not needed until Chapter 8.

To learn mathematics you must work out problems, and for this reason the exercises in the text are important. They vary in their difficulty, and cover most of the topics in the chapter. Some of the answers are available, and can be found at www.holmes.rpi.edu. This web page also contains a typos list.

I would like to express my gratitude to the many students who have taken my FOAM course at Rensselaer. They helped me immeasurably in understanding the subject, and provided much-needed encouragement to write this book. It is also a pleasure to acknowledge the suggestions of John Ringland, and his students, who read an early version of the manuscript.

Troy, New York *Mark H. Holmes*
March, 2009

Contents

Preface vii

1 Dimensional Analysis 1
1.1 Introduction 1
1.2 Examples of Dimensional Reduction 3
1.2.1 Maximum Height of a Projectile 5
1.2.2 Drag on a Sphere 6
1.2.3 Toppling Dominoes 13
1.2.4 Endnotes 15
1.3 Theoretical Foundation 16
1.3.1 Pattern Formation 19
1.4 Similarity Variables 22
1.5 Nondimensionalization and Scaling 25
1.5.1 Projectile Problem 26
1.5.2 Weakly Nonlinear Diffusion 30
1.5.3 Endnotes 32
Exercises 33

2 Perturbation Methods 43
2.1 Regular Expansions 43
2.2 How to Find a Regular Expansion 48
2.2.1 Given a Specific Function 48
2.2.2 Given an Algebraic or Transcendental Equation 51
2.2.3 Given an Initial Value Problem 53
2.3 Introduction to Singular Perturbations 58
2.4 Introduction to Boundary Layers 60
2.4.1 Endnotes 66
2.5 Multiple Boundary Layers 68
2.6 Multiple Scales and Two-Timing 72
Exercises 79

3 Kinetics . . . 87
3.1 Introduction . . . 87
3.1.1 Radioactive Decay . . . 87
3.1.2 Predator-Prey . . . 88
3.1.3 Epidemic Model . . . 88
3.2 Kinetic Equations . . . 89
3.2.1 The Law of Mass Action . . . 91
3.2.2 Conservation Laws . . . 92
3.2.3 Steady-States . . . 94
3.2.4 Examples . . . 94
3.2.5 End Notes . . . 96
3.3 General Mathematical Formulation . . . 97
3.4 Michaelis-Menten Kinetics . . . 100
3.4.1 Numerical Solution . . . 102
3.4.2 Quasi-Steady-State Approximation . . . 103
3.4.3 Perturbation Approach . . . 105
3.5 Assorted Applications . . . 111
3.5.1 Elementary and Nonelementary Reactions . . . 111
3.5.2 Reverse Mass Action . . . 113
3.6 Steady-States and Stability . . . 114
3.6.1 Reaction Analysis . . . 115
3.6.2 Geometric Analysis . . . 115
3.6.3 Perturbation Analysis . . . 118
3.7 Oscillators . . . 126
3.7.1 Stability . . . 128
Exercises . . . 132

4 Diffusion . . . 141
4.1 Introduction . . . 141
4.2 Random Walks and Brownian Motion . . . 142
4.2.1 Calculating $w(m, N)$ . . . 145
4.2.2 Large N Approximation . . . 148
4.3 Continuous Limit . . . 149
4.3.1 What Does D Signify? . . . 151
4.4 Solving the Diffusion Equation . . . 153
4.4.1 Point Source . . . 154
4.4.2 Fourier Transform . . . 157
4.5 Continuum Formulation of Diffusion . . . 169
4.5.1 Balance Law . . . 169
4.5.2 Fick's Law of Diffusion . . . 171
4.5.3 Reaction-Diffusion Equations . . . 177
4.6 Random Walks and Diffusion in Higher Dimensions . . . 179
4.6.1 Diffusion Equation . . . 182
4.7 Langevin Equation . . . 185
4.7.1 Properties of the Forcing . . . 188

4.7.2 Endnotes . . . 194
Exercises . . . 194

5 Traffic Flow . . . 205
5.1 Introduction . . . 205
5.2 Continuum Variables . . . 206
5.2.1 Density . . . 207
5.2.2 Flux . . . 208
5.3 Balance Law . . . 209
5.3.1 Velocity Formulation . . . 210
5.4 Constitutive Laws . . . 211
5.4.1 Constant Velocity . . . 212
5.4.2 Linear Velocity . . . 213
5.4.3 General Velocity Formulation . . . 214
5.4.4 Flux and Velocity . . . 216
5.4.5 Reality Check . . . 217
5.5 Constant Velocity . . . 218
5.5.1 Characteristics . . . 221
5.6 Nonconstant Velocity . . . 225
5.6.1 Small Disturbance Approximation . . . 226
5.6.2 Method of Characteristics . . . 229
5.6.3 Rankine-Hugoniot Condition . . . 233
5.6.4 Expansion Fan . . . 236
5.6.5 Shock Waves . . . 241
5.6.6 Return of Phantom Traffic Jams . . . 245
5.6.7 Summary . . . 247
5.7 Cellular Automata Modeling . . . 248
Exercises . . . 254

6 Continuum Mechanics: One Spatial Dimension . . . 265
6.1 Introduction . . . 265
6.2 Coordinate Systems . . . 265
6.2.1 Material Coordinates . . . 266
6.2.2 Spatial Coordinates . . . 267
6.2.3 Material Derivative . . . 270
6.2.4 End Notes . . . 272
6.3 Mathematical Tools . . . 273
6.4 Continuity Equation . . . 275
6.4.1 Material Coordinates . . . 276
6.5 Momentum Equation . . . 277
6.5.1 Material Coordinates . . . 279
6.6 Summary of the Equations of Motion . . . 279
6.7 Steady-State Solution . . . 280
6.8 Constitutive Law for an Elastic Material . . . 282
6.8.1 Derivation of Strain . . . 284

6.8.2 Material Linearity . . . 286
6.8.3 End Notes . . . 289
6.9 Morphological Basis for Deformation . . . 290
6.9.1 Metals . . . 290
6.9.2 Elastomers . . . 293
6.10 Restrictions on Constitutive Laws . . . 294
6.10.1 Frame-Indifference . . . 295
6.10.2 Entropy Inequality . . . 298
6.10.3 Hyperelasticity . . . 302
Exercises . . . 304

7 Elastic and Viscoelastic Materials . . . 311
7.1 Linear Elasticity . . . 311
7.1.1 Method of Characteristics . . . 313
7.1.2 Laplace Transform . . . 316
7.1.3 Geometric Linearity . . . 327
7.2 Viscoelasticity . . . 328
7.2.1 Mass, Spring, Dashpot Systems . . . 329
7.2.2 Equations of Motion . . . 331
7.2.3 Integral Formulation . . . 335
7.2.4 Generalized Relaxation Functions . . . 337
7.2.5 Solving Viscoelastic Problems . . . 338
Exercises . . . 342

8 Continuum Mechanics: Three Spatial Dimensions . . . 351
8.1 Introduction . . . 351
8.2 Material and Spatial Coordinates . . . 352
8.2.1 Deformation Gradient . . . 353
8.3 Material Derivative . . . 356
8.4 Mathematical Tools . . . 358
8.4.1 General Balance Law . . . 361
8.5 Continuity Equation . . . 362
8.5.1 Incompressibility . . . 362
8.6 Linear Momentum Equation . . . 363
8.6.1 Stress Tensor . . . 364
8.6.2 Differential Form of Equation . . . 367
8.7 Angular Momentum . . . 367
8.8 Summary of the Equations of Motion . . . 368
8.9 Constitutive Laws . . . 368
8.9.1 Representation Theorem and Invariants . . . 372
8.10 Newtonian Fluid . . . 374
8.10.1 Pressure . . . 374
8.10.2 Viscous Stress . . . 375
8.11 Equations of Motion for a Viscous Fluid . . . 378
8.11.1 Incompressibility . . . 379

8.11.2 Boundary and Initial Conditions 380
8.12 Material Equations of Motion 383
8.12.1 Frame-Indifference 385
8.12.2 Elastic Solid 387
8.12.3 Linear Elasticity 389
8.13 Energy Equation 390
8.13.1 Incompressible Viscous Fluid 391
8.13.2 Elasticity 391
Exercises 394

9 Fluids 403
9.1 Newtonian Fluids 403
9.2 Steady Flow 404
9.2.1 Plane Couette Flow 405
9.2.2 Poiseuille Flow 408
9.3 Vorticity 411
9.3.1 Vortex Motion 412
9.4 Irrotational Flow 414
9.4.1 Potential Flow 417
9.5 Ideal Fluid 419
9.5.1 Circulation and Vorticity 420
9.5.2 Potential Flow 423
9.5.3 End Notes 426
9.6 Boundary Layers 427
9.6.1 Impulsive Plate 427
9.6.2 Blasius Boundary Layer 429
Exercises 434

A Taylor's Theorem 441
A.1 Single Variable 441
A.2 Two Variables 441
A.3 Multivariable Versions 442

B Fourier Analysis 445
B.1 Fourier Series 445
B.2 Fourier Transform 447

C Stochastic Differential Equations 449

D Identities 451
D.1 Trace 451
D.2 Determinant 451
D.3 Vector Calculus 452

E Equations for a Newtonian Fluid 453
E.1 Cartesian Coordinates 453
E.2 Cylindrical Coordinates 453

References 455

Index 463

Chapter 1
Dimensional Analysis

1.1 Introduction

Before beginning the material on dimensional analysis, it is worth considering a simple example that demonstrates what we are doing. One that qualifies as simple is the situation of when a object is thrown upwards. The resulting mathematical model for this is an equation for the height $x(t)$ of the projectile from the surface of the Earth at time t. This equation is determined using Newton's second law, $F = ma$, and the law of gravitation. The result is

$$\frac{d^2x}{dt^2} = -\frac{gR^2}{(R+x)^2}\,, \quad \text{for } \ 0 < t, \tag{1.1}$$

where g is the gravitational acceleration constant and R is the radius of the Earth. Finding the solution x of this equation requires two integrations. Each will produce an integration constant, and we need more information to find these constants. This is done by specifying the initial conditions. Assuming the projectile starts at the surface with velocity v_0 then the initial conditions are as follows

$$x(0) = 0, \tag{1.2}$$

$$\frac{dx}{dt}(0) = v_0. \tag{1.3}$$

The resulting initial value problem for x consists in finding the solution of (1.1) that satisfies (1.2) and (1.3). Mathematically, the problem is challenging because it involves solving a second-order nonlinear differential equation. One option for finding the solution is simply to use a computer. However, the limitation with this is that it does not provide much insight into how the solution depends on the terms in the equation. One of the primary objectives of this text is to use mathematics to derive a fundamental understanding of how and why things work the way they do, and so, we are very interested in

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_1,
© Springer Science+Business Media, LLC 2009

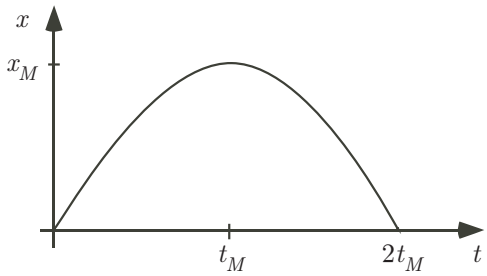


Figure 1.1 The solution (1.5) of the projectile problem in a uniform gravitational field.

obtaining at least an approximate solution of this problem. This is the same point-of-view taken in most physics books and it is worth looking at how they might address this issue. Adopting for the moment the typical Physics I approach, in looking at the equation in (1.1) it is not unreasonable to assume R is significantly larger than even the largest value of x. If true then we should be able to replace the $x+R$ term with just R. In this case the problem reduces to solving

$$\frac{d^2x}{dt^2} = -g, \quad \text{for} \;\; 0 < t. \tag{1.4}$$

Integrating and then using the two initial conditions yields

$$x(t) = -\frac{1}{2}gt^2 + v_0 t. \tag{1.5}$$

This solution is shown schematically in Figure 1.1. We have what we wanted, a relatively simple expression that serves as an approximation to the original nonlinear problem. To complete the derivation we should check that the assumption made in the derivation is satisfied, namely x is much smaller than R. Now, the maximum height for (1.5) occurs when

$$\frac{dx}{dt} = 0. \tag{1.6}$$

Solving this equation yields $t = v_0/g$ and from this it follows that the maximum height is

$$x_M = \frac{v_0^2}{2g}\,. \tag{1.7}$$

Therefore, we must require that $v^2/(2g)$ is much less than R, which we write as $v_0^2/(2g) << R$.

It is now time to critique the above derivation. The first criticism is that the approach is heuristic. The reason is that even though the argument for replacing $x + R$ with R seems plausible, we simply ignored a particular term in the equation. The projectile problem is not particularly complicated so

dropping a term as we did is straightforward. However, in the real world where problems can be quite complicated, dropping a term in one part of the problem can lead to inconsistencies in another part. A second criticism can be made by asking a question. Specifically, what exactly is the effect of the nonlinearity on the projectile? Our reduction replaced the nonlinear gravitational force, which is the right-hand side of (1.1), with a uniform gravitational field given by $-g$. Presumably if gravity decreases with height then the projectile will be going higher than we would expect based on our approximation in (1.5). It is of interest to understand quantitatively what this nonlinear effect is and whether it might interfere with our reduction.

Based on the comments of the previous paragraph we need to make the reduction process more systematic. The procedure that is used to simplify the problem should enable us to know exactly what is large or small in the problem, and it should also enable us to construct increasingly more accurate approximations to the problem. Explaining what is involved in a systematic reduction occurs in two steps. The first, which is the objective of this chapter, involves the study of dimensions and how these can be used to simplify the mathematical formulation of the problem. After this, in Chapter 2, we develop techniques to construct accurate approximations of the resulting equations.

1.2 Examples of Dimensional Reduction

The first idea that we explore will, on the surface, seem to be rather simple, but it is actually quite profound. It has to do with the dimensions of the physical variables, or parameters, in a problem. To illustrate, suppose we know that the speed s of a ball is determined by its radius r and the length of time t it has been moving. Implicit in this statement is the assumption that the speed does not depend on any other physical variable. In mathematical terms we have that $s = f(r, t)$. The function f is not specified and all we know is that there is some expression that connects the speed with r and t. The only possible way to combine these two quantities to produce the dimension of speed is through their ratio r/t. For example, it is impossible to have $s = \alpha r + \beta t$ without α and β having dimensions. This would mean α and β are physical parameters, and we have assumed there are no others in the problem. This observation enables us to conclude that based on the original assumptions that the only function we can have is $s = \alpha r/t$, where α is a number.

What we are seeing in this example is that the dimensions of the variables in the problem end up dictating the form of the function. This is very useful information and we will spend some time exploring how to exploit this idea. To set the stage we need to introduce some of the terminology. The first is the concept of a fundamental dimension. As is well known, physical variables such as force, density, and velocity can be broken down into length L, time

Quantity	**Dimensions**	**Quantity**	**Dimensions**
Acceleration	LT^{-2}	Enthalpy	ML^2T^{-2}
Angle	1	Entropy	$ML^2T^{-2}\theta^{-1}$
Angular Acceleration	T^{-2}	Gas Constant	$L^2T^{-2}\theta^{-1}$
Angular Momentum	ML^2T^{-1}	Internal Energy	ML^2T^{-2}
Angular Velocity	T^{-1}	Specific Heat	$L^2T^{-2}\theta^{-1}$
Area	L^2	Temperature	θ
Energy, Work	ML^2T^{-2}	Thermal Conductivity	$MLT^{-3}\theta^{-1}$
Force	MLT^{-2}	Thermal Diffusivity	L^2T^{-1}
Frequency	T^{-1}	Heat Transfer Coefficient	$MT^{-3}\theta^{-1}$
Concentration	L^{-3}		
Length	L	Capacitance	$M^{-1}L^{-2}T^4I^2$
Mass	M	Charge	TI
Mass Density	ML^{-3}	Charge Density	$L^{-3}TI$
Momentum	MLT^{-1}	Conductivity	$M^{-1}L^{-3}T^3I^2$
Power	ML^2T^{-3}	Electric Current Density	$L^{-2}I$
Pressure, Stress, Elastic Modulus	$ML^{-1}T^{-2}$	Electric Current	I
Surface Tension	MT^{-2}	Electric Displacement	$L^{-2}TI$
Time	T	Electric Potential	$ML^2T^{-3}I^{-1}$
Torque	ML^2T^{-2}	Electric Field Intensity	$MLT^{-3}I^{-1}$
Velocity	LT^{-1}	Inductance	$ML^2T^{-2}I^{-2}$
Viscosity (Dynamic)	$ML^{-1}T^{-1}$	Magnetic Field Intensity	$L^{-1}I$
Viscosity (Kinematic)	L^2T^{-1}	Magnetic flux	$L^2MT^{-2}I^{-1}$
Volume	L^3	Permeability	$MLT^{-2}I^{-2}$
Wave Length	L	Permittivity	$M^{-1}L^{-3}T^4I^2$
Strain	1	Electric Resistance	$ML^2T^{-3}I^{-2}$

Table 1.1 Fundamental dimensions for commonly occurring quantities. A quantity with a one in the dimensions column is dimensionless.

T, and mass M (see Table 1.1). Moreover, length, time, and mass are independent in the sense that one of them cannot be written in terms of the other two. For these two reasons we will consider L, T, and M as *fundamental dimensions*. For problems involving thermodynamics we will expand this list to include temperature (θ) and for electrical problems we add current (I). In conjunction with this, given a physical variable x we will designate the fundamental dimensions of x using the notation $[x]$. For example, $[velocity] = L/T$, $[force] = ML/T^2$, $[g] = L/T^2$, and $[density] = M/L^3$.

It is important to understand that nothing is being assumed about which specific system of units is used to determine the values of the variables or parameters. Dimensional analysis requires that the equations be independent of the system of units. For example, both Newton's law $F = ma$ and the differential equation (1.1) do not depend on the specific system one selects. For this reason these equations are said to be *dimensionally homogeneous*. If one were to specialize (1.1) to SI units and set $R = 6378\,\mathrm{km}$ and $g = 9.8\,\mathrm{m/sec}^2$ they would end up with an equation that is not dimensionally homogeneous.

1.2.1 Maximum Height of a Projectile

The process of dimensional reduction will be explained by applying it to the projectile problem To set the stage, suppose we are interested in the maximum height x_M of the projectile. Based on Newton's second law, and the initial conditions in (1.2) and (1.3), it is assumed that the only physical parameters that x_M depends on are g, v_0, and the mass m of the projectile. Mathematically this assumption is written as $x_M = f(g, m, v_0)$. The function f is unknown but we are going to see if the dimensions can be used to simplify the expression. The only way to combine g, m, v_0 to produce the correct dimensions is through a product or ratio. So, our start-off hypothesis is that there are numbers a, b, c so that

$$[x_M] = [m^a v_0^b g^c]. \tag{1.8}$$

Using the fundamental dimensions for these variables the above equation is equivalent to

$$\begin{aligned} L &= M^a (L/T)^b (L/T^2)^c \\ &= M^a L^{b+c} T^{-b-2c}. \end{aligned} \tag{1.9}$$

Equating the exponents of the respective terms in this equation we conclude

$$\begin{aligned} L: &\quad b + c = 1, \\ T: &\quad -b - 2c = 0, \\ M: &\quad a = 0. \end{aligned}$$

Solving these equations we obtain $a = 0$, $b = 2$, and $c = -1$. This means the only way to produce the dimensions of length using m, v_0, g is through the ratio v_0^2/g. Given our start-off assumption (1.8), we conclude that x_M is proportional to v_0^2/g. In other words, the original assumption that $x_M = f(g, m, v_0)$ dimensionally reduces to the expression

$$x_M = \alpha \frac{v_0^2}{g}\,, \tag{1.10}$$

where α is an arbitrary number. With (1.10) we have come close to obtaining our earlier result (1.7) and have done so without solving a differential equation or using calculus to find the maximum value. Based on this rather minimal effort we can make the following observations:

- If the initial velocity is increased by a factor of 2 then the maximum height will increase by a factor of 4. This observation offers an easy method for experimentally checking on whether the original modeling assumptions are correct.
- The constant α can be determined by running one experiment. Namely, for a given initial velocity $v_0 = \bar{v}_0$ we measure the maximum height $x_M = \bar{x}_M$. With these known values, $\alpha = g\bar{x}_M/\bar{v}_0^2$. Once this is done, the formula in (1.10) can be used to determine x_M for any v_0.

The steps we have used are the basis for the method of dimensional reduction, where an expression is simplified based on the fundamental dimensions of the quantities involved. Given how easy it was to obtain (1.10) the method is very attractive as an analysis tool. It does have limitations and one is that we do not know the value of the number α. It also requires us to be able to identify at the beginning what parameters are needed. The importance of this and how this relies on understanding the physical laws underlying the problem will be discussed later.

The purpose of the above example is to introduce the idea of dimensional reduction. What it does not show is how to handle problems with several parameters and this is the purpose of the next two examples.

1.2.2 Drag on a Sphere

In the design of automobiles, racing bicycles, and aircraft there is an overall objective to keep the drag on the object as small as possible. It is interesting to see what insight dimensional analysis might provide in such a situation, but since we are beginners it will be assumed the object is very simple and is a sphere (see Figure 1.2). The modeling assumption that is made is that the drag force D_F on the sphere depends on the radius R of the sphere, the velocity v of the sphere, the density ρ of the air, and the dynamic viscosity μ of the air. The latter is a measure of the resistance force of the air to motion and we will investigate this in Chapter 8. For the moment all we need is its fundamental dimensions and these are given in Table 1.1. In mathematical terms the modeling assumption is

$$D_F = f(R, v, \rho, \mu), \tag{1.11}$$

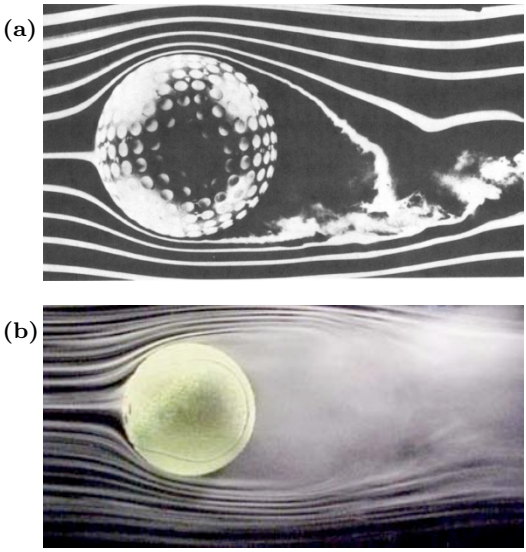


Figure 1.2 Air flow around an object can be visualized using smoke. The flow around a golf ball is shown in (a) (Brown [1971]), and around a tennis ball in (b) (Bluck [2000]). In both cases the air is moving from left to right.

and we want to use dimensional reduction to find a simplified version of this expression. Similar to the last example, the first question is whether we can find numbers a, b, c, d so that

$$[D_F] = [R^a v^b \rho^c \mu^d]. \tag{1.12}$$

Expressing these using fundamental dimensions yields

$$\begin{aligned} MLT^{-2} &= L^a (L/T)^b (M/L^3)^c (M/LT)^d \\ &= L^{a+b-3c-d} T^{-b-d} M^{c+d}. \end{aligned}$$

As before we equate the respective terms and conclude

$$\begin{aligned} L: &\quad a + b - 3c - d = 1, \\ T: &\quad -b - d = -2, \\ M: &\quad c + d = 1. \end{aligned} \tag{1.13}$$

We have four unknowns and three equations, so it is anticipated that in solving the above system of equations one of the unknowns will be undetermined. From the T equation we have $b = 2 - d$, and from the M equation $c = 1 - d$. The L equation then gives us $a = 2 - d$. With these solutions, and based on our assumption in (1.12), we have that

$$\begin{aligned} D_F &= \alpha\rho R^{2-d} v^{2-d} \rho^{1-d} \mu^d \\ &= \alpha\rho R^2 v^2 \left(\frac{\mu}{Rv\rho}\right)^d, \end{aligned}$$

where α is an arbitrary number. This can be written as

$$D_F = \alpha\rho R^2 v^2 \Pi^d, \tag{1.14}$$

where

$$\Pi = \frac{\mu}{Rv\rho}, \tag{1.15}$$

and d, α are arbitrary numbers. This is the *general product solution* for how D_F depends on the given variables. The quantity Π is dimensionless, and it is an example of what is known as a dimensionless product. Physically, it can be thought of as the ratio of the viscous force (μ) to the inertial force ($Rv\rho$) in the air. Calling it a product is a bit misleading as Π involves both multiplications and divisions. Some avoid this by calling it a dimensionless group. We will use both expressions in this book.

The formula for D_F in (1.14) is not the final answer. What remains is to determine the consequence of the arbitrary exponent d. The key observation is that given any two sets of values for (α, d), say (α_1, d_1) and (α_2, d_2), then

$$\begin{aligned} D_F &= \alpha_1\rho R^2 v^2 \Pi^{d_1} + \alpha_2\rho R^2 v^2 \Pi^{d_2} \\ &= \rho R^2 v^2 (\alpha_1\rho\Pi^{d_1} + \alpha_2\Pi^{d_2}) \end{aligned}$$

is also a solution. Extending this observation we conclude that another solution is

$$D_F = \rho R^2 v^2 (\alpha_1\Pi^{d_1} + \alpha_2\Pi^{d_2} + \alpha_3\Pi^{d_3} + \cdots), \tag{1.16}$$

where $d_1, d_2, d_3, \ldots$ are arbitrary numbers as are the coefficients $\alpha_1, \alpha_2, \alpha_3, \ldots$. To express this in a more compact form, note that the expression within the parentheses in (1.16) is simply a function of Π. From this observation we obtain the *general solution*, which is

$$D_F = \rho R^2 v^2 F(\Pi), \tag{1.17}$$

where F is an arbitrary function of the dimensionless product Π. We have, therefore, been able to use dimensional analysis to reduce (1.11), which involves an unknown of four variables, down to an unknown function of one variable. Although this is a significant improvement, the result is perhaps not as satisfying as the one obtained for the projectile example, given in (1.10), because we have not been able to determine F. However, there are various ways to address this issue, and some of them will be considered below.

Representation of Solution

Now that the derivation is complete a few comments are in order. First, it is possible for two people to go through the above steps and come to what looks to be very different conclusions. For example, the general solution can also be written as

$$D_F = \frac{\mu^2}{\rho} H(\Pi), \tag{1.18}$$

where H is an arbitrary function of Π. The proof that this is equivalent to (1.17) comes from the requirement that the two expressions must produce the same result. In other words, it is required that

$$\frac{\mu^2}{\rho} H(\Pi) = \rho R^2 v^2 F(\Pi).$$

Solving this for H yields

$$H(\Pi) = \frac{1}{\Pi^2} F(\Pi).$$

The fact that the right-hand side of the above equation only depends on Π shows that (1.18) is equivalent to (1.17). As an example, if $F(\Pi) = \Pi$ in (1.17), then $H(\Pi) = 1/\Pi$ in (1.18).

Another representation for the general solution is

$$D_F = \rho R^2 v^2 G(Re), \tag{1.19}$$

where

$$Re = \frac{Rv\rho}{\mu}, \tag{1.20}$$

and G is an arbitrary function of Re. This form is the one usually used in fluid dynamics, where the dimensionless product Re is known as the Reynolds number. The function G is related to the drag coefficient C_D, through the equation $G = \frac{\pi}{2} C_D$. Because of its importance in fluids, G has been measured for a wide range of Reynolds numbers, producing the curve shown in Figure 1.3. To transform between the representation in (1.19), and the one in (1.17), note $Re = 1/\Pi$. From the requirement

$$\rho R^2 v^2 G(Re) = \rho R^2 v^2 F(\Pi),$$

we obtain

$$G(Re) = F(1/Re).$$

The reason for the different representations is that there are four unknowns in (1.12) yet only three equations. This means one of the unknowns is used in the general solution and, as expressed in (1.14), we used d. If you were to use one of the others then a different looking, but mathematically equivalent,

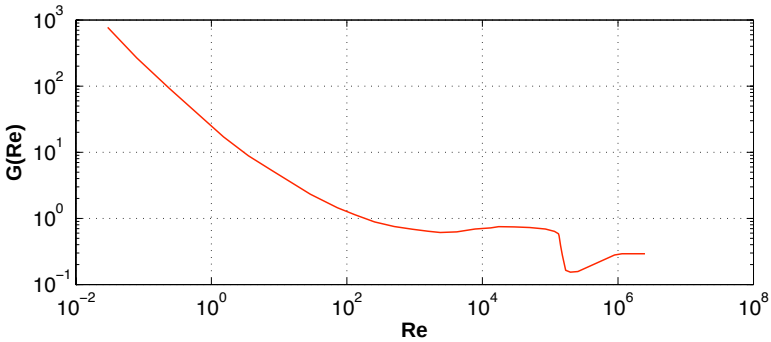


Figure 1.3 The measured values of the function $G(Re)$ that arises in the formula for the drag on a sphere, as given in (1.19).

expression would be obtained. The fact that there are multiple ways to express the solution can be used to advantage. For example, if one is interested in the value of D_F for small values of the velocity then (1.19) would be a bit easier to use. The reason is that to investigate the case of small v it is somewhat easier to determine what happens to G for Re near zero than to expand F for large values of Π. For the same reason, (1.17) is easier to work with for studying large velocities. One last comment to make is that even though there are choices on the form of the general solution, they all have exactly the same number of dimensionless products.

Determining F

A more challenging question concerns how to determine the function F in (1.17). The mathematical approach would be to solve the equations for fluid flow around a sphere and from this find F. This is an intriguing idea and one that will be used from time to time in this book. There is, however, another more applied approach that makes direct use of (1.17). Specifically, a sequence of experiments is run to measure $F(r)$ for $0 < r < \infty$. To do this a sphere with a given radius R_0, and a fluid with known density ρ_0 and viscosity μ_0, are selected. In this case (1.17) can be written as

$$F(r) = \frac{\gamma D_F}{v^2} \tag{1.21}$$

where $\gamma = 1/(\rho_0 R_0^2)$ is known and fixed. The experiment consists of taking various values of v and then measuring the resulting drag force D_F on the sphere. To illustrate, suppose our choice for the sphere and fluid give $R = 1$, $\rho_0 = 2$, and $\mu_0 = 3$. Also, suppose that running the experiment using $v = 4$ produces a measured drag of $D_F = 5$. In this case $r = \mu_0/(R_0 v \rho_0) = 3/8$ and $\gamma D_F/v^2 = 5/32$. Our conclusion is therefore that $F(3/8) = 5/32$. In this

way, picking a wide range of v values we will be able to determine the values for the function $F(r)$. This approach is used extensively in the real world and the example we are considering has been a particular favorite for study. The data determined from such experiments are shown in Figure 1.3.

A number of conclusions can be drawn from Figure 1.3. For example, there is a range of Re values where G is approximately constant. Specifically, if $10^3 < Re < 10^5$ then $G \approx 0.7$. This is the reason why in the fluid dynamics literature you will occasionally see the statement that the drag coefficient $C_D = \frac{2}{\pi}G$ for a sphere has a constant value of approximately 0.44. For other Re values, however, G is not constant. Of particular interest, is the dependence of G for small values of Re. This corresponds to velocities v that are very small, what is known as Stokes flow. The data in Figure 1.3 show that G decreases linearly with Re in this region. Given that this is a log-log plot, then this means that $\log(G) = a - b\log(Re)$, or equivalently, $G = \alpha/Re^b$ where $\alpha = 10^a$. Curve fitting this function to the data in Figure 1.3 it is found that $\alpha \approx 17.6$ and $\beta \approx 1.07$. These are close to the exact values of $\alpha = 6\pi$ and $\beta = 1$, which are obtained by solving the equations of motion for Stokes flow. Inserting these values into (1.19), the conclusion is that the drag on the sphere for small values of the Reynolds number is

$$D_F \approx 6\pi\mu R v. \tag{1.22}$$

This is known as Stokes formula for the drag on a sphere, and we will have use for it in Chapter 4 when studying diffusion.

Scale Models

Why all the work to find F? Well, knowing this function allows for the use of scale model testing. To explain, suppose it is required to determine the drag on a sphere with radius R_f for a given velocity v_f when the fluid has density ρ_f and viscosity μ_f. Based on (1.17) we have $D_F = \rho_f R_f^2 v_f^2 F(\Pi_f)$, where

$$\Pi_f = \frac{\mu_f}{R_f v_f \rho_f}. \tag{1.23}$$

Consequently, we can determine D_F if we know the value of F at Π_f. Also, suppose that this cannot be measured directly as R_f is large and our experimental equipment can only handle small spheres. We can still measure $F(\Pi_f)$ using a small value of R if we change one or more of the parameters in such a way that the value of Π_f does not change. If R_m, μ_m, ρ_m and v_m are the values used in the experiment then we want to select them so that

$$\frac{\mu_m}{R_m v_m \rho_m} = \frac{\mu_f}{R_f v_f \rho_f}, \tag{1.24}$$

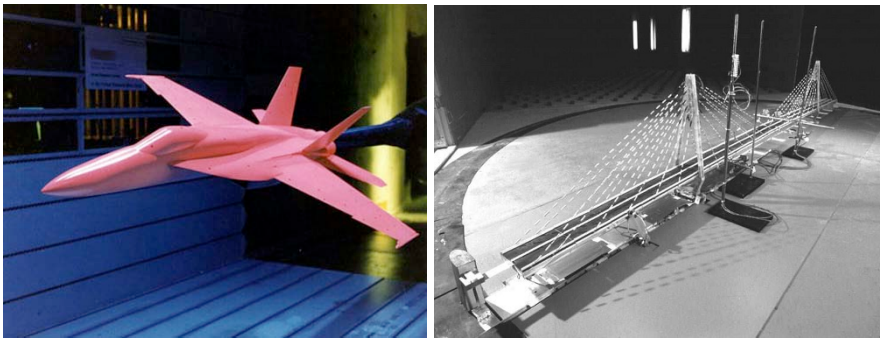


Figure 1.4 Scale model testing. Dimensional analysis is used in the development of scale models used in wind tunnels. On the left there is a flight test of an F-18 model in NASA's 11 ft transonic wind tunnel (NASA [2008]), and on the right a wind tunnel test of a 1:160 scale model of the Owensboro Bridge (Hsu [2009]).

or equivalently

$$v_m = \frac{\mu_m R_f \rho_f}{\mu_f R_m \rho_m} v_f. \tag{1.25}$$

This equation relates the values for the full-scale ball (subscript f) to those for the model used in the experiment (subscript m). As an example, suppose we are interested in the drag on a very large sphere, say $R_f = 100\,m$, but our equipment can only handle smaller values, say $R_m = 2\,m$. If the fluid for the two cases is the same, so $\rho_m = \rho_f$ and $\mu_m = \mu_f$, then according to (1.25), in our experiment we should take $v_m = 50 v_f$. If the experimental apparatus is unable to generate velocities 50 times the value of v_f then it would be necessary to use a different fluid to reduce this multiplicative factor.

The result in the above example is the basis of scale model testing used in wind tunnels (see Figure 1.4). Usually these tests involve more than just keeping one dimensionless product constant as we did in (1.24). Moreover, it is evident in Figure 1.4 that the models look like the originals, they are just smaller. This is the basis of geometric similarity, where the lengths of the model are all a fraction of the original. For example, the bridge in Figure 1.4 is a $\frac{1}{20}$th scale model of the Owensboro Bridge. Other scalings are sometimes used and the most common are kinematic similarity, where velocities are scaled, and dynamic similarity, where forces are scaled.

Endnotes

One question that has not been considered so far is, how do you know to assume that the drag force depends on the radius, velocity, density, and dynamic viscosity? The assumption comes from knowing the laws of fluid dynamics, and identifying the principal terms that contribute to the drag.

For the most part, in this chapter the assumptions will be stated explicitly, as they were in this example. Later in the text, after the basic physical laws are developed, it will be possible to construct the assumptions directly. However, one important observation can be made, and that is the parameters used in the assumption should be independent. For example, even though the drag on a sphere likely depends on the surface area and volume of the sphere it is not necessary to include them in the list. The reason is that it is already assumed that D_F depends on the radius R and both the surface area and volume are determined using R.

The problem of determining the drag on a sphere is one of the oldest in fluid dynamics. Given that the subject is well over 150 years old, you would think that whatever useful information can be derived from this particular problem was figured out long ago. Well, apparently not, as research papers still appear regularly on this topic. A number of them come from the sports industry, where there is interest in the drag on soccer balls (Asai et al. [2007]), golf balls (Smits and Ogg [2004]), tennis balls (Goodwill et al. [2004]), as well as nonspherical-shaped balls (Mehta [1985]). Others have worked on how to improve the data in Figure 1.3, and an example is the use of a magnetic suspension system to hold the sphere (Sawada and Kunimasu [2004]). A more novel idea is to drop different types of spheres down a deep mine shaft, and then use the splash time as a means to determine the drag coefficient (Maroto et al. [2005]). The point here is that even the most studied problems in science and engineering still have interesting questions that remain unanswered.

1.2.3 Toppling Dominoes

Domino toppling refers to the art of setting up dominoes, and then knocking them down. The current world record for this is 4,000,000 plus dominoes for a team, and 300,000 plus for an individual. One of the more interesting aspects of this activity is that as the dominoes fall it appears as if a wave is propagating along the line of dominoes. The objective of this example is to examine what dimensional analysis might be able to tell us about the velocity of this wave. A schematic of the situation is shown in Figure 1.5. The assumption is that the velocity v depends on the spacing d, height h, thickness t, and the gravitational acceleration constant g. Therefore, the modeling assumption is $v = f(d, h, t, g)$ and we want to use dimensional reduction to find a simplified version of this expression. As usual, the first step is to find numbers a, b, c, e so that

$$[v] = [d^a h^b t^c g^e].$$

Expressing these using fundamental dimensions yields

$$\begin{aligned} LT^{-1} &= L^a L^b L^c (L/T^2)^e \\ &= L^{a+b+c+e} T^{-2e}. \end{aligned}$$

Equating the respective terms we obtain

$$\begin{aligned} L: &\quad a+b+c+e = 1, \\ T: &\quad -2e = -1. \end{aligned}$$

Solving these two equations gives us that $e = \frac{1}{2}$ and $b = \frac{1}{2} - a - c$. With this we have that

$$\begin{aligned} v &= \alpha d^a h^{1/2-a-c} t^c g^{1/2} \\ &= \alpha \sqrt{hg} \left(\frac{d}{h}\right)^a \left(\frac{t}{h}\right)^c \\ &= \alpha \sqrt{hg} \Pi_1^a \Pi_2^c, \end{aligned} \tag{1.26}$$

where α is an arbitrary number, and the two dimensionless products are

$$\begin{aligned} \Pi_1 &= \frac{d}{h}, \\ \Pi_2 &= \frac{t}{h}. \end{aligned}$$

The expression in (1.26) is the general product solution. Therefore, the general solution for how the velocity depends on the given parameters is

$$v = \sqrt{hg}\, F(\Pi_1, \Pi_2), \tag{1.27}$$

where F is an arbitrary function of the two dimensionless products. The proof of how (1.27) follows from (1.26) is very similar to the method used to derive (1.17) from (1.15).

Dimensional analysis has been able to reduce the original assumption involving a function of four-dimensional parameters down to one involving two dimensionless products. This example is also informative as it demonstrates how to obtain the general solution when more than one dimensionless product is involved. The question remains, however, if this really applies to toppling

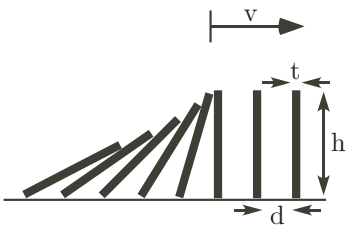


Figure 1.5 Schematic of toppling dominoes, creating a wave that propagates with velocity v.

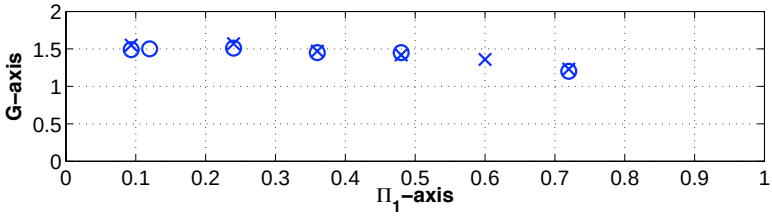


Figure 1.6 Data for two different types of toppling dominoes (Stronge and Shu [1988]). In these experiments, $t = 0.12h$, so the thin domino approximation is appropriate.

dominoes. It does, but in using this formula it is usually assumed the dominoes are very thin, or more specifically that $t << h$. This means that it is possible to assume $\Pi_2 = 0$, and (1.27) simplifies to

$$v = \sqrt{hg}\, G(\Pi_1), \tag{1.28}$$

where G is an arbitrary function. Some effort has been made to measure G, and the measurements for two different types of dominoes are given in Figure 1.6. Although the data show that G decreases with Π_1, it is approximately constant over the range of Π_1 values used in the experiments. Therefore, as an approximation we conclude that the speed at which dominoes topple is $v \approx 1.5\sqrt{hg}$. A typical domino has $h = 5$ cm, which results in a velocity of $v \approx 1$ m/s. To obtain a more explicit formula for G, however, requires the solution of a challenging mathematical problem, and an expanded discussion of this can be found in Efthimiou and Johnson [2007].

1.2.4 Endnotes

Based on the previous examples, the benefits of using dimensional reduction are apparent. However, a word of caution is needed here as the method gives the impression that it is possible to derive useful information without getting involved with the laws of physics or potentially difficult mathematical problems. One consequence of this is that the method is used to comment on situations and phenomena that are simply inappropriate (e.g., to study psychoacoustic behavior). The method relies heavily on knowing the fundamental laws for the problem under study, and without this whatever conclusions made using dimensional reduction are limited. For example, we earlier considered the drag on a sphere and in the formulation of the problem we assumed that the drag depends on the dynamic viscosity. Without knowing the equations of motion for fluids it would not have been possible to know that this term needed to be included or what units it might have. By not in-

cluding it we would have concluded that $d = 0$ in (1.14) and instead of (1.17) we would have $D_F = \alpha\rho R^2 v^2$ where α is a constant. In Figure 1.3 it does appear that D_F is approximately independent of Re when $10^3 < Re < 10^5$. However, outside of this interval, D_F is strongly dependent on Re, and this means ignoring the viscosity would be a mistake. Another example illustrating the need to know the underlying physical laws arises in the projectile problem when we included the gravitational constant. Again, this term is essential and without some understanding of Newtonian mechanics it would be missed completely. The point here is that dimensional reduction can be a very effective method for simplifying complex relationships, but it is based heavily on knowing what the underlying laws are that govern the systems being studied.

1.3 Theoretical Foundation

The theoretical foundation for dimensional reduction is contained in the Buckingham Pi Theorem. To derive this result assume we have a physical quantity q that depends on physical parameters or variables $p_1, p_2, \ldots, p_n$. In this context, the word physical means that the quantity is measurable. Each can be expressed in fundamental dimensions and we will assume that the L, T, M system is sufficient for this task. In this case we can write

$$[q] = L^{\ell_0} T^{t_0} M^{m_0}, \tag{1.29}$$

and

$$[p_i] = L^{\ell_i} T^{t_i} M^{m_i}. \tag{1.30}$$

Our modeling assumption is that $q = f(p_1, p_2, \ldots, p_n)$. To dimensionally reduce this expression we will determine if there are numbers $a_1, a_2, \ldots, a_n$ so that

$$[q] = [p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n}]. \tag{1.31}$$

Introducing (1.29) and (1.30) into the above expression, and then equating exponents, we obtain the equations

$$\begin{aligned} L: &\quad \ell_1 a_1 + \ell_2 a_2 + \cdots + \ell_n a_n = \ell_0, \\ T: &\quad t_1 a_1 + t_2 a_2 + \cdots + t_n a_n = t_0, \\ M: &\quad m_1 a_1 + m_2 a_2 + \cdots + m_n a_n = m_0. \end{aligned}$$

This can be expressed in matrix form as

$$\mathbf{A}\mathbf{a} = \mathbf{b}, \tag{1.32}$$

where

$$\mathbf{A} = \begin{pmatrix} \ell_1 & \ell_1 & \cdots & \ell_n \\ t_1 & t_2 & \cdots & t_n \\ m_1 & m_2 & \cdots & m_n \end{pmatrix}, \tag{1.33}$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} \ell_0 \\ t_0 \\ m_0 \end{pmatrix}. \tag{1.34}$$

The matrix $\mathbf{A}$ is known as the dimension matrix. As expressed in (1.33) it is $3 \times n$ but if we were to have used L, T, M, θ as the fundamental system then it would be $4 \times n$. In other words, the number of rows in the dimension matrix equals the number of fundamental units needed, and the number of columns equals the number of parameters that q is assumed to depend on.

With (1.32) we have transformed the dimensional reduction question into a linear algebra problem. To determine the consequences of this we first consider the situation that (1.32) has no solution. In this case the assumption that q depends on $p_1, p_2, \ldots, p_n$ is incomplete and additional parameters are needed. This situation motivates the following definition.

Definition 1.1. The set $p_1, p_2, \ldots, p_n$ is dimensionally incomplete for q if it is not possible to combine the p_i's to produce a quantity with the same dimension as q. If it is possible, the set is dimensionally complete for q.

From this point on we will assume the p_i's are complete and there is at least one solution of (1.32). To write down the general solution we consider the associated homogeneous equation, namely $\mathbf{Aa} = \mathbf{0}$. The set of solutions of this equation form a subspace $K(\mathbf{A})$, known as the kernel of $\mathbf{A}$. Letting k be the dimension of this subspace then the general solution of $\mathbf{Aa} = \mathbf{0}$ can be written as $\mathbf{a} = \gamma_1 \mathbf{a}_1 + \gamma_2 \mathbf{a}_2 + \cdots + \gamma_k \mathbf{a}_k$, where $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ is a basis for $K(\mathbf{A})$ and $\gamma_1, \gamma_2, \ldots, \gamma_k$ are arbitrary. It is understood here that if $k = 0$ then $\mathbf{a} = \mathbf{0}$. With this, the general solution of (1.32) can be written as

$$\mathbf{a} = \mathbf{a}_p + \gamma_1 \mathbf{a}_1 + \gamma_2 \mathbf{a}_2 + \cdots + \gamma_k \mathbf{a}_k, \tag{1.35}$$

where $\mathbf{a}_p$ is any vector that satisfies (1.32) and $\gamma_1, \gamma_2, \ldots, \gamma_k$ are arbitrary numbers.

Example: Drag on a Sphere

To connect the above discussion with what we did earlier consider the drag on a sphere example. Writing (1.13) in matrix form we obtain

$$\begin{pmatrix} 1 & 1 & -3 & -1 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

This is the matrix equation (1.32) for this particular example. Putting this in augmented form, and row reducing, yields the following

$$\left(\begin{array}{rrrr|r} 1 & 1 & -3 & -1 & 1 \\ 0 & -1 & 0 & -1 & -2 \\ 0 & 0 & 1 & 1 & 1 \end{array}\right) \to \left(\begin{array}{rrrr|r} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 \end{array}\right).$$

From this we conclude that $a = 2 - d$, $b = 2 - d$, and $c = 1 - d$. To be consistent with the notation in (1.35), set $d = \gamma$, so the solution is

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \end{pmatrix},$$

where γ is arbitrary. Comparing this with (1.35) we have that $k = 1$,

$$\mathbf{a}_p = \begin{pmatrix} 2 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{a}_1 = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \end{pmatrix}. \quad \blacksquare$$

It is now time to take our linear algebra conclusions and apply them to the dimensional reduction problem. Just as the appearance of d in (1.14) translated into the appearance of a dimensionless product in the general solution given in (1.17), each of the γ_i's in (1.35) gives rise to a dimensionless product in the general solution for the problem we are currently studying. To be specific, writing the ith basis vector $\mathbf{a}_i$ in component form as

$$\mathbf{a}_i = \begin{pmatrix} \alpha \\ \beta \\ \vdots \\ \gamma \end{pmatrix}, \tag{1.36}$$

then the corresponding dimensionless product is

$$\Pi_i = p_1^\alpha p_2^\beta \cdots p_n^\gamma. \tag{1.37}$$

Moreover, because the $\mathbf{a}_i$'s are independent vectors, the dimensionless products $\Pi_1, \Pi_2, \ldots, \Pi_k$ are independent.

As for the particular solution $\mathbf{a}_p$ in (1.35), assuming it has components

$$\mathbf{a}_p = \begin{pmatrix} a \\ b \\ \vdots \\ c \end{pmatrix}, \tag{1.38}$$

then the quantity

$$Q = p_1^a p_2^b \cdots p_n^c \tag{1.39}$$

has the same dimensions as q.

Based on the conclusions of the previous two paragraphs, the general product solution is $q = \alpha Q \Pi_1^{\kappa_1} \Pi_2^{\kappa_2} \cdots \Pi_k^{\kappa_k}$, where $\alpha, \kappa_1, \kappa_2, \ldots, \kappa_k$ are arbitrary constants. From this we obtain the following theorem.

Theorem 1.1. *Assuming the formula $q = f(p_1, p_2, \ldots, p_n)$ is dimensionally homogeneous and dimensionally complete, then it is possible to reduce it to one of the form $q = Q\,F(\Pi_1, \Pi_2, \ldots, \Pi_k)$, where $\Pi_1, \Pi_2, \ldots, \Pi_k$ are independent dimensionless products of $p_1, p_2, \ldots, p_n$. The quantity Q is a dimensional product of $p_1, p_2, \ldots, p_n$ with the same dimensions as q.*

According to this theorem, the original formula for q can be reduced from a function of n variables down to one with k. The value of k, which equals the nullity of the dimension matrix, ranges from 0 to $n-1$ depending on the given quantities $p_1, p_2, \ldots, p_n$. In the case that $k = 0$ the function F reduces to a constant and the conclusion is that $q = \alpha Q$, where α is an arbitrary number.

The importance of this theorem is that it establishes that the process used to reduce the drag on a sphere and toppling dominoes examples can be applied to much more complex problems. It also provides insight into how the number of dimensionless products is determined. There are still, however, fundamental questions left unanswered. For example, those with a more mathematical bent might still be wondering if this result can really be true no matter how discontinuous the original function f might be. Others might be wondering if the fundamental units used here, particularly length and time, are really independent. This depth of inquiry, although quite interesting, is beyond the scope of this text. Those wishing to pursue further study of these and related topics should consult Penrose [2007] and Bluman and Anco [2002].

1.3.1 Pattern Formation

The mechanism responsible for the colorful patterns on seashells, butterfly wings, zebras, and the like has intrigued scientists for decades. An experiment that has been developed to study pattern formation involves pouring chemicals into one end of a long tube, and then watching what happens as they interact while moving along the tube. This apparatus is called a plug-flow

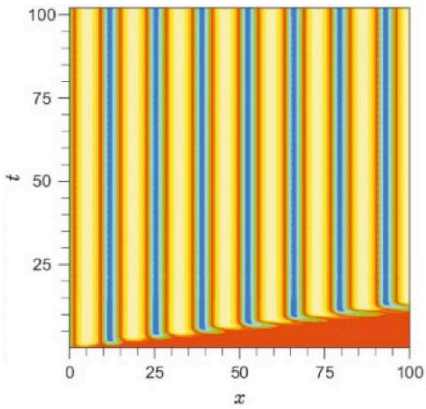


Figure 1.7 Spatial pattern created in a plug-flow reactor (Bamforth et al. [2000]). The tube occupies the interval $0 \leq x \leq 100$, and starting at $t = 0$ the chemicals are poured into the left end. As they flow along the tube a striped pattern develops.

reactor and the outcome of one such experiment is shown in Figure 1.7. It was found in these experiments that patterns appear only for certain pouring velocities v. According to what is known as the Lengyel-Epstein model, this velocity depends on the concentration U of the chemical used in the experiment, the rate k_2 at which the chemicals interact, the diffusion coefficient D of the chemicals, and a parameter k_3 that has the dimensions of concentration squared. The model is therefore assuming

$$v = f(U, k_2, D, k_3). \tag{1.40}$$

From Table 1.1 we have that $[v] = L/T$, $[U] = 1/L^3$, $[D] = L^2/T$, and $[k_3] = 1/L^6$. Also, from the Lengyel-Epstein model one finds that $[k_2] = L^3/T$. Using dimensional reduction we require

$$[v] = [U^a k_2^b D^c k_3^d]. \tag{1.41}$$

Expressing these using fundamental dimensions yields

$$\begin{aligned} LT^{-1} &= (L^{-3})^a (L^3 T^{-1})^b (L^2 T^{-1})^c (L^{-6})^d \\ &= L^{-3a+3b+2c-6d} T^{-b-c}. \end{aligned}$$

As before we equate the respective terms and conclude

$$\begin{aligned} L &: -3a + 3b + 2c - 6d = 1 \\ T &: \qquad\qquad\quad -b - c = -1. \end{aligned}$$

These equations will enable us to express two of the unknowns in terms of the other two. There is no unique way to do this, and one choice yields

$b = -1 + 3a + 6d$ and $c = 2 - 3a - 6d$. From this it follows that the general product solution is

$$\begin{aligned} v &= \alpha U^a k_2^{3a+6d-1} D^{2-3a-6d} k_3^d \\ &= \alpha k_2^{-1} D^2 (U k_2^3 D^{-3})^a (k_2^6 D^{-6} k_3)^d. \end{aligned}$$

This can be rewritten as

$$v = \alpha k_2^{-1} D^2 \Pi_1^a \Pi_2^d, \tag{1.42}$$

where

$$\Pi_1 = \frac{U k_2^3}{D^3}, \tag{1.43}$$

and

$$\Pi_2 = \frac{k_2^6 k_3}{D^6}. \tag{1.44}$$

The dimensionless products Π_1 and Π_2 are independent, and this follows from the method used to derive these expressions. Independence is also evident from the observation that Π_1 and Π_2 do not involve exactly the same parameters. From this result it follows that the general form of the reduced equation is

$$v = k_2^{-1} D^2 F(\Pi_1, \Pi_2). \tag{1.45}$$

It is of interest to compare (1.45) with the exact formula obtained from solving the differential equations coming from the Lengyel-Epstein model. It is found that

$$v = \sqrt{k_2 D U} G(\beta), \tag{1.46}$$

where $\beta = k_3/U^2$ and G is a rather complicated square root function (Bamforth et al. [2000]). This result appears to differ from (1.45). To investigate this, note that $\beta = \Pi_2/\Pi_1^2$. Equating (1.45) and (1.46) it follows that

$$\begin{aligned} F(\Pi_1, \Pi_2) &= \frac{k_2^{3/2} U^{1/2}}{D^{3/2}} G(\beta) \\ &= \sqrt{\Pi_1} G(\Pi_2/\Pi_1^2). \end{aligned}$$

Because the right-hand side is a function of only Π_1 and Π_2 then (1.45) does indeed reduce to the exact result (1.46). Dimensional reduction has therefore successfully reduced the original unknown function of four variables in (1.40) down to one with only two variables. However, the procedure is not able to reduce the function down to one dimensionless variable, as given in (1.46). In this problem that level of reduction requires information only available from the differential equations, something that dimensional arguments are not able to discern.

1.4 Similarity Variables

Dimensions can be used not just to reduce formulas, they can be also used to simplify complex mathematical problems. The degree of simplification depends on the parameters, and variables, in the problem. One of the more well-known examples is the problem of finding the density $u(x,t)$ of a chemical over the interval $0 < x < \infty$. In this case the density satisfies the diffusion equation

$$D\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \tag{1.47}$$

where the boundary conditions are

$$u|_{x=0} = u_0, \quad u|_{x\to\infty} = 0, \tag{1.48}$$

and the initial condition is

$$u|_{t=0} = 0. \tag{1.49}$$

The constant D is called the diffusion coefficient, and its dimensions can be determined from the terms in the differential equation. In particular, the dimensions of the left and right sides of (1.47) must be the same, and this means $[Du_{xx}] = [u_t]$. Because $[u] = M/L^3$ then $[u_{xx}] = [u]/L^2 = M/L^5$ and $[u_t] = [u]/T = M/(TL^3)$. From this we have $[D]M/L^5 = M/(TL^3)$, and therefore $[D] = L^2/T$. In a similar manner, in boundary condition (1.48), $[u_0] = [u] = M/L^3$. As a final comment, the physical assumptions underlying the derivation of (1.47) are the subject of Chapter 4. In fact, the solution we are about to derive is needed in Section 4.5.2 to solve the diffusion equation.

Dimensional Reduction

The conventional method for solving the diffusion equation on a semi-infinite spatial interval is to use an integral transform, and this will be considered in Chapter 4. It is also possible to find u using dimensional reduction. The approach is based on the observation that the only dimensional variables, and parameters, appearing in the problem are u, u_0, D, x, and t. In other words, it must be true that $u = f(x,t,D,u_0)$. With this we have the framework for dimensional reduction, and the question is whether we can find numbers a,b,c,d so that

$$[u] = [x^a t^b D^c (u_0)^d]. \tag{1.50}$$

Using fundamental dimensions yields

$$\begin{aligned} ML^{-3} &= L^a T^b (L^2/T)^c (M/L^3)^d \\ &= L^{a+2c-3d} T^{b-c} M^d, \end{aligned}$$

and then equating the respective terms gives us

$$\begin{aligned} L &: a + 2c - 3d = -3, \\ T &: \qquad b - c = 0, \\ M &: \qquad\qquad d = 1. \end{aligned} \tag{1.51}$$

The solution of the above system can be written as $d = 1$ and $b = c = -a/2$. Given the assumption in (1.50), we conclude that the general product solution is

$$u = \alpha u_0 \left(\frac{x}{\sqrt{Dt}} \right)^a .$$

The general solution therefore has the form

$$u = u_0 F(\eta), \tag{1.52}$$

where

$$\eta = \frac{x}{\sqrt{Dt}} . \tag{1.53}$$

In this case, η is called a *similarity variable* as it is a dimensionless product that involves the independent variables in the problem.

When working out the drag on a sphere example, we discussed how it is possible to derive different representations of the solution. For the current example, when solving (1.51), instead of writing $b = c = -a/2$, we could just as well state that $a = -2b$ and $c = b$. In this case (1.52) is replaced with $u = u_0 G(\xi)$ where $\xi = Dt/x^2$. Although the two representations are equivalent, in the sense that one can be transformed into the other, it does make a difference which one is used when deriving a similarity solution. The reason is that we will be differentiating the solution, and (1.52) leads to much simpler formulas than the other representation. The rule of thumb here is that you want x in the numerator of the similarity variable. If you would like a hands on example of why this is true, try working out the steps below using the representation $u = u_0 G(\xi)$ instead of (1.52).

Similarity Solution

Up to this point we have been using a routine dimensional reduction argument. Our result, given in (1.52), is interesting as it states that the solution has a very specific dependence on the independent variables x and t. Namely, u can be written as a function of a single intermediate variable η. To determine F we substitute (1.52) back into the problem and find what equation F satisfies. With this in mind note, using the chain rule,

$$\begin{aligned}\frac{\partial u}{\partial t} &= u_0 F'(\eta)\frac{\partial \eta}{\partial t}\\ &= u_0 F'(\eta)\left(-\frac{x}{2D^{1/2}t^{3/2}}\right)\\ &= -u_0 F'(\eta)\frac{\eta}{2t}\,.\end{aligned}$$

In a similar manner one can show that

$$\frac{\partial^2 u}{\partial x^2} = u_0 F''(\eta)\frac{1}{Dt}\,.$$

Substituting these into (1.47) yields

$$F'' = -\frac{1}{2}\eta F', \quad \text{for } \ 0 < \eta < \infty. \tag{1.54}$$

We must also transform the boundary and initial conditions. The boundary condition at $x = 0$ takes the form

$$F(0) = 1, \tag{1.55}$$

while the condition as $x \to \infty$ and the one at $t = 0$ both translate into

$$F(\infty) = 0. \tag{1.56}$$

With this we have transformed a problem involving a partial differential equation (PDE) into one with an ordinary differential equation (ODE). As required, the resulting problem for F is only in terms of η. All of the original dimensional quantities, including the independent variables x and t, do not appear anywhere in the problem. This applies not just to the differential equation, but also to the boundary and initial conditions.

The problem for F is simpler than the original diffusion problem and, by itself, makes the use of dimensional analysis worthwhile. In this particular problem it is so simple that it is possible to solve for F. This can be done by letting $G = F'$, so the equation takes the form $G' = -\frac{1}{2}\eta G$. The general solution of this is $G = \alpha \exp(-\eta^2/4)$. Because $F' = G$, we conclude that the general solution is

$$F(\eta) = \beta + \alpha \int_0^\eta e^{-s^2/4} ds. \tag{1.57}$$

From (1.55) we have that $\beta = 1$ and from (1.56) we get

$$1 + \alpha \int_0^\infty e^{-s^2/4} ds = 0. \tag{1.58}$$

Given that $\int_0^\infty e^{-s^2/4} ds = \sqrt{\pi}$, then

$$F(\eta) = 1 - \frac{1}{\sqrt{\pi}} \int_0^\eta e^{-s^2/4} ds. \tag{1.59}$$

Expressions like this arise so often that they have given rise to a special function known as the complementary error function $\operatorname{erfc}(\eta)$. This is defined as

$$\operatorname{erfc}(\eta) = 1 - \frac{2}{\sqrt{\pi}} \int_0^\eta e^{-r^2} dr. \tag{1.60}$$

Therefore, we have found that the solution of the diffusion problem is

$$u(x,t) = u_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right). \tag{1.61}$$

As the above example demonstrates, using similarity variables and dimensional analysis provides a powerful tool for solving PDEs. It is, for example, one of the very few methods known that can be used to solve nonlinear problems. Its limitation is that the problem must have a specific form to work. We were able to solve the above diffusion problem because dimensional analysis reduced the form of the solution down to a function of one variable. This does not always happen and in such cases the method provides no insight into how the problem can be solved. As an example, if the spatial interval in the above diffusion problem is changed to one that is finite, so $0 < x < \ell$, then dimensional analysis will show that there are two independent similarity variables. This represents no improvement as we already know it is a function of two independent variables, so a reduction is not possible. Even with these limitations, however, similarity variables and their use in solving differential equations is a thriving area and a good introduction of the material can be found in Bluman and Cole [1974].

1.5 Nondimensionalization and Scaling

Another use we will have for dimensional analysis is to transform a problem into dimensionless form. The reason for this is that the approximation methods that are used to reduce difficult problems are based on comparisons. For example, in the projectile problem we simplified the differential equation by assuming that x was small compared to R. In contrast there are problems where the variable of interest is large, or it is slow or that it is fast compared to some other term in the problem. Whatever the comparison, it is important to know how all of the terms in the problem compare and for this we need the concept of scaling.

1.5.1 Projectile Problem

The reduction of the projectile equation (1.1) was based on the assumption that x is not very large, and so $x + R$ could be replaced with just R. We will routinely use arguments like this to find an approximate solution and it is therefore essential we take more care in making such reductions. The way this is done is by first scaling the variables in the problem using characteristic values. The best way to explain what this means is to work out an example and the projectile problem is an excellent place to start.

Change Variables

The first step in nondimensionalizing a problem is to introduce a change of variables, which for the projectile problem will have the form

$$\begin{aligned} t &= t_c s, \\ x &= x_c u. \end{aligned}$$

In the above formula, x_c is a constant and it is a characteristic value of the variable x. It is going to be determined using the physical parameters in the problem, which for the projectile problem are g, R, and v_0. In a similar manner, t_c is a constant that has the dimensions of time and it represents a characteristic value of the variable t. In some problems it will be clear at the beginning how to select x_c and t_c. However, it is assumed here that we have no clue at the start what to choose and will not select them until the problem is studied a bit more. All we know at the moment is that whatever the choice, the new variables u, s are dimensionless. To make the change of variables note that from the chain rule

$$\begin{aligned} \frac{d}{dt} &= \frac{ds}{dt}\frac{d}{ds} \\ &= \frac{1}{t_c}\frac{d}{ds}\,, \end{aligned} \tag{1.62}$$

and

$$\frac{d^2}{dt^2} = \frac{d}{dt}\left(\frac{d}{dt}\right) = \frac{1}{t_c^2}\frac{d^2}{ds^2}\,. \tag{1.63}$$

With this the projectile equation (1.1) takes the form

$$\frac{1}{t_c^2}\frac{d^2}{ds^2}(x_c u) = -\frac{gR^2}{(R + x_c u)^2}\,. \tag{1.64}$$

The method requires us to collect the parameters into dimensionless groups. There is no unique way to do this, and this can cause confusion when first

learning the procedure as there is no fixed method or answer. For example, to nondimensionalize the denominator in (1.64) one can factor it as either $R(1 + x_c u/R)$ or as $x_c(R/x_c + u)$. The first has the benefit of enabling us to cancel the R in the numerator. Making this choice yields

$$\frac{x_c}{gt_c^2}\frac{d^2u}{ds^2} = -\frac{1}{(1 + x_c u/R)^2}\,, \tag{1.65}$$

where the initial conditions (1.2), (1.3) are

$$u(0) = 0, \tag{1.66}$$

$$\frac{du}{ds}(0) = \frac{t_c}{x_c}v_0. \tag{1.67}$$

Find the Dimensionless Groups

Our change of variables has resulted in three dimensionless groups appearing in the transformed problem. They are

$$\Pi_1 = \frac{x_c}{gt_c^2}\,, \tag{1.68}$$

$$\Pi_2 = \frac{x_c}{R}\,, \tag{1.69}$$

$$\Pi_3 = \frac{t_c v_0}{x_c}\,. \tag{1.70}$$

There are a few important points that need to made here. First, the Π's do not involve the variables u, s and only depend on the parameters in the problem. Second, they are dimensionless and to accomplish this it was necessary to manipulate the projectile problem so the parameters end up grouped together to form dimensionless ratios. The third, and last, point is that the above three dimensionless groups are independent in the sense that it is not possible to write any one of them in terms of the other two. For example, Π_1 is the only one that contains the parameter g while Π_2 is the only one containing R. It is understood that in making the statement that the three groups are independent that x_c and t_c can be selected, if desired, independently of any of the parameters in the problem.

Before deciding on how to select x_c and t_c, it is informative to look a little closer at the above dimensionless groups. We begin with Π_2. In physical terms it is a measure of a typical, or characteristic, height of the projectile compared to the radius of the Earth. In comparison, Π_3 is a measure of a typical, or characteristic, velocity x_c/t_c compared to the velocity the projectile starts with. Finally, the parameter group Π_1 measures a typical, or characteristic, acceleration x_c/t_c^2 in comparison to the acceleration due to

gravity in a uniform field. These observations can be helpful when deciding on how to nondimensionalize a problem as will be shown next.

Use Dimensionless Groups to Determine Scaling

It is now time to actually decide on what to take for x_c and t_c. There are whole papers written on what to consider as you select these parameters, but we will take a somewhat simpler path. For our problem we have two parameters to determine, and we will do this by setting two of the above dimensionless groups equal to one. What we need to do is decide on which two to pick, and we will utilize what might be called rules of thumb.

Rule of Thumb 1: Pick the Π's that appear in the initial and/or boundary conditions.

We only have initial conditions in our problem, and the only dimensionless group involved with them is Π_3. So we set $\Pi_3 = 1$ and conclude

$$x_c = v_0 t_c. \tag{1.71}$$

Rule of Thumb 2: Pick the Π's that appear in the reduced problem.

To use this rule it is first necessary to explain what the reduced problem is. This comes from the earlier assumption that the object does not get very high in comparison to the radius of the Earth, in other words, Π_2 is small. The reduced problem is the one obtained in the extreme limit of $\Pi_2 \to 0$. Taking this limit in (1.65)-(1.67), and using (1.71), the reduced problem is

$$\Pi_1 \frac{d^2u}{ds^2} = -1\,,$$

where

$$u(0) = 0, \quad \text{and} \quad \frac{du}{ds}(0) = 1.$$

According to the stated rule of thumb, we set $\Pi_1 = 1$, and so

$$x_c = v_0^2/g. \tag{1.72}$$

This choice for x_c seems reasonable based on our earlier conclusion that the maximum height for the uniform field case is $v_0^2/(2g)$.

Combining (1.71) and (1.72), we have that $x_c = v_0^2/g$ and $t_c = v_0/g$. With this scaling then (1.65) - (1.67) take the form

$$\frac{d^2u}{ds^2} = -\frac{1}{(1+\epsilon u)^2}\,, \tag{1.73}$$

where

$$u(0) = 0, \tag{1.74}$$

$$\frac{du}{ds}(0) = 1. \tag{1.75}$$

The dimensionless parameter appearing in the above equation is

$$\epsilon = \frac{v_0^2}{gR}\,. \tag{1.76}$$

This parameter will play a critical role in our constructing an accurate approximation of the solution of the projectile problem. This will be done in the next chapter but for the moment recall that since $R \approx 6.4 \times 10^6\,\mathrm{m}$ and $g \approx 9.8\,\mathrm{m/s}^2$ then $\epsilon \approx 1.6 \times 10^{-8} v_0^2$. Consequently for baseball bats, sling shots, BB-guns, and other everyday projectile-producing situations, where v_0 is not particularly large, the parameter ϵ is very small. This observation is central to the subject of the next chapter.

Changing Your Mind

Before leaving this example it is worth commenting on the nondimensionalization procedure by asking a question. Namely, how bad is it if different choices would have been made for x_c and t_c? For example, suppose for some reason one decides to take $\Pi_2 = 1$ and $\Pi_3 = 1$. The resulting projectile problem is

$$\epsilon \frac{d^2u}{ds^2} = -\frac{1}{(1+u)^2}\,, \tag{1.77}$$

where $u(0) = 0$, $\frac{du}{ds}(0) = 1$, and ϵ is given in (1.76). No approximation has been made here and therefore this problem is mathematically equivalent to the one given in (1.73)-(1.75). Based on this, the answer to the question would be that using this other scaling is not so bad. However, the issue is amenability and what properties of the solution one is interested in. To explain, earlier we considered how the solution behaves if v_0 is not very large. With the scaling that produced (1.77), small v_0 translates into looking at what happens when ϵ is near zero. Unfortunately, the limit of $\epsilon \to 0$ results in the loss of the highest derivative in the differential equation and (1.77) reduces to $0 = -1$. How to handle such singular limits will be addressed in the next chapter but it requires more work than is necessary for this problem. In comparison, letting ϵ approach zero in (1.73) causes no such complications and for this reason it is more amenable to the study of the small v_0 limit. The point here is that if there are particular limits, or conditions, on the parameters that it is worth taking them into account when constructing the scaling.

1.5.2 Weakly Nonlinear Diffusion

To explore possible extensions of the nondimensionalization procedure we consider a well-studied problem involving nonlinear diffusion. The problem consists of finding the concentration $c(x,t)$ of a chemical over an interval $0 < x < \ell$. The concentration satisfies

$$D\frac{\partial^2 c}{\partial x^2} = \frac{\partial c}{\partial t} - \lambda(\gamma - c)c, \tag{1.78}$$

where the boundary conditions are

$$c|_{x=0} = c|_{x=\ell} = 0, \tag{1.79}$$

and the initial condition is

$$c|_{t=0} = c_0 \sin(5\pi x/\ell). \tag{1.80}$$

The nonlinear diffusion equation (1.78) is known as Fisher's equation, and it arises in the study of the movement of genetic traits in a population. A common simplifying assumption made when studying this equation is that the nonlinearity is weak, which means that the term λc^2 is small in comparison to the others in the differential equation. This assumption will be accounted for in the nondimensionalization.

Before starting the nondimensionalization process we should look at the fundamental dimensions of the variables and parameters in the problem. First, c is a concentration, which corresponds to the number of molecules per unit volume, and so $[c] = L^{-3}$. The units for the diffusion coefficient D were determined earlier, and it was found that $[D] = L^2/T$. As for γ, the $\gamma - c$ term in the differential equation requires these two quantities to have the same dimensions, and so $[\gamma] = [c]$. Similarly, from the differential equation we have $[\lambda(\gamma - c)c] = [\frac{\partial c}{\partial t}]$, and from this it follows that $[\lambda] = L^3T^{-1}$. Finally, from the initial condition we have that $[c_0] = [c]$. It is important to make an observation related to dimensions, and this will be done by asking a question: is it possible to replace the initial condition (1.80) with $c|_{t=0} = c_0 \sin(5\pi x)$ or with $c|_{t=0} = c_0 \sin(x)$? The answer in both cases is no, and the reason is that the argument of the sine function must be dimensionless. For exactly the same reason it is not possible to use $c|_{t=0} = c_0 e^x$. It is possible, however, to use $c|_{t=0} = c_0 x$ or $c|_{t=0} = c_0 x^2$, although the dimensions of c_0 differ from what we found earlier.

Now, to nondimensionalize the problem we introduce the change of variables

$$x = x_c y, \tag{1.81}$$

$$t = t_c s, \tag{1.82}$$

$$c = c_c u. \tag{1.83}$$

In this context, x_c has the dimensions of length and is a characteristic value of the variable x. Similar statements apply to t_c and c_c. Using the chain rule as in (1.62) the above differential equation takes the form

$$\frac{Dc_c}{x_c^2}\frac{\partial^2 u}{\partial y^2} = \frac{c_c}{t_c}\frac{\partial u}{\partial s} - \lambda c_c(\gamma - c_c u)u.$$

It is necessary to collect the parameters into dimensionless groups, and so in the above equation we rearrange things a bit to obtain

$$\frac{Dt_c}{x_c^2}\frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial s} - \lambda t_c c_c(\gamma/c_c - u)u. \tag{1.84}$$

In conjunction with this we have the boundary conditions

$$u|_{y=0} = u|_{y=\ell/x_c} = 0, \tag{1.85}$$

and the initial condition is

$$u|_{s=0} = (c_0/c_c)\sin(5\pi x_c y/\ell). \tag{1.86}$$

The resulting dimensionless groups are

$$\Pi_1 = \frac{Dt_c}{x_c^2}, \tag{1.87}$$

$$\Pi_2 = \lambda t_c c_c, \tag{1.88}$$

$$\Pi_3 = \gamma/c_c, \tag{1.89}$$

$$\Pi_4 = \ell/x_c, \tag{1.90}$$

$$\Pi_5 = c_0/c_c. \tag{1.91}$$

It is important to note that the five dimensionless groups given above are independent in the sense that it is not possible to write one of them in terms of the other four. As before this statement is based on our ability to select, if desired, the scaling parameters x_c, t_c, c_c independently of each other and the other parameters in the problem. Also, in counting the dimensionless groups one might consider adding a sixth. Namely, in the initial condition (1.86) there is $\Pi_6 = 5\pi x_c/\ell$. The reason it is not listed above is that it is not independent of the others because $\Pi_6 = 5\pi/\Pi_4$. The 5π is a number and does not play a role in determining dimensional independence.

We have three scaling parameters to specify, namely x_c, t_c, c_c. Using Rule of Thumb 1, the Π's that appear in the boundary and initial conditions are

set equal to one. In other words, we set $\Pi_4 = 1$ and $\Pi_5 = 1$, from which it follows that $x_c = \ell$ and $c_c = c_0$.

To use Rule of Thumb 2, we need to investigate what it means to say that the nonlinearity is weak. The equation (1.84) is nonlinear due to the term $\lambda t_c c_c u^2$, and the coefficient $\lambda t_c c_c$ is the associated strength of the nonlinearity. For a weakly nonlinear problem one is interested in the solution for small values of $\lambda t_c c_c$. Taking the extreme limit we set $\lambda t_c c_c = 0$ in (1.84) to produce the reduced equation. The only group that remains in this limit is Π_1, and for this reason this is the group we select. So, setting $\Pi_1 = 1$ then we conclude $t_c = \ell^2/D$.

The resulting nondimensional diffusion equation is

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial s} - \epsilon(b-u)u, \tag{1.92}$$

with boundary conditions

$$u(0,s) = u(1,s) = 0, \tag{1.93}$$

and the initial condition

$$u(y,0) = \sin(5\pi y). \tag{1.94}$$

The dimensionless parameters appearing in the above equation are $\epsilon = \lambda c_0 \ell^2/D$ and $b = \gamma/c_0$. With this, weak nonlinearity corresponds to assuming that ϵ is small.

1.5.3 Endnotes

As you might have noticed, the assumption of a weak nonlinearity was used in the projectile problem, although it was stated in more physical terms. In both examples the reduced problem, obtained setting $\epsilon = 0$, is linear. It is certainly possible that a physical problem is not weakly nonlinear but involves some other extreme behavior. As an example, in nonlinear diffusion problems you come across situations involving weak diffusion. What this means for (1.84) is that Dt_c/x_c^2 has a small value. In the extreme limit that this term is zero then the only group that remains in the reduced problem is Π_2. Setting $\Pi_2 = 1$ then $t_c = c_0/\lambda$. With this, (1.84) becomes

$$\epsilon\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} + (b-u)u, \tag{1.95}$$

where $\epsilon = Dc_0/(\lambda\ell^2)$ and $b = \gamma/c_0$. With this, weak diffusion corresponds to assuming that ϵ is small.

For those keeping track of the rules of thumb used to nondimensionalize a problem we have two. The first we ran across is the rule that the dimensionless groups in the initial and boundary conditions are set to one. The second rule arose when setting the dimensionless groups in the reduced problem to one. Although these can be effective rules, it is certainly possible to find problems where another scaling should be considered, and examples are given in Exercises 1.17 and 3.8. The overall objective in all cases is that the nondimensionalization is based on characteristic values of the variables.

Exercises

1.1. The amount of noise permitted from the large rollers used in road construction was recently limited by changes in the environmental laws. Rather than build multiple full-sized rollers in an attempt to find one that satisfied the new law a manufacturer decided that dimensional analysis could be used. The assumption they made was that the frequency f of the sound coming off the roller depends on the elastic modulus E and the density ρ of the steel used to construct the roller as well as on the length ℓ of the roller.

(a) Find a dimensionally reduced form for f.

(b) In building a scale model for testing the manufacturer selected the parameters so that

$$\frac{f_m}{f_f} = \frac{\ell_f}{\ell_m}\sqrt{\frac{\rho_f E_m}{\rho_m E_f}},$$

where the subscript f designates full-sized and the subscript m designates scale model. Explain why this was done.

1.2. For a pendulum that starts from rest, the period p depends on the length ℓ of the rod, on gravity g, on the mass m of the ball, and on the initial angle θ_0 at which the pendulum is started.

(a) Use dimensional analysis to determine the functional dependence of p on these four quantities.

(b) For the largest pendulum ever built, the rod is 70 ft and the ball weighs 900 lbs. Assuming that $\theta_0 = \pi/6$ explain how to use a pendulum that fits on your desk to determine the period of this largest pendulum.

(c) Suppose it is found that p depends linearly on θ_0, with $p = 0$ if $\theta_0 = 0$. What does your result in part (a) reduce to in this case?

1.3. The velocity v at which flow in a pipe will switch from laminar to turbulent depends on the diameter d of the pipe as well as on the density ρ and dynamic viscosity μ of the fluid.

(a) Find a dimensionally reduced form for v.

(b) Suppose the pipe has diameter $d = 100$ and for water (where $\rho = 1$ and $\mu = 10^{-2}$) it is found that $v = 0.25$. What is v for olive oil (where $\rho = 1$ and $\mu = 1$)? The units here are in cgs.

1.4. The luminosity of certain giant and supergiant stars varies in a periodic manner. It is hypothesized that the period p depends upon the star's average radius r, its mass m, and the gravitational constant G.

(a) Newton's law of gravitation asserts that the attractive force between two bodies is proportional to the product of their masses divided by the square of the distance between them, that is,

$$F = \frac{Gm_1m_2}{d^2},$$

where G is the gravitational constant. From this determine the (fundamental) dimensions of G.

(b) Use dimensional analysis to determine the functional dependence of p on m, r, and G.

(c) Arthur Eddington used the theory for thermodynamic heat engines to show that

$$p = \sqrt{\frac{3\pi}{2\gamma G\rho}},$$

where ρ is the average density of the star and γ is the ratio of specific heats for stellar material. How does this differ from your result?

(d) In Figure 1.8 the data for a pulsating star are given. Explain how you could use data like this to complete the formula you derived in part (b).

1.5. When a drop of liquid hits a wetted surface a crown formation appears, as shown in Figure 1.9(a). It has been found that the number of points N on the crown depends on the speed U at which the drop hits the surface, the radius r and density ρ of the drop, and the surface tension σ of the liquid making up the drop. How N depends on these quantities has been studied extensively and some of the reasons why are given in Rioboo et al. [2003].

(a) Use dimensional reduction to determine the functional dependence of N on U, r, ρ, and σ. Express your answer in terms of the Weber number $W_e = \rho U^2 r/\sigma$.

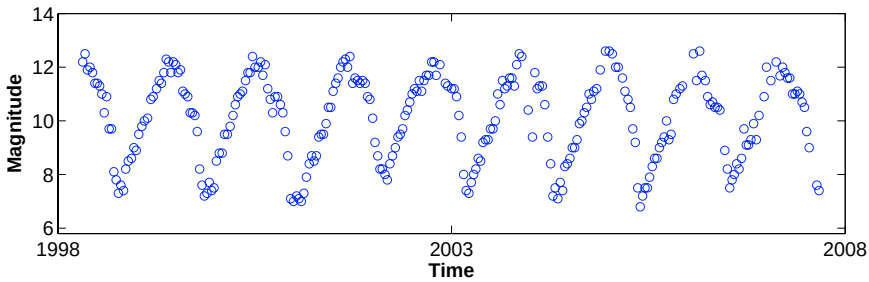


Figure 1.8 Luminosity of a Mira type variable star, 1621+19 U Herculis (AAVSO [2009]).

(a)

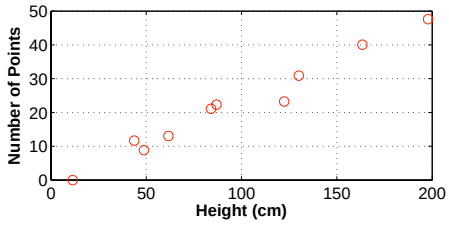
(b) 

Figure 1.9 (a) Formation of a crown when a liquid drop hits a wetted surface. (b) The measured values of the number of points N (Hobbs and Kezweent [1967]).

(b) The value of N has been measured as a function of the initial height h of the drop and the results are shown in Figure 1.9(b). Express your answer in part (a) in terms of h by writing U in terms of h and g. Assume the drop starts with zero velocity.

(c) The data in Figure 1.9(b) show a piecewise linear dependence on h, specifically, N can be described as a continuous function made up of two linear segments. Use this, and your result from part (b), to find the unknown function in part (a). In the experiments, $r = 3.6$ mm, $\rho = 1.1014$ gm/cm^3, and $\sigma = 50.5$ dyn/cm.

(d) According to your result from part (c), what must the initial height of the drop be to produce at least 80 points?

(e) According to your result from part (c), how many points are generated for a drop of mercury when $h = 200$ cm? Assume $r = 3.6$ mm, $\rho = 13.5$ gm/cm^2, and $\sigma = 435$ dyn/cm.

1.6. The frequency ω of waves on a deep ocean is found to depend on the wavelength λ of the wave, the surface tension σ of the water, the density ρ of the water, and gravity.

(a) Use dimensional reduction to determine the functional dependence of ω on λ, σ, ρ, and g.

(b) In fluid dynamics it is shown that

$$\omega = \sqrt{gk + \frac{\sigma k^3}{\rho}}\,,$$

where $k = 2\pi/\lambda$ is the wavenumber. How does this differ from your result in (a)?

1.7. A ball is dropped from a height h_0 and it rebounds to a height h_r. The rebound height depends on the elastic modulus E, radius R, and the mass density ρ of the ball. It also depends on the initial height h_0 and the gravitational constant g.

(a) Find a dimensionally reduced form for h_r.

(b) Suppose it is found that h_r depends linearly on h_0, with $h_r = 0$ if $h_0 = 0$. What does your formula from part (a) reduce to in this case?

(c) Suppose the density of the ball is doubled. Use the result in (a) to explain how to change E so the rebound height stays the same.

1.8. A ball, when released underwater, will rise towards the surface with velocity v. This velocity depends on the density ρ_b and radius R of the ball, on gravity g, and on the density ρ_f and kinematic viscosity ν of the water.
(a) Find a dimensionally reduced form for v.
(b) In fluid mechanics, using Stokes' Law, it is found that

$$v = \frac{2gR^2(\rho_b - \rho_f)}{9\nu\rho_f}\,.$$

How does this differ from your result from part (a)? It is interesting to note that this formula is used by experimentalists to determine the viscosity of fluids. They do this by measuring the velocity in an apparatus called a falling ball viscometer, and then solving for ν in the above formula.

1.9. In electric image tomography the objective is to determine the properties inside an object and this is done by applying a potential U to the surface. What is measured is the resulting electric current j on the surface. Suppose that it is found that the electric potential u within a spherical body depends on the object's radius R and conductivity σ as well as depends on U and j.
(a) Find a dimensionally reduced form for u.
(b) Suppose that given a particular object that doubling the applied potential U causes the internal potential u to increase by a factor of four. How does this help simply your result in (a)?
(c) Suppose it is necessary to know the internal potential u when using a large applied potential, say $U = 2500V$. However, for legal reasons it is required that only applied potentials less than $250V$ can be used. Explain, using your result from (a), how to legally determine the large applied potential value.

1.10. The velocity v of water through a circular pipe depends on the pressure difference p between the two ends of the pipe, the length ℓ and radius r of the pipe, as well as on the dynamic viscosity μ and density ρ of the water.
(a) Use dimensional analysis to determine the functional dependence of v on the above quantities.
(b) Suppose it is found that v depends linearly on p, with $v = 0$ if $p = 0$. What does your formula from part (a) reduce to in this case?
(c) Your formula from part (b) should contain a general function of one, or more, dimensionless products. Explain how to experimentally determine this function. Be specific about which parameters are fixed, and which are varied, in the experiment. Also, your experiment should vary as few of the parameters as possible in determining this function.

1.11. In a high energy explosion there is a very rapid release of energy E that produces an approximately spherical shock wave that expands in time (Figure 1.10).

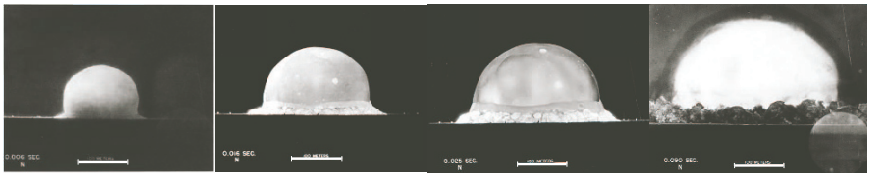


Figure 1.10 Shock wave produced by a nuclear explosion, at 6 msec, 16 msec, 25 msec, and 90 msec. The width of the white bar in each figure is $100m$ (Brixner [2009]).

(a) Assuming the radius R of the shock wave depends on E, the length of time t since the explosion, and the density ρ of the air, use dimensional reduction to determine how the radius depends on these quantities. This expression is known as the Taylor-Sedov formula.
(b) It was shown by G. I. Taylor that if $E = 1J$ and $\rho = 1\,kg/m^3$ then $R = t^{2/5}\,m/s^{2/5}$. Use this information and the result from (a) to find the exact formula for R.
(c) Use the photographs in Figure 1.10, and your result from (b) to estimate the energy released. The air density is $\rho = 1\,kg/m^3$.
(d) The blast wave from a supernova can be modeled using the Taylor-Sedov formula. Explain how this can be used to estimate the date the supernova took place, using your result from part (b). As an example, use Tycho, which currently has a radius of about 33.2 light years, an estimated energy of $10^{44}J$, and density $\rho = 2 \times 10^{-21}$ kg/m^3.

1.12. The vertical displacement $u(x)$ of an elastic string of length ℓ satisfies the boundary value problem

$$\tau \frac{d^2u}{dx^2} + \mu u = p, \quad \text{for } \ 0 < x < \ell,$$

where $u(0) = 0$, $u(\ell) = U$. Also, p is a constant and has the dimensions of force per length.
(a) What are the dimensions for the constants τ and μ?
(b) Show how it is possible to nondimensionalize this problem so it takes the form

$$\frac{d^2v}{ds^2} + \alpha v = \beta, \quad \text{for } \ 0 < s < 1,$$

where $v(0) = 0$, $v(1) = 1$. Make sure to state what α, β are.

1.13. From Newton's second law, the displacement $y(t)$ of the mass in a mass, spring, dashpot system satisfies

$$m\frac{d^2y}{dt^2} = F_s + F_d, \quad \text{for } \ 0 < t,$$

where m is the mass, F_s is the restoring force in the spring, and F_d is the damping force. To have a compete IVP we need to state the initial conditions,

and for this problem assume

$$y(0) = 0, \quad \frac{dy}{dt}(0) = v_0.$$

(a) Suppose there is no damping, so $F_d = 0$, and the spring is linear, so $F_s = -ky$. What are the dimensions for the spring constant k? Nondimensionalize the resulting IVP. Your choice for y_c and t_c should result in no dimensionless products being left in the IVP.
(b) Now, in addition to a linear spring, suppose linear damping is included, so,

$$F_d = -c\frac{dy}{dt}.$$

What are the dimensions for the damping constant c? Using the same scaling as in part (a), nondimensionalize the IVP. Your answer should contain a dimensionless parameter ϵ that measures the strength of the damping. In particular, if c is small then ϵ is small. The system in this case is said to have weak damping.

1.14. The velocity $v(t)$ of the waves on a deep ocean satisfies the equation

$$\frac{dv}{dt} + kv^2 = \ell v, \quad \text{for } 0 < t,$$

where $v(0) = V$.
(a) What are the dimensions of the constants k, ℓ, and V?
(b) Assuming a weak nonlinearity, use the Rules of Thumb given in Section 1.5 to nondimensionalize this problem.

1.15. The equation for an elastic beam is

$$EI\frac{\partial^4 u}{\partial x^4} + \rho\frac{\partial^2 u}{\partial t^2} = 0,$$

where the boundary conditions are $u = u_0 \sin(\omega t)$ and $\frac{\partial u}{\partial x} = 0$ at $x = 0$, while $u = \frac{\partial u}{\partial x} = 0$ at $x = \ell$. Assume the initial conditions are $u = 0$ and $\frac{\partial u}{\partial t} = 0$ at $t = 0$. Here E is the elastic modulus, I is the moment of inertia, and ρ is the mass per unit length of the beam. Nondimensionalize the problem in such a way that the resulting boundary conditions contain no nondimensional groups.

1.16. When an end of a slender strip of paper is put into a cup of water, because of absorption, the water rises up the paper. The density ρ of the water along the strip satisfies the differential equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial J}{\partial x} = 0,$$

where J is known as the flux.

(a) What are the dimensions of J?
(b) The flux J depends on the gravitational constant g, the strip width d, the density gradient $\frac{\partial \rho}{\partial x}$, and the surface tension σ of the water. Find a dimensionally reduced form for J.
(c) What does your result in (b) reduce to if it is found that J depends linearly on the density gradient, with $J = 0$ if $\rho_x = 0$? What is the resulting differential equation?
(d) If the strip has length h the boundary conditions are $\rho = \rho_0$ at $x = 0$ and $J = 0$ at $x = h$. The initial condition is $\rho = 0$ at $t = 0$. With this, and your differential equation from (c), nondimensionalize the problem for ρ in such a way that no nondimensional groups appear in the final answer.

1.17. A thermokinetic model for the concentration u and temperature q of a mixture consists of the following equations (Gray and Scott [1994])

$$\frac{du}{dt} = k_1 - k_2 u e^{k_3 q},$$
$$\frac{dq}{dt} = k_4 u e^{k_3 q} - k_5 q.$$

The initial conditions are $u = 0$ and $q = 0$ at $t = 0$.
(a) What are the dimensions of the k_i's?
(b) Explain why the rule of thumb for scaling used in the projectile problem does not help here.
(c) Find the steady-state solution, that is, the solution of the differential equations with $u' = 0$ and $q' = 0$.
(d) Nondimensionalize the problem using the steady-state solution from (c) to scale u and q. Make sure to explain how you selected the scaling for t.

1.18. The equations that account for the relativistic motion of a planet around the sun are

$$\frac{d^2 r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{Gm}{r^2} + \frac{b}{r^3},$$
$$\frac{d}{dt}\left(r^2 \frac{d\theta}{dt}\right) = 0,$$

where b is a constant. Assume the initial conditions are $r = r_0$, $r' = 0$, and $\theta = 0$ at $t = 0$.
(a) What are the dimensions of r_0, b?
(b) Nondimensionalize the problem. The scaling should be chosen so the only nondimensional group appearing in the problem involves b.

1.19. Suppose you are given a dimensionless function $f(\Pi)$ where Π is a dimensionless group. Also, suppose $\Pi = A^a B^b C^c$ where A, B, C are dimensional parameters and the exponents a, b, c are nonzero numbers.

(a) Show that if $f(\Pi)$ is found to be linear in A then it must be that $f(\Pi) = \alpha \Pi^{1/a} + \beta$ where α, β are arbitrary numbers.
(b) What can you conclude if it is found that $\sqrt{AB} f(\Pi)$ is linear in A?
(c) Suppose it is found that if A is doubled that the value of F increases by a factor of four. Can this be used to determine F?

1.20. This problem explores some consequences of dimensional quantities.
(a) If g is the gravitational acceleration constant, explain why $\sin(g)$ and e^g make no sense.
(b) Explain why density, volume, and velocity can be used in place of length, mass, and time as fundamental units.
(c) Explain why volume, velocity, and acceleration cannot be used in place of length, mass, and time as fundamental units.

1.21. In quantum chromodynamics three parameters that play a central role are the speed of light c, Planck's constant $\hbar$, and the gravitational constant G.
(a) Explain why it is possible to use $[c], [\hbar], [G]$ as fundamental units.
(b) The distance ℓ_p at which the strong, electromagnetic and weak forces become equal depends on $c, \hbar, G$. Find a dimensionally reduced form for how ℓ_p depends on these three parameters. Based on this result, if the speed of light were to double what happens to ℓ_p?
(c) The Bohr radius a of an electron depends on $\hbar$, the electron's charge e, and the mass m_e of the electron. Find a dimensionally reduced form for a.

1.22. The speed c_m at which magnetonsonic waves travel through a plasma depends on the intensity B of the magnetic field, the permeability μ_0 of free space, and the density ρ and pressure p of the plasma.
(a) Use dimensional reduction to determine the functional dependence of c_m on B, μ_0, ρ, and p.
(b) From the basic laws for plasmas it is shown that

$$c_m = \sqrt{V_A^2 + c_s^2}\,,$$

where $V_A = B/\sqrt{\mu_0 \rho}$ is the Alfven speed and $c_s = \sqrt{\gamma p/\rho}$ is the sound speed in the gas. In the latter expression, γ is a number. How does this differ from your result in (a)?

1.23. In the study of the motion of particles moving along the x-axis one comes across the problem of finding the velocity u that satisfies the nonlinear partial differential equation

$$u_t + u u_x = 0, \tag{1.96}$$

where

$$u(x,0) = \begin{cases} 0 & \text{if } x < 0 \\ u_0 & \text{if } 0 < x. \end{cases} \tag{1.97}$$

Assume that u_0 is a positive constant. The equation (1.96) is derived in Chapter 5, and it is known as the inviscid Burgers' equation. It, along with the jump condition in (1.97), form what is known as a Riemann problem.

(a) What three physical quantities does u depend on?

(b) Use dimensional reduction, and a similarity variable, to reduce this problem to a nonlinear ordinary differential equation with two boundary conditions.

(c) Use the result from part (b) to solve the Riemann problem. The solution, which is known as an expansion fan, must be continuous for $t > 0$.

(d) What is the solution if the initial condition (1.97) is replaced with $u(x, 0) = u_0$?

(e) Suppose that, rather than velocity, the variable u is displacement. Explain why it is not possible for u to satisfy (1.96).

1.24. Consider the partial differential equation

$$u_t + Du_{xxxx} = 0,$$

where $u = u_0$ at $x = 0$, $u \to 0$ as $x \to \infty$, and $u = 0$ at $t = 0$. Use dimensional reduction, and a similarity variable, to reduce this problem to an ordinary differential equation.

1.25. The equation of the concentration c, on an interval of length ℓ, is

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} + \mu c,$$

where the boundary conditions are $c(x, 0) = 0$, $c(0, t) = c_0$, and $c(\ell, t) = 0$.

(a) What are the dimensions of D, c_0, and μ?

(b) Nondimensionalize the problem so it has the form

$$\frac{\partial u}{\partial s} = \frac{\partial^2 u}{\partial y^2} + \alpha u,$$

where the boundary conditions are $u(y, 0) = 0$, $u(0, s) = 1$, and $u(1, s) = 0$.

1.26. One of the standard experimental tests used in the study of fluid motion through porous materials consists of determining the displacement u when the material is given a constant load. The governing differential equation in this case is

$$H\left[1 + \left(\frac{\partial u}{\partial x}\right)^3\right]\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}.$$

The boundary conditions are

$$\frac{\partial u}{\partial x} = -1, \quad \text{at} \quad x = 0,$$

and

$$u = 0, \quad \text{as} \quad x \to \infty.$$

The initial condition is

$$u = 0, \quad \text{at} \quad t = 0.$$

(a) What are the dimensions of the constant H?
(b) Find a dimensionally reduced form for the solution and then use this to transform the above diffusion problem into one involving a nonlinear ordinary differential equation. Make sure to state what happens to the boundary and initial conditions. You do not need to solve this problem.
(c) In the experiment the surface displacement $u(0,t)$ is measured. Without solving the problem use your results from (b) to sketch $u(0,t)$ as a function of t.
(d) Suppose the experimental data show that $u(0,t) = 16t$ cm/sec. Using your result from part (c), explain why the mathematical model is incorrect. Also, explain why changing the differential equation to either $Hu_{xx} = u_t$ or to $H[1 + (u_x)^5]u_{xx} = u_t$ will also produce an incorrect model.

1.27. Consider the problem of solving the diffusion equation

$$D\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t},$$

where the boundary conditions are

$$u = 0, \quad \text{as} \quad x \to \pm\infty.$$

Instead of an initial condition, assume the solution satisfies

$$\int_{-\infty}^{\infty} u dx = \gamma, \quad \forall t > 0.$$

(a) What are the dimensions of γ?
(b) Find a dimensionally reduced form for the solution and then use this to transform the above diffusion equation into an ordinary differential equation. How do the boundary conditions transform? The integral condition should be considered in the dimensional reduction but its conversion using the similarity variable will wait until part (d).
(c) Find the solution of the problem from part (b). You can assume $F' \to 0$ and $\eta F \to 0$ as $\eta \to \pm\infty$. As a hint, you might want to look for the expression $(\eta F)'$ in your equation.
(d) The solution from part (c) should contain an arbitrary constant. Find its value using the given integral condition and with this show that

$$u = \frac{\gamma}{\sqrt{\pi D t}} e^{-x^2/(4Dt)}.$$

This is known as the fundamental, or point source, solution of the diffusion equation.

Chapter 2
Perturbation Methods

2.1 Regular Expansions

To introduce the ideas underlying perturbation methods and asymptotic approximations, we will begin with an algebraic equation. The problem we will consider is how to find an accurate approximation of the solution x of the quadratic equation

$$x^2 + 2\epsilon x - 1 = 0, \tag{2.1}$$

in the case of when ϵ is a small positive number. The examples that follow this one are more complex and, unlike this equation, we will not necessarily know at the start how many solutions the equation has. A method for determining the number of real-valued solutions involves sketching the terms in the equation. With this in mind, we rewrite the equation as $x^2 - 1 = -2\epsilon x$. The left- and right-hand sides of this equation are sketched in Figure 2.1. Based on the intersection points, it is seen that there are two solutions. One is a bit smaller than $x = 1$ and the other is just to the left of $x = -1$. Another observation is that the number of solutions does not change as $\epsilon \to 0$. The fact that the reduced problem, which is the one obtained when setting $\epsilon = 0$, has the same number of solutions as the original problem is a hallmark of what are called regular perturbation problems.

Our goal is to derive approximations of the solutions for small ϵ, and for this simple problem we have a couple of options on how to do this.

Method 1: Solve then Expand

It is an easy matter to find the solution using the quadratic formula. The result is

$$x = -\epsilon \pm \sqrt{1 + \epsilon^2}. \tag{2.2}$$

This completes the solve phase of the process. To obtain an approximation for the two solutions, for small ϵ, we first use the binomial expansion (see Table 2.1) to obtain

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_2,
© Springer Science+Business Media, LLC 2009

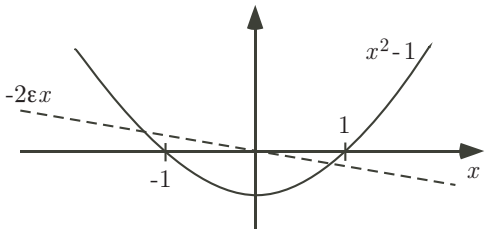


Figure 2.1 Sketch of the functions appearing in the quadratic equation in (2.1).

$$\sqrt{1+\epsilon^2} = 1 + \frac{1}{2}\epsilon^2 - \frac{1}{8}\epsilon^4 + \cdots . \tag{2.3}$$

A comment needs to be made here about the equal sign in this expression. The right hand side is an infinite series and by stating it is equal to $\sqrt{1+\epsilon^2}$ it is meant that given a value of ϵ that the series converges to $\sqrt{1+\epsilon^2}$. Said another way, given a value of ϵ, the more terms that are added together in the series the closer the sum gets to the value of $\sqrt{1+\epsilon^2}$. For this to be true it is necessary to require that $\epsilon^2 < 1$, but we are assuming ϵ is close to zero so this is not a restriction in this problem.

Substituting (2.3) into (2.2) yields

$$\begin{aligned} x &= -\epsilon \pm \left(1 + \frac{1}{2}\epsilon^2 - \frac{1}{8}\epsilon^4 + \cdots\right) \\ &= \pm 1 - \epsilon \pm \frac{1}{2}\epsilon^2 \mp \frac{1}{8}\epsilon^4 + \cdots . \end{aligned} \tag{2.4}$$

In the last step the terms are listed in order according to their power of ϵ. With this we can list various levels of approximation of the solutions, as follows

$x \approx \pm 1$	1 term approximation
$x \approx \pm 1 - \epsilon$	2 term approximation
$x \approx \pm 1 - \epsilon \pm \frac{1}{2}\epsilon^2$	3 term approximation.

So, we have accomplished what we set out to do, which is to derive an approximation of the solution for small ϵ. The procedure is straightforward but it has a major drawback because it requires us to be able to first solve the equation before constructing the approximation. For most problems this is simply impossible, so we need another approach.

Method 2: Expand then Solve

This approach requires us to first state what we consider to be the general form of the approximation for x. This requires a certain amount of experience and a reasonable place to start is with Taylor's theorem. We know that the solution depends on ϵ, we just don't know how. Emphasizing this dependence by writing $x(\epsilon)$, then using Taylor's theorem for ϵ near zero, we obtain

$$
\begin{aligned}
f(x) &= f(0) + xf'(0) + \frac{1}{2}x^2 f''(0) + \frac{1}{3!}x^3 f'''(0) + \cdots \\
(a+x)^\gamma &= a^\gamma + \gamma x a^{\gamma-1} + \frac{1}{2}\gamma(\gamma-1)x^2 a^{\gamma-2} + \frac{1}{3!}\gamma(\gamma-1)(\gamma-2)x^3 a^{\gamma-3} + \cdots \\
\frac{1}{1-x} &= 1 + x + x^2 + x^3 + \cdots \\
\frac{1}{(1-x)^2} &= 1 + 2x + 3x^2 + 4x^3 + \cdots \\
\frac{1}{\sqrt{1-x}} &= 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \frac{5}{16}x^3 + \cdots \\
e^x &= 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \cdots \\
a^x &= e^{x\ln(a)} = 1 + x\ln(a) + \frac{1}{2}(x\ln(a))^2 + \frac{1}{3!}(x\ln(a))^3 + \cdots \\
\sin(x) &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots \\
\cos(x) &= 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 + \cdots \\
\sin(a+x) &= \sin(a) + x\cos(a) - \frac{1}{2}x^2\sin(a) + \cdots \\
\cos(a+x) &= \cos(a) - x\sin(a) - \frac{1}{2}x^2\cos(a) + \cdots \\
\ln(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots \\
\ln(a+x) &= \ln(a) + \ln(1+x/a) = \ln(a) + \frac{x}{a} - \frac{1}{2}\left(\frac{x}{a}\right)^2 + \frac{1}{3}\left(\frac{x}{a}\right)^3 + \cdots
\end{aligned}
$$

Table 2.1 Taylor series expansions, about $x = 0$, for some of the more commonly used functions.

$$x(\epsilon) = x(0) + \epsilon x'(0) + \frac{1}{2}\epsilon^2 x''(0) + \cdots . \tag{2.5}$$

This implies that x can be expanded using integer powers of ϵ. With nonlinear equations, there is no guarantee that the powers have to be integers. An example is the equation $x^2 - \epsilon = 0$. A reasonable assumption is that the solution can be expanded in powers of ϵ, although not necessarily in integer powers. For this reason we will assume that the general form of the expansion is

$$x \sim x_0 + \epsilon^\alpha x_1 + \epsilon^\beta x_2 + \cdots . \tag{2.6}$$

The values of $x_0, x_1, x_2, \ldots$ and $\alpha, \beta, \ldots$ will be determined when solving the equation. It is assumed that the terms are listed in order according to their power of ϵ, which means we assume $0 < \alpha < \beta < \cdots$. This requirement is known as a well-ordering assumption and we will make it every time we write down such an expression. It is also assumed that $x_0, x_1, x_2, \ldots$ do not depend on ϵ.

We will need to be able to identify the coefficients in an expansion and the big O notation will be used for this. So, in (2.6) the $O(\epsilon^\alpha)$ coefficient is x_1 and the $O(\epsilon^\beta)$ coefficient is x_2. For the same reason x_0 is the coefficient of the $O(1)$ term.

You might be wondering why we don't just assume at the start that $\alpha = 1$ and $\beta = 2$. After all, this is what we found earlier using the "Solve then Expand" approach. One reason for this has already been given, namely the solutions of nonlinear problems do not necessarily involve integer powers and we want a method that can handle such situations. A second reason is that we will learn something important by having the equation tell us that $\alpha = 1$. The expansion in (2.6) is really nothing more than an educated guess. It is quite possible that it is incorrect and it is important to see how the equation will tell us we have made an incorrect assumption.

We will substitute (2.6) into (2.1), but before doing so note

$$\begin{aligned} x^2 &\sim (x_0 + \epsilon^\alpha x_1 + \epsilon^\beta x_2 + \cdots)(x_0 + \epsilon^\alpha x_1 + \epsilon^\beta x_2 + \cdots) \\ &\sim x_0^2 + 2\epsilon^\alpha x_0 x_1 + 2\epsilon^\beta x_0 x_2 + \epsilon^{2\alpha} x_1^2 + \cdots. \end{aligned} \tag{2.7}$$

To start we will concentrate on finding the first two terms in the expansion for x. In this case, with (2.7), (2.1) takes the form

$$x_0^2 + 2\epsilon^\alpha x_0 x_1 + \cdots + 2\epsilon(x_0 + \epsilon^\alpha x_1 + \cdots) - 1 = 0. \tag{2.8}$$

We are constructing an approximation for small ϵ. In letting $\epsilon \to 0$ in the above equation we obtain the equation for the $O(1)$ term.

$O(1)$ $x_0^2 - 1 = 0$

The solutions are $x_0 = \pm 1$.

With this (2.8) reduces to

$$2\epsilon^\alpha x_0 x_1 + \cdots + 2\epsilon(x_0 + \epsilon^\alpha x_1 + \cdots) = 0. \tag{2.9}$$

Now, with the given values of x_0 we are left with a $2\epsilon x_0$ term in the above equation. There are no $O(\epsilon)$ terms on the right-hand side so there can be no $O(\epsilon)$ terms on the left-hand side. The only term available to cancel, or balance, out $2\epsilon x_0$ is $2\epsilon^\alpha x_0 x_1$, and for this to happen it is necessary that $\alpha = 1$. So, the equation has told us exactly what value this exponent must have. This gives us the following problem.

$O(\epsilon)$ $2x_0 x_1 + 2x_0 = 0$

The solution is $x_1 = -1$.

We have determined the first two terms in the expansion, but we could easily continue and find more. For example, to find the next term note that the equation in (2.8), using (2.7), reduces to

$$2\epsilon^{\beta}x_0x_2 + \epsilon^2x_1^2 + \cdots + 2\epsilon(\epsilon x_1 + \epsilon^{\beta}x_2 + \cdots) = 0. \tag{2.10}$$

We now have $\epsilon^2x_1^2$ and $2\epsilon^2x_1$ on the left, with no $O(\epsilon^2)$ terms on the right. The only term available to eliminate them is $2\epsilon^{\beta}x_0x_2$ and, therefore, $\beta = 2$. With this we obtain the equation for the $O(\epsilon^2)$ terms in the equation.

$O(\epsilon^2)$ $2x_0x_2 + x_0^2 - 2x_1 = 0$

The solutions are $x_2 = \pm\frac{1}{2}$.

The above procedure can be used to find the successively higher order terms in the expansion. Rather than do that it is more worthwhile to consider what we have done to get to this point. Our conclusion is that one of the solutions is

$$x \sim 1 - \epsilon + \frac{1}{2}\epsilon^2, \tag{2.11}$$

and the other is

$$x \sim -1 - \epsilon - \frac{1}{2}\epsilon^2. \tag{2.12}$$

These approximations hold for small ϵ, and for this reason they are said to be asymptotic expansions of the solutions as $\epsilon \to 0$.

The formal definition of what it means to be an asymptotic expansion states that the difference between x and the expansion goes to zero faster than the last term included in the expansion. For (2.11) this means that

$$\lim_{\epsilon\to 0} \frac{x - (1 - \epsilon + \frac{1}{2}\epsilon^2)}{\epsilon^2} = 0.$$

For the same reason, $x \sim 1-\epsilon$ is also an asymptotic expansion of the solution because

$$\lim_{\epsilon\to 0} \frac{x - (1 - \epsilon)}{\epsilon} = 0.$$

This is the basis for what is known as the limit-process definition of an asymptotic expansion. This is important for those interested in the theoretical foundations of the subject. For us, the critical point is that the asymptotic expansion is determined by how the function, or solution, behaves as $\epsilon \to 0$. The definition does not say anything about what happens when more terms are used in the expansion for a given value of ϵ. If we were to calculate every term in the expansion, and produce an infinite series in the process, the fact that it is an asymptotic expansion does not mean the series has to converge. In fact, some of the more interesting asymptotic expansions diverge. For this

reason it is inappropriate to use an equal sign in (2.6) and why the symbol $\sim$ is used instead.

2.2 How to Find a Regular Expansion

The ideas used to construct asymptotic expansions of the solutions of a quadratic equation are easily extended to more complex problems. Exactly how one proceeds depends on how the problem is stated, and the following three situations are the most common.

2.2.1 Given a Specific Function

The expansion in (2.3) is an example of this situation. For these problems Taylor's theorem is most often used to construct the expansion, and it is not unusual to have to use it more than once. Typical examples are used below to illustrate how this is done.

Example 1

$$f(\epsilon) = \sin(e^{\epsilon}).$$

This is a compound function. To find, say, a three-term expansion of this for small ϵ one starts with the innermost function, which in this case is e^{ϵ}. To find a three-term expansion of this we can use the Taylor expansion of e^x given in Table 2.1, because ϵ small is equivalent in this case to x near zero. So, we have $e^{\epsilon} \sim 1 + \epsilon + \frac{1}{2}\epsilon^2 + \cdots$ and from this we conclude

$$\sin(e^{\epsilon}) \sim \sin\left(1 + \epsilon + \frac{1}{2}\epsilon^2 + \cdots\right).$$

The next observation to make is that the argument of the sine function on the right hand side has the form $\sin(1+y)$ where $y = \epsilon + \frac{1}{2}\epsilon^2 + \cdots$ is close to zero for small ϵ. This means the expansion given in Table 2.1 for $\sin(a+x)$ is applicable, where $a = 1$ and $x = y$. Using this fact we obtain

$$
\begin{aligned}
\sin(e^{\epsilon}) &\sim \sin(1+y) \\
&\sim \sin(1) + \cos(1)y - \frac{1}{2}\sin(1)y^2 + \cdots \\
&\sim \sin(1) + \cos(1)\left(\epsilon + \frac{1}{2}\epsilon^2 + \cdots\right) \\
&\qquad - \frac{1}{2}\sin(1)\left(\epsilon + \frac{1}{2}\epsilon^2 + \cdots\right)^2 + \cdots \\
&\sim \sin(1) + \epsilon\cos(1) + \frac{1}{2}\epsilon^2\left[\cos(1) - \sin(1)\right] + \cdots . \quad \blacksquare
\end{aligned} \tag{2.13}
$$

One might argue that the above calculation is not necessary because (2.13) can be obtained easily, and directly, from Taylor's theorem applied to $f(\epsilon) = \sin(e^{\epsilon})$. This is correct, and it is worthwhile to have multiple methods available for constructing an expansion. However, the direct approach only works on certain functions. It is easy to find examples when the direct approach does not work, one is given below, and another is given in Exercise 2.2.

Example 2

$$
f(\epsilon) = \frac{1}{[1-\cos(\epsilon)]^3} .
$$

To find a two-term expansion of this for small ϵ, we start with the inner most function, which is $\cos(\epsilon)$. Using the Taylor expansion of $\cos(x)$ given in Table 2.1, we have

$$
\cos(\epsilon) \sim 1 - \frac{1}{2}\epsilon^2 + \frac{1}{24}\epsilon^4 + \cdots .
$$

With this

$$
\begin{aligned}
\frac{1}{1-\cos(\epsilon)} &\sim \frac{1}{\frac{1}{2}\epsilon^2 - \frac{1}{24}\epsilon^4 + \cdots} \\
&= \frac{2}{\epsilon^2}\frac{1}{1 - \frac{1}{12}\epsilon^2 + \cdots} .
\end{aligned}
$$

The last term can be expanded using the binomial expansion, which is the second entry in Table 2.1. In particular, with $a = 1$, $x = -\frac{1}{12}\epsilon^2 + \cdots$, and $\gamma = -3$,

$$\begin{aligned}\frac{1}{(1-\frac{1}{12}\epsilon^2+\cdots)^3} &= \left(1-\frac{1}{12}\epsilon^2+\cdots\right)^{-3}\\ &\sim 1-3\left(-\frac{1}{12}\epsilon^2+\cdots\right)+6\left(-\frac{1}{12}\epsilon^2+\cdots\right)^2+\cdots\\ &\sim 1+\frac{1}{4}\epsilon^2+\cdots.\end{aligned}$$

The resulting two-term expansion is

$$\frac{1}{[1-\cos(\epsilon)]^3} = \frac{8}{\epsilon^6}\left(1+\frac{1}{4}\epsilon^2+\cdots\right). \quad \blacksquare$$

A comment is warranted about the expansions obtained in the last two examples. A general form of the expansion obtained in (2.13) is

$$f \sim f_0 + \epsilon^\alpha f_1 + \epsilon^\beta f_2 + \cdots, \tag{2.14}$$

where $0 < \alpha < \beta < \cdots$. In comparison, a general form of the expansion obtained in the second example is

$$f \sim \epsilon^\alpha f_0 + \epsilon^\beta f_1 + \epsilon^\gamma f_2 + \cdots, \tag{2.15}$$

where $\alpha < \beta < \gamma < \cdots$. The reason for pointing this out is that in the problems to follow it is necessary to guess, at the start, what form the expansion has. Our default assumption will be (2.14). The more general version given in (2.15) will only be used if (2.14) fails, or else there is some indication that such a general form is necessary.

Taylor's theorem is the most used method for expanding functions, but this should not be interpreted that one always ends up with a power series. An example is the function $e^{-1/\epsilon}$. Assuming it can be expanded using a power series, so $e^{-1/\epsilon} \sim x_0 + \epsilon x_1 + \epsilon^2 x_2 + \cdots$, then the coefficients must satisfy the following limits

$$\begin{aligned}x_0 &= \lim_{\epsilon\to 0^+} e^{-1/\epsilon},\\ x_1 &= \lim_{\epsilon\to 0^+} \frac{e^{-1/\epsilon}-x_0}{\epsilon},\\ x_2 &= \lim_{\epsilon\to 0^+} \frac{e^{-1/\epsilon}-x_0-\epsilon x_1}{\epsilon^2}.\end{aligned}$$

Using l'Hospital's rule one finds that each limit is zero, and so $x_0 = 0, x_1 = 0, x_2 = 0, \ldots$. In other words, as far as the functions $1, \epsilon, \epsilon^2, \ldots$ are concerned, $e^{-1/\epsilon}$ is just zero. This function certainly has rather small values but it is not identically zero. What is happening is that $e^{-1/\epsilon}$ goes to zero so quickly that the power functions are not able to describe it other than just conclude the function is zero. In this case $e^{-1/\epsilon}$ is said to be transcendentally small

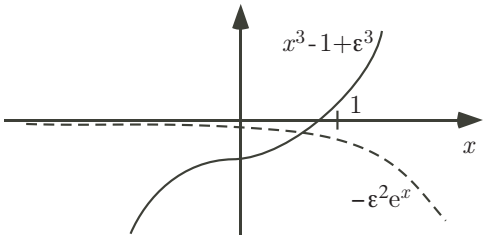


Figure 2.2 Sketch of the functions appearing in the transcendental equation in (2.16).

relative to the power functions. Even so, Taylor's theorem can be used with such functions and an example is $\sqrt{1+e^{-1/\epsilon}}$. Given the small value of $e^{-1/\epsilon}$ then we can think of the function as $\sqrt{1+y}$ where y is close to zero. Using Taylor's theorem, $\sqrt{1+y} \sim 1 + \frac{1}{2}y - \frac{1}{8}y^2 + \cdots$ and from this we conclude $\sqrt{1+e^{-1/\epsilon}} \sim 1 + \frac{1}{2}e^{-1/\epsilon} - \frac{1}{8}e^{-2/\epsilon} + \cdots$. This result shows that the appropriate scale functions in this case are not power functions but the functions $1, e^{-1/\epsilon}, e^{-2/\epsilon}, e^{-3/\epsilon}, \ldots$.

2.2.2 Given an Algebraic or Transcendental Equation

The idea here is that we are given an algebraic or transcendental equation and we want to construct an approximation for the solution(s). This is exactly what we did for the quadratic equation example (2.1). To use the method on a slightly more difficult problem consider solving

$$x^3 + \epsilon^2 e^x - 1 + \epsilon^3 = 0. \tag{2.16}$$

Our goal is to derive a two-term approximation of the solution. It is recommended that the first step in the construction is to assess how many solutions there are and, if possible, their approximate location. The reason for this is that we will have to guess the form of the expansion and any information we might have about the solution can be helpful. With this in mind the functions involved in this equation are sketched in Figure 2.2. There is one real-valued solution that is located slightly to the left of $x = 1$ for small values of ϵ. In other words, the expansion for the solution should not start out as $x \sim \epsilon x_0 + \cdots$ because this would be assuming that the solution goes to zero as $\epsilon \to 0$. Similarly, we should not assume $x \sim \frac{1}{\epsilon}x_0 + \cdots$ as the solution does not become unbounded as $\epsilon \to 0$. For this reason we will assume that the appropriate expansion has the form

$$x \sim x_0 + \epsilon^\alpha x_1 + \cdots . \tag{2.17}$$

This is going to be substituted into (2.16) and this requires us to expand e^x. Using (2.17), and Table 2.1,

$$\begin{aligned} e^x &\sim e^{x_0+\epsilon^\alpha x_1+\cdots} \\ &= e^{x_0} e^{\epsilon^\alpha x_1+\cdots}. \end{aligned}$$

Setting $y = \epsilon^\alpha x_1 + \cdots$, and noting that y is close to zero for small ϵ, then

$$\begin{aligned} e^x &\sim e^{x_0} e^y \\ &= e^{x_0}\left(1 + y + \frac{1}{2}y^2 + \cdots\right) \\ &= e^{x_0}\left(1 + (\epsilon^\alpha x_1 + \cdots) + \frac{1}{2}(\epsilon^\alpha x_1 + \cdots)^2 + \cdots\right) \\ &= e^{x_0}\left(1 + \epsilon^\alpha x_1 + \cdots\right). \end{aligned}$$

Using the binomial expansion, given in Table 2.1, and (2.17) we also have that

$$x^3 \sim x_0^3 + 3\epsilon^\alpha x_0^2 x_1 + \cdots .$$

With this, the original equation given in (2.16) takes the form

$$x_0^3 + 3\epsilon^\alpha x_0^2 x_1 + \cdots + \epsilon^2 e^{x_0}\left(1 + \epsilon^\alpha x_1 + \cdots\right) - 1 + \epsilon^3 = 0. \tag{2.18}$$

The first problem to solve is obtained by simply setting $\epsilon = 0$, which gives us the following $O(1)$ problem.

$O(1)$ $x_0^3 - 1 = 0$

The real-valued solution is $x_0 = 1$. With this, the next term in (2.18) that must be considered is $\epsilon^2 e^{x_0}$. Given that there are no ϵ^2 terms on the right-hand side of the equation then one of the other terms on the left must balance with $\epsilon^2 e^{x_0}$. The only one available is $3\epsilon^\alpha x_0^2 x_1$, and for this to happen we get $\alpha = 2$. This gives us the following problem.

$O(\epsilon^2)$ $3x_0^2 x_1 + e^{x_0} = 0$

The solution is $x_1 = -\frac{1}{3}e$.

We have therefore found that a two-term expansion of the solution is

$$x \sim 1 - \frac{1}{3}\epsilon^2 e + \cdots . \tag{2.19}$$

This expansion is plotted in Figure 2.3 along with the numerical solution. The asymptotic nature of the approximation is evident as $\epsilon \to 0$.

The procedure used to find x_0 and x_1 can be continued without difficulty to find the higher-order terms x_2, x_3, etc. It is also very easy to extend

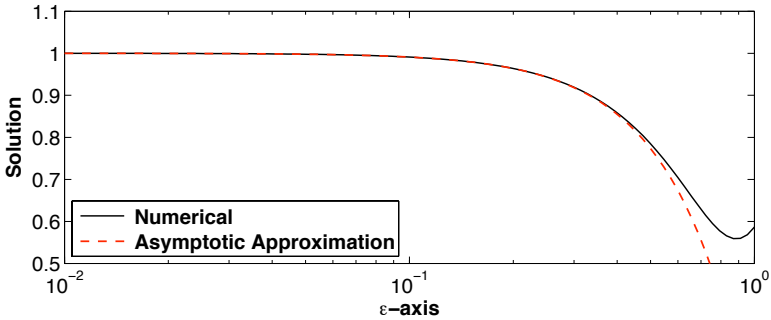


Figure 2.3 Comparison between the numerical solution of (2.16) and the asymptotic expansion (2.19).

the procedure to find the complex-valued solutions. What might not have been noticed is that not all of the terms in the original equation contribute to the approximation of the solution in (2.19). Namely, the ϵ^3 term does not contribute and the reason is that we have only computed the expansion through ϵ^2. In fact, if ϵ^3 were to be replaced with ϵ^4 or $\sin(\epsilon^3)$ the expansion in (2.19) still holds. It would not hold, however, if ϵ^3 were to be changed to ϵ or $\cos(\epsilon)$.

2.2.3 Given an Initial Value Problem

The next stage in the development is to apply regular expansions to problems involving differential equations. We will work out two examples, the first involves a single equation, and the second a system.

Example 1

The projectile problem furnishes an excellent example. Using (1.65) - (1.67) the problem to solve is

$$\frac{d^2x}{dt^2} = -\frac{1}{(1+\epsilon x)^2}, \quad \text{for} \quad 0 < t, \tag{2.20}$$

where

$$x(0) = 0, \tag{2.21}$$

$$\frac{dx}{dt}(0) = 1. \tag{2.22}$$

It is important to note that we are using the nondimensional problem and not the original given in (1.65) - (1.67). The use of an asymptotic expan-

sion is predicated on one or more parameters taking on an extreme value. In the projectile problem the assumption is that the initial velocity v_0 is small. As we saw in the last chapter, the solution depends on the parameters through a combination of products, both dimensional and nondimensional. Consequently, the study of small v_0 is actually a study of what happens when the parameter group containing v_0 takes on an extreme value. Based on the scaling we used the specific limit is $\epsilon \to 0$.

The procedure for constructing the expansion will mimic what was done earlier. We start by stating what we believe to be the appropriate form for the expansion. Generalizing (2.6), our assumption is

$$x \sim x_0(t) + \epsilon^\alpha x_1(t) + \cdots . \tag{2.23}$$

The expansion is suppose to identify how the solution depends on ϵ. The terms in the expansion can, and almost inevitability will, depend on the other variables and parameters in the problem. For the projectile problem this means that each term in the expansion depends on time and this dependence is included in (2.23).

In preparation for substituting (2.23) into (2.20) note

$$\begin{aligned}\frac{1}{(1+\epsilon x)^2} &= 1 - 2\epsilon x + 3\epsilon^2 x^2 + \cdots \\ &\sim 1 - 2\epsilon(x_0 + \epsilon^\alpha x_1 + \cdots) + 3\epsilon^2(x_0 + \cdots)^2 + \cdots \\ &= 1 - 2\epsilon x_0 + \cdots .\end{aligned}$$

With this, the differential equation (2.20) becomes

$$x_0'' + \epsilon^\alpha x_1'' + \cdots = -1 + 2\epsilon x_0 + \cdots . \tag{2.24}$$

It is critical that the initial conditions are also included, and for these we have

$$x_0(0) + \epsilon^\alpha x_1(0) + \cdots = 0, \tag{2.25}$$

$$x_0'(0) + \epsilon^\alpha x_1'(0) + \cdots = 1. \tag{2.26}$$

As usual we break the above equations down into problems depending on the power of ϵ.

$O(1)$ $x_0'' = -1$
$x_0(0) = 0$, $x_0'(0) = 1$

The solution of this problem is $x_0 = t(1 - \frac{1}{2}t)$. With this the next highest term left in (2.24) is $2\epsilon x_0$. The term available to balance with this is $\epsilon^\alpha x_1''$, and from this we conclude $\alpha = 1$. This gives us the following problem.

$O(\epsilon)$ $x_1'' = 2x_0$
$x_1(0) = 0,\ x_1'(0) = 0$
The solution of this problem is $x_1 = \frac{1}{12}t^3(4-t)$.

We have therefore found that a two-term expansion of the solution is

$$x \sim t(1 - \frac{1}{2}t) + \frac{1}{12}\epsilon t^3(4-t) + \cdots . \tag{2.27}$$

This rather simple-looking expression is a two-term asymptotic expansion of the nonlinear projectile problem. Physically, the first term, $t(1-\frac{1}{2}t)$, gives the displacement of the projectile for a uniform gravitational field, and is the nondimensional version of (1.5). The second term, $\epsilon t^3(4-t)/12$, gives us the correction due to the nonlinear gravitational field.

To determine how well we have done in approximating the solution, a comparison is shown in Figure 2.4 for $\epsilon = 0.1$ and $\epsilon = 0.01$. It is seen that the one-term approximation, $x \sim t(1-\frac{1}{2}t)$, produces a reasonably accurate approximation for $\epsilon = 0.01$, but not when $\epsilon = 0.1$. In contrast, the two-term approximation (2.27) does very well for both values. To put this into perspective, if the object's initial velocity is the speed of sound then $\epsilon \approx 0.002$, while if it is equal to the Earth's escape velocity then $\epsilon \approx 2$. Figure 2.4 shows

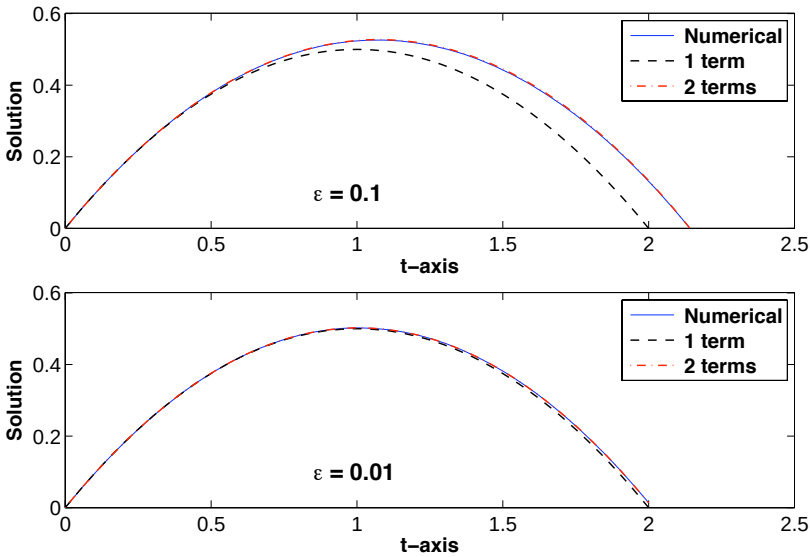


Figure 2.4 Comparison between the numerical solution of the projectile problem and the asymptotic expansion (2.27). In the upper graph $\epsilon = 0.1$, and in the lower graph $\epsilon = 0.01$. In both graphs the curves for the exact solution and two-term expansion are almost indistinguishable.

that at subsonic velocities the uniform gravitational field approximation is adequate. If the initial velocity is a bit larger, more than three times the speed of sound, then the nonlinear correction is needed. Finally, we expect the approximation to improve as ϵ gets closer to zero, and the graphs in Figure 2.4 are consistent with that expectation. ■

Example 2

The ideas used to find an approximation for a single equation are easily extended to systems. As an example, consider the thermokinetic model of Exercise 1.17. In nondimensional variables, the equations are

$$\frac{du}{dt} = 1 - ue^{\epsilon(q-1)}, \tag{2.28}$$

$$\frac{dq}{dt} = ue^{\epsilon(q-1)} - q. \tag{2.29}$$

The initial conditions are $u(0) = q(0) = 0$. We are assuming here that the nonlinearity is weak, which means that ϵ is small. Also, to simplify the problem, the other parameters that appear in the nondimensionalization have been set to one.

Generalizing (2.23), we expand both functions using our usual assumption, which is that

$$\begin{aligned} u &\sim u_0(t) + \epsilon u_1(t) + \cdots, \\ q &\sim q_0(t) + \epsilon q_1(t) + \cdots. \end{aligned}$$

In writing down the above expansions, it is assumed that the second terms in the expansions are $O(\epsilon)$, rather than $O(\epsilon^\alpha)$. This is done to simplify the calculations to follow.

Before substituting the expansions into the differential equations, note that

$$\begin{aligned} e^{\epsilon(q-1)} &\sim 1 + \epsilon(q-1) + \frac{1}{2}\epsilon^2(q-1)^2 + \cdots \\ &\sim 1 + \epsilon(q_0 + \epsilon q_1 + \cdots - 1) + \frac{1}{2}\epsilon^2(q_0 + \epsilon q_1 + \cdots - 1)^2 + \cdots \\ &\sim 1 + \epsilon(q_0 - 1) + \cdots, \end{aligned}$$

and

$$\begin{aligned} ue^{\epsilon(q-1)} &\sim (u_0 + \epsilon u_1 + \cdots)[1 + \epsilon(q_0 - 1) + \cdots] \\ &\sim u_0 + \epsilon\,[u_0(q_0 - 1) + u_1] + \cdots. \end{aligned}$$

With this, (2.28), (2.29) take the form

$$u_0' + \epsilon u_1' + \cdots = 1 - u_0 - \epsilon\left(u_0(q_0 - 1) + u_1\right) + \cdots,$$
$$q_0' + \epsilon q_1' + \cdots = u_0 - q_0 + \epsilon\left(u_0(q_0 - 1) + u_1 - q_1\right) + \cdots.$$

As usual we break the above equations down into problems depending on the power of ϵ.

$O(1)$ $u_0' = 1 - u_0$
$q_0' = u_0 - q_0$

The solution of this problem that satisfies the initial conditions $u_0(0) = q_0(0) = 0$ is $u_0 = 1 - e^{-t}$, and $q_0 = 1 - (1+t)e^{-t}$.

$O(\epsilon)$ $u_1' = -u_1 - u_0(q_0 - 1)$
$q_1' = -q_1 + u_1 + u_0(q_0 - 1)$

The initial conditions are $u_1(0) = q_1(0) = 0$. The equation for u_1 is first order, and the solution can be found using an integrating factor. Once u_1 is determined then the q_1 equation can be solved using an integrating factor. Carrying out the calculation one finds that $u_1 = \frac{1}{2}(t^2 + 2t - 4)e^{-t} + (2+t)e^{-2t}$, $q_1 = \frac{1}{6}(t^3 - 18t + 30)e^{-t} - (2t+5)e^{-2t}$.

We have therefore found that a two-term expansion of the solution is

$$u(t) \sim 1 - e^{-t} + \epsilon\left(\frac{1}{2}(t^2 + 2t - 4)e^{-t} + (2+t)e^{-2t}\right), \tag{2.30}$$

$$q(t) \sim 1 - (1+t)e^{-t} + \epsilon\left(\frac{1}{6}(t^3 - 18t + 30)e^{-t} - (2t+5)e^{-2t}\right). \tag{2.31}$$

A comparison of the numerical solution for $q(t)$, and the above asymptotic approximation for $q(t)$ is shown in Figure 2.5 for $\epsilon = 0.1$. It is seen that even the one-term approximation, $q \sim 1 - (1+t)e^{-t}$, produces a reasonably accurate approximation, while the two-term approximation is indistinguish-

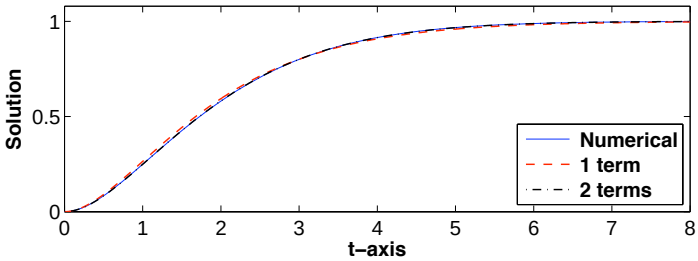


Figure 2.5 Comparison between the numerical solution for $q(t)$, and the asymptotic expansion (2.31). In the calculation, $\epsilon = 0.1$.

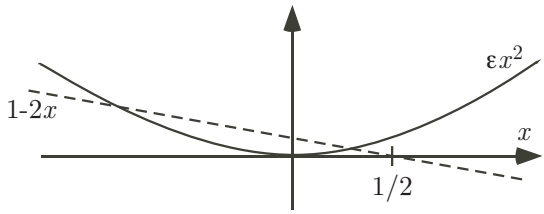


Figure 2.6 Sketch of the functions appearing in the quadratic equation in (2.32).

able from the numerical solution. The approximations for $u(t)$, which are not shown, are also as accurate. ■

2.3 Introduction to Singular Perturbations

All of the equations considered up to this point produce regular expansions. This means, roughly, that the expansions can be found without having to transform the problem. We now turn our attention to those that are not regular, what are known as singular perturbation problems. The first example considered is the quadratic equation

$$\epsilon x^2 + 2x - 1 = 0. \tag{2.32}$$

A tip-off that this is singular is that ϵ multiplies the highest-order term in the equation. Setting $\epsilon = 0$ drops the order down to linear, and this has dramatic effects on the number and types of solutions.

The best place to begin is to sketch the functions in the equation to get an idea of the number and location of the solutions. This is done in Figure 2.6, which shows that there are two real-valued solutions. One is close to $x = \frac{1}{2}$ and for this reason it is expected that an expansion of the form $x \sim x_0 + \epsilon^\alpha x_1 + \cdots$ will work. The second solution is far to the left on the negative x-axis, and the smaller ϵ the farther to the left it is located. Consequently, we should not be shocked later when we find that the expansion for this solution has the form $x \sim \frac{1}{\epsilon} x_0 + \cdots$, where x_0 is negative.

We start out as if this were a regular perturbation problem and assume the solutions can be expanded as

$$x \sim x_0 + \epsilon^\alpha x_1 + \cdots . \tag{2.33}$$

Substituting this into (2.32) we obtain

$$\epsilon(x_0^2 + 2\epsilon^\alpha x_0 x_1 + \cdots) + 2(x_0 + \epsilon^\alpha x_1 + \cdots) - 1 = 0. \tag{2.34}$$

Equating like powers of ϵ produces the following problems.

$O(1)$ $2x_0 - 1 = 0$

The solution is $x_0 = \frac{1}{2}$. With this, to balance the term $\epsilon^2 x_0^2$ we take $\alpha = 1$. This gives us the following problem.

$O(\epsilon^2)$ $x_0^2 + 2x_1 = 0$

The solution is $x_1 = -\frac{1}{8}$.

We therefore have

$$x \sim \frac{1}{2} - \frac{1}{8}\epsilon + \cdots . \tag{2.35}$$

This expansion is consistent with the conclusions we derived earlier from Figure 2.6 for one of the solutions. It is also apparent that no matter how many terms we calculate in the expansion (2.33) we will not obtain the second solution.

The failure of the regular expansion to find all of the solutions is typical of a singular perturbation problem. The method used to remedy the situation is to introduce a scaling transformation. Specifically, we will change variables and let

$$\bar{x} = \frac{x}{\epsilon^\gamma} . \tag{2.36}$$

With this, (2.32) takes the form

$$\underset{①}{\epsilon^{1+2\gamma}\bar{x}^2} + \underset{②}{2\epsilon^\gamma \bar{x}} - \underset{③}{1} = 0. \tag{2.37}$$

The reason for not finding two solutions earlier was that the quadratic term was lost when $\epsilon = 0$. Given the fact that this term is why there are two solutions in the first place we need to determine how to keep it in the equation as $\epsilon \to 0$. In other words, term ① in (2.37) must balance with one of the other terms and this must be the first problem solved as $\epsilon \to 0$. For example, suppose we assume the balance is between terms ① and ③, while term ② is of higher or equal order. For this to occur, we need $O(\epsilon^{1+2\gamma}) = O(1)$ and this would mean $\gamma = -\frac{1}{2}$. With this ①, ③ $= O(1)$ and ② $= O(\epsilon^{-1/2})$. This result is inconsistent with our original assumption that ② is higher order. Therefore, the balance must be with another term. This type of argument is central to singular problems and we will use a table format to present the steps used to determine the correct balance.

Balance	Condition on γ	Consistency Check	Conclusion
① $\sim$ ③ with ② higher order	$1 + 2\gamma = 0$ $\Rightarrow \gamma = -1/2$	①, ③ $= O(\epsilon)$ and ② $= O(\epsilon^{-1/2})$	Inconsistent with balance
① $\sim$ ② with ③ higher order	$1 + 2\gamma = \gamma$ $\Rightarrow \gamma = -1$	①, ② $= O(\epsilon^{-1})$ and ③ $= O(1)$	Consistent with balance

Based on the above analysis, $\gamma = -1$ and with this the equation takes the form

$$\bar{x}^2 + 2\bar{x} - \epsilon = 0. \tag{2.38}$$

With this we assume our usual expansion, which is

$$\bar{x} \sim \bar{x}_0 + \epsilon^\alpha \bar{x}_1 + \cdots . \tag{2.39}$$

The equation in this case becomes

$$\bar{x}_0^2 + 2\epsilon^\alpha \bar{x}_0 \bar{x}_1 + \cdots + 2(\bar{x}_0 + \epsilon^\alpha \bar{x}_1 + \cdots) - \epsilon = 0. \tag{2.40}$$

This gives us the following problems.

$O(1)$ $\bar{x}_0^2 + 2\bar{x}_0 = 0$

The solutions are $\bar{x}_0 = -2$ and $\bar{x}_0 = 0$. With this, to balance the $-\epsilon$ term in (2.40), we take $\alpha = 1$. This gives us the following problem.

$O(\epsilon)$ $2\bar{x}_0\bar{x}_1 + 2\bar{x}_1 - 1 = 0$

If $\bar{x}_0 = -2$ then $\bar{x}_1 = -\frac{1}{2}$, while if $\bar{x}_1 = 0$ then $\bar{x}_1 = \frac{1}{2}$.

It might appear that we have somehow produced three solutions, the one in (2.35) along with the two found above. However, it is not hard to show that the solution corresponding to $\bar{x}_0 = 0$ is the same one that was found earlier using a regular expansion. Consequently, the sought-after second solution is

$$x \sim \frac{1}{\epsilon}(-2 - \frac{1}{2}\epsilon + \cdots). \tag{2.41}$$

The procedure used to derive this result contained many of the ideas we used to find regular expansions. The most significant difference is the introduction of a scaled variable, (2.36), and the subsequent balancing used to determine how the highest-order term participates in the problem. As we will see shortly, these will play a critical role when analyzing similar problems involving differential equations.

2.4 Introduction to Boundary Layers

As our introductory example of a singular perturbation problem involving a differential equation we will consider solving

$$\epsilon y'' + 2y' + 2y = 0, \quad \text{for } \ 0 < x < 1, \tag{2.42}$$

where the boundary conditions are

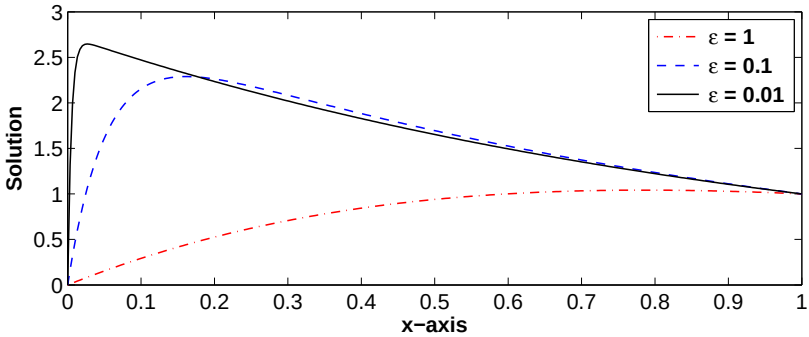


Figure 2.7 Graph of the exact solution of the boundary value problem (2.42)-(2.44), for various values of ϵ. Note the appearance of the boundary layer near $x = 0$ as ϵ decreases.

$$y(0) = 0, \tag{2.43}$$

and

$$y(1) = 1. \tag{2.44}$$

This is a boundary value problem and it has the telltale signs of a singular perturbation problem. Namely, ϵ is multiplying the highest derivative so setting $\epsilon = 0$ results in a lower order problem.

This problem has been selected as the introductory problem because it can be solved exactly, and we will be able to use this to evaluate the accuracy of our approximations. To find the exact solution one assumes $y = e^{rx}$ and from the differential equation concludes that $r_\pm = (-1 \pm \sqrt{1-2\epsilon})/\epsilon$. With this the general solution is $y = c_1 e^{r_+ x} + c_2 e^{r_- x}$. Imposing the two boundary conditions one finds that

$$y = \frac{e^{r_+ x} - e^{r_- x}}{e^{r_+} - e^{r_-}}. \tag{2.45}$$

This function is plotted in Figure 2.7, for various values of ϵ. It is seen that as ϵ deceases the solution starts to show a rapid transition in the region near $x = 0$. Also, if you look at the graphs you will notice that the rapid change takes place over a spatial interval that has a length about equal to the size of ϵ. The reason for making this observation is that our approximation will consist of two pieces, one for x near zero and the other that applies to the rest of the interval. The fact that we end up having to split the interval is not unexpected given what is occurring in Figure 2.7.

Step 1. Outer Solution

The first step is simply to use a regular expansion and see what results. Similar to what we did with the earlier projectile problem, it is assumed

$$y \sim y_0(x) + \epsilon y_1(x) + \cdots . \tag{2.46}$$

Introducing this into (2.42) we obtain

$$\epsilon(y_0'' + \epsilon y_1'' + \cdots) + 2(y_0' + \epsilon y_1' + \cdots) + 2(y_0 + \epsilon y_1 + \cdots) = 0, \tag{2.47}$$

where

$$y_0(0) + \epsilon y_1(0) + \cdots = 0, \tag{2.48}$$

and

$$y_0(1) + \epsilon y_1(1) + \cdots = 1. \tag{2.49}$$

Proceeding in the usual manner yields the following problems.

$O(1)$ $2y_0' + 2y_0 = 0$
$y_0(0) = 0$, $y_0(1) = 1$

The general solution of the differential equation is $y_0 = ae^{-x}$, where a is an arbitrary constant. This is where the singular nature of the problem starts to have an affect. We have one constant but there are two boundary conditions. Can we satisfy at least one of them? For some problems the answer is no, and we will consider such an example later. For this problem we can and we need to determine which one. To help with this decision, the solution is sketched in Figure 2.8 in the case of when $a > 0$ and when $a < 0$. The two boundary conditions are also shown in the figure. It is apparent that of the two possibilities, the $a > 0$ curve is the only one capable of satisfying one of the boundary conditions, and it is the one at $x = 1$. Assuming this is the case then $a = e$ and $y_0(x) = e^{1-x}$.

$O(\epsilon)$ $y_0'' + 2y_1' + 2y_1 = 0$
$y_1(1) = 0$

Note that only the boundary condition at $x = 1$ is listed here as this is the only one we believe this expansion is capable of satisfying. The general solution of the differential equation is $y_1 = (b - x/2)e^{1-x}$, where b is an arbitrary constant. With the given boundary condition we obtain $y_1(x) = (1 - x)e^{1-x}/2$.

Our regular expansion has yielded

$$y \sim e^{1-x} + \cdots . \tag{2.50}$$

Only the first term has been included here as this is all we are going to determine in this example. The second term was calculated earlier to demonstrate that it is easy to find, and also to show that including the second term does not help us satisfy the boundary condition at $x = 0$. It is this fact that will require us to scale the problem and this brings us to the next step.

STEP 2. Inner, or Boundary Layer, Solution
We will now construct an approximation of the solution in the neighborhood of $x = 0$, which corresponds to the interval where the function undergoes a rapid increase as shown in Figure 2.7. Given its location, the approximation is called the boundary layer solution. It is also known as an inner solution, and correspondingly, the approximation in (2.50) is the outer solution. The width of this layer shrinks as $\epsilon \to 0$, so we must make a change of variables to account for this. With this in mind we introduce the boundary layer coordinate

$$\bar{x} = \frac{x}{\epsilon^{\gamma}} \,. \tag{2.51}$$

The exact value of γ will be determined shortly but we already have some inkling what it might be. We saw in Figure 2.7 that the rapid change in the solution near $x = 0$ takes place over an interval that has width of about ϵ. So, it should not be too surprising that we will find that $\gamma = 1$. In any case, using the chain rule

$$\frac{d}{dx} = \frac{d\bar{x}}{dx}\frac{d}{d\bar{x}} = \frac{1}{\epsilon^{\gamma}}\frac{d}{d\bar{x}} \,, \tag{2.52}$$

and

$$\frac{d^2}{dx^2} = \frac{1}{\epsilon^{2\gamma}}\frac{d^2}{d\bar{x}^2} \,. \tag{2.53}$$

We will designate the solution as $Y(\bar{x})$ when using $\bar{x}$ as the independent variable. With this the differential equation becomes

$$\underset{①}{\epsilon^{1-2\gamma}Y''} + \underset{②}{2\epsilon^{-\gamma}Y'} + \underset{③}{2Y} = 0. \tag{2.54}$$

We determine γ by balancing the terms in the above equation. Our goal is for the highest derivative to remain in the equation as $\epsilon \to 0$. This gives us the following two possibilities.

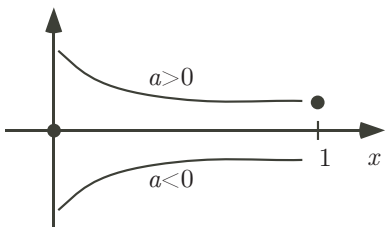


Figure 2.8 Sketch of the possible form of the outer solution $y_0 = ae^x$, depending on the sign of a. Also shown are the two given boundary conditions.

Balance	Condition on γ	Consistency Check	Conclusion
① $\sim$ ③ with ② higher-order	$1-2\gamma=0$ $\Rightarrow \gamma=1/2$	①, ③ $=O(1)$ and ② $=O(\epsilon^{-1/2})$	Inconsistent with balance
① $\sim$ ② with ③ higher-order	$1-2\gamma=-\gamma$ $\Rightarrow \gamma=1$	①, ② $=O(\epsilon^{-1})$ and ③ $=O(1)$	Consistent with balance

Based on the above analysis we take $\gamma = 1$ and with this the differential equation takes the form

$$Y'' + 2Y' + 2\epsilon Y = 0. \tag{2.55}$$

Assuming $Y(\bar{x}) \sim Y_0(\bar{x}) + \cdots$ the differential equation becomes

$$(Y_0'' + \cdots) + 2(Y_0' + \cdots) + 2\epsilon(Y_0 + \cdots) = 0, \tag{2.56}$$

and from this we obtain the following problem.

$O(1)$ $Y_0'' + 2Y_0' = 0$
$Y_0(0) = 0$

Note that the boundary condition at $x = 0$ has been included here but not the one at $x = 1$. The reason is that we are building an approximation of the solution in the immediate vicinity of $x = 0$ and it is incorrect to assume it can satisfy the condition at the other end of the interval. Now, the general solution of the differential equation is $Y_0 = A + Be^{-2\bar{x}}$, where A, B are arbitrary constants. With the given boundary condition this reduces to $Y_0 = A(1 - e^{-2\bar{x}})$.

The approximation of the solution in the boundary layer is

$$Y(\bar{x}) \sim A(1 - e^{-2\bar{x}}) + \cdots . \tag{2.57}$$

We will determine A by connecting this result with the approximation we have for the outer region, and this brings us to the next step.

Step 3. Matching
We have made several assumptions about the solution and it is now time to prove that they are correct. To explain what this means, our approximation consists of two different expansions, and each applies to a different part of the interval. The situation we find ourselves in is sketched in Figure 2.9. This indicates that when coming out of the boundary layer the approximation in (2.57) approaches a constant value A. Similarly, the outer solution approaches a constant value, e, as it enters the boundary layer. There is a transition region, what is usually called an overlap domain, where the two

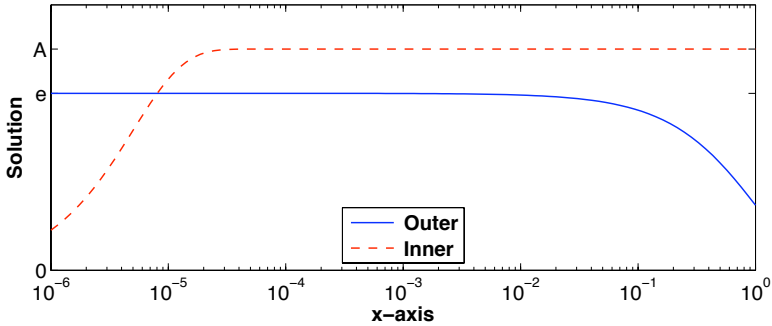


Figure 2.9 Graph of the inner approximation (2.57), and the outer approximation (2.50), before matching.

approximations are both constant. Given that they are approximations of the same function then we need to require that the inner and outer expansions are equal in this region. In more mathematical terms, the requirement we will impose on these two expansions is

$$\lim_{\bar{x}\to\infty} Y_0 = \lim_{x\to 0} y_0. \tag{2.58}$$

This is called the matching condition. With this we conclude $A = e$ and the resulting functions are plotted in Figure 2.10 for $\epsilon = 10^{-4}$. The overlap domain is clearly seen in this figure.

Step 4. Composite Expansion
The approximation of the solution we have comes in two pieces, one that applies near $x = 0$ and another that works everywhere else. Because neither can be used over the entire interval we say that they are not uniformly valid for $0 \le x \le 1$. The question we consider now is whether we can combine them

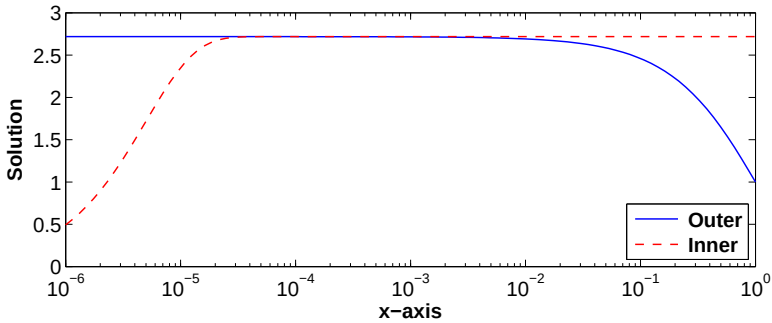


Figure 2.10 Graph of the inner approximation (2.57), and the outer approximation (2.50), after matching in the particular case of when $\epsilon = 10^{-4}$. Note the overlap region where the two approximations produce, approximately, the same result.

in some way to produce a uniform approximation, that is, one that works over the entire interval. The position we are in is summarized in Figure 2.11. The inner and outer solutions are constant outside the region where they are used to approximate the solution, and the constant is the same for both solutions. The value of the constant can be written as either $y_0(0)$ or as $Y_0(\infty)$, and the fact that they are equal is a consequence of the matching condition (2.58). This observation can be used to construct a uniform approximation. Namely, we just add the approximations together and then subtract the constant. The result is

$$\begin{aligned} y &\sim y_0(x) + Y_0(\bar{x}) - y_0(0) \\ &= e^{1-x} - e^{1-2x/\epsilon}. \end{aligned} \tag{2.59}$$

This function is known as a composite expansion and it is valid for $0 \le x \le 1$. To demonstrate its effectiveness it is plotted in Figure 2.12 along with the exact solution for $\epsilon = 10^{-1}$ and for $\epsilon = 10^{-2}$. It is evident from this figure that we have constructed a relatively simple expression that is a very good approximation of the solution over the entire interval.

2.4.1 Endnotes

One of the characteristics of a boundary layer is that its width goes to zero as $\epsilon \to 0$, yet the change in the solution across the layer does not go to zero. This type of behavior occurs in a wide variety of problems, although the terminology changes depending on the application and particular type of problem. For example, there are problems where the jump occurs in the interval $0 < x < 1$, a situation known as an interior layer. They are also not limited to BVPs and arise in IVPs, PDEs, etc. A example of this is shown in Figure 2.13. The boundary layer is the thin white region on the surface of the object. In this layer the air velocity changes rapidly, from zero on the object

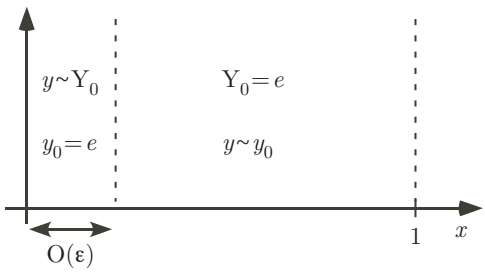


Figure 2.11 Sketch of the inner and outer regions and the values of the approximations in those regions.

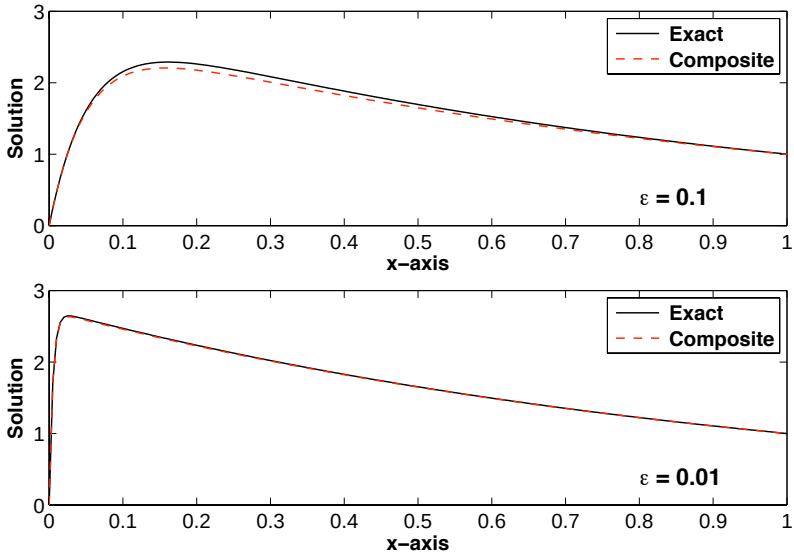


Figure 2.12 Graph of the exact solution (2.45) and composite approximation (2.59) for two values of ϵ.

to the large value in the outer flow. The parabolic curve that appears to be attached to the front of the object is a shock wave. The pressure undergoes a rapid change across the shock, and for this reason it is an example of an interior layer. The presence of a boundary layer is an issue when finding the numerical solution. As an example, Figure 2.14 shows the grid system used to solve the equations for the air flow over an object, in this case an airplane. The presence of a boundary layer necessitates the use of a large number of grid points near the surface, which greatly adds to the computational effort needed to solve the problem.

Another important comment to make concerns the existence of a boundary layer in the solution. In particular, an ϵ multiplying the highest derivative is not a guarantee of a boundary, or interior, layer. A simple example is $\epsilon y'' + y = 0$, for which the general solution is $y = a\sin(x/\sqrt{\epsilon}) + b\cos(x/\sqrt{\epsilon})$. In this case, instead of containing a rapidly decaying exponential function

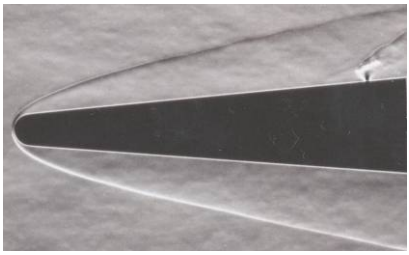


Figure 2.13 Image of high speed flow, from left to right, over a fixed, wedge-shaped object. The thin white region on the surface of the object is the boundary layer. The parabolic curve is a shock wave, a topic which is studied in Chapter 5.

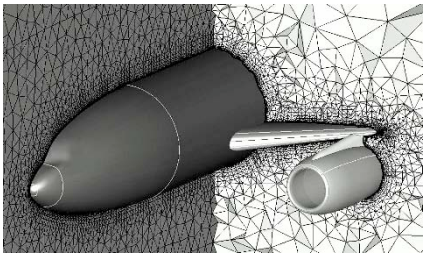


Figure 2.14 Grid refinement needed near the boundary to numerically calculate the air flow over an airplane (Steinbrenner and Abelanet [2007]).

that is characteristic of a boundary layer, the solution consists of rapidly varying oscillatory functions. The approximation method most often used in such situations is known as the WKB method. We will only scratch the surface of this subject, and a more extensive study of this can be found in Holmes [1995].

2.5 Multiple Boundary Layers

As a second boundary layer example we will consider the boundary value problem

$$\epsilon^2 y'' + \epsilon x y' - y = -e^x, \quad \text{for} \;\; 0 < x < 1, \tag{2.60}$$

where the boundary conditions are

$$y(0) = 2, \tag{2.61}$$

and

$$y(1) = 1. \tag{2.62}$$

When given a problem with a small parameter it is worthwhile to quickly check to see what might happen when $\epsilon = 0$. Setting $\epsilon = 0$ in (2.60), we lose all the derivative terms and simply end up with $y = e^x$. This function is incapable of satisfying either boundary condition, so we will find two boundary layers for this solution, one at each end of the interval. This is one of the reasons for considering this particular equation. Another is that it has variable coefficients and it is worth working out an example to see how to handle such situations. The procedure used to construct an asymptotic approximation of the solution will follow the steps we used in the last example, and for this reason there will be fewer explanations of what is being done.

Step 1. Outer Solution
Assuming $y \sim y_0(x) + \epsilon y_1(x) + \cdots$ one finds from the differential equation that $y_0 = e^x$. As stated above, this cannot satisfy either boundary condition, and this brings us to the next step.

Steps 2 and 3. Boundary Layer Solutions and Matching
Given that there is a layer at each end we need to split this step into two parts.

a) Layer at $x = 0$. In this region we will denote the solution as $Y(\bar{x})$. The boundary coordinate is the same as before. Setting $\bar{x} = x/\epsilon^{\gamma}$ and using the formulas in (2.52), (2.53) the differential equation (2.60) becomes

$$\underset{①}{\epsilon^{2-2\gamma}Y''} + \underset{②}{\epsilon\bar{x}Y'} - \underset{③}{Y} = \underset{④}{-e^{\epsilon^{\gamma}\bar{x}}}. \tag{2.63}$$

We will balance the terms in the usual manner but note that terms ③ and ④ are the same order. This is because $e^{\epsilon^{\gamma}\bar{x}} \sim 1 + \epsilon^{\gamma}\bar{x} + \cdots$. Consequently, when deciding on what balance we need in (2.63), term ④ will not be considered.

Balance	Condition on γ	Consistency Check	Conclusion
① $\sim$ ② with ③ higher-order	$2 - 2\gamma = 1$ $\Rightarrow \gamma = 1/2$	①, ② $= O(\epsilon)$ and ③ $= O(1)$	Inconsistent with balance
① $\sim$ ③ with ② higher-order	$2 - 2\gamma = 0$ $\Rightarrow \gamma = 1$	①, ③ $= O(1)$ and ② $= O(\epsilon)$	Consistent with balance

Consequently, with $\gamma = 1$, the differential equation becomes

$$Y'' + \epsilon\bar{x}Y' - Y = -e^{\epsilon\bar{x}}. \tag{2.64}$$

Assuming $Y(\bar{x}) \sim Y_0(\bar{x}) + \cdots$ we obtain the following problem to solve.

$O(1)$ $Y_0'' - Y_0 = -1$
$Y_0(0) = 2$

The general solution of the differential equation is $Y_0 = 1 + Ae^{\bar{x}} + Be^{-\bar{x}}$, where A, B are arbitrary constants. With the given boundary condition this reduces to $Y_0 = 1 + Ae^{\bar{x}} + (1 - A)e^{-\bar{x}}$.

As before, this boundary layer solution must match with the outer solution calculated earlier. The requirement is

$$\lim_{\bar{x}\to\infty} Y_0 = \lim_{x\to 0} y_0. \tag{2.65}$$

Given that $\lim_{\bar{x}\to\infty} e^{\bar{x}} = \infty$, for Y_0 to be able to match with the outer solution we must set $A = 0$. With this our first term approximation in this boundary layer is

$$Y_0(\bar{x}) = 1 + e^{-\bar{x}}. \tag{2.66}$$

b) Layer at $x = 1$. In this region we will denote the solution as $\widetilde{Y}(\widetilde{x})$. The boundary layer in this case is located at $x = 1$, and so the coordinate will be centered at this point. In particular, we let

$$\widetilde{x} = \frac{x-1}{\epsilon^{\gamma}}\,. \tag{2.67}$$

The differentiation formulas are similar to those in (2.52), (2.53). Also, we have that $x = 1 + \epsilon^{\gamma}\widetilde{x}$. With this the differential equation (2.60) becomes

$$\underset{①}{\epsilon^{2-2\gamma}\widetilde{Y}''} + \underset{②}{\epsilon^{1-\gamma}(1+\epsilon^{\gamma}\widetilde{x})\widetilde{Y}'} - \underset{③}{\widetilde{Y}} = \underset{④}{-e^{1+\epsilon^{\gamma}\widetilde{x}}}. \tag{2.68}$$

Similar to what happened earlier, terms ③ and ④ are the same order, so term ④ will not be considered in the balancing.

Balance	Condition on γ	Consistency Check	Conclusion
① $\sim$ ③ with	$2 - 2\gamma = 0$	①, ③ $= O(1)$	Consistent
② higher-order	$\Rightarrow \gamma = 1$	and ② $= O(1)$	with balance

Consequently, with $\gamma = 1$, the differential equation becomes

$$\widetilde{Y}'' + (1+\epsilon\widetilde{x})\widetilde{Y}' - \widetilde{Y} = -e^{1+\epsilon\widetilde{x}}. \tag{2.69}$$

Assuming $\widetilde{Y}(\widetilde{x}) \sim \widetilde{Y}_0(\widetilde{x}) + \cdots$ we obtain the following problem to solve.

$O(1)$ $\widetilde{Y}_0'' + \widetilde{Y}_0' - \widetilde{Y}_0 = -e$
$\widetilde{Y}_0(0) = 1$

In the boundary condition, $\widetilde{Y}_0$ is evaluated at $\widetilde{x} = 0$ because $x = 1$ corresponds to $\widetilde{x} = 0$. The general solution of the differential equation is $\widetilde{Y}_0 = e + Ae^{r_+\widetilde{x}} + Be^{r_-\widetilde{x}}$, where $r_{\pm} = (-1 \pm \sqrt{5})/2$ and A, B are arbitrary constants. With the given boundary condition this reduces to $\widetilde{Y}_0 = e + Ae^{r_+\widetilde{x}} + (1 - e - A)e^{r_-\widetilde{x}}$.

This boundary layer solution must match with the outer solution calculated earlier. The requirement is

$$\lim_{\widetilde{x}\to-\infty} \widetilde{Y}_0 = \lim_{x\to 1} y_0. \tag{2.70}$$

This expression appears different from the one used earlier for the layer at $x = 0$. The reason is that the position of the layer has changed, but the matching principle is the same. Namely, for the approximations to match it is necessary that when you come out of the boundary layer into the outer

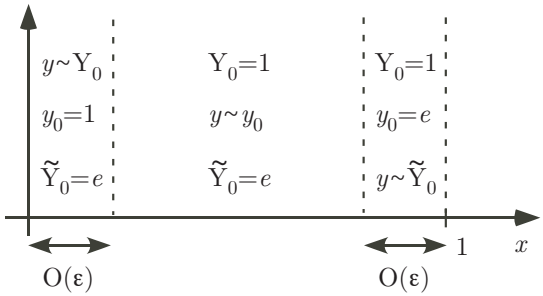


Figure 2.15 Sketch of the three regions and the values of the approximations in those regions.

region (i.e., $\widetilde{x} \to -\infty$) that you get the same value as when you enter the boundary layer from the outer region (i.e., $x \to 1$). Given that $r_+ > 0$ and $r_- < 0$ then $\lim_{\widetilde{x}\to-\infty} e^{r_-\widetilde{x}} = \infty$ and $\lim_{\widetilde{x}\to-\infty} e^{r_+\widetilde{x}} = 0$. For $\widetilde{Y}_0$ to be able to match with the outer solution we must set $1 - e - A = 0$. With this our first term approximation in this boundary layer is

$$\widetilde{Y}_0(\widetilde{x}) = e + (1-e)e^{r_+\widetilde{x}}. \tag{2.71}$$

Step 4. Composite

In a similar manner as in the last example, it is possible to combine the three approximations we have derived to produce a uniform approximation. The situation is shown schematically in Figure 2.15. It is seen that in each region the two approximations not associated with that region add to $1 + e$. This means we simply add the three approximations together and subtract $1 + e$. In other words,

$$\begin{aligned} y &\sim y_0(x) + Y_0(\bar{x}) + \widetilde{Y}_0(\widetilde{x}) - y_0(0) - y_0(1) \\ &= e^x + e^{-x/\epsilon} + (1-e)e^{r_+(x-1)/\epsilon}. \end{aligned} \tag{2.72}$$

This function is a composite expansion of the solution and it is valid for $0 \le x \le 1$. To demonstrate its effectiveness the composite approximation is plotted in Figure 2.16 along with the numerical solution for $\epsilon = 10^{-1}$ and for $\epsilon = 10^{-2}$. The approximations are not very accurate for $\epsilon = 10^{-1}$, but this is not unexpected given that ϵ is not particularly small. In contrast, for $\epsilon = 10^{-2}$ the composite approximation is quite good over the entire interval, and it is expected to get even better for smaller values of ϵ.

2.6 Multiple Scales and Two-Timing

As the last two examples have demonstrated, the presence of a boundary layer limits the region over which an approximation can be used. Said another way, the inner and outer approximations are not uniformly valid over the entire interval. The tell-tale sign that this is going to happen is that when $\epsilon = 0$ the highest derivative in the problem is lost. However, the lack of uniformity can occur in other ways and one investigated here relates to changes in the solution as a function of time. It is easier to explain what happens by working out a typical example. For this we use the pendulum problem. Letting $\theta(t)$ be the angular deflection made by the pendulum, as shown in Figure 2.17, the problem is

$$\theta'' + \sin(\theta) = 0, \tag{2.73}$$

where

$$\theta(0) = \epsilon, \tag{2.74}$$

and

$$\theta'(0) = 0. \tag{2.75}$$

The equation of motion (2.73) comes from Newton's second law, $F = ma$, where the external forcing F is gravity. It is assumed the initial angle is small,

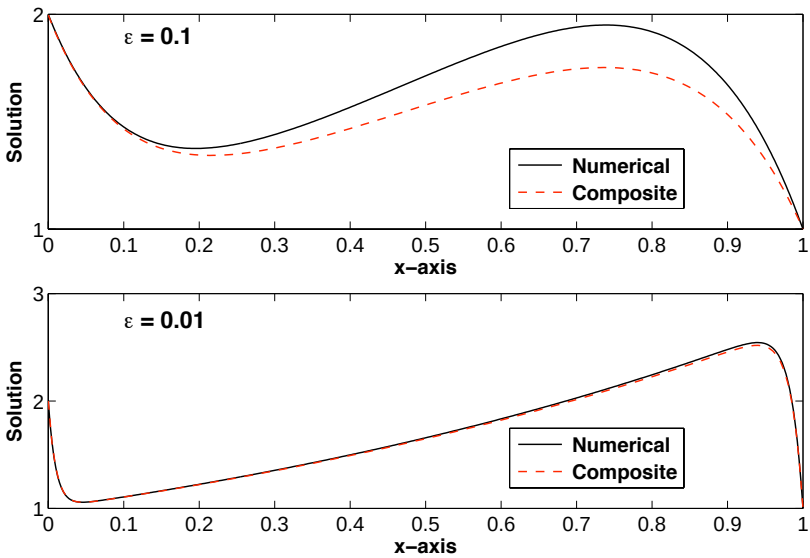


Figure 2.16 Graph of the numerical solution of the boundary value problem (2.60)-(2.62) and the composite approximation of the solution (2.72). In the upper plot $\epsilon = 10^{-1}$ and in the lower plot $\epsilon = 10^{-2}$.

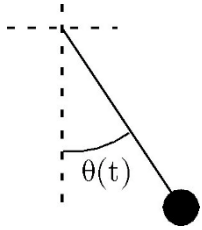


Figure 2.17 Pendulum example.

and this is the reason for the initial condition (2.74). It is also assumed that the pendulum starts from rest, so the initial velocity (2.75) is zero.

Although the problem is difficult to solve we have at least some idea of what the solution looks like because of everyday experience with a pendulum (e.g., watching a grandfather clock or using a swing). Starting out with the given initial conditions, the pendulum should simply oscillate back and forth. A real pendulum will eventually stop due to damping, but we have not included this in the model so our pendulum should go forever.

The fact that the small parameter is in the initial condition, and not in the differential equation, is a bit different from what we had in the last two examples but we are still able to use our usual approximation methods. The appropriate expansion in this case is

$$\theta(t) \sim \epsilon(\theta_0(t) + \epsilon^\alpha \theta_1(t) + \cdots). \tag{2.76}$$

The ϵ multiplying the series is there because of the initial condition. If we did not have it, and tried $\theta = \theta_0 + \epsilon^\alpha \theta_1 + \cdots$, we would find that $\theta_0 = 0$ and $\alpha = 1$. The assumption in (2.76) is made simply to avoid all the work to find that the first term in the expansion is just zero. Before substituting (2.76) into the problem recall $\sin(x) = x - \frac{1}{6}x^3 + \cdots$ when x is close to zero. This means, because the θ in (2.76) is close to zero,

$$\begin{aligned}\sin(\theta) &\sim \sin(\epsilon(\theta_0 + \epsilon^\alpha \theta_1 + \cdots)) \\ &\sim (\epsilon\theta_0 + \epsilon^{\alpha+1}\theta_1 + \cdots) - \frac{1}{6}(\epsilon\theta_0 + \ \cdots)^3 + \cdots \\ &\sim \epsilon\theta_0 + \epsilon^{\alpha+1}\theta_1 - \frac{1}{6}\epsilon^3\theta_0^3 + \cdots .\end{aligned} \tag{2.77}$$

With this the equation of motion (2.73) becomes

$$\epsilon\theta_0'' + \epsilon^{\alpha+1}\theta_1'' + \cdots + \epsilon\theta_0 + \epsilon^{\alpha+1}\theta_1 - \frac{1}{6}\epsilon^3\theta_0^3 + \cdots = 0, \tag{2.78}$$

and the initial conditions are

$$\epsilon\theta_0(0) + \epsilon^{\alpha+1}\theta_1(0) + \cdots = \epsilon, \tag{2.79}$$

and

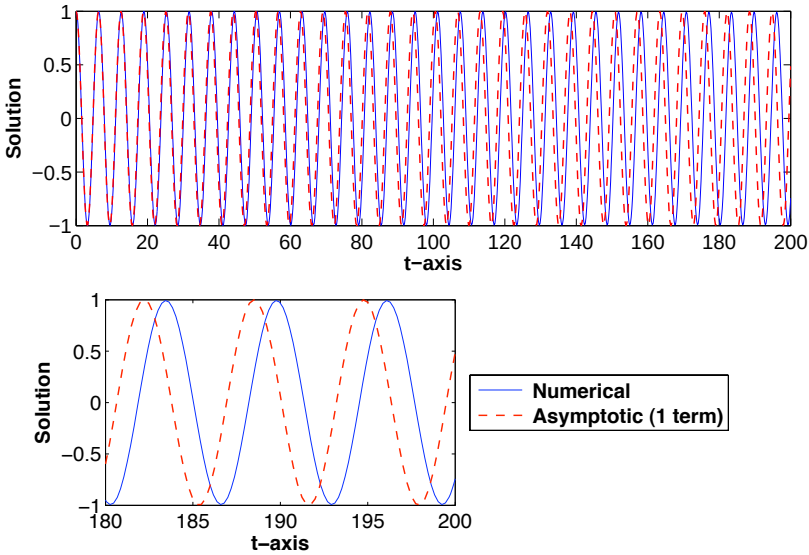


Figure 2.18 Graph of the numerical solution of the pendulum problem (2.60)-(2.62) and the first term in the regular perturbation approximation (2.76). Shown are the solutions over the entire time interval, as well as a close up of the solutions near $t = 200$. In the calculation $\epsilon = \frac{1}{3}$ and both solutions have been divided by $\epsilon = \frac{1}{3}$.

$$\epsilon\theta_0'(0) + \epsilon^{\alpha+1}\theta_1'(0) + \cdots = 0. \tag{2.80}$$

Proceeding in the usual manner yields the following problem.

$O(\epsilon)$ $\theta_0'' + \theta_0 = 0$
$\theta_0(0) = 1,\ \theta_0'(0) = 0$

The general solution of the differential equation is $\theta_0 = a\cos(t) + b\sin(t)$, where a, b are arbitrary constants. It is possible to write this solution in the more compact form $\theta_0 = A\cos(t + B)$, where A, B are arbitrary constants. As will be explained later, there is a reason for why the latter form is preferred in this problem. With this, and the initial conditions, it is found that $\theta_0 = \cos(t)$.

The plot of the one-term approximation, $\theta \sim \epsilon\cos(t)$, and the numerical solution are shown in Figure 2.18. The asymptotic approximation describes the solution accurately at the start, and reproduces the amplitude very well over the entire time interval. What it has trouble with is matching the phase and this is evident in the lower plot in Figure 2.18. One additional comment to make is the value for ϵ used in Figure 2.18 is not particularly small, so getting an approximation that is not very accurate is no surprise. However, if a smaller value is used the same difficulty arises. The difference is that the

first term approximation works over a longer time interval but eventually the phase error seen in Figure 2.18 occurs.

In looking to correct the approximation to reduce the phase error we calculate the second term in the expansion. With the given θ_0 there is an $\epsilon^3\theta_0^3$ term in (2.78). To balance this we use the θ_1 term in the expansion and this requires $\alpha = 2$. With this we have the following problem to solve.

$O(\epsilon^3)$ $\theta_1'' + \theta_1 = \frac{1}{6}\theta_0^3$
$\theta_1(0) = 0,\ \theta_1'(0) = 0$

The method of undetermined coefficients can be used to find a particular solution of this equation. This requires the identity $\cos^3(t) = \frac{1}{4}(3\cos(t) + 3\cos(3t))$, in which case the differential equation becomes

$$\theta_1'' + \theta_1 = \frac{1}{24}(3\cos(t) + 3\cos(3t)). \tag{2.81}$$

With this the general solution is found to be $\theta_1 = a\cos(t) + b\sin(t) - \frac{1}{16}t\sin(t)$, where a, b are arbitrary constants. From the initial conditions this reduces to $\theta_1 = -\frac{1}{16}t\sin(t)$.

The plot of the two term approximation,

$$\theta \sim \epsilon\cos(t) - \frac{1}{16}\epsilon^3 t\sin(t), \tag{2.82}$$

and the numerical solution is shown in Figure 2.19. It is clear from this that we have been less than successful in improving the approximation. The culprit here is the $t\sin(t)$ term. As time increases its contribution grows, and it eventually gets as large as the first term in the expansion. Because of this it is called a secular term, and it causes the expansion not to be uniformly valid for $0 \le t < \infty$. This problem would not occur if time were limited to a finite

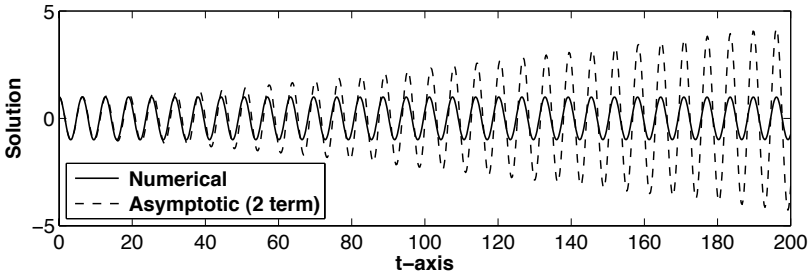


Figure 2.19 Graph of the numerical solution of the pendulum problem (2.60)-(2.62) and the regular perturbation approximation (2.72). In the calculation $\epsilon = \frac{1}{3}$ and the solution has been divided by $\epsilon = \frac{1}{3}$.

interval, as happened in the projectile problem. However, for the pendulum there is no limit on time and this means the expansion is restricted to when it can be used. One last comment to make concerns how this term ended up in the expansion in the first place. In the differential equation for θ_1, given in (2.81), the right hand side contains $\cos(t)$ and this is a solution of the associated homogeneous equation. It is this term that produces the $t\sin(t)$ in the expansion and it is this term we would like to prevent from appearing in the problem.

What is happening is that there is a slow change in the solution that the first term approximation is unable to describe. In effect there are two time scales acting in this problem. One is the basic period of oscillation, as seen in Figure 2.18, and the other is a slow time scale over which the phase changes. Our approximation method will be based on this observation. We will explicitly assume there are two concurrent time scales, given as

$$t_1 = t, \tag{2.83}$$

$$t_2 = \epsilon^\gamma t. \tag{2.84}$$

The value of γ is not known yet, and we will let the problem tell us the value as we construct the approximation. Based on this assumption it is not surprising that the method is called two-timing, or the method of multiple scales.

To illustrate the idea underlying two-timing, consider the function

$$u = e^{-3\epsilon t}\sin(5t).$$

This consists of an oscillatory function, with a slowly decaying amplitude. This can be written using the two-timing variables as

$$u = e^{-3t_2}\sin(5t_1),$$

where $\gamma = 1$.

The change of variables in (2.83), (2.84) is reminiscent of the boundary layer problems in the previous section. The difference here is that we are not separating the time axis into separate regions but, rather, using two time scales together. As we will see, this has a profound effect on how we construct the approximation.

To determine how the change of variables affects the time derivative, we have, using the chain rule,

$$\begin{aligned}\frac{d}{dt} &= \frac{dt_1}{dt}\frac{\partial}{\partial t_1} + \frac{dt_2}{dt}\frac{\partial}{\partial t_2}\\ &= \frac{\partial}{\partial t_1} + \epsilon^\gamma\frac{\partial}{\partial t_2}\,.\end{aligned} \tag{2.85}$$

The second derivative is

$$\begin{aligned}\frac{d^2}{dt^2} &= \left(\frac{\partial}{\partial t_1} + \epsilon^\gamma \frac{\partial}{\partial t_2}\right)\left(\frac{\partial}{\partial t_1} + \epsilon^\gamma \frac{\partial}{\partial t_2}\right) \\ &= \frac{\partial^2}{\partial t_1^2} + 2\epsilon^\gamma \frac{\partial^2}{\partial t_1 \partial t_2} + \epsilon^{2\gamma}\frac{\partial^2}{\partial t_2^2}\,. \end{aligned} \tag{2.86}$$

The steps used to construct an asymptotic approximation of the solution will closely follow what we did earlier. It should be kept in mind during the derivation that the sole reason for introducing t_2 is to prevent a secular term from appearing in the second term.

With the introduction of a second time variable, the expansion is assumed to have the form

$$\theta \sim \epsilon(\theta_0(t_1, t_2) + \epsilon^\alpha \theta_1(t_1, t_2) + \cdots). \tag{2.87}$$

The only difference between this and the regular expansion (2.76) used earlier is that the terms are allowed to depend on both time variables. When this is substituted into the differential equation we obtain an expression similar to (2.78), except the time derivatives are given in (2.85) and (2.86). Specifically, we get

$$\epsilon\frac{\partial^2}{\partial t_1^2}\theta_0 + \epsilon^{\alpha+1}\frac{\partial^2}{\partial t_1^2}\theta_1 + 2\epsilon^{\gamma+1}\frac{\partial^2}{\partial t_1 \partial t_2}\theta_0 + \cdots + \epsilon\theta_0 + \epsilon^{\alpha+1}\theta_1 - \frac{1}{6}\epsilon^3\theta_0^3 + \cdots = 0, \tag{2.88}$$

and the initial conditions are

$$\epsilon\theta_0(0,0) + \epsilon^{\alpha+1}\theta_1(0,0) + \cdots = \epsilon, \tag{2.89}$$

and

$$\epsilon\frac{\partial}{\partial t_1}\theta_0(0,0) + \epsilon^{\alpha+1}\frac{\partial}{\partial t_1}\theta_1(0,0) + \epsilon^{\gamma+1}\frac{\partial}{\partial t_2}\theta_0(0,0) + \cdots = 0. \tag{2.90}$$

Proceeding in the usual manner yields the following problem.

$O(\epsilon)$ $\frac{\partial^2}{\partial t_1^2}\theta_0 + \theta_0 = 0$
$\theta_0(0,0) = 1,\ \frac{\partial}{\partial t_1}\theta_0(0,0) = 0$

The general solution of the differential equation is $\theta_0 = A(t_2)\cos(t_1 + B(t_2))$, where A, B are arbitrary functions of t_2. The effects of the second time variable are seen in this solution, because the coefficients are now functions of the second time variable. To satisfy the initial conditions we need $A(0)\cos(B(0)) = 1$ and $A(0)\sin(B(0)) = 0$. From this we have that $A(0) = 1$ and $B(0) = 0$.

In the differential equation (2.88), with the $O(\epsilon)$ terms out of the way, the next term to consider is $\epsilon^3\theta_0^3$. The only terms we have available to balance

with this have order $\epsilon^{\alpha+1}$ and $\epsilon^{\gamma+1}$. To determine which terms to use we can use a balance argument, similar to what was done for boundary layers. It is found that both terms are needed and this means $\alpha+1 = \gamma+1 = 3$. This is an example of a distinguished balance. A somewhat different way to say this is, the more components of the equation you can include in the approximation the better. In any case, our conclusion is that $\alpha = \gamma = 2$ and this yields the next problem to solve.

$O(\epsilon^3)$ $\frac{\partial^2}{\partial t_1^2}\theta_1 + \theta_1 + 2\frac{\partial^2}{\partial t_1 \partial t_2}\theta_0 = \frac{1}{6}\theta_0^3$
$\theta_1(0,0) = 0,\ \frac{\partial}{\partial t_1}\theta_1(0,0) + \frac{\partial}{\partial t_2}\theta_0(0,0) = 0$

The method of undetermined coefficients can be used to find a particular solution. To be able to do this we first substitute the solution for θ_0 into the differential equation and then use the identity $\cos^3(t) = \frac{1}{4}(3\cos(t) + 3\cos(3t))$. The result is

$$\begin{aligned}\theta_1'' + \theta_1 = \frac{1}{24}&[3\cos(t_1 + B) + 3\cos(3(t_1 + B))] \\ &+2A'\sin(t_1 + B) + 2AB'\cos(t_1 + B).\end{aligned} \tag{2.91}$$

We are at a similar point to what occurred using a regular expansion, as given in (2.81). As before, the right-hand side of the differential equation contains functions that are solutions of the associated homogeneous equation, namely, $\cos(t_1 + B)$ and $\sin(t_1 + B)$. If they are allowed to remain they will produce a solution containing either $t_1\cos(t_1 + B)$ or $t_1\sin(t_1 + B)$. Either one will cause a secular term in the expansion and for this reason we will select A and B to prevent this from happening. To lose $\sin(t_1 + B)$ we take $A' = 0$ and to eliminate $\cos(t_1 + B)$ we take $2AB' = -\frac{1}{8}$. With the earlier determined initial conditions $A(0) = 1$ and $B(0) = 0$, we conclude that $A = 1$ and $B = -t_2/16$.

In the above analysis we never actually determined θ_1. It is enough to know that the problem for θ_1 will not result in a secular term in the expansion. We did find A and B, and with them the expansion is

$$\theta \sim \epsilon\cos(t - \epsilon^2 t/16) + \cdots . \tag{2.92}$$

To investigate the accuracy of this approximation, it is plotted in Figure 2.20 using the same values as for Figure 2.19. Clearly we have improved the first term approximation, and now do well with both amplitude and phase.

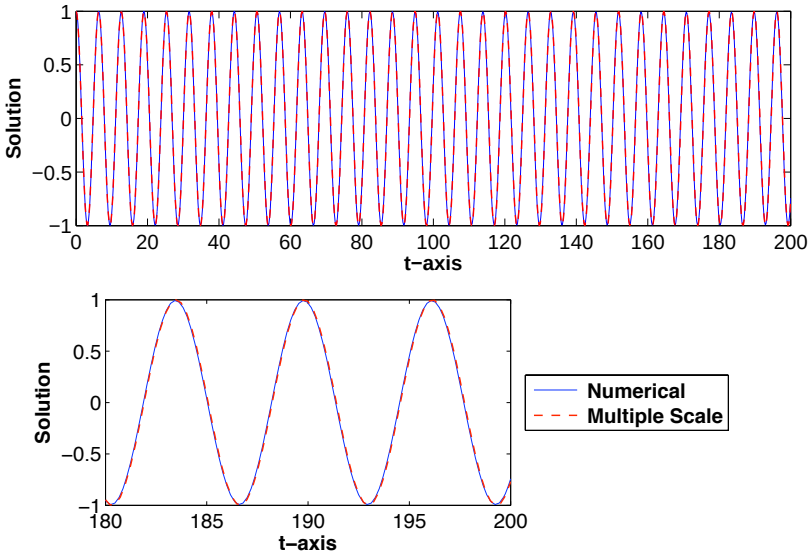


Figure 2.20 Graph of the numerical solution of the pendulum problem (2.60)-(2.62) and the multiple scale approximation (2.92). Shown are the solutions over the entire time interval, as well as a close up of the solutions near $t = 200$. In the calculation $\epsilon = \frac{1}{3}$ and the solution has been divided by $\epsilon = \frac{1}{3}$.

Exercises

2.1. Assuming $f \sim a_1\epsilon^\alpha + a_2\epsilon^\beta + \cdots$ find α, β (with $\alpha < \beta$), and nonzero a_1, a_2, for the following:
(a) $f = e^{\sin(\epsilon)}$.
(b) $f = \sqrt{1 + \cos(\epsilon)}$.
(c) $f = 1/\sqrt{\sin(\epsilon)}$.
(d) $f = 1/(1 - e^\epsilon)$.
(e) $f = \sin(\sqrt{1 + \epsilon x})$, for $0 \le x \le 1$.
(f) $f = \epsilon \exp(\sqrt{\epsilon} + \epsilon x)$, for $0 \le x \le 1$.

2.2. Let $f(\epsilon) = \sin(e^\epsilon)$.
(a) According to Taylor's theorem, $f(\epsilon) = f(0) + \epsilon f'(0) + \frac{1}{2}\epsilon^2 f''(0) + \cdots$. Show that this gives (2.13).
(b) Explain why the formula used in part (a) can not be used to find an expansion of $f(\epsilon) = \sin(e^{\sqrt{\epsilon}})$. Also, show that the method used to derive (2.13) still works, and derive the expansion.

2.3. Consider the equation

$$x^2 + (1 - 4\epsilon)x - \sqrt{1 + 4\epsilon} = 0.$$

(a) Sketch the functions in this equation and then use this to explain why there two real-valued solutions.
(b) Find a two-term asymptotic expansion, for small ϵ, of each solution.

2.4. Consider the equation

$$\ln(x) = \epsilon x.$$

(a) Sketch the functions in this equation and then use this to explain why there are two real-valued solutions.
(b) Find a two-term asymptotic expansion, for small ϵ, of the solution near $x = 1$.

2.5. Consider the equation

$$xe^x = \epsilon.$$

(a) Sketch the functions in this equation and then use this to explain why there is one real-valued solution.
(b) Find a two-term asymptotic expansion, for small ϵ, of the solution.

2.6. Consider the equation

$$x^3 = \epsilon e^{-x}.$$

(a) Sketch the functions in this equation and then use this to explain why there is one real-valued solution.
(b) Find a two-term asymptotic expansion, for small ϵ, of the solution.

2.7. Consider the equation

$$\sin(x + \epsilon) = x.$$

(a) Sketch the functions in this equation and then use this to explain why there is one real-valued solution.
(b) Find a two-term asymptotic expansion, for small ϵ, of the solution.

2.8. Consider the equation

$$\frac{x^3}{1 + x} = \epsilon.$$

(a) Sketch the functions in this equation and then use this to explain why there is only one real-valued solution and describe where it is located for small values of ϵ. Use this to explain why you might want to use an expansion of the form $x \sim \epsilon^\alpha x_0 + \epsilon^\beta x_1 + \cdots$ rather than the one in (2.17).
(b) Find a two-term asymptotic expansion, for small ϵ, of each solution.

2.9. Consider the equation

$$x(x + 2) = \epsilon(x - 1).$$

(a) Sketch the functions in this equation and then use this to explain why there are two real-valued solutions and describe where they are located for small values of ϵ.
(b) Find a two-term asymptotic expansion, for small ϵ, of each solution.

2.10. Consider the equation

$$x(x-1)(x+2) = \epsilon e^x.$$

(a) Sketch the functions in this equation. Use this to explain why there are three real-valued solutions, and describe where they are located for small values of ϵ. Use this to explain why you might want to use an expansion of the form $x \sim \epsilon^\alpha x_0 + \epsilon^\beta x_1 + \cdots$ for one of the solutions, while (2.17) should work for the other two.
(b) Find a two-term asymptotic expansion, for small ϵ, of each solution.

2.11. Consider the equation $\epsilon x^4 - x - 1 = 0$.
(a) Sketch the functions in this equation and then use this to explain why there are only two real-valued solutions to this equation and describe where they are located for small values of ϵ.
(b) Find a two-term asymptotic expansion, for small ϵ, of each solution.

2.12. Consider the equation

$$\frac{1}{1+x^2} = \epsilon x^2.$$

(a) Sketch the functions in this equation and then use this to explain why there are two real-valued solutions and describe where they are located for small values of ϵ. Use this to explain why an expansion of the form given in (2.17) is not a good idea.
(b) Find a two-term asymptotic expansion, for small ϵ, of the solutions.

2.13. Find a two-term expansion of the solution of

$$\frac{dv}{dt} + \epsilon v^2 + v = 0, \quad \text{for } \ 0 < t,$$

where $v(0) = 1$.

2.14. The projectile problem that includes air resistance is

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} = -\frac{1}{(1+\epsilon x)^2},$$

where $x(0) = 0$, and $\frac{dx}{dt}(0) = 1$. For small ϵ, find a two-term expansion of the solution.

2.15. Air resistance is known to depend nonlinearly on velocity, and the dependence is often assumed to be quadratic. Assuming gravity is constant, the equations of motion are

$$\frac{d^2y}{dt^2} = -\epsilon\frac{dy}{dt}\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2},$$
$$\frac{d^2x}{dt^2} = -1 - \epsilon\frac{dx}{dt}\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}.$$

Here x is the vertical elevation of the object, and y is its horizontal location. The initial conditions are $x(0) = y(0) = 0$, and $\frac{dx}{dt}(0) = \frac{dy}{dt}(0) = 1$. The assumption is that air resistance is weak, and so ϵ is small and positive.

(a) For small ϵ, find the first terms in the expansions for x and y.
(b) Find the second terms in the expansions for x and y.

2.16. Consider the nonlinear boundary value problem

$$\frac{d}{dx}\left(\frac{y_x}{1+\epsilon y_x^2}\right) - y = 0, \quad \text{for } \ 0 < x < 1,$$

where $y(0) = 1$ and $y(1) = e^{-1}$. This type of nonlinearity arises in elasticity, a topic taken up in Chapter 6.

(a) Explain why a boundary layer is not expected in the problem, and a regular expansion should work.
(b) For small ϵ, find a two-term expansion of the solution.

2.17. The Friedrichs' (1942) model problem for a boundary layer in a viscous fluid is

$$\epsilon y'' = a - y', \quad \text{for } \ 0 < x < 1,$$

where $y(0) = 0$ and $y(1) = 1$ and a is a given positive constant.

(a) After finding a first term of the inner and outer expansions, derive a composite expansion of the solution.
(b) Taking $a = 1$, plot the exact and composite solutions, on the same axes, for $\epsilon = 10^{-1}$. Do the same thing for $\epsilon = 10^{-2}$ and for $\epsilon = 10^{-3}$. Comment on the effectiveness, or non-effectiveness, of the expansion in approximating the solution.
(c) Suppose you assume the boundary layer is at the other end of the interval. Show that the resulting first term approximations from the inner and outer regions do not match.

2.18. Given that the boundary layer is at $x = 0$, find a composite expansion of

$$\epsilon y'' + 3y' - y^4 = 0, \quad \text{for } \ 0 < x < 1,$$

where $y(0) = 1$ and $y(1) = 1$.

2.19. Given that the boundary layer is at $x = 0$, find a composite expansion of

$$\epsilon^2 y'' + y' + \epsilon y = x, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 1$ and $y(1) = 1$.

2.20. Find a composite expansion of

$$\epsilon y'' + y' - y = 0, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 0$, $y(1) = -1$.

2.21. Find a composite expansion of

$$\epsilon y'' + 2y' - y^3 = 0, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 0$ and $y(1) = 1$.

2.22. Find a composite expansion of

$$\epsilon y'' + y' + \epsilon y = 0, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 1$, $y(1) = 2$.

2.23. Given that the boundary layer is at $x = 1$, find a composite expansion of

$$\epsilon y'' - 3y' - y^4 = 0, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 1$ and $y(1) = 1$.

2.24. Find a composite expansion of

$$\epsilon y'' - \frac{1}{2} y' - xy = 0, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 1$ and $y(1) = 1$.

2.25. Find a composite expansion of

$$\epsilon y'' - (2 - x^2) y = -1, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 0$, $y(1) = 2$.

2.26. Find a composite expansion of

$$\epsilon y'' - (1 + x) y = 2, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = 0$, $y(1) = 0$.

2.27. As found in Exercise 1.13, the equation for a weakly damped oscillator is

$$y'' + \epsilon y' + y = 0, \quad \text{for} \;\; 0 < t,$$

where $y(0) = 1$ and $y(0) = 0$.

(a) For small ϵ, find a two-term regular expansion of the solution.
(b) Explain why the expansion in (a) is not well-ordered for $0 \le t < \infty$. What requirement is needed on t so it is well-ordered?
(c) Use two-timing to construct a better approximation to the solution.

2.28. The weakly nonlinear Duffing equation is

$$y'' + y' + \epsilon y^3 = 0, \quad \text{for} \;\; 0 < t,$$

where $y(0) = 0$ and $y'(0) = 1$.
(a) For small ϵ, find a two-term regular expansion of the solution.
(b) Explain why the expansion in (a) is not well-ordered for $0 \le t < \infty$. What requirement is needed on t so it is well-ordered?
(c) Use two-timing to construct a better approximation to the solution.

2.29. This problem derives additional information from the projectile problem.
(a) Let t_M be the time at which the projectile reaches its maximum height. Given that the solution depends on ϵ, it follows that t_M depends on ϵ. Use (2.27) to find a two-term expansion of t_M for small ϵ. What is the resulting two-term expansion for the maximum height x_M?
(b) Let t_E be the time at which the projectile hits the ground. Given that the solution depends on ϵ, it follows that t_E depends on ϵ. Use (2.27) to find a two-term expansion of t_E for small ϵ.
(c) Based on your results from parts (a) and (b), describe the effects of the nonlinear gravitational field on the motion of the projectile.

2.30. In the study of reactions of chemical mixtures one comes across the following problem

$$\frac{d^2y}{dx^2} = -\epsilon e^y, \quad \text{for} \;\; 0 < x < 1,$$

where $y(0) = y(1) = 0$. This is known as Bratu's equation, and it illustrates some of the difficulties one faces when solving nonlinear equations.
(a) Explain why a boundary layer is not expected in the problem and find the first two terms in a regular expansion of the solution.
(b) The function

$$y = -2 \ln\left[\frac{\cosh(\beta(1-2x))}{\cosh(\beta)}\right],$$

where β satisfies

$$\cosh(\beta) = 2\beta\sqrt{\frac{2}{\epsilon}}, \tag{2.93}$$

is a solution of the Bratu problem. By sketching the functions in (2.93), as functions of β, explain why there is an ϵ_0 where if $0 < \epsilon < \epsilon_0$ then there are exactly two solutions, while if $\epsilon_0 < \epsilon$ then there are no solutions.

(c) Comment on the conclusion drawn from part (b) and your result in part (a). Explain why the regular expansion does not fail in a manner found in a boundary layer problem but that it is still not adequate for this problem.

Chapter 3
Kinetics

3.1 Introduction

We now investigate how to model, and analyze, the interactions of multiple species and how these interactions produce changes in their populations. Examples of such problems are below.

3.1.1 Radioactive Decay

A radioactive isotope is unstable, and will decay by emitting a particle, transforming into another isotope. As an example, tritium ${}^3H_1$ is a radioactive form of hydrogen that occurs when cosmic rays interact with the atmosphere. It decays by emitting an electron e and antineutrino ν to produce a stable helium isotope ${}^3He_2$. The conventional way to express this conversion is

$$ {}^3H_1 \rightarrow {}^3He_2 + e + \nu. \tag{3.1} $$

The assumption used to model such situations is that the rate of decrease in the amount of radioactive isotope is proportional to the amount currently present. To translate this into mathematical terms let $N(t)$ designate the amount of the radioactive material present at time t. In this case we obtain the rate equation

$$ \frac{dN}{dt} = -kN, \quad \text{for } \ 0 < t, \tag{3.2} $$

where

$$ N(0) = N_0. \tag{3.3} $$

In the above equation k is the proportionality constant and is assumed to be positive.

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_3,
© Springer Science+Business Media, LLC 2009

3.1.2 Predator-Prey

This involves two species and a typical situation is a population of predators, wolves, which survives by eating another species, rabbits. To write down a model for their interaction, let $R(t)$ and $W(t)$ denote the number of rabbits and wolves, respectively. In this case, we have

$$\frac{dR}{dt} = aR - bRW, \tag{3.4}$$

$$\frac{dW}{dt} = -cW + dRW. \tag{3.5}$$

In the above equations a, b, c, d are proportionality constants. To obtain the first equation, it has been assumed that the population of rabbits, with wolves absent, increases at a rate proportional to their current population (aR). When the wolves are present it is assumed the rabbit population decreases at a rate proportional to both populations ($-bRW$). Similarly, for the second equation, the number of wolves, with rabbits absent, decreases at a rate proportional to their current population ($-cW$), but increases at a rate proportional to both the rabbit and wolf populations when rabbits are available (dRW). To complete the formulation we need the initial concentrations, given as $R(0) = R_0, W(0) = W_0$.

3.1.3 Epidemic Model

Epidemics, such as the black death and cholera, have come and gone throughout human history. Given the catastrophic nature of these events there is a long history of scientific study trying to predict how and why they occur. One of particular prominence is the Kermack-McKendrick model for epidemics. This assumes the population can be separated into three groups. One is the population $S(t)$ of those susceptible to the disease, another is the population $I(t)$ that is ill, and the third is the population $R(t)$ of individuals that have recovered. A model that accounts for the susceptible group getting sick, the subsequent increase in the ill population, and the eventual increase in the recovered population is the following set of equations

$$\frac{dS}{dt} = -k_1 SI, \tag{3.6}$$

$$\frac{dI}{dt} = -k_2 I + k_1 SI, \tag{3.7}$$

$$\frac{dR}{dt} = k_2 I, \tag{3.8}$$

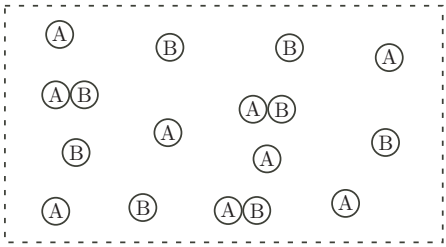


Figure 3.1 Sample domain illustrating assumptions underlying the Law of Mass Action, where two species combine to form a third.

where $S(0) = S_0, I(0) = I_0, R(0) = R_0$. In the above equations k_1, k_2 are proportionality constants. Given the three groups, and the letters used to designate them, this is an example of what is known as a SIR model in mathematical epidemiology. This model does not account for births or deaths, and for this reason the total population stays constant. This can be seen in the above equations because

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0,$$

or in other words $\frac{d}{dt}(S + I + R) = 0$. The fact that $S + I + R$ is constant is an example of a conservation law, and these will play a prominent role in this chapter.

3.2 Kinetic Equations

The common thread in the above examples is that one or more species combine, or transform, to form new or additional species. This is a situation common in chemistry and we will extend the theory developed in chemical kinetics to describe interacting populations or species. The main result is the Law of Mass Action and to motivate how it is derived consider a region containing a large number of two species, labeled as A and B. A small portion of this region is shown in Figure 3.1. As indicated in the figure, both species are assumed to be distributed throughout the region. It is also assumed that they are in motion, and when an A and B come into contact they combine to form a new species C. The C's are shown in the figure with an A and B stuck together. The symbolism for this is

$$A + B \to C. \tag{3.9}$$

The question is, can we use this information to determine the concentrations of the three species? The reaction in (3.9) states that one A and one B

are used to construct one C. This means that the rate of change of the concentrations of A and B are the same, and they are the negative of the change in C. In other words,

$$\frac{dA}{dt} = \frac{dB}{dt} = -\frac{dC}{dt}\,. \tag{3.10}$$

In the above expressions there is a mild case of notation abuse in the sense that we are letting A, B, and C also designate the concentrations of the respective species. This dual usage of having the letters designate individual molecules as in (3.9) and concentrations as in (3.10) is common in kinetics and should not cause problems in the development.

The equalities in (3.10) can be rewritten as

$$\frac{dA}{dt} = -r, \tag{3.11}$$

$$\frac{dB}{dt} = -r, \tag{3.12}$$

$$\frac{dC}{dt} = r, \tag{3.13}$$

where r is known as the rate of the reaction. Now, r depends on the collision frequency of A and B, and this means it depends on the concentrations of A and B. Also, if there are no A's, or if there are no B's, then the rate is zero. We are therefore assuming that $r = r(A, B)$, where $r(A, 0) = r(0, B) = 0$. To obtain a first term approximation of this function we use Taylor's theorem to obtain

$$r = r_{00} + r_{10}A + r_{01}B + r_{20}A^2 + r_{11}AB + r_{02}B^2 + \cdots\,.$$

In this expression

$$\begin{aligned} r_{00} &= r(0,0), \\ r_{10} &= \frac{\partial r}{\partial A}(0,0), \qquad & r_{01} &= \frac{\partial r}{\partial B}(0,0), \\ r_{20} &= \frac{1}{2}\frac{\partial^2 r}{\partial A^2}(0,0), \quad & r_{02} &= \frac{1}{2}\frac{\partial^2 r}{\partial B^2}(0,0). \end{aligned}$$

All of these terms are zero. For example, because $r(A, 0) = 0$ it follows that

$$\frac{\partial r}{\partial A}(A,0) = 0 \quad \text{and} \quad \frac{\partial^2 r}{\partial A^2}(A,0) = 0.$$

Similarly, because $r(0, B) = 0$, it follows that $r_{01} = r_{02} = 0$. What is not necessarily zero is the mixed derivative term

$$r_{11} = \frac{\partial^2 r}{\partial A \partial B}(0,0).$$

Therefore, the first nonzero term in the Taylor series is $r_{11}AB$, and from this we have

$$r = kAB, \tag{3.14}$$

where k is known as the rate constant. This expression, along with the rate equations in (3.11)-(3.13), is the Law of Mass Action as applied to the reaction in (3.9). This will be generalized to more complicated reactions in the next section. Before doing so it is of interest to note what happens if A and B are the same species. In this case (3.9) becomes $2A \to C$ and (3.14) takes the form $r = kA^2$. Also, we no longer have an equation for B, but because two A's are now lost every time a C is produced then (3.11) becomes $A' = -2r$. The equation for C stays the same. This shows that the coefficients in the reaction play a role in both the formula for r as well as in the rates for the respective species in the reaction.

3.2.1 The Law of Mass Action

To state the general form of the Law of Mass Action certain terms need to be defined. For this we generalize the above example and consider the reaction

$$\alpha A + \beta B \to \gamma C + \delta D. \tag{3.15}$$

The coefficients α, β, γ, δ are nonnegative constants known as the *stoichiometric coefficients* for the reaction. In effect, this reaction states that α of the A's combine with β of the B's to form γ of the C's and δ of the D's. Said this way, the implication is that the stoichiometric coefficients are integers. The fact is that they generally are, although we will not make this assumption explicitly. The species on the left, A and B, are the *reactants* and those on the right, C and D, are the *products* for this particular reaction. The order of the reaction is the total number of reactants, which in this case is $\alpha + \beta$.

The Law of Mass Action, which will be given shortly, states that the rate r of the reaction in (3.15) is

$$r = kA^{\alpha}B^{\beta}, \tag{3.16}$$

where k is the rate constant or the reaction rate coefficient. In writing down this formula the notation has been corrupted a bit. As happened in the earlier example, we started off letting A, B designate the reactants but in the rate formula (3.16) these same letters have been used to designate their concentrations.

We are now is position to state the assumptions underlying the Law of Mass Action.

Definition 3.1. The Law of Mass Action consists of the following three assumptions:

1. The rate, r, of the reaction is proportional to the product of the reactant concentrations, with each concentration raised to the power equal to its respective stoichiometric coefficient.
2. The rate of change of the concentration of each species in the reaction is the product of its stoichiometric coefficient with the rate of the reaction, adjusted for sign (+ if product and − if reactant).
3. For a system of reactions, the rates add.

To illustrate, consider the reaction in (3.15). Part 1 of the definition is simply the formula (3.16) put into words. As for Part 2, the rate of change $\frac{dA}{dt}$ is equal to $-\alpha r$, while $\frac{dC}{dt}$ is equal to γr. Combining this information, from the Law of Mass Action the kinetic equations for the concentrations are

$$\begin{aligned}
\frac{dA}{dt} &= -\alpha r \\
&= -\alpha k A^{\alpha} B^{\beta}, \\
\frac{dB}{dt} &= -\beta k A^{\alpha} B^{\beta}, \\
\frac{dC}{dt} &= \gamma k A^{\alpha} B^{\beta}, \\
\frac{dD}{dt} &= \delta k A^{\alpha} B^{\beta}.
\end{aligned} \tag{3.17}$$

To complete the formulation, it is assumed that the initial concentrations are known, and so, $A(0) = A_0, B(0) = B_0, C(0) = C_0, D(0) = D_0$ are given.

The specific units of the terms in the above equations depend on the application. For example, if the species are chemicals then concentration, using SI units, is measured in moles per decimeter (mol/dm^3). It is not unusual, however, to find that when using liquids that concentrations are measured using molarity (M) where $1M = 6.02 \times 10^{23}$ molecules per liter. In applications involving gases the units that are often used are moles per cubic centimeter. If the application involves populations then population density (e.g., number per area) is used. Whatever the application, the units for the rate constant depend on the specific reaction. This can be seen from (3.17) because $[A'] = [k][A^{\alpha}B^{\beta}]$. If A and B are concentrations then $[k] = T^{-1}L^{3(\alpha+\beta-1)}$. Consequently, the units for the rate coefficient for $A + B \to C$ are different than they are for the reaction $A + 2B \to C$.

3.2.2 Conservation Laws

We have produced four equations for the four species involved in the example reaction in (3.15). Although they are not particularly easy to solve there is one significant simplification we are able to make. To explain what this is, note that it is possible to combine the first two equations to produce

zero on the right-hand side. Specifically, $\frac{d}{dt}(\beta A - \alpha B) = 0$ and this means $\beta A - \alpha B = constant$. Using the stated initial conditions it follows that

$$\beta A - \alpha B = \beta A_0 - \alpha B_0. \tag{3.18}$$

In a similar manner, by combining the C and A equations we obtain

$$\gamma A - \alpha C = \gamma A_0 - \alpha C_0, \tag{3.19}$$

and from the D and A equations

$$\delta A - \alpha D = \delta A_0 - \alpha D_0. \tag{3.20}$$

The equations (3.18)-(3.20) are conservation laws. These will play an essential role in our study of kinetic equations, so it is important to define exactly what this means.

Definition 3.2. Given species $A, B, C, \ldots, Z$ and numbers $a, b, c, \ldots, z$ then $aA + bB + cC + \cdots + zZ$ is said to be conserved if

$$\frac{d}{dt}(aA + bB + cC + \cdots + zZ) = 0. \tag{3.21}$$

It is required that at least one of the numbers $a, b, c, \ldots, z$ is nonzero, and (3.21) holds irrespective of the values for the initial conditions and rate constants. The corresponding conservation law is $aA + bB + cC + \cdots + zZ = constant$.

One particularly useful application of conservation laws is to reduce the number of equations that need to be solved. For example, we have

$$B = B_0 + \beta(A - A_0)/\alpha, \tag{3.22}$$
$$C = C_0 + \gamma(A_0 - A)/\alpha, \tag{3.23}$$
$$D = D_0 + \delta(A_0 - A)/\alpha. \tag{3.24}$$

Therefore, once we know A we will then be able to determine the other three concentrations. The equation for A now takes the form

$$\frac{dA}{dt} = -\alpha k A^{\alpha}(a + bA)^{\beta}, \tag{3.25}$$

where $a = B_0 - bA_0$ and $b = \beta/\alpha$. This is still a formidable equation but we only have to deal with one rather than four as was originally stated.

One thing to keep in mind when looking for conservation laws is that they are not unique. For example, because $\frac{d}{dt}(\beta A - \alpha B) = 0$ and $\frac{d}{dt}(\gamma A - \alpha C) = 0$ then given any two numbers x, y we have that $\frac{d}{dt}[x(\beta A - \alpha B) + y(\gamma A - \alpha C)] = 0$. Therefore, $x(\beta A - \alpha B) + y(\gamma A - \alpha C) = constant$ is a conservation law. The objective is not to find all possible combinations but, rather, the minimum

number from which all others can be found. Most of the reactions considered in this book are simple enough that it will be evident what the minimum number is. It is possible, however, to develop a theory for determining this number and this will be discussed in Section 3.3.

3.2.3 Steady-States

In addition to the conservation laws we will also be interested in the steady-state solutions. To be a steady-state the concentration must be constant and it must satisfy the kinetic equations. From (3.25) there are two steady-states, one is $A = 0$ and the second is $A = -a/b$. The corresponding steady-state values for the other species in the reaction are determined from (3.22)-(3.24). The one restriction we impose is that the concentrations are non-negative. Because of this, if $a \geq 0$ then the only one physically relevant steady-state solution of (3.25) is $A = 0$.

3.2.4 Examples

1. Consider the reaction $A \to 2C$. The rate of the reaction is $r = kA$, and so, the kinetic equations are

$$\frac{dA}{dt} = -kA,$$
$$\frac{dC}{dt} = 2kA.$$

The reaction is first order, and the conservation law is obtained by noting $\frac{d}{dt}(2A + C) = 0$. From this it follows that $C = C_0 + 2A_0 - 2A$. Also, there is only one possible steady-state for this system, which is $A = 0, C = C_0 + 2A_0$. It is worth noting that this steady-state can be obtained directly from the reaction. Namely, if one starts out with A_0 molecules of A and each of these is transformed into two molecules of C then the reaction will continue until A is exhausted, so $A = 0$, and the amount of C has increased by $2A_0$. ■

2. If the reaction is $A + B \to 3C$ then the rate is $r = kAB$ and the kinetic equations are

$$\frac{dA}{dt} = -kAB, \tag{3.26}$$
$$\frac{dB}{dt} = -kAB, \tag{3.27}$$
$$\frac{dC}{dt} = 3kAB. \tag{3.28}$$

The reaction is second order. To find the conservation laws, note that (3.26) and (3.27) can be subtracted to yield $\frac{dA}{dt} - \frac{dB}{dt} = 0$. Writing this as $\frac{d}{dt}(A-B) = 0$, it follows that $A - B = A_0 - B_0$. Similarly, (3.26) and (3.28) can be combined to yield $\frac{d}{dt}(3A + C) = 0$, from which it follows that $3A + C = 3A_0 + C_0$. The conclusion is that $B = B_0 + A - A_0$ and $C = C_0 + 3(A_0 - A)$. The resulting reduced equation is therefore

$$\frac{dA}{dt} = -kA(B_0 - A_0 + A).$$

This is known as the logistic equation and it can be solved using separation of variables. The steady-states are $A = 0$ and $A = B_0 - A_0$. The latter is physically relevant only if $B_0 - A_0 \geq 0$. As before the steady-states are evident directly from the reaction. Because A and B combine to form three molecules of C then the reaction will continue until you run out of either A or B. If it is A then $A = 0$, $B = B_0 - A_0$, $C = C_0 + 3A_0$ while if it is B then $B = 0$, $A = A_0 - B_0$, $C = C_0 + 3B_0$. ■

3. As a third example we consider a system of three reactions given as

$$A \rightleftharpoons C + D, \tag{3.29}$$
$$A + B \rightarrow 2A + C. \tag{3.30}$$

We need to explain exactly what is written here. First, (3.29) is a compact way to write $A \rightarrow C+D$ and $C+D \rightarrow A$. In this case the reaction is said to be reversible. Each gets its own rate constant and we will use k_1 for the first and k_{-1} for the second. Secondly, (3.30) is an example of an autocatalytic reaction because A is being used to produce more of itself (i.e., there is more A at the end of the reaction even though it is one of the reactants). We will use k_2 for its rate constant. The corresponding rates are $r_1 = k_1A$, $r_{-1} = k_{-1}CD$, and $r_2 = k_2AB$. Now, the Law of Mass Action applies to each reaction and the rates are added to construct the kinetic equation for each species (Part 3 of Definition 3.1). For example, the kinetic equation for A is

$$\begin{aligned}\frac{dA}{dt} &= -r_1 + r_{-1} - r_2 + 2r_2 \\ &= -k_1A + k_{-1}CD - k_2AB + 2k_2AB \\ &= -k_1A + k_{-1}CD + k_2AB.\end{aligned}$$

Note that for the reaction in (3.30), A is treated as both a reactant ($-r_2$) and a product ($+2r_2$) as specified by the reaction. In a similar manner the kinetic equations for the other species are

$$\frac{dB}{dt} = -k_2 AB, \tag{3.31}$$

$$\frac{dC}{dt} = k_1 A - k_{-1} CD, \tag{3.32}$$

$$\frac{dD}{dt} = k_1 A - k_{-1} CD. \tag{3.33}$$

The useful conservation laws in this case are $\frac{d}{dt}(A + B + C) = 0$ and $\frac{d}{dt}(C - D) = 0$. From this we get $B = B_0 + A_0 + C_0 - A - C$ and $D = D_0 + C - C_0$. This enables us to reduce the system to the two equations

$$\frac{dA}{dt} = k_{-1} C(\beta + C) + k_2 A(\alpha - A - C), \tag{3.34}$$

$$\frac{dC}{dt} = k_1 A - k_{-1} C(\beta + C), \tag{3.35}$$

where $\alpha = -k_1/k_2 + B_0 + A_0 + C_0$ and $\beta = D_0 - C_0$. Finding the possible steady-states is an interesting exercise for these reactions. From (3.31) we have that for a steady-state either $A = 0$ or $B = 0$. For example, if $A = 0$ then from (3.32) we have either $C = 0$ or $D = 0$. For the case of $C = 0$, from the conservation laws, we get $B = B_0 + A_0 + C_0$ and $D = D_0 - C_0$. One can calculate the other solutions in the same way. What is interesting is that the reactions paint a slightly different picture. The two reactions in (3.29) are no help in finding the steady-states as A simply converts back and forth with C and D. The reaction in (3.30), however, will stop when B is exhausted. In fact, $B = 0$ is the only apparent species with a steady-state as there is no reason for the two reactions in (3.29) to stop. The fact that the reactions give us a different conclusion from what we derived from the differential equations has to do with stability. The differential equations give all mathematically possible steady-states irrespective of whether they can be achieved physically. The reactions contain this information by the way they are stated, although they are limited in what they can tell us (e.g., what happens to the other concentrations). What is needed is to introduce the mathematical tools to study stability and this will be done later in the chapter. ■

3.2.5 End Notes

It was stated that the coefficient k in the rate of a reaction is constant. In reality, k can depend on the conditions under which the reaction takes place. For example, the rate of a chemical reaction depends strongly on the

temperature. The most widely used assumption concerning this dependence is the Arrhenius equation, which states

$$k = k_0 e^{-E/RT},$$

where k_0, E, R are parameters and T is temperature measured in Kelvin units. The complication here is that chemical reactions can release, or absorb, heat, and for this reason the temperature can change as the reaction proceeds. It is assumed in this text that the reactions take place in a medium that allows for maintaining a constant temperature.

As a second comment, one might conclude from the physical interpretation of (3.15) that a reaction involving three reactants is rare as it requires three molecules to collide simultaneously. However, it is quite common to find models that contain reactions involving three or more reactants. It is also not unusual to find models with fractional coefficients. This is one of the reasons for introducing the idea of an elementary reaction. These are reactions in which the molecular steps are exactly as stated in the reaction statement. In this case the stoichiometric coefficients equal the number of molecules involved in the reaction. In chemical applications all elementary reactions are either first- or second-order. The fact is, however, that for most reactions the elementary steps are not known. There are multiple reasons for this, but typically it is due to the small concentrations and short life times of the intermediate species, which makes measuring them difficult. Consequently, non-elementary reactions are used and they should be thought of as an approximation of the actual reaction mechanism.

Finally, even though the Law of Mass Action is based on a collection of physically motivated assumptions, the formulation is heuristic. For example, in explaining the dependence of the reaction rate in (3.14) on the species concentrations, we introduced the idea of collision frequency. The fact is that two molecules do not necessarily combine when colliding, and the actual event depends on the collision energy, collision angle, etc. This has led to research into using molecular dynamics to derive the Law of Mass Action from more fundamental principles. This is outside the scope of this textbook and those interested should consult Houston [2006] or Henriksen and Hansen [2008].

3.3 General Mathematical Formulation

For the general form of the schemes considered here we assume there are n reactions involving m distinct species $X_1, X_2, \ldots, X_m$. The scheme is composed of a set of reactions of the form

$$\sum_{i=1}^{m} \alpha_{ij} X_i \to \sum_{i=1}^{m} \beta_{ij} X_i \quad \text{for } j = 1, 2, \ldots, n. \tag{3.36}$$

In this setting a species can appear as just a reactant, so $\beta_{ij} = 0$, or as just a product, so $\alpha_{ij} = 0$, or both. Also, the stoichiometric coefficients α_{ij}, β_{ij} are assumed to be non-negative, with at least one of the α's and one of the β's nonzero in each reaction. The reaction rate r_j for the jth reaction is

$$r_j = k_j \prod_{i=1}^{m} X_i^{\alpha_{ij}}, \tag{3.37}$$

where k_j is the rate constant for the jth reaction. With this, the kinetic equation for the time evolution of the concentration of X_i is

$$\frac{d}{dt} X_i = \sum_{j=1}^{n} (\beta_{ij} - \alpha_{ij})\, r_j. \tag{3.38}$$

This can be written in matrix form as follows,

$$\frac{d}{dt} \mathbf{X} = \mathbf{S}\mathbf{r}, \tag{3.39}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_m)^T$ is the vector of concentrations and $\mathbf{r} = (r_1, r_2, \ldots, r_n)^T$ is the rate vector. The $m \times n$ matrix $\mathbf{S}$ is called the stoichiometric matrix and $\mathbf{S}_{ij} = \beta_{ij} - \alpha_{ij}$.

Example

For the reactions

$$\begin{aligned} X_1 &\to X_2, \\ X_2 &\to X_1 + X_3, \end{aligned}$$

the stoichiometric matrix is

$$\mathbf{S} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix},$$

and the rate vector is

$$\mathbf{r} = \begin{pmatrix} k_1 X_1 \\ k_2 X_2 \end{pmatrix}.$$

The resulting matrix problem is

$$\frac{d}{dt} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} k_1 X_1 \\ k_2 X_2 \end{pmatrix}.$$

This is the matrix equation given in (3.39) for this particular example. ■

A conservation law for the general reaction scheme in (3.36) satisfies

$$\frac{d}{dt}(a_1X_1 + a_2X_2 + \ldots + a_mX_m) = 0,$$

where the a_j's are constants that will be determined later. Integrating this equation we obtain

$$a_1X_1 + a_2X_2 + \ldots + a_mX_m = a_1X_{10} + a_2X_{20} + \ldots + a_mX_{m0},$$

where X_{j0} is the initial concentration of X_j. It is convenient to express this in vector form, which is

$$\mathbf{a} \cdot \mathbf{X} = \mathbf{a} \cdot \mathbf{X}_0, \tag{3.40}$$

where $\mathbf{a} = (a_1, \ldots, a_m)^T$ and $\mathbf{X}_0 = (X_{10}, X_{20}, \ldots, X_{m0})^T$. To determine the vector $\mathbf{a}$, multiply (3.39) by $\mathbf{a}$ to obtain $\mathbf{a} \cdot \mathbf{X}' = \mathbf{a} \cdot \mathbf{S}\mathbf{r}$. Given that $\mathbf{a} \cdot \mathbf{X}' = (\mathbf{a} \cdot \mathbf{X})'$ and $\mathbf{a} \cdot \mathbf{S}\mathbf{r} = \mathbf{r} \cdot \mathbf{S}^T\mathbf{a}$, then a conservation law corresponds to $\mathbf{r} \cdot \mathbf{S}^T\mathbf{a} = 0$. As stated earlier, a conservation law is independent of the values of the rate constants. Therefore, the vector $\mathbf{a}$ must satisfy

$$\mathbf{S}^T\mathbf{a} = \mathbf{0}. \tag{3.41}$$

Written this way, finding the conservation laws has been reduced to a linear algebra problem.

Recall the set of all solutions of (3.41) from a subspace known as the kernel, or null space, of $\mathbf{S}^T$. Let $K(\mathbf{S}^T)$ designate this subspace. If $K(\mathbf{S}^T)$ contains only the zero vector, so $\mathbf{a} = \mathbf{0}$ is the only solution of (3.41), then there are no conservation laws. Assuming there are nonzero solutions then let $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$ be a basis for $K(\mathbf{S}^T)$. Each basis vector produces a conservation law of the form in (3.40), and it is independent of the laws obtained from the other basis vectors. What this means is that the conservation law obtained using, say, $\mathbf{a}_1$ cannot be obtained by combining the conservation laws obtained using $\mathbf{a}_2, \mathbf{a}_3, \ldots, \mathbf{a}_k$. Moreover, because these vectors form a basis, given any conservation law we are able to write it in terms of the laws obtained from basis vectors. These observations are summarized in the following result.

Theorem 3.1. *The number of independent conservation laws of the system (3.36) is equal to the nullity of* $\mathbf{S}^T$*, and the basis vectors of the kernel of* $\mathbf{S}^T$ *correspond to the independent conservation laws of the system.*

To be specific, if (3.41) has only the zero solution then there are no conservation laws. Otherwise, a complete set of conservation laws is

$$\mathbf{a}_i \cdot \mathbf{X} = \mathbf{a}_i \cdot \mathbf{X}_0, \quad \text{for } i = 1, 2, \ldots, k, \tag{3.42}$$

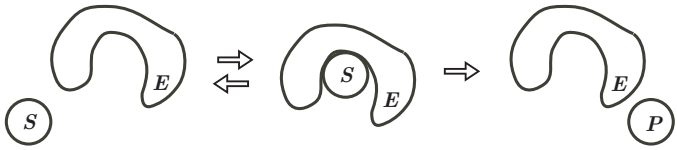


Figure 3.2 The steps in the Michaelis-Menten mechanism, where an enzyme, E, assists S in transforming into P.

where $\mathbf{X}_0 = (X_{10}, X_{20}, \ldots, X_{m0})^T$ are the initial concentrations and the vectors $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$ form a basis of the kernel of $\mathbf{S}^T$. The benefit of this is that each independent conservation law can be used to eliminate one of the differential equations in (3.39). In schemes only involving a few reactions it is not necessary to use this result as one can usually just look at the equations and determine the independent conservation laws. However, for systems containing many equations the above result is very useful as it provides a systematic method for reducing the problem.

Example: cont'd

In the previous example, (3.41) takes the form

$$\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Forming the augmented matrix and row reducing yields the following

$$\left(\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{array}\right) \to \left(\begin{array}{ccc|c} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right). \tag{3.43}$$

The solution is $a_3 = 0$ and $a_2 = a_1$. Consequently, the kernel has dimension one and a basis is $\mathbf{a}_1 = (1, 1, 0)^T$. The corresponding conservation law is $X_1 + X_2 = X_{10} + X_{20}$. ■

3.4 Michaelis-Menten Kinetics

Many chemical and biological systems depend on enzymes to catalyze one or more of their component reactions. Often the exact mechanisms are not well understood and can involve very complicated pathways with multiple enzymes and other catalysts. A relatively simple description of the mechanism is provided by the Michaelis-Menten model, in which the reaction involves an enzyme binding to a substrate that subsequently reacts to form a new product molecule. A schematic of the steps involved is shown in Figure 3.2. This is

now considered the prototype example for an enzyme-catalyzed reaction, and the specific steps in the reaction are as follows

$$S + E \rightleftharpoons C, \tag{3.44}$$

$$C \to P + E. \tag{3.45}$$

In this reaction, S is the substance that is transformed by the reaction, E is an enzyme that facilitates the conversion, C is an intermediate complex, and P is the final product produced by the reaction. Using the Law of Mass Action, the resulting kinetic equations are

$$\frac{dS}{dt} = -k_1 SE + k_{-1}C, \tag{3.46}$$

$$\frac{dE}{dt} = -k_1 SE + k_{-1}C + k_2 C, \tag{3.47}$$

$$\frac{dC}{dt} = k_1 SE - k_{-1}C - k_2 C, \tag{3.48}$$

$$\frac{dP}{dt} = k_2 C. \tag{3.49}$$

For initial conditions, it is assumed that we start with S and E and no complex or product. In other words,

$$S(0) = S_0, E(0) = E_0, C(0) = 0, P(0) = 0. \tag{3.50}$$

Two useful conservation laws for this reaction are $\frac{d}{dt}(E + C) = 0$ and $\frac{d}{dt}(S + C + P) = 0$. Using the stated initial conditions, the conservation laws give us that $E = E_0 - C$ and $P = S_0 - S - C$. Therefore, we can reduce the reactions to the two equations

$$\frac{dS}{dt} = -k_1 E_0 S + (k_{-1} + k_1 S)C, \tag{3.51}$$

$$\frac{dC}{dt} = k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C. \tag{3.52}$$

Even though we have reduced the original system down to two equations, it is still not clear how to solve the problem or what properties the solution has. There are different ways to proceed and, as in many problems, the choice depends on one's interests and background. We will consider three, one using numerical methods, one based on rates of the reactions, and then one using perturbation expansions.

As a historical note, the steps involved in the reaction were used by Brown [1902] to describe the hydrolysis of sucrose. The equations were later analyzed by Michaelis and Menten [1913]. As it turns out, the hydrolysis of sucrose produces two simpler sugars, glucose and fructose. The enzyme in this case is invertase, also known as β-fructofuranosidase, and the structure of this molecule is shown in Figure 3.3. It is clear from this figure that the represen-

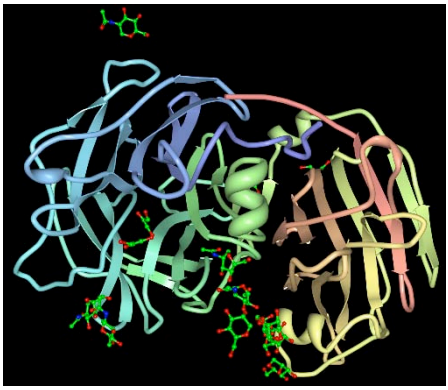


Figure 3.3 The three-dimensional structure of the enzyme invertase (Verhaest et al. [2006]).

tation in Figure 3.2 is simplistic, but it still provides an effective description of the overall reaction. What happens in these reactions is that invertase attaches itself to the sucrose and splits, or cleaves, the sugar into two smaller molecules. There are two product molecules, not one as indicated in Figure 3.2, but this has minimal effect on the resulting reaction scheme. It is also worth noting that the discovery of this particular reaction represents the beginning of enzyme kinetics as a scientific discipline, and for this reason it has become one of the standard examples in biochemistry courses. It also has commercial applications. The splitting of sucrose into simpler sugars is called inversion, and the mixture produced is called inverted sugar. Apparently, according to The Sugar Association, inverted sugar is sweeter than regular sugar and this has useful applications in the food business.

3.4.1 Numerical Solution

Solving the problem numerically is straightforward, and one only has to decide on what parameter values to use. We will use values that come from a model of the transport of P-glycoprotein (Tran et al. [2005]). They found that $k_1 = 10^9$ $\mathrm{M}^{-1}\mathrm{s}^{-1}$, $k_{-1} = 7.5 \times 10^5$ s^{-1}, $k_2 = 10^3$ s^{-1}, $E_0 = 10^{-5}$ M, and $S_0 = 100E_0$. The resulting solution curves are shown in Figure 3.4. One of the interesting aspects of this experiment is that the amount of E is small in comparison to the initial concentration of S. This is typical because enzymes are usually very effective catalysts, and this is why they are usually present in relatively small concentrations. A second interesting feature of the solution is that both C and E change relatively quickly, within the first few microseconds of the experiment. For example, as shown in the top plot of Figure 3.4, C starts at zero but increases up to approximately 0.56. In the lower plot of Figure 3.4, this change in C is not evident, and it appears that C simply starts out at $t = 0$ with a value of 0.56. This is the type of behavior that

would be expected if there were a boundary layer, although in this problem because we are looking at the time variable it is more appropriate to refer to this as an initial layer. To take advantage of this we need to first nondimensionalize the problem and then make use of our asymptotic approximation tools.

Given the numerical results in Figure 3.4 it is worth going back to the original equations (3.46)-(3.49) and calculating the steady-state solutions. These are found by assuming the concentrations are constant. One finds that the only solution is $S = 0, C = 0, E = E_0, P = S_0$. This is the solution seen in Figure 3.4 as $t \to \infty$. Said another way, the reactions are such that the original concentration of S is transformed into P, and once this is complete E is returned to its original concentration. This is the essence of a catalytic reaction.

3.4.2 Quasi-Steady-State Approximation

The rapid changes in the initial values of C and E were evident to the experimentalists who studied this reaction scheme. The physical reasoning usually

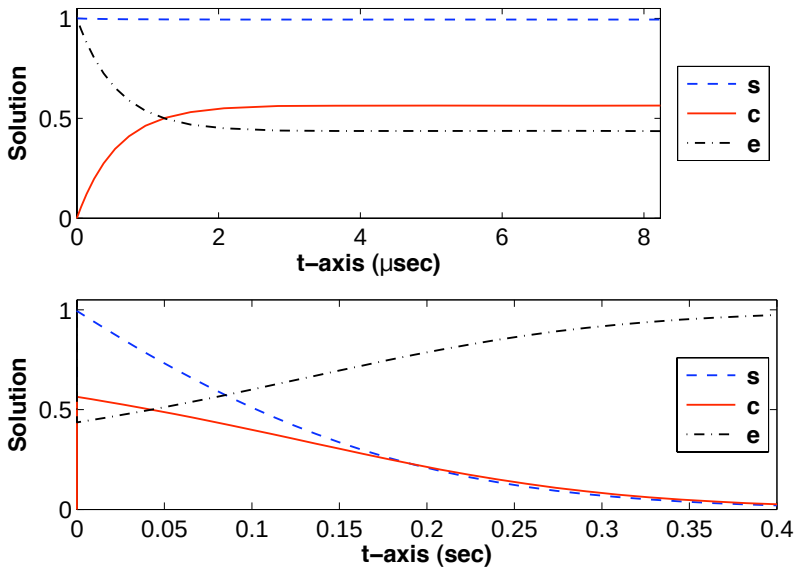


Figure 3.4 Numerical solution of (3.46)-(3.49) using parameter values for the transport of P-glycoprotein (Tran et al. [2005]). Shown are $s = S/S_0$, $c = C/E_0$, and $e = E/E_0$. The upper plot shows the solution curves during the first few moments of the experiment, while the lower plot shows the same curves over a longer time period.

given is that, assuming the concentration of S is not too small, the enzyme is so efficient that whenever an E becomes free that it immediately attaches itself to an S to form another complex C. The implication is that the concentration of C changes so quickly in response to the values of the other species that it immediately satisfies its steady-state equation. This is the basis of what is known as a quasi-steady-state assumption (QSSA), an idea first proposed by Briggs and Haldane [1928]. The argument made is that the time interval over which $C'(t) \neq 0$ is very small and the equations can be replaced with

$$\frac{dS}{dt} = -k_1 E_0 S + (k_{-1} + k_1 S)C, \tag{3.53}$$

$$0 = k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C. \tag{3.54}$$

Solving the last equation for C yields

$$C = \frac{k_1 E_0 S}{k_2 + k_{-1} + k_1 S}\,. \tag{3.55}$$

To put this approximation in the context of the transport of P-glycoprotein, as given in Figure 3.4, it is seen that the concentration of S changes on a time scale of seconds. This is why S appears to be constant in the upper plot. In comparison, C changes on a much faster time scale, measured in microseconds. This means C adjusts to the value of S so quickly, moving to what it assumes is the steady-state, that its value is determined by the formula in (3.55). The exception to this statement is what happens at the very beginning of the experiment, where C must undergo a jump to be able to satisfy (3.55).

Mathematically one should be a bit skeptical with this approximation. For one, the formula in (3.55) does not satisfy the given initial condition for C. For another, because S is time-dependent, C in (3.55) clearly depends on t, and this does not appear to be consistent with the assumption used to derive this result. These questions will be addressed once the perturbation solution is derived. For the moment we will assume all is well and in this case the equation for S, given in (3.53), can be written as

$$\frac{dS}{dt} = -\frac{v_M S}{K_M + S}\,. \tag{3.56}$$

In this equation $v_M = k_2 E_0$, and K_M is the Michaelis constant given as

$$K_M = \frac{k_{-1} + k_2}{k_1}\,. \tag{3.57}$$

Experimentalists use (3.56) to determine $v_M = k_2 E_0$ and K_M by measuring S' at $t = 0$. The specifics of how this is done are explored in Exercise 3.19. As it turns out, experimental studies that determine all three rate constants

are not common. There are technical, and mathematical, challenges in determining these constants, and an indication of what is involved can be found in Tran et al. [2005].

One of the questions that arises when measuring v_M and K_M is how it is possible to use the initial values of the concentrations. This appears to be a contradiction to the earlier assumption that enough time has passed so a quasi-steady-state has been reached. It should be pointed out that there are good reasons for using the initial values. One is that S is known at $t = 0$. Another is that the model assumes the reaction (3.45) is not reversible. In most applications it is, although the rate of the reverse reaction is so slow that it is not included. Using the values at the beginning minimizes the influence of this reaction on the measurements. This still leaves unresolved the apparent inconsistency in the assumptions, and for this we use a more mathematical argument.

3.4.3 Perturbation Approach

The QSSA is one of the standard methods used by biophysicists, and mathematicians, to reduce a reaction scheme. As pointed out in the derivation there are several mathematical questions concerning the consistency of the assumptions and for this reason we now consider a perturbation approximation. The underlying hypothesis in the analysis is that it takes very little enzyme to convert S to P. In other words, it is assumed that E_0 is much smaller than S_0.

Nondimensionalization
The first step in analyzing the solution is to nondimensionalize the problem. For S we use its initial condition $S(0) = S_0$ and set $S = S_0 s$, where s is the nondimensional version of S. The initial condition for C is not much help here, but the conservation law $E + C = E_0$ is because it indicates that the concentration of C can range up to E_0. Based on this observation, we take $C = E_0 c$, where c is the nondimensional version of C. It is not clear what to use for the time variable, and so we simply set $t = t_c \tau$, where τ is the nondimensionalized time variable. Introducing these into (3.51), (3.52) and cleaning things up a bit, produces the equations

$$\frac{1}{t_c k_1 E_0} \frac{ds}{d\tau} = -s + (\mu + s)c, \tag{3.58}$$

$$\frac{1}{t_c k_1 S_0} \frac{dc}{d\tau} = s - (\kappa + s)c, \tag{3.59}$$

where $\mu = k_{-1}/(k_1 S_0)$ and $\kappa = (k_{-1} + k_2)/(k_1 S_0)$. We are left with two dimensionless groups that involve t_c, and one of them will be set to one to determine t_c. The conventional choice is to use the group in (3.58), and with

this $t_c = 1/(k_1 E_0)$. In this case the Michaelis-Menten problem becomes

$$\frac{ds}{d\tau} = -s + (\mu + s)c, \tag{3.60}$$

$$\epsilon \frac{dc}{d\tau} = s - (\kappa + s)c, \tag{3.61}$$

where

$$s(0) = 1,\ c(0) = 0, \tag{3.62}$$

and

$$\epsilon = \frac{E_0}{S_0}\,. \tag{3.63}$$

We are assuming ϵ is small, and because it is multiplying the highest derivative in (3.61) we have a singular perturbation problem. For this reason it should be no surprise that we will find that the function c has a layer, at $t = 0$, where it undergoes a rapid transition. A consequence of this is that c quickly reaches what is called a quasi-steady-state and for all intents and purposes $s - (\kappa + s)c = 0$. Models containing fast dynamics, which in this case is the c equation, and slow dynamics, the s equation, are common in applications. In this way, the Michaelis-Menten system serves as a prototype enzymatic reaction as well as a prototype fast-slow dynamical system. Exactly how this happens, and what to do about the initial condition for c, will be derived using a perturbation argument.

The equations in (3.60), (3.61) are a relatively straightforward perturbation problem. We will concentrate on deriving the first term in the approximation and the first step is the outer expansion.

Outer Expansion

The appropriate expansions are $s \sim s_0 + \epsilon s_1 + \cdots$ and $c \sim c_0 + \epsilon c_1 + \cdots$. Inserting these into (3.60),(3.61) and collecting the order $O(1)$ terms we get

$$\frac{ds_0}{d\tau} = -s_0 + (\mu + s_0)c_0, \tag{3.64}$$

$$0 = s_0 - (\kappa + s_0)c_0. \tag{3.65}$$

Solving (3.65) for c_0 and substituting the result into (3.64) gives us

$$\frac{ds_0}{d\tau} = -\frac{\lambda s_0}{\kappa + s_0}\,, \tag{3.66}$$

where $\lambda = k_2/(k_1 S_0)$. Separating variables, and integrating, leads us to the following solution

$$\kappa \ln(s_0) + s_0 = -\lambda\tau + A, \tag{3.67}$$

$$c_0 = \frac{s_0}{\kappa + s_0}\,, \tag{3.68}$$

where A is a constant of integration that will be determined later when the layers are matched. The implicit nature of the solution in (3.67) is typical when solving nonlinear differential equations. It is possible to simplify this expression using what is known as the Lambert W function, however (3.67) is sufficient for what we have in mind.

Inner Expansion
The initial layer coordinate is $\bar{\tau} = \tau/\epsilon$ and in this region we will designate the solutions as $\bar{s}$ and $\bar{c}$. The problem is therefore

$$\frac{d\bar{s}}{d\bar{\tau}} = \epsilon(-\bar{s} + (\mu + \bar{s})\bar{c}), \tag{3.69}$$

$$\frac{d\bar{c}}{d\bar{\tau}} = \bar{s} - (\kappa + \bar{s})\bar{c}, \tag{3.70}$$

where

$$\bar{s}(0) = 1,\ \bar{c}(0) = 0. \tag{3.71}$$

The appropriate expansions in this case are $\bar{s} \sim \bar{s}_0 + \epsilon\bar{s}_1 + \cdots$ and $\bar{c} \sim \bar{c}_0 + \epsilon\bar{c}_1 + \cdots$. Inserting these into (3.69), (3.70) and collecting the first-order terms we get

$$\frac{d\bar{s}_0}{d\bar{\tau}} = 0,$$

$$\frac{d\bar{c}_0}{d\bar{\tau}} = \bar{s}_0 - (\kappa + \bar{s}_0)\bar{c}_0,$$

where

$$\bar{s}_0 = 1,\ \bar{c}_0(0) = 0.$$

Solving these equations gives us

$$\bar{s}_0 = 1, \tag{3.72}$$

$$\bar{c}_0 = \frac{1}{1+\kappa}\left(1 - e^{-(1+\kappa)\bar{\tau}}\right). \tag{3.73}$$

The solution for $\bar{s}$ indicates that, at least to first-order, it does not have an initial layer structure and is constant in this region. The variable $\bar{c}$ on the other hand does depend on the initial layer coordinate and this is consistent with our earlier numerical experiments.

Matching and Composite Expansion
The idea underlying matching is the same as used in the previous chapter, namely the solution coming out of the inner layer is the same as the solution going into the inner layer. Mathematically, the requirements are

$$\lim_{\bar{\tau}\to\infty} \bar{s}_0 = \lim_{\tau\to 0} s_0, \tag{3.74}$$

$$\lim_{\bar{\tau}\to\infty} \bar{c}_0 = \lim_{\tau\to 0} c_0. \tag{3.75}$$

From (3.72) and (3.74) we conclude that $s_0(0) = 1$. It is not hard to show that in this case (3.75) is satisfied, and we have that $A = 1$ in (3.67).

The only remaining task is to construct a composite expansion. This is done by adding the inner and outer solutions and then subtracting their common part. In other words, $s \sim s_0(\tau) + \bar{s}_0(\bar{\tau}) - s_0(0)$ and $c \sim c_0(\tau) + \bar{c}_0(\bar{\tau}) - c_0(0)$. The conclusion is that

$$s \sim s_0(\tau), \tag{3.76}$$

$$c \sim \frac{s_0}{\kappa + s_0} - \frac{1}{1+\kappa} e^{-(1+\kappa)\tau/\epsilon}, \tag{3.77}$$

where s_0 is found by solving

$$\kappa \ln(s_0) + s_0 = -\lambda\tau + 1. \tag{3.78}$$

Analysis of Solution

Although a simple closed form expression for s_0 is not available it is still possible to describe its basic behavior. First, from (3.66) we know it is monotone decreasing. For small values of τ, we have from (3.78) that $s_0 \approx s_0(0) + s_0'(0)\tau = 1 - \lambda\tau/(1+\kappa)$. In other words, it starts off by decreasing linearly with slope $-\lambda/(1+\kappa)$. For large τ values we have that $\kappa \ln(s_0) \approx -\lambda\tau$, and so, $s_0 \approx e^{-\lambda\tau/\kappa}$. Therefore, s_0 decays exponentially to zero for large τ, with a relaxation time constant κ/λ. Based on this information it is an easy matter to sketch the function.

It is also not difficult to solve the s_0 equation using, for example, Newton's method. The resulting composite approximation for $c(\tau)$ is shown in Figure 3.5 using the parameter values from the model of the transport of P-glycoprotein (Tran et al. [2005]). Also given is the solution obtained from solving the original equations (3.51), (3.52) numerically. The agreement is so good that it is difficult to distinguish between the numerical and composite solutions. This agreement is not limited to the P-glycoprotein values, and to demonstrate this another comparison is given in Figure 3.6 with $k_1 = 1.75 \times 10^{11}$ $\mathrm{M}^{-1}\mathrm{s}^{-1}$ and the other values the same as before. This particular choice was made as it also demonstrates an interesting behavior in the solution of the Michaelis-Menten equations. Namely, it appears that there is a transition layer in the solution as k_1 increases, located in the general vicinity of $\tau = 1.8 \times 10^5$. With the composite approximations it is easy to explain this behavior. Because κ is very small, we have from (3.77) that $c \approx 1$ outside the initial layer until s drops down to the value of κ. In the previous paragraph we know that the linear decrease in s can be approximated as $s \approx 1 - \lambda\tau$. Setting this equal to κ and solving we find that $\tau \approx 1/\lambda$. With the given parameter values this gives us $\tau \approx 1.75 \times 10^5$, which is very close to what is

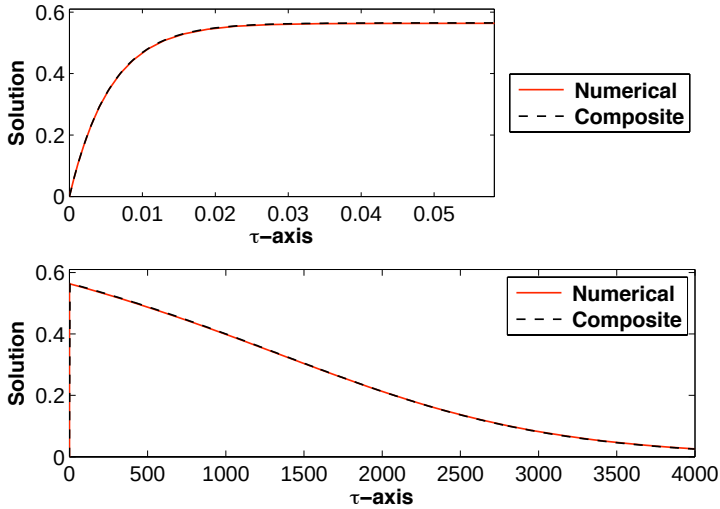


Figure 3.5 The solution for $c(\tau)$ obtained from solving the equations numerically, and from the composite approximation in (3.77), (3.78). The curves are so close it is difficult to distinguish between them. The parameter values are the same as used for Figure 3.4.

seen in Figure 3.6.

Connection with QSSA

It is informative to return to the assumptions underlying the quasi-steady-state assumption (3.53), (3.54). In the outer region, in dimensional coordinates, the equations (3.64), (3.65) become

$$\begin{aligned} \frac{dS}{dt} &= -k_1 E_0 S + (k_{-1} + k_1 S)C, \\ 0 &= k_1 E_0 S - (k_2 + k_{-1} + k_1 S)C. \end{aligned}$$

These are identical to those used in the quasi-steady-state assumption. In other words, QSSA is effectively equivalent to using an outer approximation of the solution. The actual justification for this type of reduction is the perturbation analysis carried out earlier.

Another observation concerns the resulting equation for S, which takes the form

$$\frac{dS}{dt} = -\frac{k_1 k_2 E_0 S}{k_2 + k_{-1} + k_1 S}\,.$$

What is of interest here is that this equation contains a rational function of the variables rather than the power functions expected from the Law of

Mass Action. The reason for mentioning this is that one finds models where the equations are rational functions and it is questionable whether they are derivable from mass action. This example demonstrates it is possible although determining this in general can be difficult. An example of this situation is given in the next section.

The composite approximation also suggests a possible modification of the analysis. The exponential dependence in (3.77) indicates that the approximation in the inner region holds not just for small ϵ, but also when κ is large. In contrast, we saw in Figure 3.6 that small values of κ can lead to the appearance of what looks to be a transition layer. To investigate these possibilities it is necessary to modify the nondimensionalization or the expansions used for the solution. This type of post-analysis of the expansion to gain insight into perhaps a better approximation is not uncommon, particularly for difficult problems where it is not clear at the beginning what scales should be used. An analysis of this type related to the Michaelis-Menten equations can be found in Segel and Slemrod [1989].

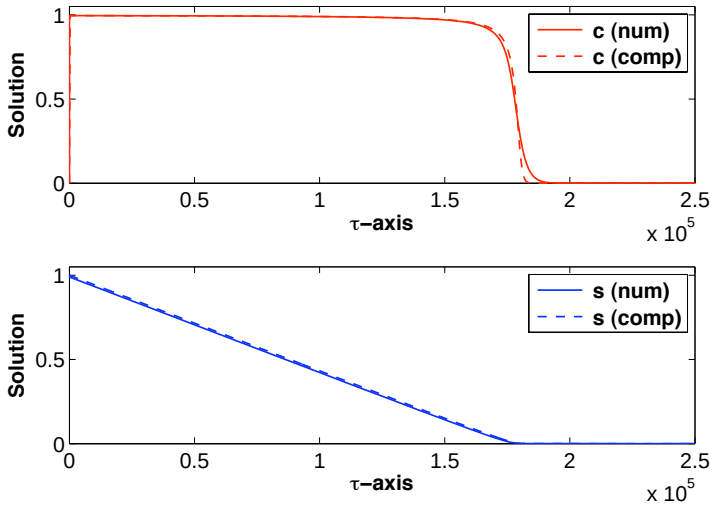


Figure 3.6 The solution of (3.60), (3.61) obtained from solving the equations numerically, and from the composite approximation in (3.77). The parameter values are the same as used for Figure 3.4 except $k_1 = 1.75 \times 10^{11}$ $\mathrm{M}^{-1}\mathrm{s}^{-1}$.

3.5 Assorted Applications

Most of the time using the Law of Mass Action is routine. Given one or more reactions the law is used to write down the kinetics equations without a great deal of trouble. In this section we consider variations of this situation that are not uncommon yet require a bit more thinking. In each case the issue is whether or not the situation under consideration is consistent with the Law of Mass Action in addition to other requirements imposed on the system.

3.5.1 Elementary and Nonelementary Reactions

The stoichiometric coefficients of most reactions are determined experimentally. As an example, suppose it is known that a product P is produced when two chemicals, A and B, are mixed together. The usual empirical-based assumption is that the rate of formation has the form $r = kA^\alpha B^\beta$, and the exponents and rate constant are determined by curve fitting this expression to the experimental measurements of r. Inevitably, such curve fits produce fractional values for α and β. This is partly due to the inherent error in the procedure, both experimental as well as numerical. It is also due to the fact that the r used in the curve fit is really an assumption about the overall rate of the reaction and not a statement about the actual sequence of molecular events through which A and B combine to form P. For example, H_2 and Br_2 combine to form hydrogen bromide, HBr. Curve fitting the rate function $r = k\mathrm{H}_2^\alpha \mathrm{Br}_2^\beta$ to the data it is found that $\alpha = 1$ and $\beta = 1/2$. Clearly, this cannot be associated with an elementary reaction, and this begs the question as to just how such an exponent can occur.

To address this question, consider the rather simple looking reaction

$$A + 2X \to P. \tag{3.79}$$

This gives rise to the kinetic equation

$$\frac{dA}{dt} = -kAX^2. \tag{3.80}$$

There are equations for the other species but for what we have in mind the above equation is enough. From a physical standpoint this type of reaction is highly unlikely, and the reason is that it requires one A and two X's to come together simultaneously to form the product P. In the real world, reactions are either unary, as in radioactive decay, or binary. The reaction in (3.79) is tertiary because it involves three reactant molecules.

As an example of this situation, one might attempt to explain the formation of water by writing $2H + O \to H_2O$. This is incorrect because the actual mechanism consists of a large number of intermediate reactions. It is common

that the sequence of elementary steps that constitute a reaction scheme is not known, and based on experimental evidence the formation of the product appears to consist of only a single reaction involving three or more reactants as in (3.79). This can happen, for example, if the amount of the intermediate species formed is very small and escapes detection, or if the speed with which the intermediate species are created and destroyed makes them difficult to detect. What we are going to consider here is, given a nonelementary reaction, is it possible to construct, at least mathematically, a sequence of elementary reactions that effectively accomplish the same task. By elementary we mean reactions that only involve one or two reactant molecules. It is understood that the expansion into elementary reactions is not necessarily unique and there are possibly multiple ways to do this. Rather, we are interested in whether it can be done at all.

As a first attempt to expand (3.79) into elementary reactions, it is not unreasonable to assume that a molecule of A combines with X to form an intermediate complex Z, and then Z combines with another X to form P. In reaction form this becomes

$$A + X \to Z, \tag{3.81}$$

$$Z + X \to P, \tag{3.82}$$

with the result that

$$\frac{dA}{dt} = -k_1 AX, \tag{3.83}$$

$$\frac{dZ}{dt} = k_1 AX - k_2 XZ. \tag{3.84}$$

The working hypothesis we will use is that the intermediate species Z plays a role similar to the intermediate species C in the Michaelis-Menten reaction. As with C, we will assume Z reaches a quasi-steady-state very quickly and, as in (3.65), this translates into the condition that $k_1 AX - k_2 XZ = 0$. Looking at this result a few seconds it is clear that it is not going to help us rewrite (3.84) so it resembles (3.80). What we are missing is the reversible reaction present in Michaelis-Menten, and so, the reaction scheme will be modified to become

$$A + X \rightleftharpoons Z, \tag{3.85}$$

$$Z + X \to P. \tag{3.86}$$

The kinetic equations in this case are

$$\frac{dA}{dt} = -k_1 AX + k_{-1} Z, \tag{3.87}$$

$$\frac{dZ}{dt} = k_1 AX - k_{-1} Z - k_2 XZ. \tag{3.88}$$

Imposing a quasi-steady-state assumption on Z yields

$$Z = \frac{k_1 AX}{k_{-1} + k_2 X} \,. \tag{3.89}$$

Substituting this into (3.87) gives us that

$$\frac{dA}{dt} = -\frac{k_1 k_2}{k_{-1} + k_2 X} AX^2 \,. \tag{3.90}$$

Assuming that $k_{-1} >> k_2 X$ then we obtain the kinetic equation in (3.80).

Clearly several assumptions went into this derivation and one should go through a more careful analysis using scaling and perturbation methods to delineate what exactly needs to be assumed. However, the fact that a non-physical reaction can be explained using elementary reactions has been established for this example.

3.5.2 Reverse Mass Action

A question that often arises is, given a set of equations, is it possible to determine if they are consistent with the Law of Mass Action. As an example, consider the predator-prey equations

$$\frac{dR}{dt} = aR - bRW, \tag{3.91}$$

$$\frac{dW}{dt} = -cW + dRW. \tag{3.92}$$

We start with the aR term in (3.91) and ask what reaction produces the equation $R' = aR$. Recall that in the original formulation of the model, the term aR is suppose to account for rabbits producing more rabbits. To construct the associated reaction, we know that the terms on the right-hand side of the differential equation are constructed from the reactants. One might therefore guess that the needed reaction is $R \to \textit{products}$. However, this, by itself, produces the equation $R' = -k_1 R$. We want to produce rabbits, not lose them, and this means we also need R as a product of the reaction. Based on this, the proposed reaction should have the form $\alpha R \to \beta R$. The equation in this case is $R' = -k_1 \alpha R^\alpha + k_1 \beta R^\alpha = k_1(\beta - \alpha)R^\alpha$. To be consistent with (3.91), we require $\alpha = 1$ and $a = k_1(\beta - 1)$. Because it is assumed that a is positive, then it is required that $\beta > 1$. For simplicity, it is assumed that $\beta = 2$.

Turning our attention to the RW terms in (3.91) and (3.92), the first guess is that the reaction is $R + W \to \textit{products}$. This agrees with (3.91) but it gives $W' = -k_2 RW$. We want to produce wolves, and this can be

accomplished by modifying the reaction to $R + W \to \gamma W$. This is consistent with the assumption that this term is responsible for the increase in the wolf population. The resulting equations are $R' = -k_2 RW$ and $W' = -k_2 RW + k_2 \gamma RW = k_2(\gamma - 1)RW$. This agrees with (3.91) and (3.92) if we let $b = k_2$ and $d = k_2(\gamma - 1)$. Because the coefficients must be positive it is required that $\gamma > 1$, and it is assumed here that $\gamma = 2$.

Working out the other term we have the following reactions,

$$R \to 2R, \tag{3.93}$$

$$R + W \to 2W, \tag{3.94}$$

$$W \to P. \tag{3.95}$$

In this last reaction P corresponds to the number of dead wolves.

Expressing the equations in reaction form provides a different viewpoint on the assumptions that were used to formulate the model. For example, the aR term in (3.91) came from the assumption that the number of rabbits increases at a rate proportional to the current population. This is commonly assumed but in looking at (3.93) it is hard to justify, at least for rabbits. For example, without some statement to the contrary, the above model applies to a population of all male rabbits just as well as to one where males and females are evenly distributed. Clearly, it must be assumed that both genders are present for the model to make any sense. Even so, the assumption that a rabbit undergoes mitosis and splits into two rabbits, as implied by (3.93), is a stretch. If one insists on only using one variable for rabbits, and not separating them into male and female, then it might be argued that a better assumption is to replace (3.93) with $R + R \to 3R$. This is still a little odd, and if used then aR in (3.91) is replaced with aR^2. Another possibility is to redefine R and assume it represents the population of only the female rabbits. This makes (3.93) somewhat easier to understand, but it raises questions about why the male rabbits do not affect the population of wolves. Clearly this is not possible.

The point of the above discussion is that models, by their very nature, are based on assumptions and it is important to have an understanding of what these assumptions are. There is nothing wrong with using a simple model to help develop an understanding of the problem, but it is essential to know what is assumed when this is done.

3.6 Steady-States and Stability

As with all time-dependent problems, one of the central questions is what happens to the solution as time increases. Specifically, if a physical system is started out with particular initial conditions, is it possible to determine if

the solution will approach a steady-state? There are various ways to address this question and we will consider three.

3.6.1 Reaction Analysis

It is possible, in some cases, to determine the steady-states and their stability properties directly from the reactions. To demonstrate how this is done we use the Michaelis-Menten reaction scheme, given as

$$S + E \rightleftharpoons C, \tag{3.96}$$

$$C \rightarrow P + E. \tag{3.97}$$

In molecular terms, the reactions $S+E \rightleftharpoons C$ state that one molecule of S combines with one molecule of E to reversibly form a complex C. By themselves, this exchange should continue indefinitely with no apparent steady-state. In comparison, in the reaction $C \rightarrow P + E$ one molecule of C breaks down, irreversibly, into P and E. The E molecule used to form C survives when P is produced, but not so with S. Based on this observation the reactions should continue until there are no more S's and C's. For each molecule of S or C we start with there will be one molecule of P produced. The only exception to this statement occurs if there are no E's available at the start, either individually or as part of C. In this case nothing happens, and the exact solution of the problem is simply $E = C = 0$, $S = S(0)$, and $P = P(0)$. Assuming $E(0)$ or $C(0)$ are nonzero, then the conclusion is that in the limit of $t \rightarrow \infty$ the solution converges to the steady-state $S = 0$, $C = 0$, $P = S(0)+P(0)+C(0)$, and $E = E(0)+C(0)$. Moreover, this limit is obtained no matter what initial values are chosen for S, C, and E. This is the property underlying the idea of an asymptotically stable steady-state. The expression used in this case is that $S = 0$, $C = 0$, $P = S(0) + P(0) + C(0)$, $E = E(0) + C(0)$ is an asymptotically stable steady-state. We will define asymptotic stability shortly, but what is significant is that we have been able to obtain this conclusion without explicitly using the kinetic equations.

3.6.2 Geometric Analysis

A second method for analyzing steady-states involves a geometric argument, and the starting point is the kinetic equations. It is a bit easier to use the nondimensional equations, and for the Michaelis Menten system they are

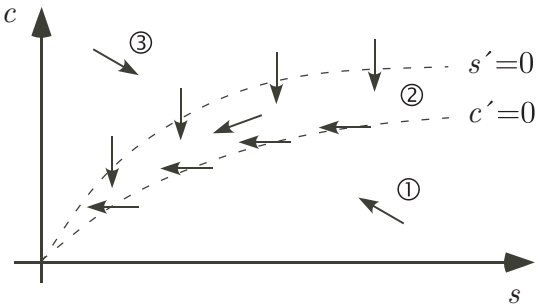


Figure 3.7 Phase plane and direction fields for the Michaelis-Menten system, in the region $0 \leq s$ and $0 \leq c$.

$$\frac{ds}{d\tau} = -s + (\mu + s)c, \tag{3.98}$$

$$\epsilon\frac{dc}{d\tau} = s - (\kappa + s)c. \tag{3.99}$$

To transform this into the phase plane we combine the above two differential equations to obtain

$$\frac{dc}{ds} = \frac{-s + (\mu + s)c}{\epsilon(s - (\kappa + s)c)}\,. \tag{3.100}$$

The idea here is that in the cs-plane the solution is a parametric curve, with the time variable τ acting as the parameter. With this viewpoint, (3.98) and (3.99) are expressions for the velocities of the respective variables. We will use these equations, along with (3.100), to sketch the solution. Before doing this note that the physically relevant solution satisfies $0 \leq s$ and $0 \leq c$. Limiting our attention to this region then the situation is sketched in Figure 3.7. The first step used to produce this figure was to sketch the two nullclines. The s-nullcline is the curve where $s' = 0$, and from (3.98) this is $c = s/(\mu + s)$. Similarly, the c-nullcline is the curve where $c' = 0$, and from (3.99) this is $c = s/(\kappa + s)$. The points where the nullclines intersect are the steady-states, and for this problem this is simply $s = c = 0$. These two curves separate the quadrant into three regions, designated ①, ②, ③. Using (3.100) we have the following cases.

Region ①. In this region $s' < 0$ and $c' > 0$. Consequently, $\frac{dc}{ds} < 0$ and this means the slope of the solution curve in this region is negative. The slope of the small arrow indicates this in the figure. The arrow points in the direction of motion, and this is determined from the inequalities $c' > 0$ and $s' < 0$. So, c is increasing and s is decreasing.

Region ②. Now $\frac{dc}{ds} > 0$, and so the slope of the solution curve in this region is positive. The small line segment indicates this in the figure. The arrow on

the line is determined by noting from (3.98) that $c' > 0$ while, from (3.98), $s' < 0$. So, c and s are decreasing.

Region ③. Because $\frac{dc}{ds} < 0$ then the slope of the solution curve in this region is negative. The small line segment indicates this in the figure. The arrow on the line is determined by noting from (3.98) that $c' < 0$ while, from (3.98), $s' > 0$. So, c is decreasing and s is increasing.

Nullclines. When $s' = 0$ the slope of the solution curve, as determined from (3.100), is vertical. When this occurs, c is decreasing and this explains the arrows. When $c' = 0$ the slope of the solution curve, as determined from (3.100), is horizontal. When this occurs, s is decreasing and this explains the arrows.

With this information we are in position to sketch the solution. The initial conditions we have been using for Michaelis-Menten are $s(0) = 1$ and $c(0) = 0$. As indicated by the direction field shown in Figure 3.7, when starting at this position, c increases and s decreases. This continues until the c-nullcline is crossed, after which the solution heads towards $c = s = 0$. This is shown in Figure 3.8. In contrast, if one starts out with $s = 0$ and $c \neq 0$, then c decreases while s increases. This continues until the s-nullcline is crossed, after which the solution converges to $c = s = 0$. It would appear that no matter where we start that the same conclusion is reached, and for this reason we conclude that $s = c = 0$ is a globally asymptotically stable steady-state. By global it is understood that the conclusion holds for every initial condition that is in the region $s \geq 0$, $c \geq 0$. As a point of interest related to this conclusion, Exercise 3.7 should be consulted.

As a check on the geometric arguments made here, the numerical solution is shown in Figure 3.9 for various values of ϵ. In each case the two nullclines are also plotted. All three graphs behave as predicted in Figure 3.8. What is most striking, however, is how the trajectory changes as ϵ decreases. The smaller the value of ϵ is, the faster the solution moves to the c-nullcline. This gives

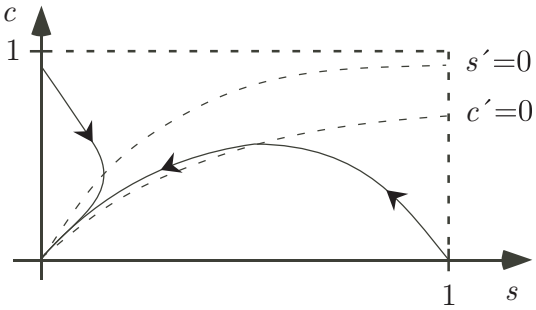


Figure 3.8 Phase plane and direction fields for the Michaelis-Menten system.

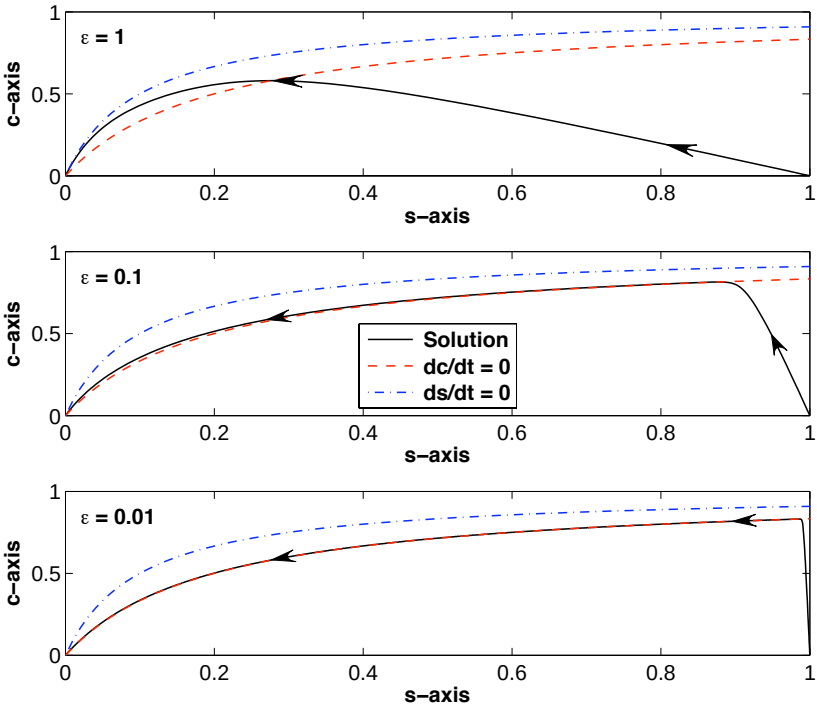


Figure 3.9 Numerical solution of the Michaelis-Menten equations for various values of ϵ, along with the s- and c-nullclines. The arrows on the solution curves show the direction of motion. Note the rapid initial rise in c for small values of ϵ.

rise to an initial layer, and we used this earlier to construct the perturbation approximation of the solution. Once the solution reaches the c-nullcline it then follows this curve into the steady-state. In geometric terminology the c-nullcline is called a slow manifold, and the solution quickly moves to this slow manifold and follows it into the steady-state. This behavior is also, essentially, the basis of the quasi-steady-state approximation described in Section 3.4.2.

3.6.3 Perturbation Analysis

The two methods we have used to study the properties of the steady-state solution have the advantage of not requiring us to solve the differential equations. This makes them very attractive but they are limited in their applicability. For example, with reaction analysis, if one has even a moderate number

of reactions it can be difficult to sort out what species reach a steady-state. Similarly, the geometric approach is most useful for systems with only two or three species. We now consider a more analytical method, one capable of resolving a large number of reactions and for this we turn to a perturbation analysis.

Before working out the Michaelis-Menten example, we consider a general version of the problem. Specifically, assume the kinetic equations can be written is vector form as

$$\frac{d}{dt}\mathbf{y} = \mathbf{F}(\mathbf{y}), \tag{3.101}$$

or in component form

$$\begin{aligned}
\frac{d}{dt}y_1 &= F_1(y_1, y_2, \ldots, y_n),\\
\frac{d}{dt}y_2 &= F_2(y_1, y_2, \ldots, y_n),\\
\vdots \qquad & \qquad \vdots\\
\frac{d}{dt}y_n &= F_n(y_1, y_2, \ldots, y_n).
\end{aligned}$$

In this case, $\mathbf{y}_s$ is a steady-state solution if $\mathbf{y}_s$ is constant and $\mathbf{F}(\mathbf{y}_s) = \mathbf{0}$. We say that $\mathbf{y}_s$ is stable if any solution that starts near $\mathbf{y}_s$ stays near it. If, in addition, initial conditions starting near $\mathbf{y}_s$ actually result in the solution converging to $\mathbf{y}_s$ as $t \to \infty$, then $\mathbf{y}_s$ is said to be *asymptotically stable.*

To investigate stability we introduce the initial condition

$$\mathbf{y}(0) = \mathbf{y}_s + \delta\mathbf{a}. \tag{3.102}$$

The idea here is that we are starting the solution close to the steady-state (see Figure 3.10), and so we assume δ is small. Now, for asymptotic stability it is required that the solution of the resulting initial value problem converges to $\mathbf{y}_s$ as time increases, irrespective of the particular values for $\mathbf{a}$. This will be determined using asymptotics, and the appropriate expansion of the solution for small δ is

$$\mathbf{y}(t) \sim \mathbf{y}_s + \delta\,\overline{\mathbf{y}}(t) + \cdots. \tag{3.103}$$

If it is found that $\lim_{t\to\infty}\overline{\mathbf{y}} = \mathbf{0}$, no matter what we pick for $\mathbf{a}$, then the steady-state is asymptotically stable (to small perturbations). This approach is called a linear stability analysis and it will be our standard method for deciding if a steady-state is stable. Note that unlike the other two methods, the perturbation approach is not capable, except for very simple problems, to determine if the steady-state is globally asymptotically stable. For this to hold we would need to prove convergence for all initial values, and not just those close to the steady-state.

Substituting (3.103) into (3.101), and using Taylor's theorem, one finds that

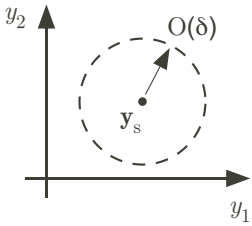


Figure 3.10 As given in (3.102), the initial conditions used in the linearized stability analysis are taken to be within a $O(\delta)$ region around the steady-state solution $\mathbf{y}_s$.

$$\begin{aligned}\frac{d}{dt}\left(\mathbf{y}_s + \delta\,\overline{\mathbf{y}}(t) + \cdots\right) &= \mathbf{F}\left(\mathbf{y}_s + \delta\,\overline{\mathbf{y}}(t) + \cdots\right)\\ &= \mathbf{F}\left(\mathbf{y}_s\right) + \mathbf{A}\left(\delta\,\overline{\mathbf{y}}(t) + \cdots\right) + \cdots\,,\end{aligned} \tag{3.104}$$

where $\mathbf{A} = \mathbf{F}'(\mathbf{y_s})$ is the Jacobian matrix for $\mathbf{F}$. More explicitly

$$\mathbf{A} = \begin{pmatrix} \dfrac{\partial F_1}{\partial y_1} & \dfrac{\partial F_1}{\partial y_2} & \cdots & \dfrac{\partial F_1}{\partial y_n} \\ \dfrac{\partial F_2}{\partial y_1} & \dfrac{\partial F_2}{\partial y_2} & \cdots & \dfrac{\partial F_2}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial F_n}{\partial y_1} & \dfrac{\partial F_n}{\partial y_2} & \cdots & \dfrac{\partial F_n}{\partial y_n} \end{pmatrix}, \tag{3.105}$$

where the derivatives in the above matrix are evaluated at $\mathbf{y}_s$. From the $O(\delta)$ terms in (3.104) and (3.102), it follows that

$$\frac{d}{dt}\overline{\mathbf{y}} = \mathbf{A}\overline{\mathbf{y}}, \tag{3.106}$$

where

$$\overline{\mathbf{y}}(0) = \mathbf{a}. \tag{3.107}$$

The solution is found by assuming that $\mathbf{y} = \mathbf{x}e^{rt}$. Substituting this into (3.106), the problem reduces to solving

$$\mathbf{A}\mathbf{x} = r\mathbf{x}. \tag{3.108}$$

This is an eigenvalue problem, where r is the eigenvalue and $\mathbf{x}$ is the associated eigenvector. With this, the values for r are determined by solving the characteristic equation

$$\det(\mathbf{A} - r\mathbf{I}) = 0, \tag{3.109}$$

where $\mathbf{I}$ is the identity matrix. Given a value of r, the eigenvector is then determined by solving $(\mathbf{A} - r\mathbf{I})\mathbf{x} = \mathbf{0}$.

In the case when $\mathbf{A}$ has n distinct eigenvalues $r_1, r_2, \ldots, r_n$, with corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, then the general solution of (3.106) has the form

$$\overline{\mathbf{y}} = \alpha_1 \mathbf{x}_1 e^{r_1 t} + \alpha_2 \mathbf{x}_2 e^{r_2 t} + \cdots + \alpha_n \mathbf{x}_n e^{r_n t}, \tag{3.110}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are arbitrary constants. The latter are determined from the initial condition (3.107). However, it is not necessary to calculate their values as we are only interested in the time-dependence of the solution. In particular, from (3.110), it follows that $\overline{\mathbf{y}} \to \mathbf{0}$ as $t \to \infty$ if $\mathrm{Re}(r_i) < 0$, $\forall i$. However, if even one eigenvalue has $\mathrm{Re}(r_i) > 0$ then it is possible to find values for $\mathbf{a}$ so $\overline{\mathbf{y}}$ is unbounded as $t \to \infty$.

If $\mathbf{A}$ does not have n distinct eigenvalues, then the general solution contains e^{rt} terms as well as those of the form $t^k e^{rt}$, where k is a positive integer. Consequently, the conclusion is the same, which is that $\overline{\mathbf{y}} \to \mathbf{0}$ as $t \to \infty$ if $\mathrm{Re}(r_i) < 0$, $\forall i$. Moreover, if there is an eigenvalue with $\mathrm{Re}(r_i) > 0$ then $\overline{\mathbf{y}}$ can become unbounded as $t \to \infty$. These conclusions have been reached without actually writing down the solution, which is possible because we only need to know the time-dependence of the solution. Those interested in the exact formula, in the case of when there are not n distinct eigenvalues, should consult Braun [1993].

The discussion in the previous paragraphs gives rise to the following result.

Theorem 3.2. *The steady-state $\mathbf{y}_s$ is asymptotically stable if all of the eigenvalues of $\mathbf{A}$ satisfy $Re(r) < 0$, and it is unstable if even one eigenvalue has $Re(r) > 0$.*

As will become evident in the following examples, the eigenvalues contain more information than just asymptotic stability.

Examples

1. We begin with the Michaelis-Menten reaction, and the equations of motion in this case are

$$\frac{ds}{d\tau} = -s + (\mu + s)c, \tag{3.111}$$

$$\epsilon \frac{dc}{d\tau} = s - (\kappa + s)c. \tag{3.112}$$

In this example, $\mathbf{y} = (s, c)^T$,

$$\mathbf{F} = \begin{pmatrix} -s + (\mu + s)c \\ \\ \dfrac{1}{\epsilon}\left(s - (\kappa + s)c\right) \end{pmatrix},$$

and, from (3.105),

$$\mathbf{A} = \begin{pmatrix} -1+c & \mu+s \\ \frac{1}{\epsilon}(1-c) & -\frac{1}{\epsilon}(\kappa+s) \end{pmatrix}. \tag{3.113}$$

Setting $s' = c' = 0$ in (3.111), (3.112) one finds that the only steady-state is $s_s = c_s = 0$. Substituting this into (3.113) yields

$$\mathbf{A} = \begin{pmatrix} -1 & \mu \\ \frac{1}{\epsilon} & \frac{\kappa}{\epsilon} \end{pmatrix}.$$

The eigenvalues are found by solving $\det(\mathbf{A} - r\mathbf{I}) = 0$, which reduces to solving the quadratic equation $\epsilon r^2 + (\kappa+\epsilon)r + \kappa - \mu = 0$. From this we get the solutions

$$r_\pm = \frac{1}{2\epsilon}\left(-(\kappa+\epsilon) \pm \sqrt{(\kappa+\epsilon)^2 - 4\epsilon(\kappa-\mu)}\right).$$

Because $\kappa - \mu > 0$ it follows that $\mathrm{Re}(r_\pm) < 0$. Therefore, from the above theorem, the steady-state $(s_s, c_s) = (0,0)$ is asymptotically stable. ■

2. Suppose the system to solve is

$$\frac{du}{dt} = v, \tag{3.114}$$

$$\frac{dv}{dt} = -v - \alpha u(1-u), \tag{3.115}$$

where α is nonzero. In this example, $\mathbf{y} = (u, v)^T$,

$$\mathbf{F} = \begin{pmatrix} v \\ -v - \alpha u(1-u) \end{pmatrix},$$

and, from (3.105),

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 2u-\alpha & -1 \end{pmatrix}.$$

Setting $u' = v' = 0$, the steady-states are found to be $(u_s, v_s) = (0,0)$ and $(u_s, v_s) = (1,0)$. In either case, the eigenvalue equation $\det(\mathbf{A} - r\mathbf{I}) = 0$ reduces to solving $r^2 + r + \alpha(1-2u_s) = 0$. From this we get the solutions

$$r_\pm = \frac{1}{2}\left(-1 \pm \sqrt{1 - 4\alpha(1-2u_s)}\right). \tag{3.116}$$

Now, for stability it is required that $\lim_{t\to\infty} e^{rt} = 0$ for both r values given in (3.116). Given that $\mathrm{Re}(r_-) \le \mathrm{Re}(r_+)$, then the stability requirement is

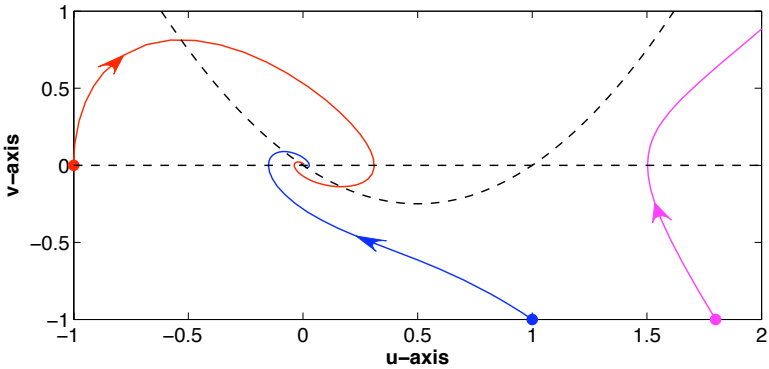


Figure 3.11 Numerical solution of (3.114), (3.115) using different starting points. In the calculations, $\alpha = 1$ so the asymptotically stable steady-state is $(u, v) = (0, 0)$. Also, the dashed curves are the two nullclines.

$\mathrm{Re}(r_+) < 0$. For the steady-state $(u_s, v_s) = (0, 0)$, $r_+ = \frac{1}{2}(-1 + \sqrt{1 - 4\alpha})$. Consequently, this steady-state is asymptotically stable if $\alpha > 0$, and it is unstable if $\alpha < 0$. For the steady-state $(u_s, v_s) = (1, 0)$, $r_+ = \frac{1}{2}(-1 + \sqrt{1 + 4\alpha})$. This is asymptotically stable if $\alpha < 0$, and it is unstable if $\alpha > 0$. These conclusions are based on the assumption that the initial condition starts near the steady-state. To illustrate what happens when this does not happen, three solution curves are shown in Figure 3.11 for $\alpha = 1$, using initial conditions that are not near the asymptotically stable steady-state $(u_s, v_s) = (0, 0)$. Although two of the solutions do end up converging to the steady-state, one does not. ■

3. As a third example we consider the system

$$\frac{du}{dt} = -v, \tag{3.117}$$

$$\epsilon \frac{dv}{dt} = u + \lambda(v - v^3/3). \tag{3.118}$$

It is assumed ϵ is positive and λ is nonzero. This is a well-known problem related to the van der Pol equation. Setting $u' = v' = 0$, the steady-state is found to be $(u_s, v_s) = (0, 0)$. With this, and carrying out the required differentiations, it is found that

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ \dfrac{1}{\epsilon} & \dfrac{\lambda}{3\epsilon} \end{pmatrix}.$$

The resulting eigenvalues are

$$r_\pm = \frac{1}{2\epsilon}\left(\lambda \pm \sqrt{\lambda^2 - 4\epsilon}\right). \tag{3.119}$$

Now, for stability it is required that $\mathrm{Re}(r_\pm) < 0$, and so the steady-state $(u_s, v_s) = (0, 0)$ is asymptotically stable if $\lambda < 0$ and unstable if $\lambda > 0$.

Up to this point this example appears to be very similar to the previous one. However, there is an important difference in how the steady-state goes unstable. In the last example the steady-state $(u_s, v_s) = (0, 0)$ went unstable because r_+ switched from negative to positive as α passed through zero. In the current example, when λ is close to zero, the solutions in (3.119) are complex valued. The steady-state goes unstable, as λ goes from negative to positive, because r_+ and r_- both move from the left half-plane, where $\mathrm{Re}(r) < 0$, into the right half-plane, where $\mathrm{Re}(r) > 0$, as λ passes through zero. Moreover, at $\lambda = 0$, $\frac{d}{d\lambda}\mathrm{Re}(r) \neq 0$. This is called a Hopf bifurcation. This has interesting repercussions in how the solution behaves when $\lambda > 0$, and exactly what it does do is considered next.

A sketch of the basic properties of the solution in the phase plane is given in Figure 3.12. The u-nullcline is a vertical line passing through the origin, while the v-nullcline is the cubic $-\lambda(v - v^3/3)$. Both are shown in the figure with dashed curves. The small arrows indicate the slope as determined from $\frac{dv}{du}$, with the arrowheads showing the direction of the solution as time increases. From this it is possible to give a rough description of the solution curve assuming $\epsilon << 1$. As an example, suppose the initial condition corresponds to the point a in the figure. Because ϵ is small, the v equation in (3.118) will move very quickly to reach a quasi-steady-state. This means the solution will move almost immediately to the v-nullcline, and given the direction of the arrows, this means it will move towards point b. Once there the solution will move upwards, following the v-nullcline very closely, until it reaches point c. It must still move upwards, but the v equation requires it to stay near the v-nullcline. The only choice is for the solution to move from c over to d. Once there it then moves down, following the v-nullcline very closely, until it reaches point e. It must continue moving downward, but will

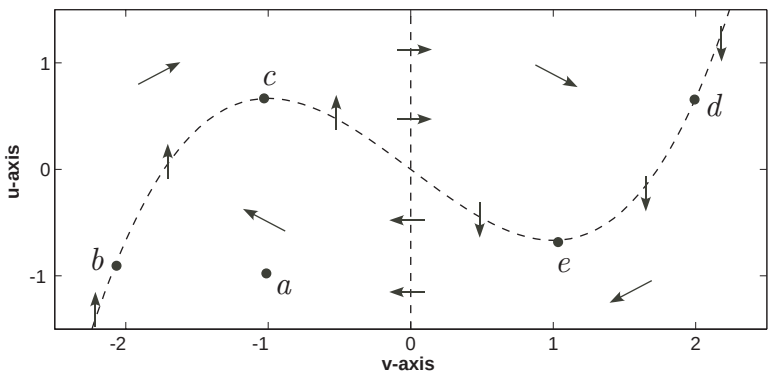


Figure 3.12 Phase plane and direction fields for Example 3, in the case of when $\lambda > 0$.

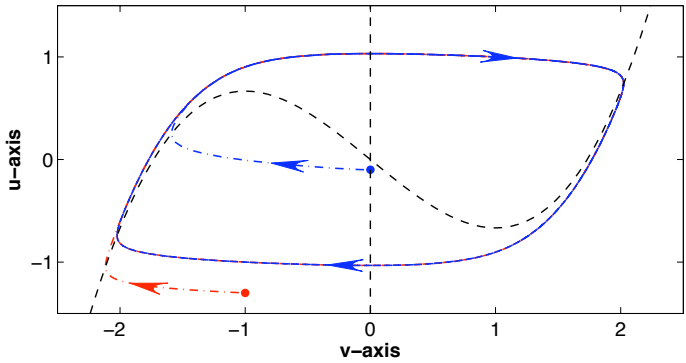


Figure 3.13 Numerical solution of (3.114), (3.115) using different starting points. In the calculations, $\lambda = 1$. Also, the dashed curves are the two nullclines.

stay near the v-nullcline. This means the solution will head towards point b, and once there the whole process repeats itself. The result is that the solution converges to a closed circuit that encloses the unstable steady-state solution. This is known as a limit cycle. To reinforce this conclusion, the numerical solution of the problem is given in Figure 3.13. Two different starting points are used, one inside and the other outside the limit cycle, and both converge to the limit cycle. A somewhat different perspective is shown in Figure 3.14, which shows the solution trajectory with the time variable given explicitly. ■

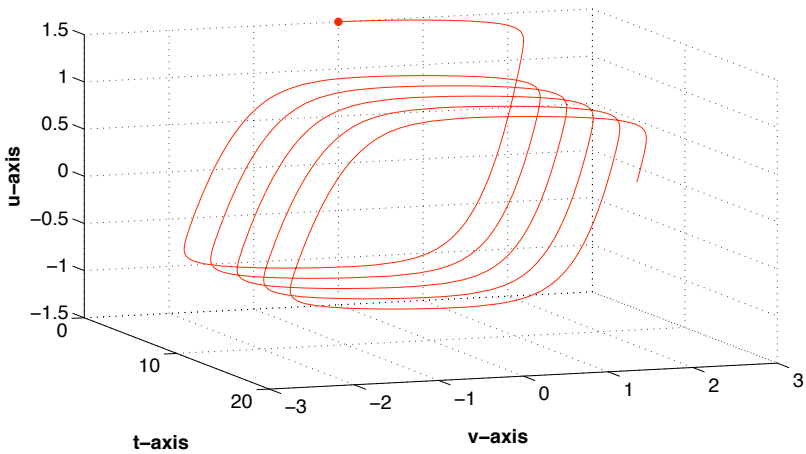


Figure 3.14 Numerical solution of (3.114), (3.115) with $\lambda = 1$. The initial condition is $(u, v) = (0, 1.5)$ and $\epsilon = 0.01$.

Theorem 3.2 leaves open what happens if there are a pair of eigenvalues that are complex conjugates with $\mathrm{Re}(r) = 0$. In such situations the nonlinear terms that were left off in the Taylor series approximation (3.104) must be considered. There are very few general formulas for such cases, and it is usually necessary to work out each problem individually. This can be done using a multiple scale analysis (Holmes [1995]), or else using more general analytical methods (Hale and Kocak [1996]).

3.7 Oscillators

The Michaelis-Menten reactions resulted in the solution converging, with little fanfare, to the steady-state solution of the system. It not uncommon for nonlinear systems to have multiple steady-states, and to have solutions that show periodic behavior. Examples of these situations were studied in the previous section. We are going to investigate an application of this material and, specifically, look at the existence of periodic solutions to a particular chemical system.

The idea of a chemical oscillator was not accepted easily, and the first paper reporting such a system generated more articles devoted to proving it wrong than trying to understand how it works. This is also true for the most well-known oscillator, the Belousov-Zhabotinskii (BZ) reaction. This was discovered by B. P. Belousov when studying the Krebs cycle. He found that a solution of citric acid in water, with acidified bromate as the oxidant and yellow ceric ions as the catalyst, alternated in color, from yellow to clear, approximately every minute and did this for an hour. His suggestion that this was a form of chemical oscillator was not accepted at the time because it

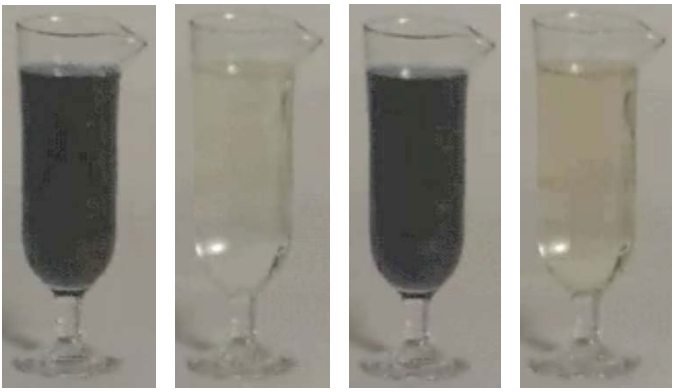


Figure 3.15 Color changes in a chemical oscillator. These are frames from a video recording of the experiment, the time of the respective frame, from left to right, is 20 sec, 28 sec, 31 sec, and 38 sec (Bodner et al. [2009]).

was believed that oscillations in closed homogeneous systems were impossible because that would imply that the reaction did not go smoothly to a thermodynamic equilibrium. About ten years later, A. M. Zhabotinskii expanded on Belousov's research and the work was presented at the 1968 Symposium on Biological and Biochemical Oscillators in Prague. Not unexpectedly, their system has become known as the BZ reaction. Since their early work many other chemical oscillators have been found, and one is shown in Figure 3.15. In this particular experiment, the system alternates between three states: clear, blue, and amber. As with many such oscillators, the exact mechanism is unknown.

The most widely accepted model for the BZ reaction is due to Field, Körös, and Noyes (Field et al. [1972], Field and Noyes [1974]). Their original formulation contained eleven reactions involving 12 chemical species. It is possible to reduce this system to five reactions, and they are

$$\begin{aligned} A + Y &\to X + P, \\ X + Y &\to 2P, \\ A + X &\to 2X + 2Z, \\ 2X &\to A + P, \\ B + Z &\to \frac{1}{2} f Y. \end{aligned}$$

The chemicals involved here are bromous acid (X), bromide (Y), cerium-4 (Z), bromate (A), an organic species (B), and a product P. The two reactions that stand out are the third because it is autocatalytic, and the last because it involves a rather unusual stoichiometric coefficient. As will be seen below, the parameter f plays an important role in producing the oscillations in the solution. This reduced model is often called the Oregonator, due to the location of Field and Noyes when they first derived it.

There are two additional simplifying assumptions made in the Oregonator description. Namely, in the experiment the concentrations of A and B are so large in comparison to the other chemicals that it is assumed they are constant during the reaction. With this the Law of Mass Action produces the following kinetic equations,

$$\begin{aligned} \frac{dX}{dt} &= k_1 AY - k_2 XY + k_3 AX - 2k_4 X^2, \\ \frac{dY}{dt} &= -k_1 AY - k_2 XY + \frac{1}{2} f k_5 BZ, \\ \frac{dZ}{dt} &= 2k_3 AX - k_5 BZ. \end{aligned}$$

To nondimensionalize the problem we take $X = X_c x$, $Y = Y_c y$, $Z = Z_c z$, and $t = t_c \tau$, where $X_c = k_3 A/(2k_4)$, $Y_c = k_3 A/k_2$, $Z_c = (k_3 A)^2/(k_4 k_5 B)$, and $t_c = 1/(k_5 B)$. In this case the above equations become

$$\epsilon x' = \alpha y - xy + x(1-x), \tag{3.120}$$
$$\delta y' = -\alpha y - xy + fz, \tag{3.121}$$
$$z' = x - z, \tag{3.122}$$

where $\epsilon = 4\times10^{-2}$, $\alpha = 8\times10^{-4}$, and $\delta = 4\times10^{-4}$. We will take advantage of the small values of these three parameters in constructing an approximation of the solution.

The objective here is to understand how the species in the reactions are able to produce sustained oscillations over a long period of time. Although we have the tools to handle all three equations, we have learned something from the Michaelis-Menten reaction that enables us to simplify the situation a bit. The very small value of δ, which multiplies y', means that this particular equation reaches a quasi-steady-state very quickly compared to the other two equations. Using this observation we have that $y = fz/(\alpha+x)$. With this the equations of motion reduce to

$$\epsilon x' = x(1-x) + \frac{\alpha - x}{\alpha + x} fz, \tag{3.123}$$
$$z' = x - z. \tag{3.124}$$

It is this system of equations that we will analyze. In this sense we will be constructing an approximation that is the outer solution to (3.120) - (3.122).

The question arises as to why the QSSA is not also applied to (3.123), which would reduce the entire system done to a single equation. This is an example that illustrates that some care is needed when using the QSSA. As we will see shortly, even though x tries to reach a quasi-steady-state, there are values for the parameter f for which the equation does not have the capability to reach a steady-state. The solution has to repeatedly reposition itself, trying to maintain a quasi-steady-state, and this gives rise to a pronounced non-steady behavior in the solution.

3.7.1 Stability

The first step is to determine the steady-states. Setting $x' = 0$ and $z' = 0$ one finds two solutions, $(x_s, z_s) = (0,0)$ and $(x_s, z_s) = (\bar{x}, \bar{x})$, where

$$\bar{x} = \frac{1}{2}\left[-\kappa + \sqrt{\kappa^2 + 4\alpha(1+f)}\right], \tag{3.125}$$

for $\kappa = \alpha + f - 1$.

To determine the stability properties of the steady-states we will use the geometric argument. The most difficult step for this is to sketch the x-nullcline, which corresponds to the curve

$$z = -x(1-x)\frac{\alpha + x}{(\alpha - x)f}\,. \tag{3.126}$$

A rough sketch can be made by making use of the fact that α is very small, while x ranges over the interval $0 \le x < \infty$. Except when x is near α, we can use the approximation $\alpha \pm x \approx \pm x$. With this (3.126) reduces to

$$z \approx \frac{1}{f}x(1-x). \tag{3.127}$$

This is simply a quadratic as shown in Figure 3.16. For x near α then (3.126) reduces to

$$z \approx -\frac{2\alpha x}{(\alpha - x)f}\,. \tag{3.128}$$

This hyperbola is also shown in Figure 3.16. With this it is possible to sketch the x-nullcline. Namely, for x near α the curve is given in (3.128), and everywhere else it is given in (3.127). A comparison between these approximations and the exact curve is given in Figure 3.16.

The above approximations for the x-nullcline make it easy to estimate where the critical points are located. For example, the local minimum is near where the two approximation curves intersect. Equating (3.127) and (3.128), we have that $x \approx 3\alpha$. Similarly, the local maximum comes from the quadratic (3.128), and this is therefore located at $x \approx \frac{1}{2}$. Both of these values are rough estimates, and more accurate approximations will be derived shortly. One last point to make here is that the value of α used in Figure 3.16 is larger than the value for the BZ system. This was done to make the characteristics of the curve more apparent in the plot because when using the actual value the three curves are so close that it is not possible to distinguish among them graphically.

To use the geometric argument the value of f needs to be specified. We will consider two cases, and the first is $f = \frac{1}{4}$. The resulting phase plane diagram

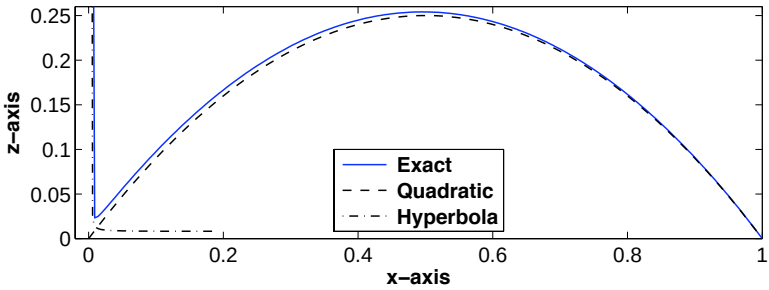


Figure 3.16 The x-nullcline (3.126) is shown along with its quadratic approximation, (3.127), and its hyperbolic approximation, (3.128). For this example, $f = 1$ and $\alpha = 4 \times 10^{-3}$.

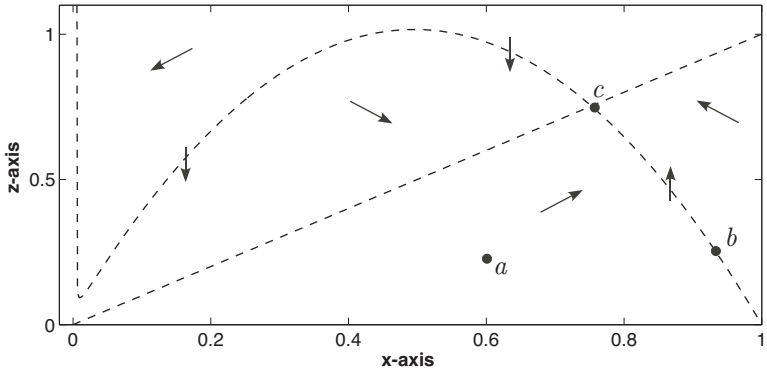


Figure 3.17 Phase plane and direction fields for (3.123), (3.124) when $f = \frac{1}{4}$.

is given in Figure 3.17. The two nullclines are shown, as are the direction fields. To describe the trajectory of the solution suppose the initial condition corresponds to the point a in the figure. Because ϵ is small, the x equation in (3.118) will move very quickly to reach a quasi-steady-state. This means the solution will move almost immediately to the x-nullcline, and given the direction of the arrows, this means it will move towards point b. Once there the solution will move upwards, following the x-nullcline very closely, until it reaches point c. Once there it is at the steady-state and will therefore remain at this location. To reinforce this conclusion, the numerical solution of the problem is given in Figure 3.18.

The convergence of the solution to the steady-state occurs when $f = \frac{1}{4}$ because the nullclines intersect at a point that the solution is able to reach. To explain what this means we consider the case of when $f = \frac{3}{4}$. The phase plane, and direction fields, are shown in Figure 3.19. As before, starting at point a, the solution will move very quickly to point b. It will then move upwards,

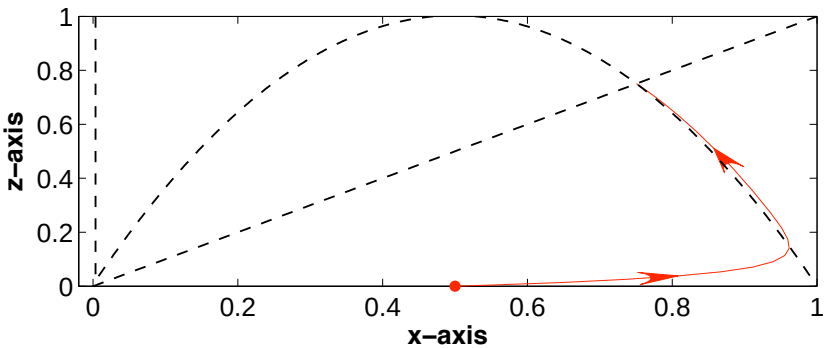


Figure 3.18 Numerical solution of (3.123), (3.124) in the case of $f = \frac{1}{4}$. The initial condition is $(x, z) = (0.5, 0)$.

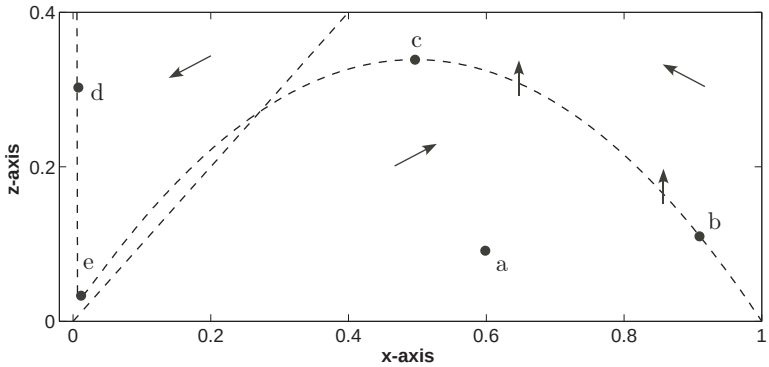


Figure 3.19 Phase plane and direction fields for (3.123), (3.124) when $f = \frac{3}{4}$.

following the x-nullcline very closely, and will continue until it reaches point c. It cannot continue following this segment of the x-nullcline because the direction field points upward. So, it leaves the nullcline and moves leftward until it again intersects the x-nullcline, which is point d. Once there it follows the nullcline down to point e. It cannot follow this segment of the x-nullcline as the direction field points downward, and so it will move quickly over to the x-nullcline near point b. Once there the whole process repeats, and what results is a limit cycle. To reinforce this conclusion, the numerical solution of the problem is given in Figure 3.20.

One of the more apparent differences between the phase planes in Figures 3.17 and 3.19 is where the two nullclines intersect. This is important as this location determines the stability of the steady-state. If the value of f is such that the nullclines intersect between the points e and c, as in Figure 3.19, then the resulting steady-state is unstable. To determine the values for f, note that the two nullclines intersect, at a nonzero value, when the following

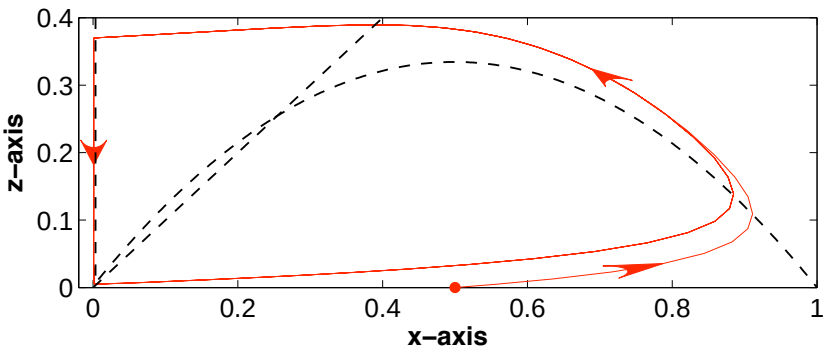


Figure 3.20 Numerical solution of (3.123), (3.124) in the case of $f = \frac{3}{4}$. The initial condition is $(x, z) = (0.5, 0)$.

holds

$$-(1-x)\frac{\alpha+x}{(\alpha-x)f}=1. \tag{3.129}$$

As shown in Exercise 3.19, for small α, the point e corresponds to $x \sim (1+\sqrt{2})\alpha$. Consequently the two nullclines intersect at the point e when $f \sim 1+\sqrt{2}$. Using a similar analysis one finds that they intersect at point c when $f \sim \frac{1}{2}$. Therefore, our conclusion is that the nonzero steady-state is unstable if $\frac{1}{2} < f < 1+\sqrt{2}$, and for these values of f the solution forms a limit cycle similar to the one shown in Figure 3.20. For the other positive values of f the steady-state is asymptotically stable.

Exercises

3.1. This problem considers various aspects of chemical reactions.

(a) What is the simplest reaction that has rate $r = kAB^3$, and conservation law $A - 3B - 4C = constant$?

(b) What is the simplest reaction that has rate $r = kAB^3$, and conservation law $3A + B = constant$?

(c) Suppose that the rate constants for two different reactions have the same dimensions. What relationship must hold between the stoichiometric coefficients of the two reactions?

3.2. The equations below come from applying the Law of Mass Action to two reactions.

$$\begin{aligned} X' &= aX + bYZ, \\ Y' &= cX + bYZ, \\ Z' &= cX + dYZ. \end{aligned}$$

(a) Find the two reactions and determine how the coefficients a, b, c, d are related to each other, if at all. Assume a, b, c, d are nonzero, but they can be positive or negative.

(b) Find the conservation law(s) for these reactions.

3.3. The equations below come from applying the Law of Mass Action to two reactions.

$$\begin{aligned} X' &= aXY, \\ Y' &= bYZ + cZ, \\ Z' &= dYZ + eZ. \end{aligned}$$

(a) Find the two reactions and determine how the coefficients a, b, c, d are related to each other, if at all. Assume a, b, c, d, e are nonzero, but they can be positive or negative.

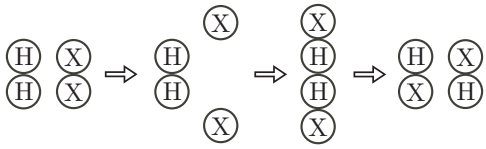


Figure 3.21 Figure for Exercise 3.4.

(b) Find the conservation law(s) for these reactions.

3.4. A well-studied family of reactions involves molecular hydrogen H_2 combining with a diatomic halogen X_2 to produce two molecules of HX. For example, one can have X be fluorine (F) or iodine (I). A sequence of steps that lead to the combination of H_2 with X_2 to produce two HX's is shown in Figure 3.21.

(a) Write down three reactions corresponding to these steps. Assume the steps are irreversible.
(b) Derive the rate equations for the five species involved in this sequence of steps.
(c) Find two independent conservation laws using your equations from part (b). Provide a reason why the laws are independent.
(d) Using reaction analysis and your answer from part (a), determine the steady-state(s). Make sure to explain your reasoning. Also, assume that only H_2 and X_2 have nonzero concentrations at the start.

3.5. Some enzymes work by sequentially binding molecules, and an example is shown in Figure 3.22. The idea here is that the enzyme E has a location that has a high affinity for binding X, and the resulting molecule then binds Y. The last step is the dissociation of the new product molecule YX from E.

(a) Write down three reactions corresponding to these steps. Assume the steps are irreversible.
(b) Derive the rate equations for the six species involved in this sequence of steps.
(c) Find three independent conservation laws using your equations from part (b). Provide a reason why the laws are independent.
(d) Using your reactions in part (a) determine the steady-state(s). Assume that only X, Y, and E have nonzero concentrations at the start. Make sure to explain your reasoning.

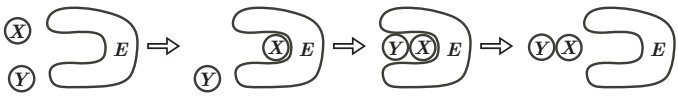


Figure 3.22 Figure for Exercise 3.5.

3.6. The conclusion in Section 3.6.2 concerning the steady-state does not appear to agree exactly with the conclusions in Section 3.6.1. Explain what is missing, or assumed, in Section 3.6.2 that is the cause of this disagreement.

3.7. The surface of a solid can act as a catalyst for certain gases. In the Eley-Rideal mechanism it is assumed that a species A in the gas attaches to a site on the surface, forming a complex C on the surface. The assumed reaction is $A + S \rightleftharpoons C$. Another species B in the gas then reacts with a surface complex to release a product P into the gas. The reaction is $B + C \to P$.

(a) Write down the kinetic equation for each species.
(b) Find three independent conservation laws and reduce the kinetic equations down to two, for S and C.

3.8. One method to reduce the spread of a disease is to vaccinate those who are susceptible. A model that attempts to account for the vaccination of the susceptible group is

$$\begin{aligned} S + I &\to 2I, \\ I &\to R, \\ S &\to R, \\ R &\to S. \end{aligned}$$

Here S, I, and R are the same groups used in the SIR model described in Section 3.1.3. Assume in the problem that $S(0) = S_0, I(0) = I_0, R(0) = 0$, where S_0 and I_0 are positive.

(a) For each reaction give a sentence or two that explains what assumptions were made to produce the given reaction. Make sure to identify the reaction(s) used to account for vaccinations.
(b) Using the Law of Mass Action, write down the initial value problem that comes from the above reactions. After this use a conservation law to reduce this to a problem for just S and I.
(c) Nondimensionalize the reduced problem in part (b), using $N_0 = S_0 + I_0$ to scale both S and I. Use s and i as the nondimensional dependent variables. The final problem, including the initial conditions, should only contain four nondimensional parameters. Identify which one is the vaccination parameter.
(d) What are the steady-states for the scaled problem?
(e) Explain why the solution must satisfy $0 \le s \le 1$ and $0 \le i \le 1$. What restrictions, if any, do you need to impose on the nondimensional parameters so the steady-states you found in part (d) satisfy these inequalities?
(f) One of the steady-states you found has $i_s = 0$. Under what conditions on the nondimensional parameters is this steady-state asymptotically stable?
(g) One of the steady-states you found has $i_s \neq 0$, what is called an epidemic equilibrium. Under what conditions on the nondimensional parameters is this steady-state asymptotically stable?

(h) With the long-term objective of keeping the number of infected individuals down to a minimum, what conditions, if any, should be imposed on the vaccination rate constant?

3.9. The Rozenzweig-MacArthur predator-prey model is

$$\frac{dS}{dt} = \lambda S - \eta S - \nu S^2 - \frac{\mu SP}{1+\alpha S},$$
$$\frac{dP}{dt} = \frac{\beta SP}{1+\alpha S} - \gamma P.$$

Because of the $1 + \alpha S$ term these equations do not appear to be the direct application of the Law of Mass Action. However, this term is similar to what is produced for Michaelis-Menten after analyzing the initial layer. Derive a reaction scheme that produces the above equations when one of the equations is assumed to be at a quasi-steady-state.

3.10. A generalization of the SIR model described in Section 3.1.3 is

$$\frac{dS}{dt} = -aSI + eR,$$
$$\frac{dI}{dt} = -bI + cSI,$$
$$\frac{dR}{dt} = dI - fR.$$

Assume that $S(0) = S_0, I(0) = I_0, R(0) = 0$, where S_0 and I_0 are positive. Also, assume that the coefficients a, b, c, d, e, f are positive constants.

(a) Write the above equations in reaction form. Explain why the coefficients a, b, c, d, e, f are not independent, and state what conditions they must satisfy.

(b) For each reaction in (a) give a sentence or two that explains what physical assumptions correspond to the given reaction.

3.11. This problem considers the dynamics of measles, and what strategy to use for vaccinations. Using the same three groups as in the SIR model, the model is

$$\frac{dS}{dt} = m(S + I + R) - (\beta I + m)S,$$
$$\frac{dI}{dt} = \beta IS - (m + g)I,$$
$$\frac{dR}{dt} = gI - mR.$$

This is known as the standard SIR model with vital dynamics. Assume that $S(0) = S_0, I(0) = I_0, R(0) = 0$, where S_0 and I_0 are positive.

(a) Show that the above system of equations can be derived from the Law of Mass Action. Do this by finding reactions that give rise to these equations.

Also, assumptions are made about the rate constants to obtain these equations and you should make sure to state what these are. Finally, for each reaction give a sentence or two that explains what assumptions were made about the population to produce the given reaction.

(b) Use a conservation law to reduce the above system to a problem for just S and I.

(c) Nondimensionalize the reduced problem in part (b), using $N_0 = S_0 + I_0$ to scale both S and I. The final problem should only contain three nondimensional parameters, one of which will be in the initial conditions. Also, use s and i as the nondimensional dependent variables. Explain why the solution must satisfy $0 \leq s \leq 1$ and $0 \leq i \leq 1$.

(d) What are the steady-states for the problem in (c)? One of them is obtained only if the parameters satisfy a certain inequality. Write this inequality as $v > 1$ and relate v to the nondimensional parameters in your problem.

(e) One of the steady-states you found has $i_s = 0$. Under what conditions on the parameters is this steady-state asymptotically stable?

(f) One of the steady-states you found has $i_s \neq 0$, what is called an epidemic equilibrium. Under what conditions on the parameters is this steady-state asymptotically stable?

(g) For measles $m = 0.02, \beta = 1800, g = 100$ (Engbert and Drepper [1994]). Show that in this case the epidemic equilibrium is asymptotically stable.

(h) One strategy for eliminating measles is to vaccinate newborns. If p is the proportion that are vaccinated then the S and R equations need to be modified and they become $S' = (1-p)m(S+I+R) - (\beta I + m)S, R' = pm(S+I+R) + gI - mR$. By making the appropriate changes in your earlier calculations determine what p must be so the epidemic equilibrium is unstable. Note you do not need to rewrite everything you did in (b)-(d) to answer this question.

(i) Use the result from (h) to explain why measles is so hard to eliminate.

3.12. In most diseases a person can be sick but not able to infect others. The SIR model can be generalized to account for this by replacing the original I group with E, which is the number of individuals who are sick but not infectious, and I, which is now the number of individuals who are infectious. This is known as the SIER model and the equations are

$$\begin{aligned}
\frac{dS}{dt} &= \mu(S+E+I+R) - (\beta I + \mu)S, \\
\frac{dE}{dt} &= \beta IS - (\mu+\sigma)E, \\
\frac{dI}{dt} &= \sigma E - (\mu+\gamma)I, \\
\frac{dR}{dt} &= \gamma I - \mu R.
\end{aligned}$$

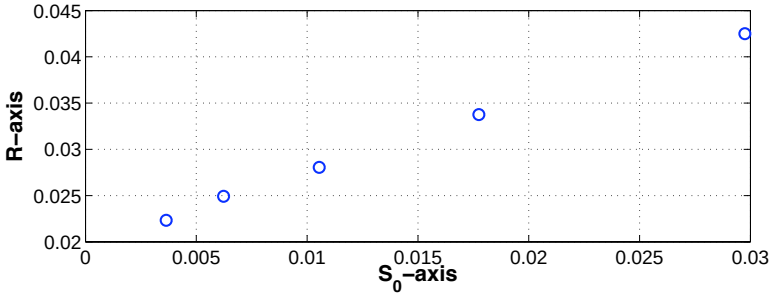


Figure 3.23 Data for the hydrolysis of urea by the enzyme urease (Kryatov et al. [2000]), where $R = -S_0/S'(0)$. In this graph, concentrations are measured in moles and time is in seconds.

Assume that $S(0) = S_0, E(0) = E_0, I(0) = 0, R(0) = 0$, where S_0 and I_0 are positive.

(a) Show that the above system of equations can be derived from the Law of Mass Action. Do this by finding reactions that give rise to these equations. Also, assumptions are made about the rate constants to obtain these equations and you should make sure to state what these are. Finally, for each reaction give a sentence or two that explains what assumptions were made about the population to produce the given reaction.

(b) Use a conservation law to reduce the above system to a problem for S, E, and I.

(c) Nondimensionalize the reduced problem in part (b), using $N_0 = S_0 + E_0$ to scale S, E, and I. The final problem should contain three nondimensional parameters in the differential equations and one nondimensional parameter in the initial conditions. Also, use s, e, and i as the nondimensional dependent variables. Explain why the solution must satisfy $0 \le s \le 1$, $0 \le e \le 1$, and $0 \le i \le 1$.

(d) What are the steady-states for the problem in (c)? One of them is obtained only if the parameters satisfy a certain inequality. Express this inequality in terms of the dimensional parameters of the problem.

(e) One of the steady-states you found has $i = 0$. Under what conditions on the parameters, if any, is this steady-state asymptotically stable?

3.13. Enzymatic reactions are characterized using v_M and K_M, as given in (3.56). For example, values of these constants are standard entries in biochemistry tables, such as Schomburg and Stephan [1997]. This problem examines how they are used in conjunction with experimental data.

(a) Assuming the QSSA is valid, show that

$$\frac{1}{v_M}\left(S_0 + K_M\right) = -\frac{S_0}{S'(0)}\,.$$

The graph of the left-hand side of the above equation, as a function of S_0, produces what is known as a Hanes-Woolf plot. Given the experimental measurement of $S_0/S'(0)$ then linear regression can be used to determine $1/v_M$ and K_M.

(b) One method for determining $S'(0)$ is to measure S at time $t = t_1$ and then use the approximation

$$S'(0) \approx \frac{S(t_1) - S_0}{t_1}.$$

In calculus this is known as a secant line approximation of the derivative at $t = 0$. Ideally t_1 should be small to guarantee the above approximation is accurate. Explain why there is a lower limit on t_1, which limits the accuracy of the approximation. Use the perturbation solution to obtain an approximation for this lower limit.

(c) Use the data in Figure 3.23 to estimate v_M and K_M.

3.14. Although trimolecular reactions are rare in the real world, it is not uncommon to find trimerizations. These are reactions in which a product is constructed using three reactant molecules of the same species, with an effective overall reaction $3A \to \textit{products}$. This exercise explores how to obtain this result using elementary reactions.

(a) One possible mechanism is

$$A + A \rightleftharpoons B,$$
$$A + B \rightleftharpoons C.$$

What are the resulting kinetics equations?

(b) Using the QSSA, and extreme parameter values, show how to reduce the equations in part (a) to obtain the approximate equation $C' = -kA^3$.

3.15. In applying the QSSA to Michaelis-Menten one finds that the product's concentration satisfies an equation of the form

$$\frac{dP}{dt} = \frac{aS}{b + S},$$

where a and b are constants. It has been observed that in some reactions the product appears to follow a rate equation more of the form

$$\frac{dP}{dt} = \frac{aS^n}{b + S^n}.$$

In biochemistry this is known as Hill's equation and n is the Hill coefficient. This exercise explores how to obtain this result using the Law of Mass Action.

(a) One explanation is that n substrate molecules must get together with the enzyme to construct the intermediate complex C. This is the idea underlying cooperativity, and the reactions are

$$nS + E \rightleftharpoons C,$$
$$C \rightleftharpoons P + E.$$

What are the resulting kinetic equations?

(b) Using the QSSA, and extreme parameter values, show how to reduce the equations in part (a) to obtain Hill's equation.

(c) The reactions in (a) are commonly assumed, but they require the unrealistic assumption that $n+1$ molecules collide to form C. A more plausible explanation is they interact sequentially. For $n = 3$ the reactions are

$$S + E \rightleftharpoons C_1,$$
$$S + C_1 \rightleftharpoons C_2,$$
$$S + C_2 \rightleftharpoons P + E.$$

Using the QSSA, and extreme parameter values, show how to reduce the kinetic equations to obtain Hill's equation.

3.16. It is found experimentally that in the hydrogen-bromine reaction, the rate for the overall reaction of producing HBr from H_2 and Br_2 is $r = kH_2Br_2^{1/2}$. The implication is that the reaction is $H_2 + \frac{1}{2}Br_2 \rightarrow HBr$. This exercise explores one of the proposals for how this reaction proceeds as a sequence of elementary reactions. The assumption is

$$Br_2 \rightleftharpoons 2Br,$$
$$Br + H_2 \rightleftharpoons HBr + H,$$
$$Br_2 + H \rightarrow HBr + Br.$$

(a) What are the resulting kinetics equations?

(b) Using the QSSA, and extreme parameter values, show how to reduce the equations in part (a) to $r = kH_2Br_2^{1/2}$.

(c) It is found, using SI units, that $k_1 = 3.8 \times 10^{-8}$, $k_{-1} = 4.2 \times 10^{-13}$, $k_2 = 380$, $k_{-2} = 7.2 \times 10^9$, and $k_3 = 9.6 \times 10^{10}$. Are your assumptions in part (b) consistent with these values?

3.17. For the systems below, find the steady-states and determine if they are asymptotically stable.

(a) $s' = c - s^2$
$c' = 1 + sc$

(b) $u' = v - u$
$v' = (2 - u - v)(1 + v^2)$

(c) $u' = v$
$v' = -\alpha(1 - u^2)v + \beta u,$
where α and β are positive constants.

3.18. The Schnakenberg model for a chemical oscillator consists of the following two rate equations

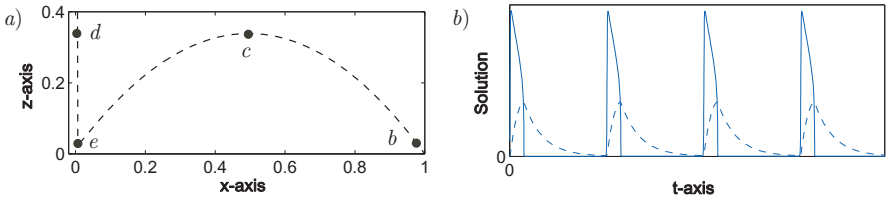


Figure 3.24 Graphs for Exercise 3.19.

$$U' = \mu^* - k_1 U V^2,$$
$$V' = -k_2 V + k_1 U V^2,$$

where μ^* is a positive constant.

(a) The k_i terms come from the Law of Mass Action. Find the two reactions that give rise to these three terms. For the record, the μ^* term accounts for a constant influx of U into the system.

(b) Show that the equations can be nondimensionalized to have the form

$$u' = \mu - uv^2,$$
$$v' = -v + uv^2,$$

where μ is a positive constant.

(c) Using the equations from part (b), find the steady-state and show that it is asymptotically stable if $\mu > 1$ and it is unstable if $\mu < 1$.

(d) Explain why the change in the stability as μ decreases, and passes through $\mu = 1$, has the properties of a Hopf bifurcation as described in Example 3 of Section 3.6.3.

3.19. The z-nullcline for (3.123), (3.124) is shown in Figure 3.24(a), and the solution curves are shown in Figure 3.24(b).

(a) Assuming small α, find first term approximations for the coordinates of the points e and c. Assume that f is independent of α.

(b) As in part (a), find first term approximations for the coordinates of the points b and d. Note that the z-coordinate for d and c are the same, and the z-coordinate for b and e are the same.

(c) Explain why the closer ϵ gets to zero, the more the points b, c, d, and e determine the limit cycle, assuming there is a limit cycle solution.

(d) The solution curves in Figure 3.24(b) are for a small ϵ and α. Identify which is $x(t)$ and which is $z(t)$. Also, locate the points b, c, d, and e in this graph. With this derive approximations for the amplitudes of the two functions.

(e) Explain why the x-coordinate of points c and e does not depend on f. Find a two-term expansion, for small α, of the x coordinate for each of these points. For each point, use (3.129) to find a two-term expansion for f that results in the two nullclines intersecting. Use this to find a two-term expansion for the interval for f that produces a limit cycle solution.

Chapter 4
Diffusion

4.1 Introduction

In the last chapter we examined how to use the kinetics of reactions to model the rate of change of populations, or concentrations. We did not consider the consequences of the motion or spatial transport of these populations. There are multiple mechanisms involved with transport, and in this chapter we will examine one of them, and it is the process of diffusion. A simple example of diffusion arises when a perfume bottle is opened. Assuming the air is still, the perfume molecules move through the air because of molecular diffusion.

One mechanism responsible for diffusion is Brownian motion. Although the random microscopic movements associated with Brownian motion were observed as early as 1785, the first significant scientific study began with Robert Brown. In the summer of 1827 he made microscopic observations of pollen granules suspended in water. What he saw surprised him as the tiny granules were in constant motion, never appearing to slow or stop, and following irregular paths much like the one in Figure 4.1. Moreover, he found that this motion was not caused by external influences such as light or convection currents. He also quickly ruled out his first idea, which was that the granules were somehow alive. However, the underlying reasons for the movement remained elusive. It was not until the early 1900's that the theoretical work of Einstein, and experimental work of Perrin, explained the motion. What is happening is that the pollen granules, which are approximately $6\,\mu$m in length, are under constant bombardment by the surrounding water molecules. Although the latter are much smaller, having a diameter of approximately $3 \times 10^{-4}\,\mu$m, there are many of them and they are responsible for a very large number of random impacts on each granule. The irregular nature of this forcing gives rise to the randomness of the motion. It is now known that Brownian fluctuations are essential to widely diverse phenomena, from passive transport of ions and nutrients for biological cells to models of the stock market (Figure 4.2).

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_4,
© Springer Science+Business Media, LLC 2009

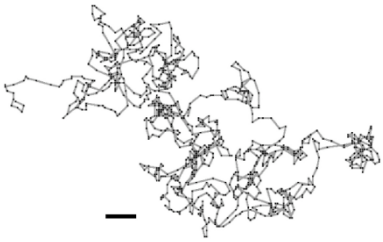


Figure 4.1 The path taken by a micron-sized particle due to Brownian motion over a 2.2 sec interval. The black bar is 10 μm (Blum et al. [2006]).

If one starts with a large number of particles, each undergoing Brownian motion, then over time the particles will tend to be spread throughout the medium. This motion will result in particles in regions of higher concentration to move into regions of lower concentration, and this process is called diffusion. Thus, diffusion is a macroscopic manifestation of the Brownian motion that is taking place on the microscopic level. This observation also identifies our approach in modeling diffusion. We will start with random walks at the microscopic scale and show how they can give rise to diffusion on the macroscopic scale. This differs from the more classic approach, considered later in the chapter, of deriving the diffusion equation using a balance law. The random walk approach provides an explanation of the possible underlying mechanisms in diffusion, and so we will begin with it.

4.2 Random Walks and Brownian Motion

The rapid fluctuating movement of a molecule is the result of impacts with the surrounding atoms and molecules. To construct a mathematical model for this situation we will consider the motion to be one-dimensional. Specifically, the molecules move back and forth along the x-axis. To account for the randomness of the motion consider a single molecule that starts out at $x = 0$. After a time step Δt, the molecule moves a distance Δx either to the right or left and it moves in either direction with equal probability. One way

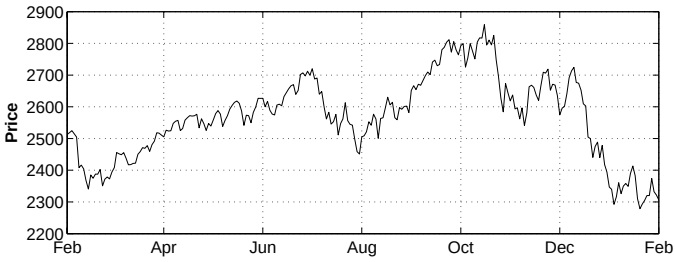


Figure 4.2 The composite index for NASDAQ over a 12-month period.

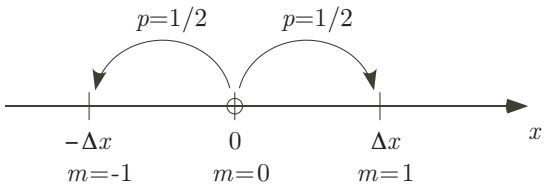


Figure 4.3 The first step in a random walk. In going from time step $N = 0$ to $N = 1$, the molecule moves one unit to the right or left with equal probability.

to think of this is that the molecule has a coin, flips it, and based on whether the outcome is heads or tails it moves left or right. A diagram illustrating the choices, and outcomes, is shown in Figure 4.3. At time step $N = 1$ the coin is again flipped and this determines whether the molecule will step left or right at time level $N = 2$. The various positions that are possible when starting at $(x, t) = (0, 0)$ are shown in Figure 4.4. Also included in this figure is the number of paths that are available to reach the respective point. For example, to reach $m = -1$ at $N = 3$ there are three possible paths. Letting L designate a left move, and R a right move, the three paths are LLR, LRL, and RLL. It is also evident that this number is the sum of the paths for the two adjacent positions, $m = -2$ and $m = 0$, at the previous time step.

Carrying out the above procedure the molecule moves back and forth on the x-axis and the path it follows is called a random walk. One observation concerning Figure 4.3 is that at any given time level not all spatial positions are possible. For example, it is impossible for the molecule to be at $x = 0$ when $N = 1, 3, 5, \ldots$. Also, each step in the random walk is independent of the preceding one. This lack of memory is characteristic of what is known as the Markov property. Finally, note that the number of paths shows more than a passing resemblance to Pascal's triangle. This connection will not be used in what follows because we will be interested in generalizations of this problem that do not have a Pascal's triangle structure.

We want to keep track of the molecule's position and given the way it is determined it should not be unexpected that probabilistic methods are needed. With this in mind, let $w(m, N)$ be the probability that the molecule is at $x = m\Delta x$ after N time steps. The time steps have a fixed value Δt, so, after N steps $t = N\Delta t$. In preparation to calculating w it is worth stating a few of the more interesting properties of this function that are evident from Figure 4.4.

- Given any time level N, the points where w is nonzero are $m = -N, -N + 2, \ldots, N - 2, N$. It is zero at all other values of m.
- The number of paths available to reach $x = m\Delta x$, at time step N, is equal to the number of available paths to reach $m - 1$, at time step $N - 1$, added to the number to reach $m + 1$, at time step $N - 1$.

This conclusion is a consequence of the observation that to be able to be located at $x = m\Delta x$ at $t = N\Delta x$, it is necessary to be located to either $x = (m+1)\Delta x$ or $x = (m-1)\Delta x$ at time $t = (N-1)\Delta x$.

- Because all paths are equally likely,

$$w(m,N) = \frac{\text{number of paths from } (x,t) = (0,0) \text{ to } (x,t) = (m\Delta x, N\Delta t)}{\text{total number of paths from } t = 0 \text{ to } t = N\Delta t}. \tag{4.1}$$

- The total number of paths from $t = 0$ to $t = N\Delta t$ is 2^N.

  The reason this holds is that a particle has two potential paths to the next time level. Hence, the total number of paths doubles with each time step.

- At each time step N the molecule must, with probability one, be located somewhere along the x-axis. In other words, $\sum_{m=-N}^{N} w(m,N) = 1$.

It is not difficult to write a computer program for a random walk, and to use this to investigate what happens when one uses a rather large number of time steps. Example paths are shown in Figure 4.5. The zig-zag nature of a random walk is clearly seen in all four graphs, and as N increases the paths show a large-scale variation in conjunction with the small-scale jagged motion due to the random changes in direction. These are the same characteristics seen in the NASDAQ curve in Figure 4.2.

Another point to make is that if the experiments in Figure 4.5 were to be redone very different paths would likely appear. For example, when $N = 10,000$ there are $2^{10,000}$ different possible paths so it is very unlikely one would reproduce the curve shown in Figure 4.5. On the other hand, suppose we ran a large number of random walks, each one starting at the same location. This has been done in Figure 4.6 where 5000 different walks have

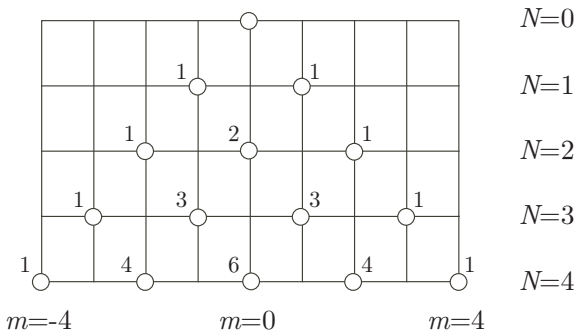


Figure 4.4 The positions that are possible in the random walk are indicated by the circles. In this diagram the molecule starts at $m = 0, N = 0$. The numbers next to each circle are the number of unique paths that are available to reach that location, when starting at $m = 0$ and $N = 0$. Also, N indicates the number of time steps and m the spatial grid location.

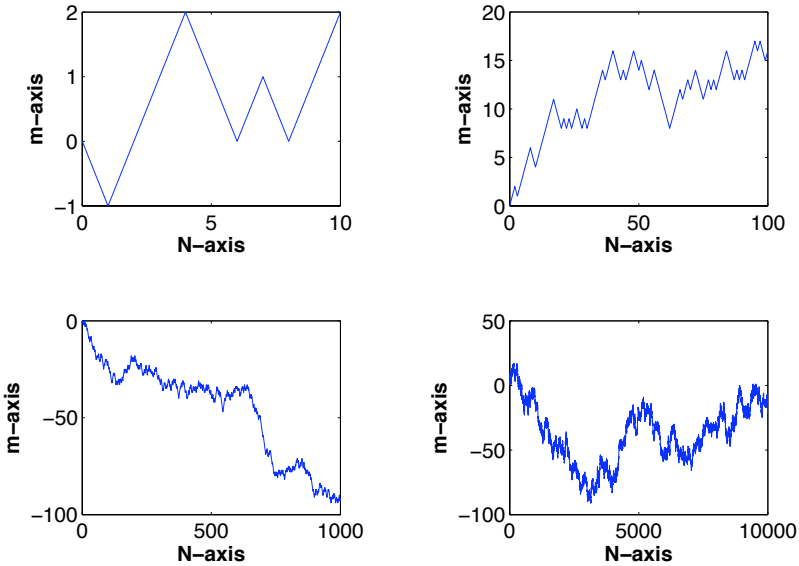


Figure 4.5 Examples of a random walk over successively longer time intervals.

been undertaken, all starting at $m = 0$. The paths, taken together, show a distinctive spatial distribution. At the last time step, $N = 100$, the paths have come together to form what looks like a bell curve (also known as a normal or Gaussian distribution). This shows that an individual path will likely be in the vicinity of $m = 0$. For example, approximately 400 paths have ended up in the histogram bin containing $m = 0$. However, there are paths that manage to be rather far from the center although none of them have reached the maximum possible distance of $m = \pm 100$. This is not really expected using only 5000 paths because the probability of being at the extreme right, or left, is 2^{-99}. Anyway, the function describing the distribution, or concentration, seen in this figure is of particular interest and below we derive a mathematical formula for it.

4.2.1 Calculating $w(m, N)$

As stated in (4.1), because all paths are equally likely, the probability of being located at $x = m\Delta x$ at time step N is equal to the number of unique paths available to reach this point divided by the total number of paths available to reach time level N. For example, in Figure 4.4, there are 6 paths that are able to reach $x = 0$ at time step $N = 4$. Because the total number of paths is 2^4

it follows that $w(0,4) = 6/2^4 = \frac{3}{8}$. We will use this observation to determine the general formula for $w(m,N)$.

To reach any point after N time steps requires a sequence of left and right spatial steps. If q is the number of left steps then the number of right steps is $N-q$. One way to think of this is that you have q of the L's, along with $N-q$ of the R's, and these are going to be arranged in an N-vector. Picking one of the $L's$, there are N positions it can be placed in the vector, while for the second L there are $N-1$ positions, etc. The result is that the total number of choices we can make is $N(N-1)(N-2)\cdots(N-q+1)$. For example, in regard to Figure 4.4, reaching $m=0$ at $N=4$ requires two L's and two R's. The unique orderings we can have are $LLRR, LRLR, LRRL, RLLR, RLLR, RRLL$. However, because $N=4$ and $q=2$ our formula states that the number should be $4\times 3 = 12$. The reason for the discrepancy is because we have considered the L's as distinct from each other. Thinking this way we would conclude that two L's, say L_1 and L_2, could produce two different paths, L_1L_2 and L_2L_1. They do not and therefore we must divide by 2!, or in the general case by q!. Because there are 2^N paths in total it therefore follows that

$$\begin{aligned} w(m,N) &= \frac{N(N-1)(N-2)\cdots(N-q+1)}{2^N q!} \\ &= \frac{N!}{2^N q!(N-q)!} \qquad \text{for } m = -N, -N+2, \ldots, N-2, N. \end{aligned} \tag{4.2}$$

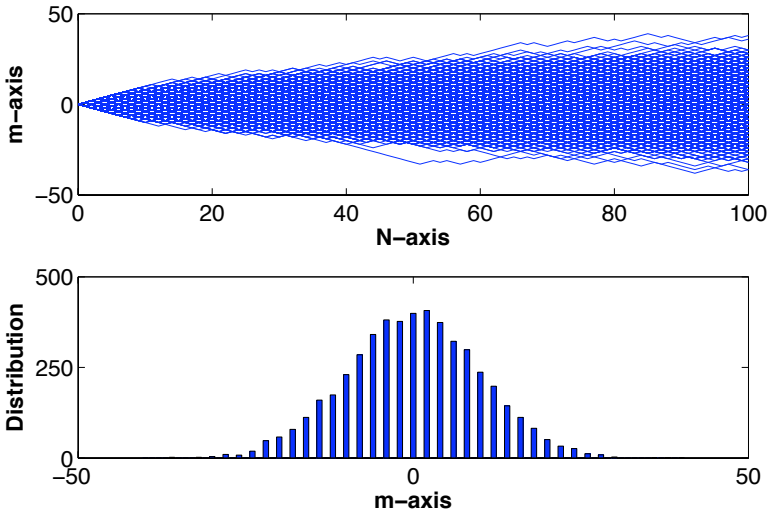


Figure 4.6 The outcome of $5,000$ random walks, all starting at $x=0$, over 100 time steps. The lower graph shows the spatial distribution of the particles at the last time step $N=100$.

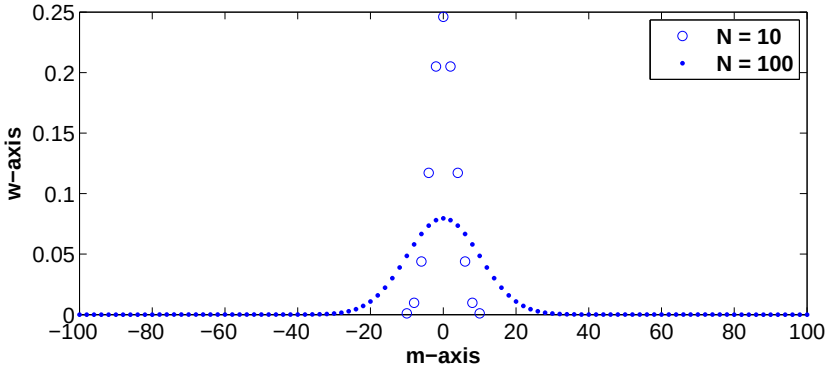


Figure 4.7 The nonzero values of $w(m, N)$, as given in (4.2), as a function of m.

To relate the value of q with the spatial position note that if there are q moves to the left, and $N-q$ moves to the right, then $x = -q\Delta x+(N-q)\Delta x$. Because we also have that $x = m\Delta x$ it follows that $m = N - 2q$, or equivalently,

$$q = \frac{1}{2}(N - m). \tag{4.3}$$

We have done what we set out to do, which is to produce a formula for $w(m, N)$. It is not obvious what a plot of w would look like although we can anticipate some of the major features. Because it is equally likely to move left or right the plot of w, as a function of m, should be symmetric about $m = 0$. For the same reason, it is expected that w decreases with distance from $m = 0$. To support these observations, w is plotted in Figure 4.7 for two values of N. It shows the expected behavior but also notice the spreading of the peak as N increases and the corresponding drop in the maximum value at $m = 0$. This is very typical for a solution that is describing a process controlled by diffusion. It is also not a coincidence that the curves in Figure 4.7 have the same structure as the distribution curve in Figure 4.6. We will return to this observation when the point source solution is discussed later in the chapter.

Random walks are used in a wide variety of applications and because of this the terminology varies a bit with the area. For example, they are used in gas dynamics to describe the motion of atoms in a gas. Such atoms do not travel in a straight line, but rather undergo random changes of direction due to frequent collisions with other atoms. This is modeled as a random walk where the spatial jump Δx is called the mean free path and it is a measure of the distance that an atom travels between two successive collisions. As an estimate of this distance, at room temperature the mean free path in air is about 10^{-7} m and a typical molecule undergoes up to 10^9 collisions per second. This small spatial scale, and the enormous number of collisions, are the basis for the continuous limit considered later.

4.2.2 Large N Approximation

Given the regular structure of the function in Figure 4.7 it would be worthwhile to see if we can simplify our formula to make it a bit easier to work with. The assumption we need to pull this off is that N is large. If this is the case we can make use of Stirling's approximation for the factorial, which is

$$n! \sim e^{-n} n^n \sqrt{2\pi n} \left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right). \tag{4.4}$$

It will be assumed that not only is N large, but the number of left moves and the number of right moves are also large. Using the first term in Stirling's approximation for each of the factorials in (4.2) we get

$$w(m, N) \sim \frac{N^N}{2^N q^q (N-q)^{N-q}} \sqrt{\frac{N}{2\pi q(N-q)}}.$$

Recalling that $q = (N-m)/2$ and $N - q = (N+m)/2$, then the above approximation can be written as

$$w(m, N) \sim \sqrt{\frac{2N}{\pi(N+m)(N-m)}}\, Q, \tag{4.5}$$

where

$$Q = \left(\frac{N}{N+m}\right)^{(N+m)/2} \left(\frac{N}{N-m}\right)^{(N-m)/2}. \tag{4.6}$$

We can simplify Q using the assumption that N is large, or more specifically that m/N is small. Both factors in (4.6), for large N, fall into the category of an indeterminate form of type 1^∞. In calculus the method used to analyze such expressions involves taking the natural log of the function. Doing this, and using the Taylor expansion of $\ln(1+x)$ given in Table 2.1, we obtain the following

$$\begin{aligned}
\ln(Q) &= \frac{N+m}{2} \ln\left(\frac{N}{N+m}\right) + \frac{N-m}{2} \ln\left(\frac{N}{N-m}\right) \\
&= -\frac{N+m}{2} \ln\left(1 + \frac{m}{N}\right) - \frac{N-m}{2} \ln\left(1 - \frac{m}{N}\right) \\
&\sim -\frac{N+m}{2} \left[\frac{m}{N} - \frac{1}{2}\left(\frac{m}{N}\right)^2 + \dots\right] - \frac{N-m}{2} \left[-\frac{m}{N} - \frac{1}{2}\left(\frac{m}{N}\right)^2 + \dots\right] \\
&= -\frac{1}{2}\frac{m^2}{N} + \dots.
\end{aligned} \tag{4.7}$$

With this we conclude that, for small m/N, the nonzero values of w can be approximated as

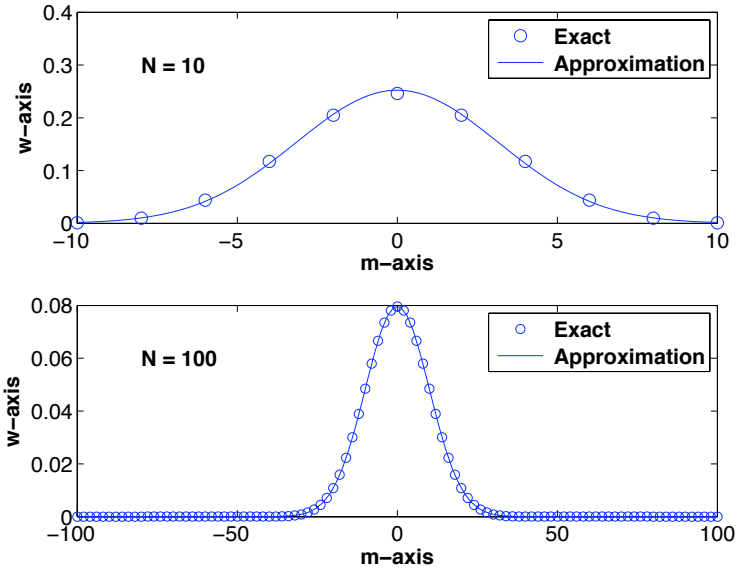


Figure 4.8 Comparison between the exact nonzero values of $w(m, N)$ calculated using (4.2), and the approximation in (4.8).

$$w(m, N) \sim \sqrt{\frac{2}{\pi N}}\, e^{-m^2/(2N)}. \tag{4.8}$$

This expression is significantly simpler than (4.2). To examine the accuracy of this approximation, in Figure 4.8 this function is plotted along with the exact values for two different values of N. It is seen that even in the case of $N = 10$ the approximation is rather good, and it improves significantly when N is larger. It is also evident that (4.8) provides a reasonable approximation on the far left and right, regions where m/N is not particularly small.

4.3 Continuous Limit

As formulated in the last section, a random walk involves discrete steps in space and time. We are now going to investigate the situation when the number of time steps becomes so large that the process is effectively a continuous function of time. As we do this it will be necessary to adjust the spatial stepsize Δx, but we will wait and let the analysis tell us just how to do this. To set the stage, we fix the time interval, and so, it is assumed $0 \le t \le T$. Of interest is what happens to the random walk solution as we increase the number of time steps from $t = 0$ to $t = T$. One way to think about this is

that the time steps become so small that the motion takes on the appearance of a continuous function, one that is a continuous function of time and not one that is making discrete jumps.

The starting point is Figure 4.4. As pointed out earlier, for this grid the number of paths available to reach $x = m\Delta x$ at time step N is equal to the number of the paths to reach $m-1$ added to the number to reach $m+1$ at time step $N-1$. Writing this as

$$\text{paths for } (m,N) = \text{paths for } (m-1,N-1) \ + \ \text{paths for } (m+1,N-1),$$

then we have that

$$\begin{aligned}\frac{\text{paths for } (m,N)}{2^N} &= \frac{\text{paths for } (m-1,N-1)}{2^N} + \frac{\text{paths for } (m+1,N-1)}{2^N} \\ &= \frac{1}{2}\frac{\text{paths for } (m-1,N-1)}{2^{N-1}} + \frac{1}{2}\frac{\text{paths for } (m+1,N-1)}{2^{N-1}}.\end{aligned}$$

Using the function $w(m,N)$, the above equation takes the form

$$w(m,N) = \frac{1}{2}w(m-1,N-1) + \frac{1}{2}w(m+1,N-1). \tag{4.9}$$

This important result gives us a formula for the probability function, and it is the basis for what is called a *master equation* for a stochastic process.

To switch from m,N to x,t recall that $x = m\Delta x$. Also, if we are using N time steps to reach $t = T$ then the size of each step is $\Delta t = T/N$. By introducing the function $u(x,t) = w(m,N)$, then (4.9) can be written as

$$2u(x,t) = u(x-\Delta x, t-\Delta t) + u(x+\Delta x, t-\Delta t). \tag{4.10}$$

It remains to use Taylor's theorem for small $\Delta x, \Delta t$. Expanding (4.10) up to the second-order yields the following

$$\begin{aligned}2u &= u - \Delta x\, u_x - \Delta t\, u_t + \frac{1}{2}\left(\Delta x^2 u_{xx} + 2\Delta x\Delta t\, u_{xt} + \Delta t^2 u_{tt}\right) + \cdots \\ &\quad + u + \Delta x\, u_x - \Delta t\, u_t + \frac{1}{2}\left(\Delta x^2 u_{xx} - 2\Delta x\Delta t\, u_{xt} + \Delta t^2 u_{tt}\right) + \cdots \\ &= 2u - 2\Delta t\, u_t + \Delta x^2 u_{xx} + \Delta t^2 u_{tt} + \cdots .\end{aligned}$$

In the above expression u and its derivatives are evaluated at (x,t). Rearranging things a bit we obtain

$$u_t - \frac{(\Delta x)^2}{2\Delta t}u_{xx} - \frac{\Delta t}{2}u_{tt} + \cdots = 0. \tag{4.11}$$

The question is, what equation is obtained for small Δx and Δt? As with the Goldilocks story, there are three possibilities and they are based on what happens to the ratio $(\Delta x)^2/\Delta t$ as Δx and Δt approach zero. If $(\Delta x)^2/\Delta t$

becomes unbounded then the first term approximation we obtain from (4.11) is $u_{xx} = 0$. Given that $u \to 0$ as $x \to \pm\infty$ we conclude that $u = 0$. For the other extreme, when $(\Delta x)^2/\Delta t \to 0$ as Δx and Δt approach zero, we obtain $u_t = 0$. This equation only applies to the steady-state and is unable to describe the time-dependent changes seen in the solution. The limit that is "just right," what mathematicians call the distinguished limit, is the case of when $(\Delta x)^2/\Delta t$ approaches a positive value as Δx and Δt approach zero. For this reason, we will assume

$$D = \frac{(\Delta x)^2}{2\Delta t} \tag{4.12}$$

remains constant in the limit. In this case, we conclude from (4.11) and (4.12) that

$$u_t = D u_{xx} \,, \tag{4.13}$$

where the constant D is known as the *diffusion coefficient* for the problem. This is the diffusion equation. As derived, $u(x,t)$ is a continuous approximation for the nonzero values of $w(m,N)$. One of the more interesting aspects of this is that it effectively provides a smooth macroscopic description of the random microscopic movements of the molecules.

4.3.1 What Does D Signify?

The only parameter appearing in the diffusion equation is D, and its value signifies the strength or weakness of the underlying diffusion process. From its definition in (4.12), it is seen that the larger the value of D the farther the molecules move in a given time step. In a medium where the molecules are more closely packed, so the random walk steps are not particularly large, the diffusion coefficient is not as big as it would be in a more dilute mixture. It is not surprising therefore that for a gas diffusing in air $D \approx 10^{-5}$ m^2/sec while for a soluble material in water $D \approx 10^{-9}$ m^2/sec.

Mean Free Path
One view of D is provided by the Einstein-Smoluchowski equation used in gas dynamics. In this formulation, the spatial jump Δx is taken to be the average distance λ that the molecule travels between collisions, what is known as the mean free path of the molecule. In conjunction with this, Δt is assumed to be equal to the average time τ between collisions. With this the diffusion coefficient (4.12) is written as

$$D = \frac{\lambda^2}{2\tau} \,. \tag{4.14}$$

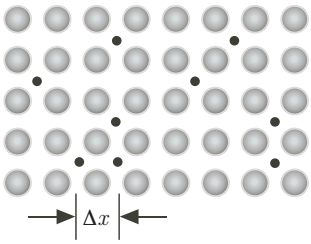


Figure 4.9 Interstitial diffusion in a solid. The atoms of the solid form a lattice, and the smaller interstitial atoms move through the lattice by undergoing a random walk.

This can be used to experimentally determine the value of D. For example, at room temperature, O_2 is found to have a mean free path of 80 nm and an average speed v of approximately 400 m/sec. Assuming $v = \lambda/\tau$ then $D = 2 \times 10^{-5}$ m^2/sec. Perhaps a more interesting observation is that $\tau = \lambda/v = 2\times10^{-10}$ sec, which means a molecule of O_2 undergoes 5×10^9 collisions per second. It should be pointed out that this is for one spatial dimension. As will be explained in Section 4.6, the three dimensional version of (4.14) is $D = \lambda^2/(6\tau)$. Therefore, although the precise value of the diffusion coefficient is affected by dimension, the order of magnitude is not.

Diffusion in Fluids

Given that D is a measure of the ability of a molecule to move through the maze created by its neighboring atoms and molecules, it should not be surprising to learn that the larger the molecule the smaller the diffusion coefficient. The formula in (4.14), however, contains no information related to the structure or state of the molecule or its surrounding medium. It is possible to derive such information, and an example is the Stokes-Einstein equation

$$D = \frac{kT}{6\pi r\eta}\,. \tag{4.15}$$

This assumes the molecules are spheres, where T is the temperature in Kelvin, r is the radius of the molecule, η is the dynamic viscosity of the solvent, and k is known as the Boltzmann constant. How it is possible to get fluid viscosity into the formula for the diffusion coefficient will be explained in Section 4.7. The interest in (4.15), at this point, is the realization that the diffusion coefficient does depend on the size of the molecule, and decreases as the molecule's radius increases. It is also interesting what information can be derived from (4.15). For example, Einstein was able to use (4.15) to calculate Avogadro's number N_A, which is the number of molecules in a mole. From the kinetic theory for gases he knew that $k = R/N_A$, where R is the universal gas constant, and from this he rewrote (4.15) as $N_A = RT/(6\pi r\eta D)$. Using independent measurements of the constants on the right-hand side of this equation Einstein obtained a simple method for finding N_A.

Diffusion in Solids
Diffusion also occurs in solids, although the process is fundamentally different from what occurs in gases and liquids. One mechanism, known as interstitial diffusion, is illustrated in Figure 4.9. Solids have a well-defined atomic structure and in metals the atoms generally form a lattice pattern. Smaller atoms are able to move through the solid by jumping between adjacent interstitial spaces. This requires the adjacent space to be unoccupied, and so this form of diffusion applies to dilute concentrations of diffusing atoms. Also, the diffusing atoms must be small enough to be able to make the jumps. For example, hydrogen, oxygen, nitrogen, and carbon are able to diffuse interstitially through metals, such as iron. However, the lattice points are relatively close so even small interstitial atoms must push their way through to the adjacent opening. This requires them to have sufficient energy to be able to squeeze through. It is possible to account for this in the diffusion coefficient by noticing that $D = p\Delta x^2/\Delta t$, where p is the probability of a jump (a more rigorous explanation of this can be found in Exercise 4.2). It is known that the probability of a successful jump depends on the thermal energy, and the higher the temperature the greater the likelihood of a successful jump. Using reaction rate theory it has been found that the specific form is

$$D = D_0 e^{-E/(kT)} \,. \tag{4.16}$$

where E is the activation energy, k is the Boltzmann constant, and T is the absolute temperature. Also, D_0 is the free solution diffusion coefficient, which is the value obtained when $T \to \infty$. This dependence on temperature is the basis for manufacturing hardened metals, where the metal is heated to allow diffusion of carbon through the lattice. For example, heating steel and allowing carbon to diffuse into the metal produces a much stronger surface, a process known as carburization. This is a slow process, as even at 900° C, the diffusion coefficient for carbon is very small, on the order of 10^{-11} m^2/sec. Interstitial diffusion is also used in the operation of fuel cells, such as those in some of the recent hybrid vehicles, and this is still an active research topic.

4.4 Solving the Diffusion Equation

Now that the diffusion equation has been derived the next question to address is how to find the solution. This equation has been studied for almost two centuries, and this has given mathematicians time to find multiple ways to construct the solution. One possibility is to use a similarity variable and an example of this was worked out in Section 1.4. Another option is to use the method of separation of variables. This is one of the first methods presented in introductory differential equations texts (e.g., Boyce and DiPrima [2004], Braun [1993], Haberman [2003]), but it is mostly limited to spatial intervals

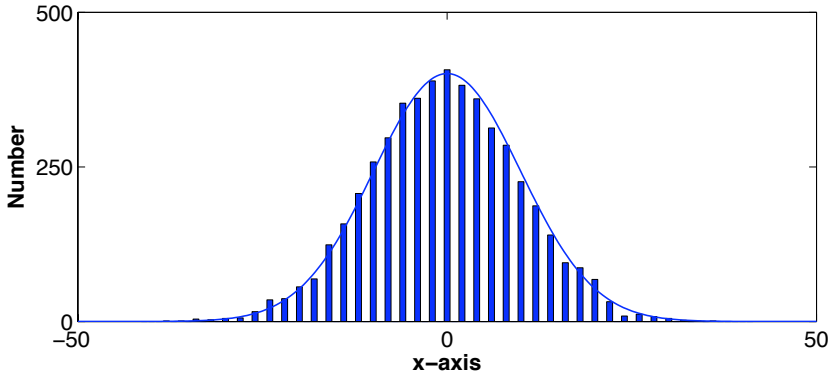


Figure 4.10 The outcome of 5000 random walks, all starting at $x = 0$, at time step $N = 100$. The solid curve is $Pu(x,t)$, where $P = 5000$ and $u(x,t)$ is given in (4.17) for $\Delta x = \Delta t = 1$.

that are bounded. We are interested in an unbounded interval, so we will use other methods.

4.4.1 Point Source

Given that the continuous approximation of the nonzero values for $w(m,N)$ produces the diffusion equation, we will investigate what happens to the large N approximation of $w(m,N)$ given in (4.8). Recalling that $u(x,t) = w(m,N)$, $x = m\Delta x$, and $t = N\Delta t$ then

$$\begin{aligned} u(x,t) &\sim \sqrt{\frac{2\Delta t}{\pi t}} e^{-x^2 \Delta t/(2t\Delta x^2)} \\ &= \frac{\Delta x}{\sqrt{\pi D t}} e^{-x^2/(4Dt)}. \end{aligned} \tag{4.17}$$

It is not hard to verify that this function satisfies the diffusion equation in (4.13). To compare it with the random walk experiment, suppose P random walks are carried out, all starting at $x = 0$. An example of this is shown in Figure 4.6, for $P = 5000$. The probability $w(m,N)$ is determined experimentally by counting the number of paths that go through the point (m,N), and then dividing this by P. Said another way, $Pu(x,t)$ approximates the total number of paths that pass through (m,N), assuming P is large and $w(m,N)$ is nonzero. A confirmation of this is given in Figure 4.10, which shows that $Pu(x,t)$ does indeed provide an excellent approximation for the number of paths.

Each path in the above experiment represents the motion of a molecule. In many applications, it is not the number of paths that are of interest but, rather, the resulting concentration of the molecules. To determine this, recall that the nonzero values of $w(m, N)$ are separated by a distance $2\Delta x$. Consequently, the concentration is $Pw(m, N)/(2\Delta x)$, or equivalently, $Pu(x, t)/(2\Delta x)$. Setting $c(x, t) = u(x, t)/(2\Delta x)$, then the approximation in (4.17) reduces to

$$c(x, t) = \frac{1}{2\sqrt{\pi D t}} e^{-x^2/(4Dt)}. \tag{4.18}$$

This function satisfies the diffusion equation and is known as the point source solution. It gets this name because if P molecules are released at the point $(x, t) = (0, 0)$, their concentration is given approximately as $Pc(x, t)$. The larger the value of P, the better the approximation. Moreover, no matter what value P has, this approximation gives the correct value for the total number of molecules. This is because

$$\int_{-\infty}^{\infty} c(x, t)dx = 1, \tag{4.19}$$

and so the total number of molecules is $\int_{-\infty}^{\infty} Pc(x, t)dx = P$.

Example: Brownian Ratchet

It is possible to take advantage of the Brownian motion of molecules. One application is based on what is known as a Brownian ratchet, where particles are moved in one direction by using a combination of diffusion and externally applied forces. A device designed with this in mind is shown in Figure 4.11, which has been used to separate solutions of DNA (Bader et al. [1999]). The view in this figure is looking down on the device. It consists of an open channel, with positive and negative electrodes placed in pairs along the bottom of the channel. The electrodes are placed close together relative to the distance between the pairs. In Figure 4.11, this corresponds to the statement that $r << \ell$. When a voltage potential is applied, the negatively charged DNA molecules move away from the negative electrodes and collect on the positive

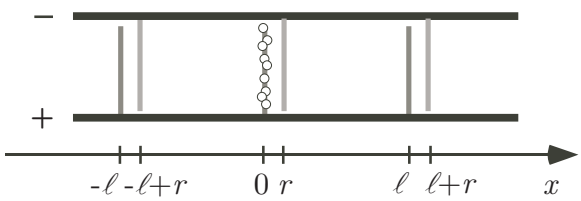


Figure 4.11 Brownian ratchet device. It consists of positive and negative electrodes placed in pairs along the bottom of a channel. At the start, the negatively charged DNA molecules are attached to the positive electrode at $x = 0$.

ones. It is assumed that at the start all of the DNA molecules are on the electrode located at $x = 0$. The objective is to get them to move in the positive x direction. With this in mind, the potential is turned off at $t = 0$. When this happens the molecules start spreading out from $x = 0$, as determined by the diffusion equation. It is assumed that the channel is relatively narrow, and the motion is only in the x-direction. Also, because the electrodes are on the bottom of the channel, the molecules are relatively free to move up and down the channel. If there are a total of P molecules, then from the point source solution we have that the concentration of DNA is given as

$$c(x,t) = \frac{P}{2\sqrt{\pi D t}} e^{-x^2/(4Dt)}. \tag{4.20}$$

After a time interval t_1 the potential is turned back on. When this happens, all of the molecules between the positive electrodes at $x = -\ell + r$ and $x = r$ move back to $x = 0$, while those between the electrodes at $x = r$ and $x = \ell + r$ move to $x = \ell$ (see Figure 4.12). Depending on t_1, it is possible that some of molecules move far enough to the left that when the potential is reapplied that they are attracted to the electrode at $x = -\ell$. To keep this to a minimum we need $c(-\ell + r, t_1)$ to be relatively small, and so it is assumed that t_1 is chosen so that $(-\ell + r)^2 >> 4Dt_1$. This also guarantees that very few molecules get past the electrode at $x = \ell + r$. Therefore, when the potential is reapplied the number of molecules that end up at $x = \ell$ is equal to the area of the shaded region in Figure 4.12. From (4.20), it follows that the number is αP, where

$$\alpha = \frac{1}{2\sqrt{\pi D t_1}} \int_r^\infty e^{-x^2/(4Dt_1)} dx. \tag{4.21}$$

Given that the total number is P then the number that move back to $x = 0$ is $(1-\alpha)P$. Once the molecules stop moving, the process is repeated, and the potential is again removed for a time period t_1. When the potential is turned back on, of the $(1-\alpha)P$ molecules that started out at $x = 0$, $(1-\alpha)^2 P$ of them will return to $x = 0$, while the others will move to $x = \ell$. Those that started at $x = \ell$ will either return to this electrode or else move to the one at $x = 2\ell$. In this way, the off-on cycles use diffusion to move the molecules to the right. In the experiments in Bader et al. [1999], $D = 1.8 \times 10^{-8}$ m^2/sec, $r = 2\,\mu$m, $\ell = 2\,\mu$m, and $t_1 = 1$ sec. For these values, $\alpha \approx 0.216$. This means that approximately 19 off-on cycles are required to be able to have only 1% of the molecules left at $x = 0$, all the rest having moved to one of the electrodes along the positive x-axis. ■

The history of Brownian ratchets is very interesting. A landmark event was Feynman's description in 1963 of a ratchet and pawl device to lift a bug, as shown in Figure 4.13. Box T_2 is filled with a gas and contains an axle with vanes attached. The Brownian collisions of gas molecules on the vanes generate random forces on the vanes. The ratchet on the other end of

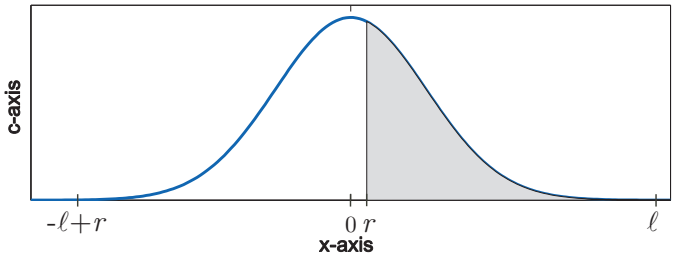


Figure 4.12 The distribution of DNA molecules right before the potential is turned back on at $t = t_1$. When this happens those in the shaded region move to the electrode at $x = \ell$, while those in the unshaded region move to $x = 0$.

the axle, however, only allows rotation in one direction, and this means the wheel rotates in one direction and in the process lifts the bug. This device generated a lot of discussion as it appears to obtain work for free, in other words it seems to be a perpetual motion machine. This is impossible because this would violate the Second Law of Thermodynamics, but exactly how this happens is still generating research papers (e.g., Gomez-Marin and Sancho [2006], den Broeck et al. [2004]). Leaving aside this intriguing bug elevator, variations on a Brownian ratchet have been found in numerous biological processes, and they have been used to develop rather unique technological devices. An introduction to this subject can be found in Hanggi et al. [2005].

4.4.2 Fourier Transform

In the original random walk derivation there were no spatial boundaries, the implication being that $-\infty < x < \infty$. The usual approach in such situations is to try a transform method. There are many to pick from, and we will consider one of the more well known, the Fourier transform. It possesses one of the distinguishing characteristics of most transforms, and that is that it converts

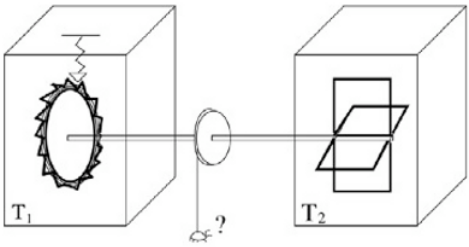


Figure 4.13 Feynman's ratchet and pawl system for lifting bugs (Feynman et al. [2005]).

differentiation into multiplication. Exactly what this comment means will be explained below.

We are interested in an unbounded interval, and the specific problem is

$$u_t = Du_{xx}, \quad \text{for} \quad \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases} \tag{4.22}$$

with the initial condition

$$u(x,0) = f(x). \tag{4.23}$$

It is assumed that $f(x)$ is piecewise continuous with $\lim_{x\to\pm\infty} f(x) = 0$.

To solve the above diffusion problem we introduce the Fourier transform of $u(x,t)$, defined as

$$U(k,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(x,t)e^{-ikx}dx. \tag{4.24}$$

Occasionally it is convenient to express the above integral in operator form and write $U = \mathcal{F}(u)$. The Fourier transform can be inverted and the formula, in the case that u is continuous at x, is

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(k,t)e^{ikx}dk. \tag{4.25}$$

In operator form this is written as $u = \mathcal{F}^{-1}(U)$. It should be noted that if u has a jump discontinuity at x, then the integral in (4.25) does not equal $u(x,t)$, but is equal to the average of the jump in u. Therefore, at a jump discontinuity

$$\frac{1}{2}\left[u(x^+,t) + u(x^-,t)\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(k,t)e^{ikx}dk, \tag{4.26}$$

where $u(x^+,t)$ is the limit from the right, and $u(x^-,t)$ is the limit from the left.

Given the improper integral in (4.24), it is evident that the definition of $\mathcal{F}(u)$ requires u to be reasonably smooth and behave appropriately as $x \to \pm\infty$. For example, the Fourier transform exists if $\int_{-\infty}^{\infty} |u|dx$ is finite and u is piecewise continuous. What is not obvious is how the integral of U in (4.25) produces the original function u. An argument similar to the one originally employed by Fourier is given in Appendix B. A more formal proof can be found in Weinberger [1995]. One observation that comes from the proof is that it is possible to extend the definition of the Fourier transform and include certain functions that do not go to zero as $x \to \pm\infty$. An example is the periodic function $f(x) = \cos(\omega x)$. This requires the introduction of what are known as generalized functions, or distributions. It is not necessary to introduce these for the applications considered here, but those interested in this should consult Friedman [2005].

	$F(k)$	$f(x)$
1.	$F(k)G(k)$	$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(s)g(x-s)ds$
2.	$aF(k)+bG(k)$	$af(x)+bg(x)$
3.	$e^{-iak}F(k)$	$f(x-a)$
4.	$F(k-a)$	$f(x)e^{iax}$
5.	$F(k+a)+F(x-a)$	$2f(x)\cos(ax)$
6.	$F(k+a)-F(x-a)$	$-2if(x)\sin(ax)$
7.	$(ik)^n F(k)$	$\frac{d^n f}{dx^n}$
8.	$\frac{d^n F}{dk^n}$	$(-ix)^n f(x)$
9.	$\frac{1}{a^2+k^2}$	$\frac{1}{a}\sqrt{\frac{\pi}{2}}e^{-a\|x\|}$ for $a>0$
10.	$\frac{k}{a^2+k^2}$	$i\sqrt{\frac{\pi}{2}}e^{-a\|x\|}\left[I_{(0,\infty)}(x)-I_{(-\infty,0)}(x)\right]$ for $a>0$
11.	$\frac{\sin(ak)}{k}$	$\sqrt{\frac{\pi}{2}}I_{(-a,a)}(x)$ for $a>0$
12.	$\frac{1}{\sqrt{a^2-k^2}}I_{(-a,a)}(k)$	$\sqrt{\frac{\pi}{2}}\ J_0(ax)$ for $a>0$
13.	$\frac{1}{a+ik}$	$\sqrt{2\pi}e^{-ax}\ I_{(0,\infty)}(x)$ for $a>0$
14.	$\frac{1}{(a+ik)^{n+1}}$	$\frac{1}{n!}\sqrt{2\pi}x^n e^{-ax}\ I_{(0,\infty)}(x)$ for $a>0$
15.	$\frac{1}{a-ik}$	$\sqrt{2\pi}e^{ax}\ I_{(-\infty,0)}(x)$ for $a>0$
16.	$\frac{1}{(a-ik)^{n+1}}$	$\frac{1}{n!}\sqrt{2\pi}(-x)^n e^{ax}\ I_{(-\infty,0)}(x)$ for $a>0$
17.	$e^{-a\|k\|}$	$\sqrt{\frac{2}{\pi}}\frac{a}{a^2+x^2}$ for $a>0$
18.	$ke^{-a\|k\|}$	$2i\sqrt{\frac{2}{\pi}}\frac{ax}{a^2+x^2}$ for $a>0$
19.	e^{-ak^2-ibk}	$\frac{1}{\sqrt{2a}}e^{-(x-b)^2/(4a)}$ for $a>0$
20.	$\frac{1}{k}(e^{-ibk}-e^{-iak})$	$-i\sqrt{2\pi}I_{(a,b)}(x)$ for $a<b$
21.	$\frac{\sin^2(ak/2)}{k^2}$	$\frac{1}{2}\sqrt{\frac{\pi}{2}}\,(a-\|x\|)\,I_{(-a,a)}(x)$ for $a>0$

Table 4.1 Inverse Fourier transforms. The indicator function $I_{(a,b)}(x)$ is defined in (4.32). The general formulas 2.-8. must be modified at a jump discontinuity, as given in (4.26). Also, the numbers a and b in this table are real-valued.

Examples

1. For the function

$$f(x) = \begin{cases} \alpha & \text{if} \quad a \le x \le b, \\ 0 & \text{otherwise,} \end{cases} \tag{4.27}$$

the Fourier transform is

$$\begin{aligned} F(k) &= \frac{1}{\sqrt{2\pi}} \int_a^b \alpha e^{-ikx} dx \\ &= \frac{1}{\sqrt{2\pi}} \frac{i\alpha}{k} \left(e^{-ikb} - e^{-ika} \right). \end{aligned} \tag{4.28}$$

This result appears as Property 20 in Table 4.1. ■

2. For the function $f(x) = e^{-\alpha|x|}$, where $\alpha > 0$, the Fourier transform is

$$\begin{aligned} F(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx-\alpha|x|} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-(ik-\alpha)x} dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-(ik+\alpha)x} dx \\ &= \sqrt{\frac{2}{\pi}} \frac{\alpha}{\alpha^2 + k^2}. \end{aligned}$$

This result appears as Property 9 in Table 4.1. ■

4.4.2.1 Transformation of Derivatives

The reason the Fourier transform will enable us to solve the diffusion equation is that it converts differentiation into multiplication. To explain what this means we use integration by parts to obtain the following result

$$\begin{aligned} \mathcal{F}(u_x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u_x e^{-ikx} dx \\ &= \frac{1}{\sqrt{2\pi}} \left(u e^{-ikx} \Big|_{x=-\infty}^{\infty} + ik \int_{-\infty}^{\infty} u e^{-ikx} dx \right) \\ &= ik\mathcal{F}(u). \end{aligned} \tag{4.29}$$

It has been assumed here that $u \to 0$ as $x \to \pm\infty$. In a similar fashion, assuming that $u_x \to 0$ as $x \to \pm\infty$, one finds that

$$\mathcal{F}(u_{xx}) = (ik)^2 \mathcal{F}(u). \tag{4.30}$$

The generalization of this to higher derivatives is given in Table 4.1. Therefore, using the Fourier transform, differentiation is transformed into multiplication by ik.

4.4.2.2 Convolution Theorem

A few of the more well-known formulas for the inverse transform are given in Table 4.1. This includes some of its general properties, which are the first eight entries. These are all derivable directly from the definition of the transform and the properties of integrals. For example, the second one, which is known as the convolution theorem, states that

$$\mathcal{F}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(s)g(x-s)ds\right) = F(k)G(k).$$

To prove this, the left-hand side of the above equation is

$$\begin{aligned}
&\mathcal{F}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(s)g(x-s)ds\right)\\
&\qquad = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)g(x-s)e^{-ikx}dsdx\\
&\qquad = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(s)\int_{-\infty}^{\infty} g(x-s)e^{-ikx}dxds\\
&\qquad = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(s)\int_{-\infty}^{\infty} g(z)e^{-ik(z+s)}dzds\\
&\qquad = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(s)e^{-iks}ds\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(z)e^{-ikz}dz\\
&\qquad = F(k)G(k).
\end{aligned}$$

In the above derivation, it is assumed that $f(x)$ and $g(x)$ decay fast enough as $x \to \pm\infty$ that the improper integrals can be interchanged.

Examples

1. Suppose

$$F(k) = \frac{\sin(3k)}{7k} - 2e^{-4|k|}.$$

To determine the original function $f(x)$, we use Property 1, Property 11 with $a = 3$, and Property 17 with $a = 4$. These are used as follows

$$
\begin{aligned}
f(x) &= \mathcal{F}^{-1}\left(\frac{\sin(3k)}{7k} - 2e^{-4|k|}\right) \\
&= \frac{1}{7}\mathcal{F}^{-1}\left(\frac{\sin(3k)}{k}\right) - 2\mathcal{F}^{-1}\left(e^{-4|k|}\right) \\
&= \frac{1}{7}I_{(-3,3)}(x) - \frac{8}{\pi}\frac{1}{16+x^2}\,, \qquad (4.31)
\end{aligned}
$$

where $I_{(a,b)}(x)$ is the indicator function and is defined as

$$
I_{(a,b)}(x) = \begin{cases} 1 & \text{if } a < x < b, \\ \frac{1}{2} & \text{if } x = a, b, \\ 0 & \text{otherwise}\ . \end{cases} \qquad (4.32)
$$

Introducing the definition of I into (4.31), then

$$
f(x) = \begin{cases} -\dfrac{8}{\pi}\dfrac{1}{16+x^2} & \text{if } \ 3 < |x|, \\ \dfrac{1}{7} - \dfrac{8}{\pi}\dfrac{1}{16+x^2} & \text{if } \ -3 < x < 3, \\ \dfrac{1}{14} - \dfrac{8}{\pi}\dfrac{1}{16+x^2} & \text{if } \ x = \pm 3. \quad \blacksquare \end{cases}
$$

2. Suppose

$$
F(k) = \frac{1}{2+ik}e^{-3k^2}.
$$

This transform is not listed in Table 4.1, however, it is a product of two that are listed. Using Property 13 with $a = 2$, the inverse of $1/(2+ik)$ is $\sqrt{2\pi}e^{-2x}I_{(0,\infty)}(x)$. Similarly, using Property 19 with $a = 3$ and $b = 0$, the inverse of e^{-3k^2} is $e^{-x^2/12}/\sqrt{6}$. Therefore, from Property 2 we obtain

$$
\begin{aligned}
f(x) &= \mathcal{F}^{-1}\left(\frac{1}{2+ik}e^{-3k^2}\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sqrt{2\pi}e^{-2s}\,I_{(0,\infty)}(s)\frac{1}{\sqrt{6}}e^{-(x-s)^2/12}ds \\
&= \frac{1}{\sqrt{6}}\int_{0}^{\infty} e^{-2s-(x-s)^2/12}ds. \qquad (4.33)
\end{aligned}
$$

It is not possible to express the integral in terms of elementary functions, so the above expression is the final answer. ■

A comment is in order about the Fourier transform and functions with jump discontinuities. As illustrated in (4.27), the transform of a function with a jump is not an issue. The inverse transform, however, is a different matter. Specifically, the inverse of the transform in (4.28) produces the orig-

inal function $f(x)$ except at the jump points. At those two points the inverse equals the average of the jump. This means that at $x = a$, and at $x = b$, the inverse transform equals $\frac{1}{2}\alpha$. This is why the indicator function $I_{(a,b)}$ in (4.32) is defined the way it is at the jump points $x = a$ and $x = b$.

4.4.2.3 Solving the Diffusion Equation

The Fourier transform will enable us to solve the diffusion equation but this brings up a dilemma common in applied mathematics. To use the transform we need to know if the solution satisfies the conditions needed to guarantee that the improper integral in (4.24) is defined. However, we do not know the solution and are therefore not able to check that the conditions are satisfied. What this means is that we will use the transform in a heuristic manner and assume that the transform can be used. Afterwards, once an answer is derived, it is possible to verify directly that it does indeed satisfy the original problem.

To use the Fourier transform to solve the diffusion equation we first take the transform of the equation and obtain

$$\mathcal{F}(u_t) = \mathcal{F}(Du_{xx}).$$

Because the transform is in x and not t, then $\mathcal{F}(u_t) = \frac{d}{dt}\mathcal{F}(u) = U_t$. With this, and using (4.30), we have that

$$U_t = -Dk^2U. \tag{4.34}$$

We also need to transform the initial condition (4.23), and this gives us

$$U(k, 0) = F(k), \tag{4.35}$$

where $F(k)$ is the Fourier transform of f. Solving (4.34), and using (4.35), yields

$$U(k, t) = F(k)e^{-Dk^2t}. \tag{4.36}$$

We now come to the step of trying to determine $u(x, t)$ given that we know its transform $U(k, t)$. One possibility is to determine this from scratch, which means using the definition of the inverse transform in (4.25) and working out the resulting integrals. The specifics of this are outlined in Exercise 4.10. The more conventional approach is to use a table of inverse Fourier transforms, and simply look up the needed formula. Our transform (4.36) is not listed in Table 4.1. However, the formula for U can be factored as a product $U = FG$, and this will enable us to find the inverse. Setting $G = e^{-Dk^2t}$ then, from Table 4.1,

$$g(x) = \frac{1}{\sqrt{2a}}e^{-x^2/(4a)}, \tag{4.37}$$

where $a = Dt$. With this, and the convolution property, we obtain

$$\begin{aligned} u(x,t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(s) g(x-s) ds \\ &= \frac{1}{2\sqrt{\pi D t}} \int_{-\infty}^{\infty} f(s) e^{-(x-s)^2/(4Dt)} ds. \end{aligned} \tag{4.38}$$

This is the sought after solution of the diffusion problem. What is interesting is that it consists of the integral of the given initial condition multiplied by the point source solution in (4.18). For those who might question some of the steps used to obtain this result, it is a simple matter to show that (4.38) does indeed satisfy the diffusion equation. What is not as straightforward is verifying that (4.38) satisfies the initial condition (4.23). Taking the limit $t \to 0^+$ requires some careful analysis of what happens in the neighborhood of $s = x$ and a proof can be found in Mikhlin [1970].

Example 1

If the initial condition is

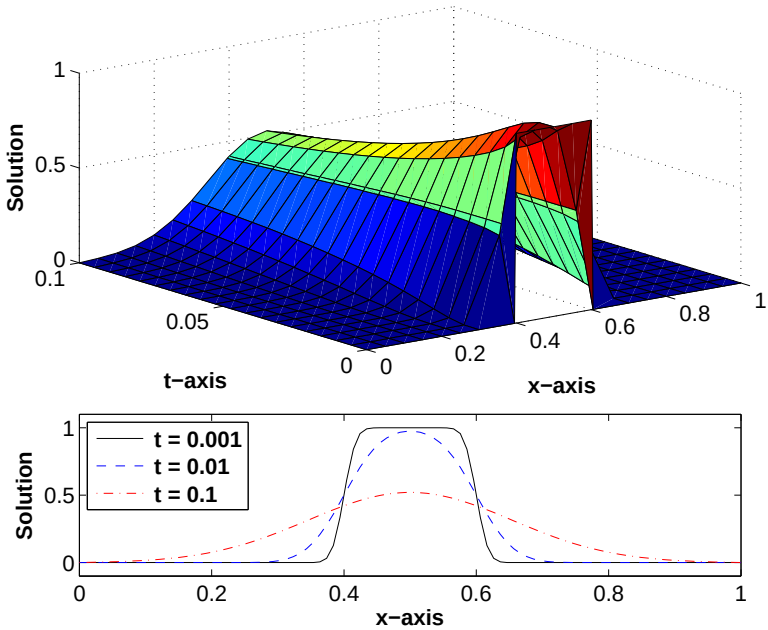


Figure 4.14 Solution (4.40) of the diffusion equation when $f(x)$ is given in (4.39), with $a = 0.4$, $b = 0.6$, and $D = 0.1$. Shown is the solution surface as well as the solution profiles at specific time values.

$$f(x) = \begin{cases} 1 & \text{if } a \le x \le b, \\ 0 & \text{otherwise,} \end{cases} \tag{4.39}$$

then, from (4.38), the solution is

$$u(x,t) = \frac{1}{2\sqrt{\pi D t}} \int_a^b e^{-(x-s)^2/(4Dt)} ds. \tag{4.40}$$

This function is shown in Figure 4.14, both as time slices as well as the solution surface for $0 \le t \le 0.1$. This illustrates several of the characteristic properties of a solution of the diffusion equation. One is that even with an initial condition that contains jumps, the solution for $0 < t$ is smooth. A second observation is that because the exponential function is positive then the solution in (4.40) is never zero for $t > 0$. This means, for example, that even though the solution starts out as zero at $x = 1000$, and the nonzero portion of $f(x)$ is far away, that the solution is nonzero at this position for any value of $t > 0$. This means diffusion occurs with infinite speed, which is physically unrealistic. The usual argument made is that the solution decays rapidly as $x \to \pm\infty$, so the consequence of this is minimal. The fact is that the diffusion equation is a mathematical model, and as such it is an approximation, albeit an effective approximation. Nevertheless, it is interesting that a random walk with steps of finite speed can give rise to macroscopic motion of infinite speed. This paradox has been the subject of numerous studies, two of the more recent being Keller [2004] and Aranovich and Donohue [2007]. ■

The three curves shown in Figure 4.14 are not surprising given the terms appearing in the diffusion equation. Recall from calculus, the curve $y = f(x)$ is concave up if $f'' > 0$, and it is concave down if $f'' < 0$. Now, for a solution of the diffusion equation $Du_{xx} = u_t$ there is a relationship between the concavity of u and the sign of u_t. Specifically, if a solution of the diffusion equation is concave down as a function of x then $u_t < 0$, and the solution decreases. Conversely, the solution increases if u is concave up. This situation

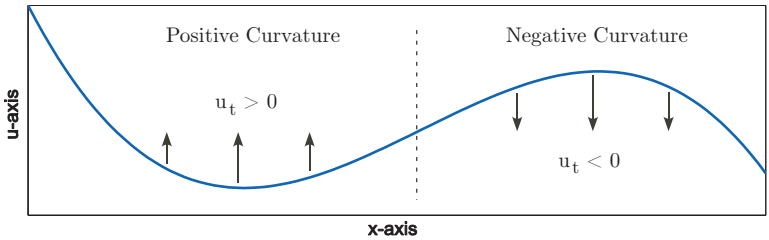


Figure 4.15 The change in the solution of the diffusion equation is determined by the local curvature in the solution. The result is the eventual straightening of the curve.

is illustrated in Figure 4.15. With this observation, given the concave down nature of u in Figure 4.14, the decrease in the solution is expected.

Example 2

Suppose the initial condition is

$$f(x) = \begin{cases} u_1 & \text{if } x \leq 0, \\ u_2 & \text{if } 0 < x. \end{cases} \tag{4.41}$$

It would seem, at first sight, that this is very similar to the previous example. What is wrong with this observation is that the above function does not satisfy the requirement that $\lim_{x\to\pm\infty} f(x) = 0$. Because of this, the Fourier transform of $f(x)$ does not exist, and so the method used to derive the solution (4.38) is not valid for this problem. However, all is not lost. Irrespective of how it was derived, as long as $f(x)$ is well behaved, (4.38) is the solution of the problem. The reason is that (4.38) satisfies the diffusion equation and the given initial condition. This happens, for example, if $f(x)$ is bounded and piecewise continuous. The function in (4.41) satisfies these conditions, and therefore the solution of the resulting diffusion problem is

$$\begin{aligned} u(x,t) &= \frac{u_1}{2\sqrt{\pi D t}} \int_{-\infty}^{0} e^{-\frac{(x-s)^2}{4Dt}} ds + \frac{u_2}{2\sqrt{\pi D t}} \int_{0}^{\infty} e^{-\frac{(x-s)^2}{4Dt}} ds \\ &= \frac{u_1}{2\sqrt{\pi D t}} \int_{0}^{\infty} e^{-\frac{(x+r)^2}{4Dt}} dr + \frac{u_2}{2\sqrt{\pi D t}} \int_{0}^{\infty} e^{-\frac{(x-s)^2}{4Dt}} ds \\ &= \frac{u_1}{\sqrt{\pi}} \int_{\eta}^{\infty} e^{-z^2} dz + \frac{u_2}{\sqrt{\pi}} \int_{-\infty}^{\eta} e^{-w^2} dw, \end{aligned}$$

where $\eta = x/(2\sqrt{Dt})$. This can be written in terms of the complementary error function $\text{erfc}(\eta)$, given in (1.60), as follows

$$u(x,t) = u_2 + \frac{1}{2}(u_1 - u_2)\text{erfc}(\eta). \quad \blacksquare \tag{4.42}$$

The situation occurring in the previous example is not unusual, and it is worth discussing a bit more. To be able to use the Fourier transform, it is necessary to impose rather strict conditions on the functions in the problem. However, the formula that is derived for the solution turns out to be defined for a much broader class of functions than originally assumed. In this case, the formula for the solution becomes the center of attention, and the method that was used to derive the formula is effectively forgotten. This is a very fortunate situation, but the caveat is that care must be taken to make sure that the formula is defined for the functions that are used.

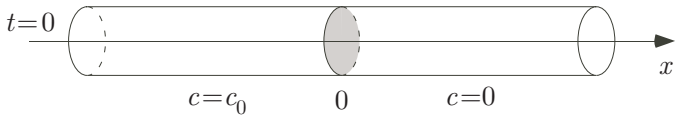


Figure 4.16 At the start of the experiment, the water in the left half of the tube contains salt, with concentration c_0, and the right half contains pure water. The separation between these two regions is removed at $t = 0$, and the movement of salt into the right side is then recorded.

Example 3: Determining D

An experimental method used to study diffusion involves compartments, and a typical example is shown in Figure 4.16. The tube is filled with water, and is separated into two compartments. The water on the left, where $x < 0$, contains salt with a constant concentration c_0. The compartment on the right, where $x \geq 0$, contains pure water and has no salt. At $t = 0$ the divider separating these two compartments is removed, and this allows the salt to move into the region $x \geq 0$. It is assumed that the tube is very long, so the interval can be taken to be $-\infty < x < \infty$. With this, assuming the motion is governed by diffusion then the concentration $c(x, t)$ of salt along the tube satisfies

$$c_t = Dc_{xx}, \quad \text{for} \quad \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

with the initial condition

$$c(x, 0) = \begin{cases} c_0 & \text{if } x < 0, \\ 0 & \text{if } 0 \geq x. \end{cases}$$

From (4.42), the solution of this problem is

$$c(x, t) = \frac{1}{2} c_0 \operatorname{erfc}(\eta), \tag{4.43}$$

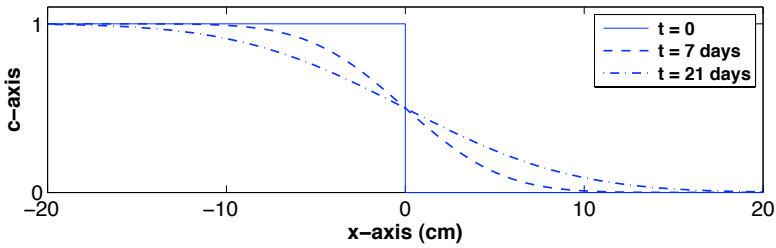


Figure 4.17 The solution (4.43) of the salt diffusion problem at three values of time. In the calculations, $D = 1.5 \times 10^{-9} \mathrm{m}^2/\mathrm{sec}$ and $c_0 = 1$.

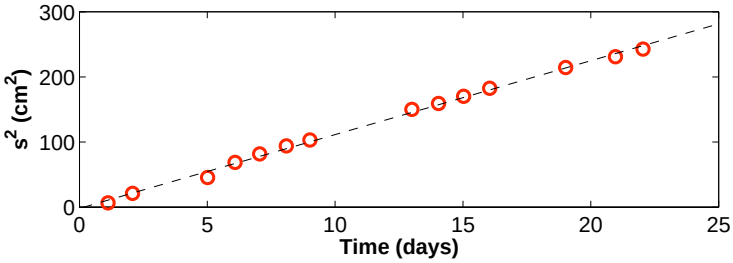


Figure 4.18 The measured values of s^2, as defined in (4.46), for the diffusion of salt in water (Booth et al. [1978]). The dashed line is a least squares fit to the data points.

where $\eta = x/(2\sqrt{Dt})$. This experiment was used by Booth et al. [1978] to investigate the diffusion of salt in water, and they found that $D = 1.5 \times 10^{-9}\text{m}^2/\text{sec}$. Using this value for the diffusion coefficient, the resulting solution (4.43) is shown in Figure 4.17. This shows that the salt does move into the region on the right. However, what is not at all clear is whether the motion is governed by diffusion, or some other transport mechanism. To check on this we need more specific information obtained from the experiment.

The apparatus they used was able to track the position where c has a specified value. For example, if they were interested in where $c = \frac{1}{10}c_0$, then their device could follow the x position where this happened. To use this with our solution in (4.43), let $\bar{c}$ be the specified concentration, and let $\bar{x}$ be the location where c has this value. From (4.43), $2\bar{c} = c_0\text{erfc}(\bar{x}/(2\sqrt{Dt})$. Solving this for $\bar{x}$ yields

$$\bar{x} = 2\alpha\sqrt{Dt}, \tag{4.44}$$

where

$$\alpha = \text{erfc}^{-1}\left(\frac{2\bar{c}}{c_0}\right). \tag{4.45}$$

In their experiments, they followed the positions $\bar{x}_1$ and $\bar{x}_2$ for two concentrations, $\bar{c}_1$ and $\bar{c}_2$, and then calculated the distance $s = \bar{x}_1 - \bar{x}_2$ between these two locations. According to the model, as given in (4.44),

$$s^2 = 4Dt(\alpha_1 - \alpha_2)^2, \tag{4.46}$$

where α_1 and α_2 are the respective values of α for $\bar{c}_1$ and $\bar{c}_2$. Therefore, the model predicts that the distance squared is linear in time. This is a very strong statement, but does this actually happen? Well, their experimental results are shown in Figure 4.18 and evidently it does. This is compelling evidence that the diffusion model applies to this system. Moreover, the model shows that the slope of this line can be used to find D. Just in case you are curious, the value computed using this data is $D = 1.5 \times 10^{-9}$ m^2/sec. ■

Something that is easy to miss in Figure 4.18 is that $s = 15$ cm when $t = 21$ days. In other words, it takes about *three weeks* for the salt to diffuse just 15 cm! Not what you would call a fast mover. Also, the value of the diffusion coefficient is typical for solutes in water. What this means is that diffusion tends to be important over short distances and short time intervals. An indication of this comes from the diffusion coefficient by noting that the value $D = 10^{-9}$ m^2/sec can also be expressed as $D = 1\,\mu$m^2/sec, the implication being that diffusion is significant over distances measured in microns and time measured in milliseconds. This is why diffusion plays such an important role in biological applications related to the function of cells. Movement over larger distances tends to be dominated by convection, which occurs when the fluid flows and in the process carries the molecules with it. The situation is a bit different for diffusion in a gas where the diffusion coefficient is larger, typically by a factor of 10^4. For example, as found in Section 4.3.1, for O_2 in air, $D = 2 \times 10^{-5}$ m^2/sec. This means that diffusion plays an important role over somewhat larger spatial and temporal intervals in a gas. Even so, convection is essential to the movement in a gas, and how to model this transport mechanism will be explored in the next chapters.

4.5 Continuum Formulation of Diffusion

The approach up to this point has been to consider the motion at the micro (or discrete) level and then consider what happens as one passes to the macro (or continuous) level. In this section we will simply start with the continuous description and not concern ourselves with what might, or might not, be happening at the micro level. This is the more conventional method used to derive the diffusion equation.

A good example illustrating the approach is the one used in the original development of the subject by Fick [1885]. He noticed than when salt is poured into water the concentration of salt slowly spreads out and eventually becomes uniformly distributed in the water. To obtain an equation for the concentration we will assume the motion is only along the x-axis. With this, let $c(x,t)$ designate the concentration of salt, which in this context has the dimensions of number of particles per unit length.

4.5.1 Balance Law

The equation for c will be derived from a balance law, and to explain how consider a small interval $a \le x \le b$. The number of particles in this interval can change for only two reasons. First, they can move along the x-axis and therefore they can move in or out of the interval. Although we do not know

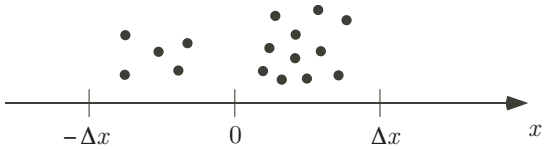


Figure 4.19 The flux at $x = 0$ depends on the difference in the number N_ℓ of particles just to the left of $x = 0$ and the number N_r just to the right of $x = 0$.

exactly how the salt is moving, it does and therefore let $J(x,t)$ designate the net number of salt particles that pass x per unit time. The second way the number of particles in the interval can change is that they are created or destroyed within the interval. This could happen, for example, through a chemical reaction. For this possibility we introduce the function $Q(x,t)$, which gives the number of particles created at x per unit time. With this our balance law has the form

$$\frac{d}{dt}\int_a^b c(x,t)dx = J(a,t) - J(b,t) + \int_a^b Q(x,t)dx. \tag{4.47}$$

In words, this equation states that the rate of change in the total number of salt particles is due to the movement of the particles across the endpoints, this is the $J(a,t)-J(b,t)$ expression, and to the creation or destruction of the particles within the interval. Using the Fundamental Theorem of Calculus, the above integral can be written as

$$\int_a^b \frac{\partial c}{\partial t}dx = -\int_a^b \frac{\partial J}{\partial x}dx + \int_a^b Q(x,t)dx.$$

This can be rewritten as

$$\int_a^b \left(\frac{\partial c}{\partial t} + \frac{\partial J}{\partial x} - Q\right) dx = 0.$$

This equation holds for any interval. As shown in analysis, if the integral of a continuous function is zero over every interval then it must be that the function is identically zero. Because of this we conclude that

$$\frac{\partial c}{\partial t} = -\frac{\partial J}{\partial x} + Q. \tag{4.48}$$

This is the balance law we were looking for. In its present form, it is very general and what we need to do is determine, or specify, the functions J and Q for the problem we are working on, namely the diffusion of salt in water.

4.5.2 Fick's Law of Diffusion

The assumption used to specify the flux is that is that particles in regions of higher concentration will tend to move toward regions of lower concentration. The situation is shown schematically in Figure 4.19. As shown, there is a small number N_ℓ of particles in the left bin, and a larger number N_r in the bin on the right. According to the rules of a random walk, in a time step, approximately half of those on the right will move into the bin on the left, and approximately half of those on the left will move to the bin on the right. The flux is the net difference over the time interval, and so $J = \frac{1}{2}(N_\ell - N_r)/\Delta t$. To express this using continuum variables, the total number of particles in the interval $a < x < a + \Delta x$ is

$$\begin{aligned} N &= \int_a^{a+\Delta x} c(s,t)dx \\ &\approx \Delta x\, c(a,t). \end{aligned}$$

With this,

$$N_\ell \approx \Delta x\, c(-\Delta x, t),$$

and

$$N_r \approx \Delta x\, c(0,t).$$

Using Taylor's theorem, we have that the flux is

$$\begin{aligned} J &\approx \frac{1}{2\Delta t}\left[\Delta x\, c(-\Delta x, t) - \Delta x\, c(0,t)\right] \\ &\approx \frac{\Delta x}{2\Delta t}\left[c(0,t) - \Delta x \frac{\partial c}{\partial x}(0,t) + \cdots - c(0,t)\right] \\ &\approx -\frac{\Delta x^2}{2\Delta t}\frac{\partial c}{\partial x}(0,t). \end{aligned}$$

The above approximation for the flux at $x = 0$ provides motivation for the assumption made in the continuum formulation. Specifically, it is assumed that the flux is given as

$$J = -D\frac{\partial c}{\partial x}\,, \tag{4.49}$$

where D is a positive constant known as the diffusion coefficient. This is known as Fick's law of diffusion, or when applied to temperature distributions it goes by the name of the Fourier law of heat conduction.

To complete the derivation, it is assumed that the particles are not created or destroyed. In this case $Q = 0$ in (4.48), and the balance equation reduces to the diffusion equation

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}. \tag{4.50}$$

The derivation of this result has required minimal effort because it uses the general balance law along with the constitutive law in (4.49). For this reason, it is favored in most derivations of the diffusion equation.

The formula for the flux given in (4.49) is an example of a constitutive law, and we will come across several of these in the later chapters. Even though we used the random walk to help motivate this assumption, it is important to understand that (4.49) does not assume that the particles are undergoing a random walk. It only assumes that the flux is proportional to the spatial derivative, and what might be going on at the molecular level is not stated.

The more typical method for determining a constitutive law is to measure J experimentally, and then use this information to specify the function. This approach is used multiple times in the chapters to follow, as evidenced by the data given in Figures 5.6, 6.5, 9.2, and 9.3. Although this approach is required when using a continuum model, a data-driven formulation does not explain why the flux depends on the specific variables appearing in the constitutive law. This was the reason for starting this chapter with the random walk analysis, because it illustrates how a microscale formulation can be used to explain macroscale motion. An active area of research addresses this issue for more complex problems, attempting to use quantum or molecular theories to derive the appropriate constitutive law. Those who are interested in this can find an introduction to this area in Balluffi et al. [2005] and Lucas [2007].

Example: Using the Flux to Find D

Fluid saturated soil, a sponge filled with water, and articular cartilage are examples of biphasic materials. They are formed from two constituents, a porous solid that is saturated with fluid. Given a sample of length ℓ, then the motion of such a material is governed by the diffusion equation

$$D\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t}, \quad \text{for} \quad \begin{cases} 0 < x < \ell, \\ 0 < t. \end{cases} \tag{4.51}$$

In the experiments described in Holmes et al. [1983], the sample is held at $x = \ell$ and the flux is prescribed at $x = 0$. The corresponding boundary conditions are

$$u(\ell, t) = 0, \tag{4.52}$$

and

$$D\frac{\partial u}{\partial x}(0, t) = -\alpha. \tag{4.53}$$

Also, the initial condition is

$$u(x, 0) = 0. \tag{4.54}$$

What is measured in the experiments is the value of $u(0, t)$, and a typical result is shown in Figure 4.20. These data are going to be used to address

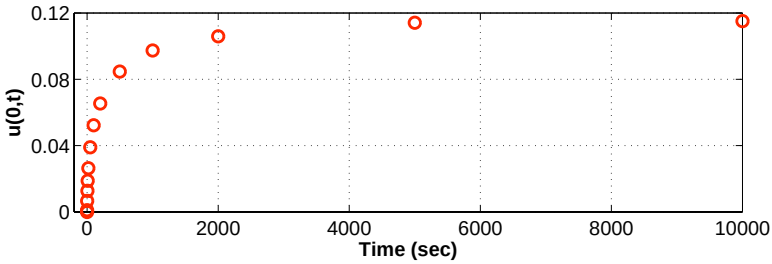


Figure 4.20 The values of $u(0,t)$ measured in response to a prescribed flux. The data are from a test on articular cartilage (Holmes et al. [1983]).

two questions. First, in looking at the values of $u(0,t)$, how do you know that the response is governed by the diffusion equation? Second, assuming that the diffusion equation is correct, can you use this information to determine D? Answering these questions will involve solving the diffusion problem, but we need to be selective in how this is done. For example, it is possible to find the solution using separation of variables. However, this is not a particularly useful method for this example because it is very difficult to develop an intuitive understanding of the solution from the Fourier series. Instead, we will derive approximations of the solution using what we know about the diffusion equation and the data in Figure 4.20.

- *Steady-State* The first approximation relates to the steady-state solution. It is seen in Figure 4.20 that $u(0,t)$ approaches a steady-state as $t \to \infty$. From (4.51), the steady-state satisfies $u_{xx} = 0$ along with the boundary conditions in (4.52) and (4.53). The corresponding solution is

$$u = \frac{\alpha}{D}(\ell - x). \tag{4.55}$$

  With this, the solution at $x = 0$ is $u = \alpha\ell/D$. Given that α and ℓ are known then we can use this equation to find D. What is needed is the steady-state value of u at $x = 0$. It appears from the data in Figure 4.20 that the steady-state has almost been reached at $t = 10{,}000$ sec. Using this approximation, then the diffusion coefficient can be calculated using the formula $D = \alpha\ell/u(0, 10{,}000)$.

The steady-state has provided the answer to the second question. This leaves the issue of how the data in Figure 4.20 can be used to help verify that the diffusion equation is the correct model. For this we consider what happens for small values of time.

- *Small Time Approximation* The solution starts out as $u = 0$, and what is responsible for causing the solution to be nonzero is the flux boundary condition (4.53). How this information moves through the interval is governed by the diffusion equation, and an indication of what happens can be

derived from Figure 4.14. Namely, it takes a certain amount of time for the nonzero part of the solution to move across the interval and appreciably affect what is happening at $x = \ell$. Up until this happens, we can assume that the sample is infinitely long, and replace (4.52) with the condition that

$$u \to 0 \text{ as } x \to \infty. \tag{4.56}$$

It is understood that in this approximation the diffusion equation is being solved, not on a finite spatial interval, but for $0 < x < \infty$. The easiest way to solve the problem in this case is using similarity variables. The only dimensional quantities appearing in this problem, other than u, are x, t, D, and α. Therefore, it follows that $u = F(x, t, D, \alpha)$. To reduce this using dimensional analysis note $[\alpha] = [Du]/L = [u]L/T$. Using an argument very similar to the one given in Section 1.4, it is found that

$$u = \alpha\sqrt{\frac{t}{D}}\, f(\eta), \tag{4.57}$$

where $\eta = x/\sqrt{Dt}$. Substituting this into the diffusion equation, and rearranging things a bit, yields

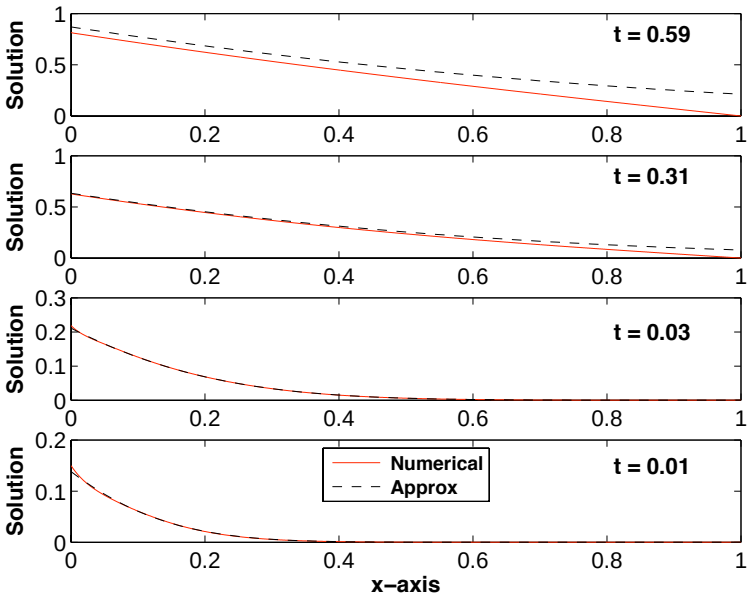


Figure 4.21 Comparison between the numerical solution of the diffusion problem and its approximate solution given in (4.60).

$$f'' + \frac{1}{2}\eta f' - \frac{1}{2}f = 0. \tag{4.58}$$

Staring at this equation for a few moments it is seen that $f = \eta$ is a solution. This enables us to use reduction of order to find the general solution. This is done by assuming $f(\eta) = \eta g(\eta)$, and using (4.58) to find g. The result is that

$$f(\eta) = a\eta + b\left(e^{-\eta^2/4} - \frac{1}{2}\eta\int_\eta^\infty e^{-s^2/4}ds\right), \tag{4.59}$$

where a and b are arbitrary constants. The condition in (4.56), and the initial condition (4.54), require that $f(\infty) = 0$. From this it follows that $a = 0$. From the flux condition (4.53) one finds that $f'(0) = -1$, and this means that $b = 2/\sqrt{\pi}$. The resulting solution is

$$\begin{aligned} u(x,t) &= 2\alpha\sqrt{\frac{t}{\pi D}}\left(e^{-\eta^2/4} - \frac{1}{2}\eta\int_\eta^\infty e^{-s^2/4}ds\right) \\ &= 2\alpha\sqrt{\frac{t}{\pi D}}\left(e^{-\eta^2/4} - \frac{\sqrt{\pi}}{2}\eta\,\mathrm{erfc}(\eta/2)\right). \end{aligned} \tag{4.60}$$

To illustrate the accuracy of this approximation, it is plotted in Figure 4.21 along with the numerical solution of the problem. For this comparison, $\ell = D = \alpha = 1$. As expected, the two solutions are very close until the disturbance effectively reaches the right endpoint. Once that happens the solution of the diffusion problem rapidly approaches the steady-state solution, which in this case is $u = 1 - x$. One of the more important conclusions that comes from (4.60) is that

$$u(0,t) = 2\alpha\sqrt{\frac{t}{\pi D}}. \tag{4.61}$$

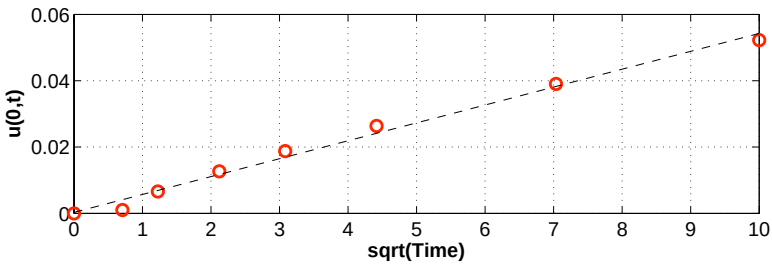


Figure 4.22 The data for $0 \le t \le 100$ from Figure 4.20 plotted as a function of $\sqrt{t}$. The data clearly show the linear dependence predicted in (4.61). The dashed line is a least squares fit to the data points.

This is the result we were looking for because this is what is measured in the experiment. It shows that if the diffusion model is correct then the deformation of the surface must increase as $\sqrt{t}$, at least for small values of t. It is difficult to see if this happens in Figure 4.20, so the data are replotted in Figure 4.22 as a function of $\sqrt{t}$, for $0 \leq t \leq 100$. Although the linear dependence seen in this figure does not prove that the diffusion model is correct, it is very compelling evidence that it is. ■

Example: Drift Diffusion

Up to this point the particle's motion is due exclusively to Brownian motion, but it can also move in response to an external force. To determine how this affects the flux, suppose the external force results in the particles having a velocity v_d, what we will refer to as the drift velocity. The flux in this case can be written as $J = J_{\text{diff}} + J_{\text{drift}}$, where J_{diff} is given in (4.49). To determine J_{drift}, we refer back to Figure 4.19. Assuming v_d is positive, then over a small time interval Δt, the particles that are able to cross $x = 0$ will be within an interval of width $v_d \Delta t$. This number is approximately $c(0,t)v_d\Delta t$, and the resulting flux, which is the number per time interval, is $J_{\text{drift}} = c(0,t)v_d$. This is for $x = 0$, and is applicable for both positive and negative v_d. Generalizing this

$$J_{\text{drift}} = v_d c,$$

and from this we obtain the following constitutive relation for the flux

$$J = v_d c - D\frac{\partial c}{\partial x}\,. \tag{4.62}$$

The corresponding equation of motion is

$$u_t = Dc_{xx} - v_d c_x\,. \tag{4.63}$$

This is an example of a convection-diffusion equation, and it can be solved in a straightforward manner using the Fourier transform (see Exercise 4.17). Rather than doing that, we will work out the steady-state solution, which means we find the function that is independent of t and satisfies $Dc_{xx} - v_d c_x = 0$. Assuming $u = e^{rx}$ one finds from the differential equation that $r = 0, v_d/D$. From this it follows that the general solution of the steady-state equation, for $v_d \neq 0$, is $c = a + be^{v_d x/D}$, where a and b are arbitrary constants. ■

A drift velocity can arise for a variety of reasons, and one example is when the particles are charged and an electrical potential is applied. The electric field will induce the particles to move, and there are various ways to determine the resulting formula for J_{drift}. One approach simply makes the assumption that the velocity is proportional to the electric field. The corresponding constitutive law for the drift velocity is $v_d = \mu E$, where E is the electric field and μ is a constant known as the mobility. It is possible

to obtain this result using a more physically based argument, and this is contained in the next example.

Example: Nernst-Planck Law

Suppose the charged particles are in solution. In this case, the motion of the particles induced by the electric field will be resisted by the viscosity of the fluid. Assuming that the motion is steady then the resulting drift velocity will correspond to when these forces balance. The electric field force is qE, where q is the charge of a particle. To determine the viscous force, it is assumed that the particle is spherical and its velocity is relatively slow. In this case, from (1.22), the viscous drag force is $D_F = 6\pi\eta r v_d$, where r is the radius of the particle and η is the dynamic viscosity of the solvent. Using (4.15) this can be rewritten as $D_F = kTv_d/D$. Equating the electric and viscous force it follows that $v_d = qED/(kT)$. The resulting constitutive law for the flux takes the form

$$J = D\left(-\frac{\partial c}{\partial x} + \frac{qE}{kT}\,c\right), \tag{4.64}$$

which is known as the Nernst-Planck law. The resulting diffusion equation is given in (4.63), where $v_d = qED/(kT)$. The steady-state solution, which produces a zero flux, is

$$c = c_0 e^{xqE/(kT)},$$

where c_0 is a constant. This is an example of what is known as a Boltzmann distribution for the concentration. ■

As a final comment on drift diffusion, although it is routinely arises in applications, there are limitations when (4.63) can be used. An assumption inherent in the derivation of this equation is that the drift velocity is constant, and, therefore, independent of c. Although this is a reasonable assumption at low concentrations and small drift velocities, it is very questionable once the concentrations and velocities increase. This observation is central to the next chapter, where the relationship between the concentration and velocity plays a central role in the analysis.

4.5.3 Reaction-Diffusion Equations

Up to this point we have assumed the particles are not created or destroyed. An example of a situation where this does not happen is when the particles are undergoing chemical reactions. To illustrate what effect this has on the diffusion equation suppose we have two species, A and B, and they are undergoing the reversible reaction

$$A \rightleftharpoons B. \tag{4.65}$$

Assuming, for the moment, that there is no diffusion, then the resulting kinetic equations are obtained using the law of mass action, and the result is

$$\begin{aligned}\frac{dA}{dt} &= -k_1 A + k_{-1} B,\\ \frac{dB}{dt} &= k_1 A - k_{-1} B.\end{aligned}$$

The right hand sides of these two equations are our source terms. Namely, for species A we have that $Q = -k_1 A + k_{-1} B$, and for species B we have $Q = k_1 A - k_{-1} B$. The resulting diffusion equations are

$$\begin{aligned}\frac{\partial A}{\partial t} &= D_a \frac{\partial^2 A}{\partial x^2} - k_1 A + k_{-1} B,\\ \frac{\partial B}{\partial t} &= D_b \frac{\partial^2 B}{\partial x^2} + k_1 A - k_{-1} B.\end{aligned}$$

It has been assumed here that the diffusion coefficients for the two species are different. The above system of equations is an example of reaction-diffusion equations.

Example: Pattern Formation

A well researched question in developmental biology is how cells in an organism arrange themselves to form patterns. One possible mechanism, first described by Turing [1952], is that two or more chemicals diffuse through an embryo and react with each other until a stable pattern of chemical concentrations is reached. A model that has been proposed for the formation of stripes involves the interaction of five chemicals across a substrate, and the equations are (Meinhardt [1982])

$$\begin{aligned}\frac{\partial g_1}{\partial t} &= D_g \frac{\partial^2 g_1}{\partial x^2} + \frac{c s_2 g_1^2}{r} - \alpha g_1 + \rho_0,\\ \frac{\partial g_2}{\partial t} &= D_g \frac{\partial^2 g_2}{\partial x^2} + \frac{c s_1 g_2^2}{r} - \alpha g_2 + \rho_0,\\ \frac{\partial s_1}{\partial t} &= D_s \frac{\partial^2 s_1}{\partial x^2} + \gamma(g_1 - s_1) + \rho_1,\\ \frac{\partial s_2}{\partial t} &= D_s \frac{\partial^2 s_2}{\partial x^2} + \gamma(g_2 - s_2) + \rho_1,\\ \frac{\partial r}{\partial t} &= c s_2 g_1^2 + c s_1 g_2^2 - \beta r.\end{aligned} \tag{4.66}$$

One way to look at this model is that g_1 identifies the cells responsible for producing the color white, while g_2 corresponds to the cells responsible for

Figure 4.23 The stripe patterns formed on a zebra obtained from numerically solving the reaction-diffusion system (4.66) (Turk [1991]).

producing the color black. To be able to function, the g_1 cells need the chemical s_2, while the g_2 cells need s_1. What prevents the two cell lines from occupying the same location is the variable r, which is known as a repressor. Given the complexity of this model, as with most reaction-diffusion systems, numerical methods are needed to find the solution. An interesting example is the zebra shown in Figure 4.23. ■

4.6 Random Walks and Diffusion in Higher Dimensions

It is interesting to consider how to extend random walks to multiple dimensions. The basic idea is that starting at $\mathbf{x}_0$ the first step in the walk produces a new position $\mathbf{x}_1$. Earlier we assumed the length of a step is fixed, and we will do the same here. If the step length is h then the formula connecting $\mathbf{x}_1$ with $\mathbf{x}_0$ is

$$\mathbf{x}_1 = \mathbf{x}_0 + h\mathbf{u}_1.$$

In one dimension it is possible to move only left or right, and in this case $\mathbf{u}_1$ is randomly chosen to be ± 1. In higher dimensions, $\mathbf{u}_1$ is a randomly selected direction, or more precisely, a randomly chosen unit vector. Once we have $\mathbf{x}_1$, the next step $\mathbf{x}_2$ in the walk is calculated in a similar fashion. The only difference is that we randomly select a new direction vector $\mathbf{u}_2$. Generalizing this procedure, the resulting formula for the position at time step n is

$$\mathbf{x}_n = \mathbf{x}_{n-1} + h\mathbf{u}_n, \tag{4.67}$$

where $\mathbf{u}_n$ is a randomly chosen unit vector.

To visualize what happens consider two dimensions, where $\mathbf{x} = (x, y)$. In this case the direction for $\mathbf{x}_n$ can be written in terms of the polar coordinate angle θ. With this, (4.67) becomes

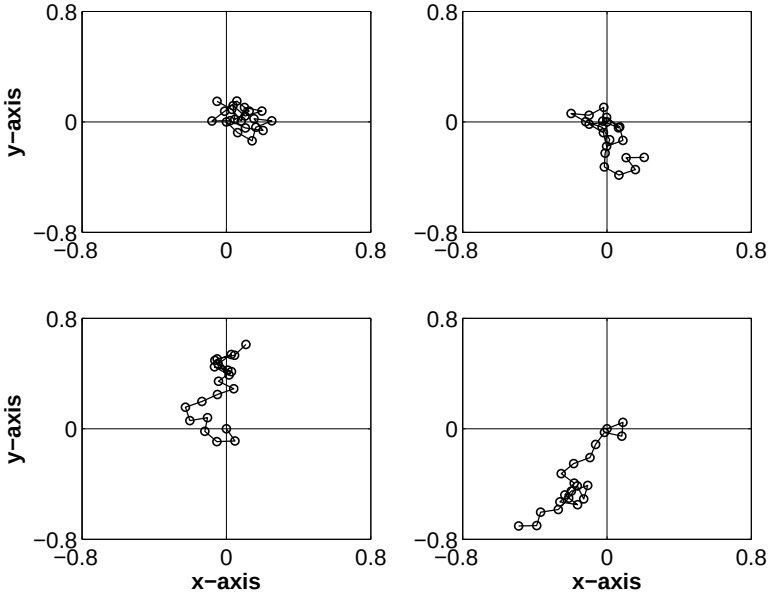


Figure 4.24 Random walks in the plane. Each starts at $x = y = 0$ and has step length $h = 0.1$.

$$x_n = x_{n-1} + h\cos(\theta_n), \tag{4.68}$$

$$y_n = y_{n-1} + h\sin(\theta_n), \tag{4.69}$$

where θ_n is randomly chosen from the interval $[0, 2\pi)$. The first 20 positions calculated using this formula are shown in Figure 4.24 using $h = 0.1$. Four random walks are shown, and not unexpectedly they are quite different from one another. Nevertheless, a few observations can be made. First, none of them has come close to reaching the maximum obtainable distance of $20h = 2$. This is not surprising because to reach the maximum distance the same angle would need to be used at each step, and this is highly unlikely. Second, there is a propensity for the positions to be close to the origin. It is natural to ask if the positions follow the Gaussian distribution found for the one-dimensional random walks shown in Figure 4.6. To determine this experimentally, suppose we take 30,000 random walks, all starting at $x = y = 0$ and look at the distribution of positions at various time levels. The results are shown in Figure 4.25 for $n = 100$, $n = 200$, and $n = 400$. It appears that at the beginning the positions are closer to the origin but as time increases they move outward. These distributions have the rough appearance of an exponential function of the radial distance from the origin, with an amplitude that decreases with n. This is the same dependence we obtained for the one-dimensional formula in

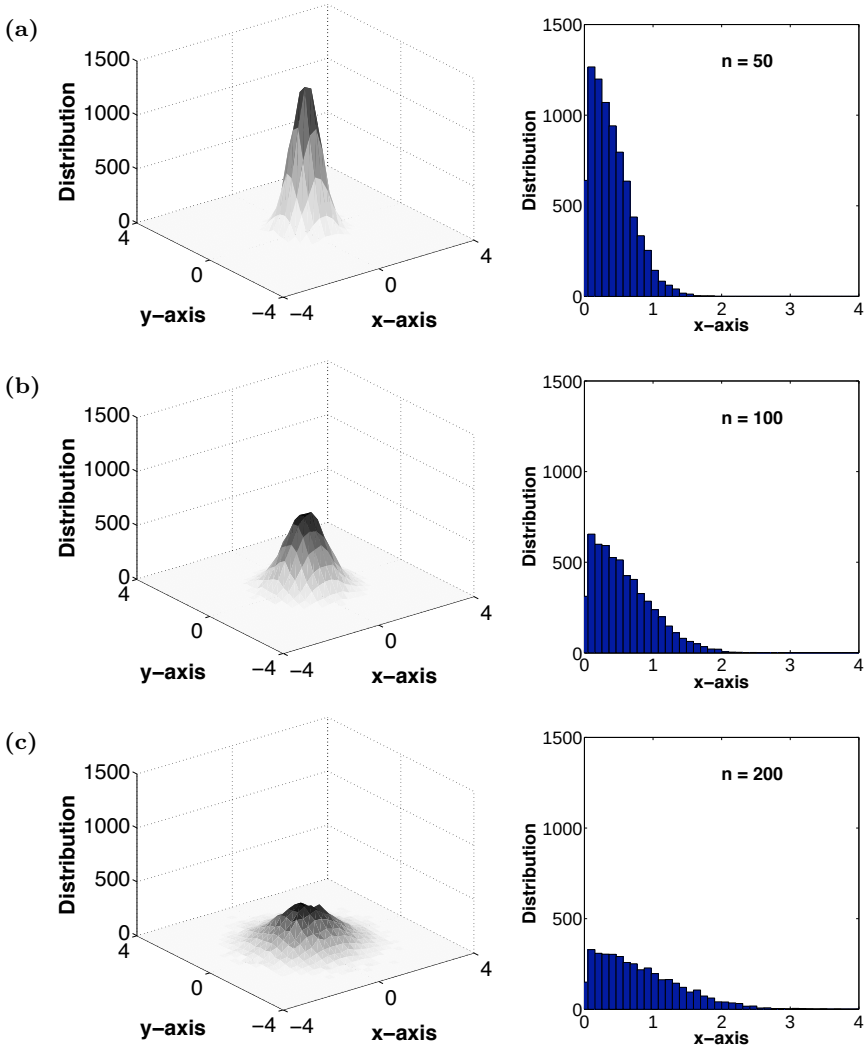


Figure 4.25 Distribution of positions for random walks that start at $x = y = 0$, using (4.68), (4.69). On the left are the distributions in the plane using 30,000 molecules, and on the right is the distribution along the positive x-axis using 100,000 molecules. The time steps are: (a) $n = 50$, (b) $n = 100$, and (c) $n = 200$.

(4.18). The generalization of the point source solution to higher dimensions is

$$\frac{1}{(4\pi Dt)^{d/2}} e^{-r^2/(4Dt)}, \tag{4.70}$$

where d equals the number of spatial dimensions and r is the radial distance to the origin. For the distributions in Figure 4.25, $d = 2$ and this means that the amplitude decreases as $1/t$. This dependence is evident in the distributions, as doubling the value of n results in a reduction in the amplitude by approximately a factor of two. This leaves open the question of how to derive (4.70), and this will be done later, after the diffusion equation has been derived.

4.6.1 Diffusion Equation

It is enough to derive the diffusion equation for two spatial dimensions. Also, we will limit the directions used in the random walk. Specifically, the angle θ_n in (4.68), (4.69) will be randomly chosen as one of the four angles $\{0, \pi/2, \pi, 3\pi/2\}$. What this means is that the positions follow a lattice pattern, and this is illustrated in Figure 4.26. The assumption is that at each time step, the molecule moves, with equal probability, to a nearest neighbor point on the lattice. Note that this is effectively the two dimensional version of interstitial diffusion shown in Figure 4.9.

In a similar manner as in the one-dimensional case, we let $w(m, p, N)$ be the probability the molecule is at $(x, y) = (m\Delta x, p\Delta y)$ after N time steps. Given that we are considering walks with a constant step size, then $h = \Delta x = \Delta y$. Suppose that at time step N the molecule is located at lattice point (m, p) (take, for example, the solid dot in Figure 4.26). The molecule's position at the previous time step has to be one of the four lattice points $(m-1, p)$, $(m+1, p)$, $(m, p-1)$, or $(m, p+1)$, which are the hollow dots in Figure 4.26. The probability of moving from each of these four points to (m, p) is $\frac{1}{4}$. We therefore have the master equation

$$\begin{aligned} w(m,p,N) = {} & \frac{1}{4}w(m-1,p,N-1) + \frac{1}{4}w(m+1,p,N-1) \\ & + \frac{1}{4}w(m,p-1,N-1) + \frac{1}{4}w(m,p+1,N-1). \end{aligned} \tag{4.71}$$

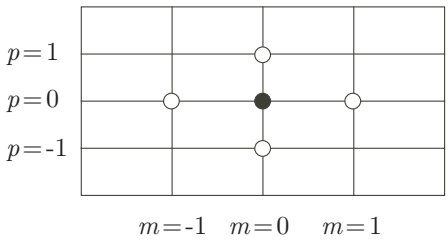


Figure 4.26 Nearest neighbor random walk on a rectangular lattice.

This is the two-dimensional equivalent of the one-dimensional master equation (4.9). The steps from this point on will follow the one-dimensional analysis very closely. To switch from (m, p, N) to (x, y, t) recall that $x = mh$, $y = ph$, and $t = N\Delta t$. Introducing the function $u(x, y, t) = w(m, p, N)$, (4.71) takes the form

$$\begin{aligned} 4u(x,t) = {} & u(x-h, y, t-\Delta t) + u(x+h, y, t-\Delta t) \\ & + u(x, y+h, t-\Delta t) + u(x, y-h, t-\Delta t). \end{aligned} \tag{4.72}$$

To continue we need the multivariable version of Taylor's theorem (see Section A). Through the quadratic terms the result is

$$\begin{aligned} & f(x+h, y+\ell, t+k) \\ & \quad = f + \left(h\frac{\partial}{\partial x} + \ell\frac{\partial}{\partial y} + k\frac{\partial}{\partial t}\right) f + \frac{1}{2}\left(h\frac{\partial}{\partial x} + \ell\frac{\partial}{\partial y} + k\frac{\partial}{\partial t}\right)^2 f + \cdots \\ & \quad = f + hf_x + \ell f_y + kf_t \\ & \qquad + \frac{1}{2}h^2 f_{xx} + \frac{1}{2}\ell^2 f_{yy} + \frac{1}{2}k^2 f_{tt} + h\ell f_{xy} + hkf_{xt} + \ell k f_{yt} + \cdots, \end{aligned} \tag{4.73}$$

where f and its derivatives on the right-hand side are evaluated at (x, y, t). Applying this to the terms in (4.72), and then simplifying, we obtain the following result

$$4u = 4u - 4(\Delta t)u_t + h^2 u_{xx} + h^2 u_{yy} + (\Delta t)^2 u_{tt} + \cdots.$$

Rearranging things a bit we obtain

$$u_t = \frac{h^2}{4\Delta t}(u_{xx} + u_{yy}) + \frac{\Delta t}{4}u_{tt} + \cdots. \tag{4.74}$$

With this, the first-order approximation for the probability satisfies

$$u_t = D(u_{xx} + u_{yy}), \tag{4.75}$$

where

$$D = \frac{h^2}{4\Delta t}. \tag{4.76}$$

The conclusion is that the resulting continuous problem is a diffusion equation. Although it was derived assuming the motion was on a rectangular lattice, as shown in Exercise 4.9, the same equation is obtained for the general random walk given in (4.68), (4.69).

The above formula for D is a factor of two smaller than what we found for one-dimensional motion. This is not surprising if one remembers that D is a measure of the spread in the molecules per time step, and the larger D the greater the spread. In one-dimensional motion the cloud can only move left or right. In contrast, in two dimensions the particles can also move around the

origin, as well as radially away from it, and this means the spreading is not as pronounced as in one dimension. In other words, the associated diffusion coefficient is less in two dimensions, and this is borne out in (4.76). Based on this, it should not be surprising that in d dimensions one still obtains a diffusion equation, with $D = h^2/(2d\Delta t)$.

Example: Point Source Solution

The symmetric solution seen in Figure 4.25 can best be described using polar coordinates. In switching from Cartesian to polar coordinates one finds that

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

Substituting these into (4.75), and simplifying, the polar coordinate form of the diffusion equation is

$$\frac{\partial u}{\partial t} = D\left(\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}\right), \tag{4.77}$$

We are interested in solutions that are symmetric about the origin, which is the case of the distributions shown in Figure 4.25. Mathematically, this means that u is independent of θ, and (4.77) reduces to

$$\frac{\partial u}{\partial t} = D\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right). \tag{4.78}$$

This is known as the radially symmetric diffusion equation. The second assumption, based on Figure 4.25, is that the total number of molecules remains constant. This means that

$$\int_0^\infty \int_0^{2\pi} ur dr = \gamma.$$

Because u is independent of θ, this reduces to

$$\int_0^\infty u\, r dr = \frac{\gamma}{2\pi}\,. \tag{4.79}$$

We want to find the solution of (4.78), that satisfies (4.79) and which also satisfies $u \to 0$ as $r \to \infty$. One of the easier ways to do this is to use similarity variables. Aside from u, the only dimensional variables or parameters in the problem are r, t, D, and γ. In other words, $u = u(r, t, D, \gamma)$. To reduce this using dimensional analysis note $[\gamma] = [u]/L^2$. Using an argument very similar to the one given in Section 1.4, it is found that

$$u = \frac{\gamma}{Dt} F(\eta), \tag{4.80}$$

where $\eta = r/\sqrt{Dt}$. Substituting this into (4.78), and rearranging things a bit, yields

$$\eta F'' + F' + \frac{1}{2}\eta^2 F' + \eta F = 0.$$

This can be rewritten as

$$\frac{d}{d\eta}(\eta F') + \frac{d}{d\eta}\left(\frac{1}{2}\eta^2 F\right) = 0.$$

Integrating, and then solving the resulting first-order differential equation for F, yields the general solution

$$F = e^{-\eta^2/4}\left(b + a\int \frac{1}{\eta} e^{\eta^2/4} d\eta\right),$$

where b and a are arbitrary constants. Now, the solution must be bounded at $\eta = 0$, and for this reason $a = 0$. The value of b is determined from (4.79), and one finds that $b = \gamma/(4\pi)$. Therefore, from (4.80), the point source solution of the diffusion equation is

$$u = \frac{\gamma}{4\pi Dt} e^{-r^2/(4Dt)}. \tag{4.81}$$

This is the function predicted in (4.70). ■

4.7 Langevin Equation

Random walks can be described as positional models of Brownian motion in the sense that they identify the locations of the molecules but they do not identify the physical reasons for the movement. To explore how to incorporate more of the physics into the modeling we need to consider what is happening to the molecule. As in Brown's original observations, the molecules involved are on the order of microns and are moving through a fluid. As such these molecules are in a sea of smaller objects, which are the atoms forming the fluid, that are undergoing thermal motions. The consequence of this is that the molecule is constantly subjected to many random impacts from these rapidly moving smaller objects. Although each atom has only a small effect on the molecule, there are many of them and together they are responsible for the molecule's random motion.

To model this we will use Newton's second law, namely $\mathbf{F} = m\mathbf{a}$. The force $\mathbf{F}$ between the molecule and the surrounding fluid will be separated into a deterministic component $\mathbf{D}$, and a random component $\mathbf{R}$, and we write

$$\mathbf{F} = \mathbf{D} + \mathbf{R}. \tag{4.82}$$

Determining $\mathbf{D}$ and $\mathbf{R}$ is based on the observation that the relevant time and space scales for the molecule and surrounding atoms are very different. The thermal motions of the atoms are rapid, and occur over very short distances, compared to those for the molecule. The force $\mathbf{F}$ is the averaged effect of these multiple individual collisions that are taking place as the molecule moves. The first term, $\mathbf{D}$, is the resistance force. As the molecule moves through the fluid there will be more collisions with the surrounding atoms on the front than on the back, and this will give rise to a resistance force. This is analogous to air resistance experienced by a object falling in air, and it is accounted for in (4.82) with $\mathbf{D}$. It is assumed that this force is proportional to the velocity. Letting $\mathbf{r}(t)$ be the position of the molecule then $\mathbf{D} = -\mu\mathbf{r}'$ where μ is a constant. The term $\mathbf{R}$ is suppose to account for everything else the atoms are doing to the molecule. As such it contains the random, and rapidly fluctuating, component of the force. The resulting equation of motion is

$$m\frac{d^2\mathbf{r}}{dt^2} = -\mu\frac{d\mathbf{r}}{dt} + \mathbf{R}(t), \tag{4.83}$$

where m is the mass of the molecule. This is known as the Langevin equation. It is an example of a stochastic differential equation due to the presence of $\mathbf{R}$. As a mathematical model it has been very influential in classical and quantum mechanics, as well as in statistical mechanics. In fact, Langevin ideas remain fundamental to contemporary scientific research in nonequilibrium statistical physics.

To continue it is necessary to specify $\mathbf{R}$. Although it does fluctuate rapidly, the amplitude or magnitude of this function is not small. In fact, to quote Langevin, "it maintains the agitation" of the molecule [Lemons and Gythiel, 1997]. What he means is that $\mathbf{R}$ is the driving force that is responsible for the observed random walk behavior of the molecule. This is interesting information, but it still leaves open the question of how to determine $\mathbf{R}$. The usual approach is to simply write down a function, providing plausibility arguments why it is appropriate. In doing so there are conditions the function is required to satisfy to be consistent with Brownian motion. One of the more basic hypotheses is that $\mathbf{R}$ is an external force and is independent of the molecule's motion, in other words, it does not depend on $\mathbf{r}$ or its derivatives. This assumption enables us to solve the equation. Introducing the velocity $\mathbf{v} = \mathbf{r}'$ then (4.83) can be written as

$$\frac{d\mathbf{v}}{dt} + \lambda\mathbf{v} = \frac{1}{m}\mathbf{R}(t), \tag{4.84}$$

where $\lambda = \mu/m$. This first-order equation can be solved using an integrating factor, and the result is

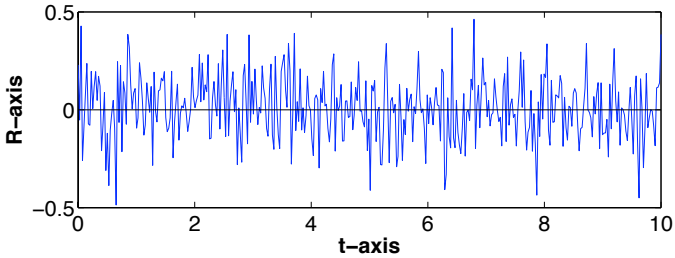


Figure 4.27 Example of the random, rapidly fluctuating, function $\mathbf{R}$ appearing in the Langevin equation (4.83).

$$\mathbf{v}(t) = \mathbf{v}(0)e^{-\lambda t} + \frac{1}{m}\int_0^t \mathbf{R}(\tau)e^{-\lambda(t-\tau)}d\tau. \tag{4.85}$$

This shows that the random forcing has a cumulative effect on the velocity because it depends on an integral of $\mathbf{R}$. How much the early values of $\mathbf{R}$ affect $\mathbf{v}$ depends on λ. The larger λ the greater the exponential decay in the integral, and the less effect the early values of $\mathbf{R}$ have on the velocity. Also note that larger values of λ reduce the contribution of the initial velocity. Said another way, the larger λ is, the quicker the molecule forgets its initial velocity and its movement is determined by Brownian randomization. Once the velocity is known, the position of the molecule can be determined by integrating (4.85), the result is

$$\mathbf{r}(t) = \mathbf{r}(0) + \frac{1}{\lambda}\mathbf{v}(0)(1 - e^{-\lambda t}) + \frac{1}{m\lambda}\int_0^t \mathbf{R}(\tau)(1 - e^{-\lambda(t-\tau)})d\tau. \tag{4.86}$$

Stating that the forcing function $\mathbf{R}$ is random does not mean that it is arbitrary. To be consistent with Brownian motion, $\mathbf{R}$ is subject to certain restrictions, and these will be derived in the next section. Before that, a comment is needed about the mathematical problem we are addressing. The example of the random forcing term $\mathbf{R}$ shown in Figure 4.27 uses 400 points along the t-axis. As will be explained later, the value of $\mathbf{R}(t_1)$ is independent of the value of $\mathbf{R}(t_2)$ if $t_1 \neq t_2$. This means that if more than 400 points are used, the graph will appear even more random than in Figure 4.27. The question that immediately arises is whether the resulting nondifferentiability of this function causes the differential equation (4.84), or its solution (4.84), to be meaningless. The answer as to why it is possible to include such a forcing function is one of the central objectives of stochastic differential equations, and how this is done is explained in Appendix C. The short answer is that differentiability is not an issue in (4.85) or (4.86), and it is these expressions that we will work with.

4.7.1 Properties of the Forcing

The solution in (4.85) is for a single molecule. We are interested in what happens when a large group of molecules are released at a point, which we will assume is the origin. Also, for simplicity, it is assumed that the molecules start out at rest, so $\mathbf{v}(0) = \mathbf{0}$. If there are K molecules in the group then the mean velocity of the group is

$$\mathbf{V} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{v}_i,$$

where $\mathbf{v}_i$ is the velocity of the ith molecule. Similarly, the mean displacement of the group is

$$\mathbf{D}_g = \frac{1}{K}\sum_{i=1}^{K}\mathbf{r}_i,$$

where $\mathbf{r}_i$ is the displacement of the ith molecule. Using (4.85) we have that

$$\mathbf{V} = \frac{1}{m}\int_0^t \mathbf{Q}(\tau)e^{-\lambda(t-\tau)}d\tau,$$

and from (4.86)

$$\mathbf{D}_g = \frac{1}{m\lambda}\int_0^t \mathbf{Q}(\tau)(1-e^{-\lambda(t-\tau)})d\tau, \tag{4.87}$$

where

$$\mathbf{Q} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{R}_i \tag{4.88}$$

is the mean random force, and $\mathbf{R}_i$ is the random forcing for the ith molecule.

Assumption 1: Zero Average.

As you might have already noticed, the molecules in the group are identical, so they have the same mass m and resistance factor μ. This brings us to the first assumption made on the random forcing. It is perhaps easiest to explain this using the two-dimensional random walk in (4.68), (4.69). All directions are equally likely. So, at time step n, if we happen to select a direction angle θ_n we could have just as likely selected the opposite direction, either $\theta_n + \pi$ or $\theta_n - \pi$. Consequently, at any given time step n, if one averages over all the displacements possible they get zero. The random forcing is assumed to be consistent with this result. In other words, it is assumed that when letting $K \to \infty$ in (4.88) they obtain $\mathbf{Q} = \mathbf{0}$. With this, taking the same limit in (4.87), we have that $\mathbf{D}_g = \mathbf{0}$. This result does not mean the group is motionless, rather it means the motion is symmetric. Given that it is equally

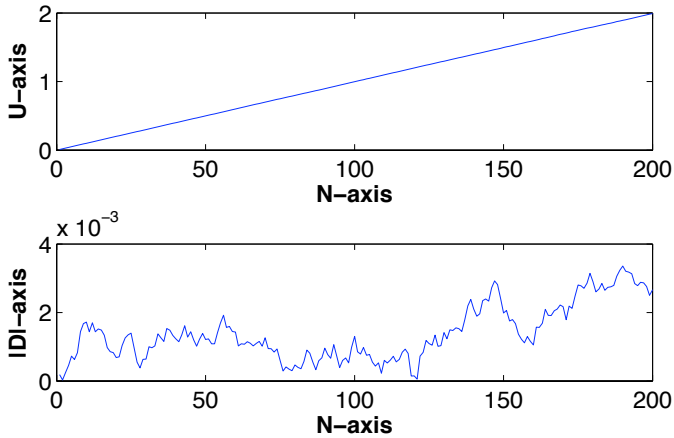


Figure 4.28 The upper graph gives the distance squared (4.89) averaged over a group of molecules moving according to the random walk (4.68), (4.69). The lower graph gives $||\mathbf{D}_g||$. In the calculation, $K = 100{,}000$ molecules were used, and the step length was 0.1.

likely to move in one direction as another, when a very large group starts out at the same point then the group will be distributed approximately symmetrically about the point as time progresses. Consequently, the resulting average displacement is approximately zero. This symmetry is evident in the distributions in Figure 4.25.

Assumption 2: Independence.

It is known that for random walks in one dimension the average displacement is zero, but the average of the displacement squared grows linearly in time (see Exercise 4.1). This same conclusion holds for multidimensional random walks, and to investigate this for the Langevin equation let

$$U = \frac{1}{K}\sum_{i=1}^{K} \mathbf{r}_i \cdot \mathbf{r}_i. \tag{4.89}$$

The values of U are given in Figure 4.28 for the two-dimensional random walk (4.68), (4.69). For completeness, the values of the magnitude of the average displacement vector (4.87) are also given. In looking at the values for $||\mathbf{D}_g||$ one might be a bit skeptical about the statement that the average displacement is zero. It should be remembered that this holds in the limit of $K \to \infty$. Also, after 200 time steps the molecules have the potential to be a distance of $200h = 20$ from the origin. Compare to this, the values for $||\mathbf{D}_g||$ in Figure 4.28 are quite small. No such qualifications, however, need

to be made about the computed values of U, which clearly show a linear dependence on time.

The question we now consider is whether the Langevin equation results in U increasing linearly in time. Substituting the solution (4.85) into (4.89), and recalling that $\mathbf{r}_i(0) = \mathbf{v}_i(0) = \mathbf{0}$, we obtain

$$U = \frac{1}{K}\sum_{i=1}^{K}\int_0^t\int_0^t \mathbf{R}_i(s)\cdot\mathbf{R}_i(\tau)f(s,\tau)d\tau ds, \tag{4.90}$$

where

$$f(s,\tau) = \frac{1}{m^2\lambda^2}(1-e^{-\lambda(t-\tau)})(1-e^{-\lambda(t-s)}). \tag{4.91}$$

This brings us to the second assumption made on the random forcing. Compared to the molecule's motion, the surrounding atoms are moving very quickly, and they are undergoing a large number of collisions with their neighbors over a very short amount of time. Consequently, the atomic events responsible for the random force at time t are effectively independent of those for the random force at a different time τ. In this case the forcing function is said to be Markovian. A consequence of this assumption is that the positive and negative values of $\mathbf{R}_i(s)\cdot\mathbf{R}_i(\tau)$ are all equally likely. It is for this reason that the average of $\mathbf{R}_i(s)\cdot\mathbf{R}_i(\tau)$, as $K\to\infty$, is zero if $s\neq\tau$. However, this does not mean that $U\to 0$ as $K\to\infty$ in (4.90) because we need to consider what happens when $s=\tau$.

Assumption 3: Concentration.

Assuming that the forcing is nonzero, the product $\mathbf{R}_i(s)\cdot\mathbf{R}_i(s)$ is positive. This means that the random forcing tends to accentuate the values in the integrals in (4.90) for $s=\tau$. The specific assumption made is that given any continuous function $f(s,\tau)$, if $0<s<t$ then

$$\lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\int_0^t \mathbf{R}_i(s)\cdot\mathbf{R}_i(\tau)f(s,\tau)d\tau = \gamma f(s,s), \tag{4.92}$$

where γ is a positive constant. As the above equation shows, $\sqrt{\gamma}$ is the amplitude of the forcing, and its value will be determined below when comparing the random walk and Langevin descriptions.

With (4.92), letting $K\to\infty$ in (4.90) we have that

$$\begin{aligned} U &= \frac{1}{m^2\lambda^2}\int_0^t \gamma\left(1-e^{-\lambda(t-s)}\right)^2 ds \\ &= \frac{\gamma}{2m^2\lambda^3}\left(2\lambda t - 3 + 4e^{-\lambda t} - e^{-2\lambda t}\right). \end{aligned} \tag{4.93}$$

For large values of time the above solution reduces to

$$U \approx \frac{\gamma}{\mu^2} t. \tag{4.94}$$

Therefore, with the stated assumptions on the random forcing, the Langevin equation gives us the expected conclusion that U increases linearly in time.

The two constants in (4.94) can be determined by introducing additional assumptions into the formulation. It is commonly assumed that the resistance term $\mu \mathbf{r}'$ in the Langevin equation is equivalent to the viscous force of a fluid. As shown in (1.22), it is known that for slow flows the drag force on a sphere of radius r is $6\pi\nu r v$, where ν is the dynamic viscosity of the fluid and v is the velocity of the sphere. This is Stokes' law for the drag on a sphere, and was introduced in Section 1.2.2. Assuming that the molecules are spheres then the conclusion is that $\mu = 6\pi\nu$. Because the viscosity of fluids such as air and water is known, then the corresponding value of μ is known. The value of γ can be determined from the theory of the kinetic theory of gases because the integral in (4.93) is associated with the thermal energy of the system. It is found that $\gamma = 6\mu k T$, where k is the Boltzmann constant and T is the absolute temperature. To relate this to the diffusion process arising from the random motion, it is known that $U = 2dDt$, where d equals the number of spatial dimensions (see Exercise 4.1). Using (4.94) we have that $D = \gamma/(2d\mu^2)$. Combining this with our values for the two constants we obtain the Stokes-Einstein equation (4.15).

Example

Because of the random forcing, each time the Langevin equation is solved a different solution is obtained. Three such solutions $r(t)$ are shown in the upper graph of Figure 4.29 for one-dimensional motion. For each solution the initial conditions are $r(0) = 0$ and $r'(0) = 0$, and the parameters are $\mu = 10$, $\gamma = 1$, and $m = 1$. The curves show the typical wandering of a Brownian motion. To check that the distance squared (4.89) is linear, the values of $U(t)$ are plotted in the lower graph of Figure 4.29. Not only is the curve linear, it is the same curve obtained using random walks shown in Figure 4.28 assuming $\Delta t = 1$. This is not a coincidence. From the Langevin formulation we have that $D = \gamma/(2d\mu^2)$, while from the random walk we have $D = h^2/(2d\Delta t)$. For these to produce the same diffusion coefficient it is therefore required that $\gamma = (\mu h)^2/\Delta t$. The given values of γ, μ, and h satisfy this equation, and that is why Figure 4.29 agrees with Figure 4.28. ■

Example: Asset Modeling

The Langevin equation was derived using ideas from molecular physics, but it has application in a wide variety of areas. One is in modeling the value of a financial asset, such as a stock. To frame this in terms of the discrete time steps used for a random walk, suppose the value of the asset at $t = n\Delta t$ is V_n

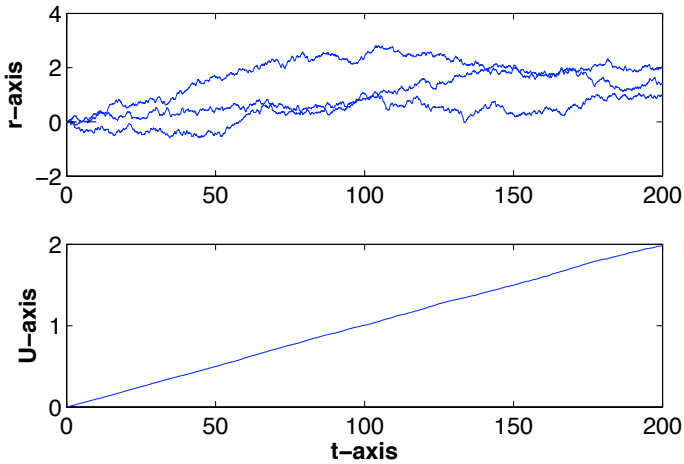


Figure 4.29 The upper graph gives three solutions of the one-dimensional Langevin equation, assuming $r(0) = 0$ and $r'(0) = 0$. The lower graph gives the distance squared (4.89), averaged over 10,000 solutions of the Langevin equation.

and we want to determine its value V_{n+1} at the next time step $t = (n+1)\Delta t$. The assumption is that the asset changes by an amount proportional to its value, and so

$$V_{n+1} = V_n + r_n V_n. \tag{4.95}$$

The coefficient r_n is the rate of return. For example, if the asset is a simple savings account, and the interest rate is μ, then $r_n = \mu\Delta t$. The value of many assets, such as stocks, are affected by external events, and their rates can vary dramatically with time. To account for this in the model, the rate is assumed to have the form

$$r_n = \mu\Delta t + \sigma\Delta W, \tag{4.96}$$

where the terms in this expression are explained below.

- *Expected Average Growth.* If external events do not affect the asset, then its value is assumed to increase at a constant rate. Just as with the savings account example, this rate is assumed to be $\mu\Delta t$. The positive constant μ is known as the drift coefficient.

- *Random Fluctuations.* The value of a stock can change due to rapidly changing external events. The $\sigma\Delta W$ term in (4.96) accounts for these fluctuations. In this expression, σ is a positive constant that depends on the particular asset under study, and is known as the volatility. The random function ΔW is time-dependent, but independent of the asset.

Combining (4.95) and (4.96) we have that

$$V_{n+1} - V_n = \mu \Delta t\, V_n + \sigma \Delta W\, V_n, \tag{4.97}$$

or equivalently

$$\frac{V_{n+1} - V_n}{\Delta t} = \left(\mu + \sigma \frac{\Delta W}{\Delta t}\right) V_n. \tag{4.98}$$

It is tempting to let $\Delta t \to 0$ in this expression, and from this conclude that $V' = (\mu + \sigma R)V$, where $R = W'$. However, the problem related to how this is possible with nondifferentiable functions comes up again, and the difficulty is compounded by the appearance of the product $\Delta W\, V_n$. One way to avoid this is to change variables and transform it into a Langevin equation. With this in mind, assume the change of variables has the form $Q = f(V)$. In this case, using Taylor's theorem for small Δt,

$$\begin{aligned} Q_{n+1} &= f(V(t_n + \Delta t)) \\ &= f\left(V_n + \Delta t\, V_n' + \frac{1}{2}(\Delta t)^2 V_n'' + \cdots\right) \\ &= f_n + \left(\Delta t\, V_n' + \frac{1}{2}(\Delta t)^2 V_n'' + \cdots\right) f_n' + \frac{1}{2}(\Delta t)^2 (V_n')^2 f_n'' + \cdots \\ &= Q_n + \Delta t\, (\mu + \sigma R_n) V_n\, f_n' + \frac{1}{2}(\Delta t)^2 (\mu + \sigma R_n)^2 V_n^2 f_n'' + \cdots \end{aligned} \tag{4.99}$$

For the random forcing associated with Brownian motion, it can be shown that for small Δt, $R^2 = 1/\sqrt{\Delta t} + \cdots$. Also, to transform (4.97) into one that resembles the Langevin equation let $Q = \ln(V)$. With this, (4.99) becomes

$$Q_{n+1} - Q_n = \Delta t \left(\mu - \frac{1}{2}\sigma^2\right) + \sigma \Delta t R_n. \tag{4.100}$$

Letting $\Delta t \to 0$ we obtain

$$\frac{dQ}{dt} = \mu - \frac{1}{2}\sigma^2 + \sigma R, \tag{4.101}$$

Letting $W = \int_0^t R(\tau) d\tau$, then the solution is

$$Q(t) = Q(0) + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t).$$

Transforming back into the original variables,

$$V(t) = V(0) e^{\lambda t + \sigma W(t)}, \tag{4.102}$$

where $\lambda = \mu - \frac{1}{2}\sigma^2$. This solution is an example of what is known as geometric Brownian motion. It gets this name because the random forcing enters through a multiplicative factor, as given in (4.95). ■

4.7.2 Endnotes

The assumption that $\mathbf{R}$ in (4.83) is a randomly varying function is an approximation. The reasoning used when introducing randomness is that this reflects the zig-zag nature of the motion. For the space and time scales we were considering, this is appropriate. However, if you were to slow time down, and look at the molecular level, the motion would appear to be smooth. As an example, if you watch a slowed down movie of billiard balls bouncing off each other, at the level of the billiard balls, the motion would appear smooth. Yet, in real-time they give the impression of changing directions instantly on impact. What this means is that the nondifferentiability of the function in Figure 4.27 is a consequence of the approximation of the forcing function. This observation has had a significant impact on the development of stochastic differential equations, and it specifically relates to how the integrals in (4.85) and (4.86) are defined. In one formulation the integrals possess important mathematical properties expected of integrals, but are not completely consistent with the physics, while other formulations do just the opposite. Exploring the ramifications of this statement is beyond the scope of this text, and those who want to learn more about this should consult Mazo [2002] and Kampen [2007].

Exercises

4.1. Suppose a total of K particles, all starting at $x = 0$, undergo a random walk. Let $x_i(N)$ be the position of the ith particle at time step N.

(a) Writing $x_i(N) = x_i(N-1) + q_i(N)$, explain why the value of q_i is either Δx or $-\Delta x$.

(b) Use the basic properties of a random walk to explain why the following holds

$$\lim_{K\to\infty} \frac{1}{K}\sum_{i=1}^{K} q_i(N) = 0.$$

Explain why the same reasoning can be used to explain why, if $N \neq M$,

$$\lim_{K\to\infty} \frac{1}{K}\sum_{i=1}^{K} q_i(N)q_i(M) = 0.$$

What is the above limit in the case of when $M = N$?

(c) The mean displacement of the group, at time step N, is

$$d_K(N) \;= \frac{1}{K}\sum_{i=1}^{K} x_i(N).$$

Relate $d_K(N)$ with $d_K(N-1)$, and from this show that $\lim_{K\to\infty} d_K = 0$. Therefore, the average displacement of a large group of particles is approximately zero.

(d) The mean-square displacement of the group is defined as

$$r_K(N) = \frac{1}{K}\sum_{i=1}^{K} x_i^2(N).$$

The value of r_K is a measure of the spread of the group. By relating $x_i(N)$ with $x_i(N-1)$, show that $\lim_{K\to\infty} r_K = N(\Delta x)^2$. Therefore, on average, for very large groups of particles the mean-square displacement of the group increases linearly with time.

(e) There are various ways to derive the formula for the diffusion coefficient. One sometimes used in physics is

$$D = \frac{< x^2 >}{2t},$$

where $< x^2 >= \lim_{K\to\infty} r_K$. Show why this agrees with the definition in (4.12).

4.2. This problem considers a random walk when the probabilities of left or right steps are not equal. In particular the probability of going right is p_r and the probability of going to the left is p_ℓ. It is assumed that p_r, p_ℓ are nonzero and $p_r + p_\ell = 1$.

(a) The grid in Figure 4.30 identifies the achievable positions a particle can reach at each time step. Letting $w(m, N)$ be the probability of each position then we know that $w(-1, 1) = p_\ell, w(0, 1) = 0$ and $w(1, 1) = p_r$. One can show that $w(0, 2) = p_r w(-1, 1) + p_\ell w(1, 1) = 2p_r p_\ell$, that is, $w(0, 2)$ is the sum of the probability of moving right from $(-1, 1)$ and moving left from $(1, 1)$. Use this principle to determine the probability for the other positions shown in the figure below. Also, show that the probabilities at

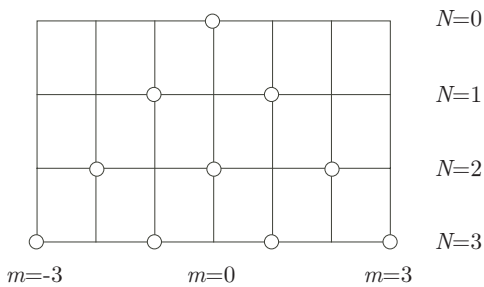


Figure 4.30 Figure for Problems 4.2 and 4.7.

each time level ($N = 0, 1, 2, 3, 4$) add to one and explain why this has to be the case.

(b) In going from time level $N = 0$ to time level $N \neq 0$, explain why to reach $x = m\Delta x$ it takes $n_r = (N + m)/2$ steps to the right and $n_\ell = (N - m)/2$ steps to the left.

(c) There are $N!/(n_r!n_\ell!)$ unique paths to reach $x = m\Delta x$ and as a consequence of this

$$w(m, N) = (p_r)^{n_r}(p_\ell)^{n_\ell}\frac{N!}{n_r!n_\ell!},$$

for $m = -N, -N+2, -N+4, \ldots, N$. Verify this formula for the positions shown in the above figure.

(d) Use Stirling's approximation to write w so it has the form $qL^{n_\ell}R^{n_r}$, where R is written in terms of p_r, N, and m while L is written in terms of p_ℓ, N, and m. Setting $Q = \ln(L^{n_\ell}R^{n_r})$ find the m that maximizes Q. After this use Taylor's theorem, through quadratic terms, to expand Q around this m value. From this show that for large N,

$$w(m, N) \sim \frac{1}{\sqrt{2\pi N p_r p_\ell}}e^{-q},$$

where

$$q = \frac{[m - N(p_r - p_\ell)]^2}{8Np_rp_\ell}.$$

(e) Use the principle in part (a) to derive a master equation, which expresses $w(m, N)$ in terms of $w(m-1, N-1)$ and $w(m+1, N-1)$. Setting $u(x, t) = w(m, N)$ then rewrite your equation in terms of u. Expand this for small Δx, Δt and derive a partial differential equation for u that involves u_x, u_t, u_{xx}. The coefficients of the equation must not depend on x or t but can depend on Δx, Δt, p_r, p_ℓ, etc. The equation must also reduce to the diffusion equation if $p_r = p_\ell$. The equation you are deriving is called the drift-diffusion equation.

(f) Show that the solution in part (d) can be written, up to a multiplicative constant, as

$$u(x, t) = \frac{1}{\sqrt{t}}e^{-(x-vt)^2/(\alpha t)}.$$

Show that this satisfies your drift-diffusion equation and in the process state how v and α are related to Δx, Δt, p_r, p_ℓ.

(g) Explain why the constant v in part (f) is known as the drift velocity.

(h) Figure 4.31 shows the result of running a biased random walk with 5,000 particles all starting at the origin. From these data, and your result in part (d), estimate the values of p_r, p_ℓ .

4.3. This problem makes use of the connection of the diffusion coefficient with the molecule's mean free path and average time between collisions, as

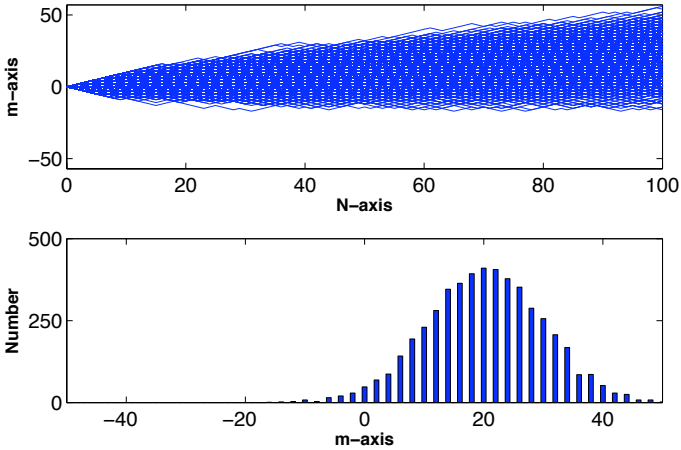


Figure 4.31 Figure for Exercise 4.2.

expressed by the Einstein-Smoluchowski equation (4.14). In what follows the average speed of the molecule is defined as $v = \lambda/\tau$.

(a) In air the mean free path is in the neighborhood of 30 times the average molecular separation distance d. In air suppose d is approximately 3×10^{-7} cm. Given that for air $D \approx 0.2 \frac{cm^2}{sec}$, approximately how long is it between collisions? Approximately how fast are the molecules traveling? How many collisions are there per second?

(b) In water the mean free path is in the neighborhood of 30 times the average molecular separation distance d. In water suppose d is approximately 3×10^{-8} cm. Given that for water $D \approx 2 \times 10^{-5} \frac{cm^2}{sec}$, approximately how long is it between collisions? Approximately how fast are the molecules traveling? How many collisions are there per second?

4.4. A lazy random walk is one that allows the particle to stay put instead of having to move left or right. For this situation assume the probability of going to the right is p_r, the probability of going to the left is p_ℓ, and the probability of not moving is p_s. As usual, $p_\ell + p_s + p_r = 1$. Also, letting Δx be the spatial stepsize and Δt the temporal stepsize then $x = m\Delta x$ and $t = N\Delta t$.

(a) The grid in Figure 4.32 identifies the achievable positions a particle can reach at each time step. Letting $w(m, N)$ be the probability the molecule is at $x = m\Delta x$ after N time steps then we know that $w(-1, 1) = p_\ell w(0, 0) = p_\ell$, $w(0, 1) = p_s w(0, 0) = p_s$, and $w(1, 1) = p_r w(0, 0) = p_r$. Determine the probability for the other positions shown in the figure. Also, show that the probabilities at each time level ($N = 0, 1, 2, 3$) add to one and explain why this has to be the case.

(b) Based on your result in part (a) what are the values of A, B, C so $w(m,N) = Aw(m-1,N-1) + Bw(m,N-1) + Cw(m+1,N-1)$.

(c) Setting $u(x,t) = w(m,N)$ then rewrite your result from part (b) in terms of u. Assuming Δx, Δt are small, derive a partial differential equation for u. The coefficients of the equation must not depend on x or t but can depend on Δx, Δt, p_ℓ, p_s, p_r, etc. Do you need to distinguish between the cases of when $p_\ell \neq p_r$ and when $p_\ell = p_r$? Also, in the case of when $p_\ell = p_r$ your result should reduce to the diffusion equation.

(d) Explain how this result differs from (4.12) and (4.13). Make explicit comparisons between the coefficient(s) of the two equations.

4.5. For the random walk we considered there was no memory of the previous step when determining the current one. An interesting modification is a correlated walk where the probability at the next time step depends upon the previous step. To examine this suppose that step $N-1$ is complete. Step N is made in the same direction with probability p and in the opposite direction with probability $1-p$. If $p > \frac{1}{2}$ it is called a persistent walk and if $p < \frac{1}{2}$ it is an anti-persistent walk. To get this procedure started a regular random walk can be used at the first step.

(a) Let $w(m,N) = f(m,N) + g(m,N)$, where $f(m,N)$ is the probability of arriving at $x = m\Delta x$, at time step N, from the left and $g(m,N)$ is the probability of arriving at $x = m\Delta x$, at time step N, from the right. Show that $f(m,N) = pf(m-1,N-1) + (1-p)g(m-1,N-1)$ and $g(m,N) = (1-p)f(m+1,N-1) + pg(m+1,N-1)$.

(b) Setting $u(x,t) = f(m,N) + g(m,N)$ and $v(x,t) = f(m,N) - g(m,N)$, expand u, v for small $\Delta x, \Delta t$. Letting $\Delta t \to 0$, with $c = \Delta x/\Delta t$ fixed and $p = 1 - \alpha\Delta t/2$, derive the following partial differential equation

$$u_{tt} + \alpha u_t = c^2 u_{xx}.$$

This is known as the telegraph equation.

(c) As in part (b), expand u, v for small $\Delta x, \Delta t$, but now let $\Delta t \to 0$ with $\Delta x^2/\Delta t$ fixed and p constant. Show that u satisfies the diffusion equation,

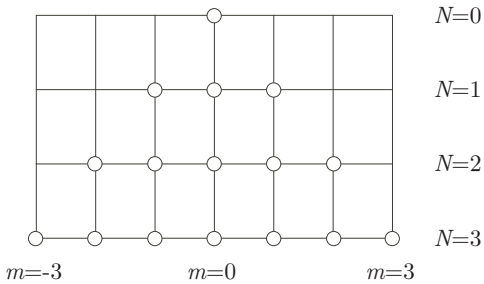


Figure 4.32 Grid for Exercise 4.4.

where the diffusion coefficient is

$$D = \frac{\Delta x^2}{2\Delta t}\frac{p}{1-p}.$$

4.6. For the diffusion problem in (4.22), (4.23) suppose

$$f(x) = \frac{\alpha}{\sqrt{\pi}}e^{-\alpha^2 x^2},$$

where σ is a positive constant.

(a) Show that

$$u(x,t) = \frac{1}{2\sqrt{\pi D(t+\tau)}}e^{-\frac{x^2}{4D(t+\tau)}},$$

where $\tau = 1/(4\alpha^2 D)$.

(b) Given that $\int_{-\infty}^{\infty} f(x)dx = 1$, show that the solution in part (a) satisfies $\int_{-\infty}^{\infty} u(x,t)dx = 1$. Explain how the latter result can also be obtained directly from the diffusion equation.

4.7. A random walk with loss is one that allows the particle to be irreversibly lost from the system at each time step. Suppose that at $t = (N-1)\Delta t$ the particle is located at $x = m\Delta x$. The assumption is that at $t = N\Delta t$ the particle will have moved to $x = (m+1)\Delta x$ with probability p_r, it will have moved to $x = (m-1)\Delta x$ with probability p_r, and it will have vanished with probability p_s. As usual, $2p_r + p_s = 1$.

(a) The grid in Figure 4.30 identifies the achievable positions a particle can reach at each time step when starting at $m = 0$. Letting $w(m,N)$ be the probability the molecule is at $x = m\Delta x$ after N time steps then we know that $w(-1,1) = p_r w(0,0) = p_r$, and $w(1,1) = p_r w(0,0) = p_r$. Determine the probability for the other positions shown in the figure.

(b) Based on your result in part (a) what are the values of A, B, C so $w(m,N) = Aw(m-1,N-1) + Bw(m,N-1) + Cw(m+1,N-1)$?

(c) Setting $u(x,t) = w(m,N)$, rewrite your result from part (b) in terms of $u(x,t)$. Assuming $\Delta x, \Delta t$ are small, derive a partial differential equation for u. In doing this assume the probability of loss is small, that is, assume $p_s = p_0\Delta t$. The coefficients of the equation you derive must not depend on x or t but can depend on Δx, Δt, p_r, and p_0. Also, in the case of when $p_0 = 0$ your result should reduce to the diffusion equation (4.13).

4.8. In two dimensions the lattice need not be square, or even rectangular. This problem examines what happens in such cases.

(a) Suppose in the lattice shown in Figure 4.26, Δx and Δy are not equal. Assuming $\lambda = \Delta x/\Delta y$ is fixed what is the resulting diffusion equation?

(b) Suppose the lattice is as shown in Figure 4.33, where $\Delta x = \Delta y$. Show that one still obtains the diffusion equation in (4.75) with D given in (4.76).

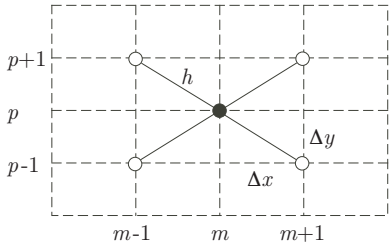


Figure 4.33 Random walk for Exercise 4.8. A molecule at the black dot will move, with equal probability, to one of the hollow dots. The step length is h.

4.9. This problem explores how to derive the diffusion equation for the general random walk in the plane, as given in (4.68), (4.69). Let $u(x, y, t)$ be the probability that the particle is located at the spatial location (x, y) at time t.

(a) Suppose that at time step $t + \Delta t$ the particle is located at (x, y). Explain why at time t the particle was located somewhere on the circle of radius h that is centered at (x, y).

(b) As an approximation to the circle in part (a), distribute N points uniformly around this circle. Specifically, take the points $(x + h\cos(j\Delta\theta), y + h\sin(j\Delta\theta))$, where $\Delta\theta = 2\pi/N$ and $j = 1, 2, \ldots, N$. Explain why the probability of the particle moving from one of these N points to (x, y) is approximately $1/N$. From this explain why

$$u(x, y, t + \Delta t) \approx \frac{1}{N}\sum_{j=1}^{N} u(x + h\cos(j\Delta\theta), y + h\sin(j\Delta\theta), t).$$

(c) Use the result from part (b) to show that for the general random walk

$$u(x, y, t + \Delta t) = \frac{1}{2\pi}\int_0^{2\pi} u(x + h\cos\theta, y + h\sin\theta, t)d\theta.$$

(d) Derive the diffusion equation from the result in part (c) by letting Δt and h approach zero.

4.10. In this problem the inverse Fourier transform for the diffusion equation is derived from scratch.

(a) Show that

$$u(x, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(k)e^{ikx - Dk^2 t}dk.$$

(b) Setting $H(k) = \mathcal{F}(e^{-x^2})$, show that $H' = -\frac{k}{2}H$. Using the fact that $H(0) = \sqrt{\pi}$ show that $\mathcal{F}(e^{-x^2}) = \sqrt{\pi}e^{-k^2/4}$. From this show that

$$\int_{-\infty}^{\infty} e^{ikq-Dk^2t}dk = \sqrt{\frac{\pi}{Dt}}\, e^{-q^2/(4Dt)}.$$

(c) Use parts (a) and (b) to derive (4.38).

4.11. This problem involves the Fourier transform and its inverse on problems involving semi-infinite intervals.

(a) Suppose $y(t)$ satisfies $y'' + 3y' + 2y = e^{-3t}$, for $t > 0$, with the initial conditions $y(0) = y'(0) = 0$. This problem determines the value of $y(t)$ for $t \geq 0$, but the Fourier transform requires the interval $-\infty < t < \infty$. Explain why it is possible to assume in this problem that $y = 0$ for $-\infty < t < 0$. Doing this, use the Fourier transform to solve the problem.

(b) Use the Fourier transform to find the function $y(t)$ that satisfies $2y'' + 7y' + 3y = -4e^{-t}$, for $t > 0$, with the initial conditions $y(0) = y'(0) = 0$.

4.12. Find the differential equation satisfied by the Fourier transform $U(k,t)$. Assume that the solution and its derivatives go to zero as $x \to \pm\infty$.

(a) $u_t + u_x = u_{xxx}$.

(b) $u_t + xu_x = 0$.

(c) $u_t + u_{xxxx} = 0$.

4.13. Explain why $\mathcal{F}(\mathcal{F}^{-1}(F)) = F$ is correct but that $\mathcal{F}^{-1}(\mathcal{F}(f)) = f$ might not be true. What assumption(s) must be made so the last statement is correct?

4.14. This problem concerns calculating the Fourier transform or its inverse.

(a) Find $f(x)$ if

$$F(k) = \frac{1}{(1+ik)(2+ik)}.$$

(b) Find $f(x)$ if

$$F(k) = \frac{1}{(2+ik)}e^{-ik}.$$

(c) Find $F(k)$ if

$$f(x) = \begin{cases} \cosh(x) & \text{if} \quad |x| \leq \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

4.15. This problem develops some of the basic properties of the Fourier transform. Assuming $F(k)$ is the Fourier transform of $f(x)$ show the following.

(a) The Fourier transform of $f(ax)$, for $a \neq 0$, is $F(k/a)/|a|$.

(b) The Fourier transform of $f(x-a)$ is $e^{-iak}F(k)$.

(c) The Fourier transform of $f(x)\cos(ax)$ is $\frac{1}{2}\left(F(k+a) + F(x-a)\right)$.

4.16. Suppose the initial condition for the diffusion problem is

$$u(x,0) = \begin{cases} u_0 & \text{if} \quad |x| \leq h, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the solution is

$$u(x,t) = \frac{1}{2}u_0\left[\operatorname{erf}\left(\frac{x+h}{2\sqrt{Dt}}\right) + \operatorname{erf}\left(\frac{x-h}{2\sqrt{Dt}}\right)\right],$$

where erf() is the error function.

4.17. This problem concerns the convection-diffusion equation

$$u_t = Du_{xx} - cu_x, \quad \text{for} \quad \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

with the initial condition

$$u(x,0) = f(x).$$

Assume c is a constant.

(a) Using the Fourier transform, find the solution of the above problem.

(b) Make the change of variables $\xi = x - ct$, $\tau = t$. Letting $v(\xi,\tau) = u(x,t)$ show that v satisfies a problem very similar to the one solved in Section 7.2.5. Use this observation, and (4.38), to write down the solution of the convection-diffusion problem.

4.18. This problem concerns the reaction-diffusion equation

$$u_t = Du_{xx} - cu, \quad \text{for} \quad \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

with the initial condition

$$u(x,0) = f(x).$$

Assume c is a positive constant.

(a) What reaction(s) give rise to this reaction-diffusion equation?

(b) Using the Fourier transform, find the solution of the above problem.

(c) Show that the problem can also be solved by first letting $u = ve^{at}$, where a is a constant of your choosing, and then using (4.38).

(d) Suppose the $-cu$ term in the differential equation is replaced with cu. What reaction(s) give rise to the resulting reaction-diffusion equation?

4.19. This problem concerns solving the wave equation

$$u_{tt} = c^2 u_{xx}, \quad \text{for} \quad \begin{cases} -\infty < x < \infty, \\ 0 < t, \end{cases}$$

with the initial conditions

$$u(x,0) = 0, \quad u_t(x,0) = g(x).$$

Assume c is a positive constant. Using the Fourier transform, find the solution of this problem.

4.20. Using the steady-state solution (4.55) to find the diffusion coefficient requires the experiment to run for almost three hours. Explain how it is possible to find D within 60 seconds of the start of the experiment.

Chapter 5
Traffic Flow

5.1 Introduction

In this chapter we again investigate the movement of objects along a one-dimensional path, but now the motion is directed rather than random. Examples of such situations include:

- Cars moving along a highway (Figure 5.1)
- Blood cells moving along a capillary (Figure 5.2)
- Molecules moving along a carbon nanotube (Figure 5.3)

Although the underlying physics of each of these is quite different they all involve the movement of objects along what is effectively a one-dimensional pathway. We will take advantage of this when developing a mathematical model for the motion, but before doing so we must first decide on how to account for the spatial and temporal variables. For example, for random walks we used discrete steps in space and time. This is also done for traffic models and it is the basis of the cellular automata description presented in Section

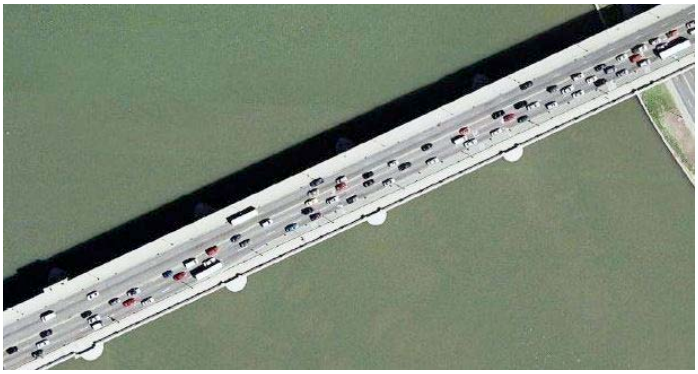


Figure 5.1 Aerial view of traffic flow (Google Maps [2007]).

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_5,
© Springer Science+Business Media, LLC 2009

Figure 5.2 Red blood cells flowing in an arteriol (Baskurt [2009]).

5.7. We will start out, however, assuming that the motion is continuous, which is the viewpoint taken when deriving the diffusion equation in Section 4.5.

5.2 Continuum Variables

We are assuming that the objects are numerous enough that it is not necessary to keep track of each one individually, and we can use an averaged value. In deriving the mathematical model, the objects here will be identified as cars and the path as a highway. There are a couple of reasons for using this particular example. One is that most everyone has experience with traffic, and is able to relate the mathematical results with the real-world application. The other reason is that the theory for traffic flow is still not complete, so there are competing ideas that can be explored. However, it should be remembered that all of this material can be applied to other systems, such as the one dimensional motion of blood cells and molecules. In fact, some of the termi-

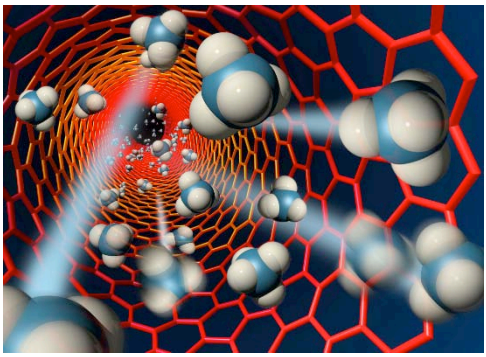


Figure 5.3 Methane molecules flowing through a carbon tubule less than 2 nanometers in diameter (Lawrence Livermore National Laboratory [2009]).

nology that is introduced comes from gas dynamics, because of its early use of the ideas developed here.

5.2.1 Density

The variable that will play a prominent role in our study is the traffic density $\rho(x,t)$. This is the number of cars per unit length, and it is instructive to consider how it might be determined experimentally. To measure ρ at $x = x_0$, for $t = t_0$, one selects a small spatial interval $x_0 - \Delta x < x < x_0 + \Delta x$ on the highway, and then counts the number of cars within this interval (see Figure 5.4). In this case

$$\rho(x_0, t_0) \approx \frac{\text{number of cars from } x_0 - \Delta x \text{ to } x_0 + \Delta x \text{ at } t = t_0}{2\Delta x}. \tag{5.1}$$

The underlying assumption here is that Δx is small enough that only cars in the immediate vicinity of x_0 are used to determine the density at this point. At the same time, Δx cannot be so small that it is on the order of the length of individual cars (and the spacing between them). In the continuum viewpoint, the cars are distributed smoothly over the entire x-axis, and the value of $\rho(x_0, t_0)$ is the limit of the right-hand side of (5.1) as $\Delta x \to 0$.

Example: Uniform Distribution

To illustrate how density is determined suppose the cars all have length ℓ, and they are evenly spaced a distance d apart (see Figure 5.5). Given a sampling interval $2\Delta x$ along the highway then the number of cars in this interval is, approximately, $2\Delta x/(\ell + d)$. Inserting this into (5.1) and letting $\Delta x \to 0$ we find that

$$\rho = \frac{1}{\ell + d}. \tag{5.2}$$

One conclusion that comes from this formula is that there is a maximum density. Because $0 \le d < \infty$ then $0 < \rho \le \rho_M$, where $\rho_M = 1/\ell$. For example, if $\ell = 17$ ft (5.2 m) and $d = 12$ ft (3.6 m) then, recalling 1 mi = 5280 ft, $\rho = 182\,\frac{\text{cars}}{\text{mi}}$ ($113\,\frac{\text{cars}}{\text{km}}$). With these dimensions then the maximum density

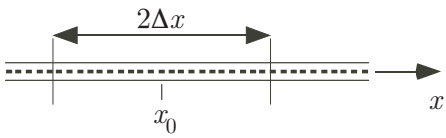


Figure 5.4 The interval along the highway used to calculate an approximate value of the density $\rho(x_0, t_0)$. It is also used to derive the balance law for traffic flow.

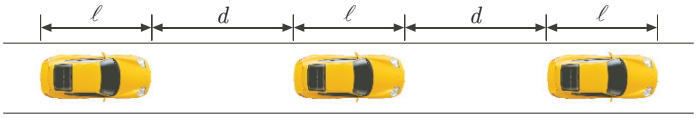


Figure 5.5 For a uniform distribution, the cars are all the same length and are evenly spaced along the highway.

that is possible, which occurs when $d = 0$, is $\rho_M = 310.6\,\frac{\text{cars}}{\text{mi}}$ $(193\,\frac{\text{cars}}{\text{km}})$. When studying traffic flow, it is useful to know the maximum merge density ρ_{mg}, which corresponds to the density that occurs when the spacing is such that exactly one car fits between two cars currently on the highway. This occurs when $d = \ell$ and for this example $\rho_{mg} = 155.3\,\frac{\text{cars}}{\text{mi}}$ $(96.5\,\frac{\text{cars}}{\text{km}})$. ■

5.2.2 Flux

The second variable we need is the flux $J(x,t)$, which has the dimensions of cars per unit time. To measure J at $x = x_0$, for $t = t_0$, one selects a small time interval $t_0 - \Delta t < t < t_0 + \Delta t$ and counts the net number of cars that pass $x = x_0$ during this time period. The convention is that a car moving to the right is counted as $+1$, while one moving to the left is counted as -1. In this case

$$J(x_0, t_0) \approx \frac{\text{net number of cars that pass } x_0 \text{ from } t = t_0 - \Delta t \text{ to } t = t_0 + \Delta t}{2\Delta t}. \tag{5.3}$$

The underlying assumption here is that Δt is small enough that only cars that are passing x_0 at, or near, $t = t_0$ are used to determine the flux at t_0. At the same time, from an experimental standpoint, Δt can not be so small that no cars are able to pass this location during this time interval. In the continuum viewpoint we are taking the cars are distributed smoothly over the entire t-axis and the value of $J(x_0, t_0)$ is the limit of the right hand side of (5.3) as $\Delta t \to 0$.

Example: Uniform Distribution (cont'd)

Returning to the previous example of uniformly distributed cars, shown in Figure 5.5, we now add in the assumption that the cars are moving with a constant positive velocity v. In this case, the cars that start out a distance $2\Delta t v$ from x_0 will pass x_0 in the time interval from $t_0 - \Delta t$ to $t_0 + \Delta t$. The corresponding number of cars is, approximately, $2v\Delta t/(\ell + d)$. Inserting this into (5.3), and letting $\Delta t \to 0$, yields

$$J = \frac{v}{\ell + d}\,. \tag{5.4}$$

For example, if $\ell = 17$ ft, $d = 51$ ft and $v = 70$ mph then $J = 5,435\ \frac{\text{cars}}{\text{hr}}$. Also, note that $J = \rho v$, which is one of the fundamental formulas in traffic flow. ■

5.3 Balance Law

To derive an equation for the density we will use what is known as a control volume argument. For this problem the control volume is a small region on the highway, from $x_0 - \Delta x$ to $x_0 + \Delta x$. This interval is shown in Figure 5.1. During the time period from $t = t_0 - \Delta t$ to $t = t_0 + \Delta$ t it is assumed that the number of cars in this interval can change only due to cars entering or leaving at the left or right ends of the interval. We are therefore assuming cars do not disappear, or pop into existence, on the highway. Actually, this could happen if we were to include an off- or onramp, but this modification will be postponed for the moment (see Exercise 5.21). As stated, our balance law for cars within the highway interval is

$$\begin{aligned}
&\{\text{number of cars in interval at } t = t_0 + \Delta t\} \\
&\qquad - \{\text{number of cars in interval at } t = t_0 - \Delta t\} \\
&= \{\text{net number of cars that cross } x_0 - \Delta x \text{ from } t_0 - \Delta t \text{ to } t_0 + \Delta t\} \\
&\qquad - \{\text{net number of cars that cross } x_0 + \Delta x \text{ from } t_0 - \Delta t \text{ to } t_0 + \Delta t\}.
\end{aligned}$$

Rewriting this using (5.1) and (5.3) yields

$$\begin{aligned}
&2\Delta x \left[\rho(x_0, t_0 + \Delta t) - \rho(x_0, t_0 - \Delta t)\right] \\
&\qquad\qquad = 2\Delta t \left[J(x_0 - \Delta x, t_0) - J(x_0 + \Delta x, t_0)\right].
\end{aligned}$$

Using Taylor's theorem, we have that

$$\begin{aligned}
&2\Delta x \Bigg(\rho + \Delta t \rho_t + \frac{1}{2}(\Delta t)^2 \rho_{tt} + \frac{1}{6}(\Delta t)^3 \rho_{ttt} + \cdots \\
&\qquad\qquad -\rho + \Delta t \rho_t - \frac{1}{2}(\Delta t)^2 \rho_{tt} + \frac{1}{6}(\Delta t)^3 \rho_{ttt} + \cdots\Bigg) \\
&= 2\Delta t \Bigg(J - \Delta x J_x + \frac{1}{2}(\Delta x)^2 J_{xx} - \frac{1}{6}(\Delta x)^3 J_{xxx} + \cdots \\
&\qquad\qquad -J - \Delta x J_x - \frac{1}{2}(\Delta x)^2 J_{xx} - \frac{1}{6}(\Delta x)^3 J_{xxx} + \cdots\Bigg),
\end{aligned}$$

where ρ and J are evaluated at (x_0, t_0). Collecting the terms in the above equation,

$$\rho_t + O\left((\Delta t)^2\right) = -J_x + O\left((\Delta x)^2\right).$$

Letting $\Delta x \to 0$ and $\Delta t \to 0$ we conclude that

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x}\,. \tag{5.5}$$

This is our balance law for motion along the x-axis. It is applicable to any continuous system in which the objects are not created or destroyed. This is why it was also obtained when deriving the model for diffusion (4.48).

5.3.1 Velocity Formulation

It is possible to express the balance law somewhat differently, by introducing the velocity $v(x,t)$ of the cars on the highway. This requires care because the velocity, like the other continuum variables, is an averaged quantity. To explain how this is done, consider a small interval on the highway as shown in Figure 5.4. One measures $v(x_0,t_0)$ experimentally by finding the average velocity of the cars in this interval. Specifically, if there are n cars in the interval, and they have velocities v_1, v_2, ... , v_n, then

$$v(x_0,t_0) \approx \frac{1}{n}\sum_{i=1}^{n} v_i.$$

In the continuum model it is assumed that the limit of this average, when letting $\Delta x \to 0$, exists, and its value is the velocity $v(x_0,t_0)$.

With the above definition, the velocity is assumed to be related to the flux through the equation

$$J = \rho v. \tag{5.6}$$

This equation was derived in the uniform distribution example discussed earlier. It is also possible to derive it for situations where the velocity is not constant (see Exercise 5.26). However, a proof for the general case is not available, and so the above formula is an assumption. Some avoid this difficulty by using (5.6) as the definition of the flux, while others use it as the definition of the velocity.

Introducing (5.6) into (5.5) gives us

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) = 0. \tag{5.7}$$

In solving this equation it will be assumed the initial density is known, that is,

$$\rho(x,0) = f(x). \tag{5.8}$$

The equation in (5.7) is the mathematical model for traffic flow that we will investigate in the first part of this chapter. Those working in traffic flow refer to this as the Lighthill-Whitham-Richards (LWR) model, naming it after those who originally derived the equation (Lighthill and Whitham [1955], Richards [1956]). However, the equation has wide applicability, and appears under different banners. For example, in continuum mechanics it is known as the continuity equation, while in electrodynamics (5.7) is the current continuity equation, where ρ is the current density and J is the current volume. Those interested in more mathematical pursuits refer to (5.7) as a scalar conservation law.

It should be kept in mind that, as with most mathematical models, (5.7) is an approximation of the true system. Not unexpectedly, there are limitations on its applicability. As a case in point, it is questionable whether the model provides an accurate description at low densities. If the objects are few and far between then the assumptions made in defining the density and flux are not valid. This will not stop us from using the model in such rarified regimes, but when this is done it should be understood that the continuum model provides more of a qualitative description of the motion. That said, in the regimes where it does apply, the continuum model has proven to be an exceptionally accurate, and mathematical interesting, description.

5.4 Constitutive Laws

Although we have derived the balance law for traffic flow, the mathematical model is incomplete. The issue is the velocity v and how it is related to the density ρ. One possibility is to investigate the physics of the problem a bit more and see if there is another equation relating these variables. This is done in mechanics, and Newton's second law is used to derive a force balance equation that can be used to find the velocity. This option is not easily adaptable to the traffic flow situation so we will take a different approach and postulate how v and ρ are related based on experimental evidence. What we will be doing is specifying a constitutive law relating the velocity and density. To do this the data for several rather different roadways are shown in Figure 5.6. The question is, what function best describes the data in this figure? The answer depends, in part, on what density and velocity intervals are of interest and what applications one has in mind. A few possible constitutive laws are discussed below.

It is worth making a couple of comments about Figure 5.6 that are unrelated to constitutive modeling. The data in the lower two graphs is what was used in the original development of the continuum traffic model, while the data in the upper two graphs is typical of more modern testing. One of the striking differences between the upper and lower graphs is the amount of data shown. This is due to the development of computerized testing systems,

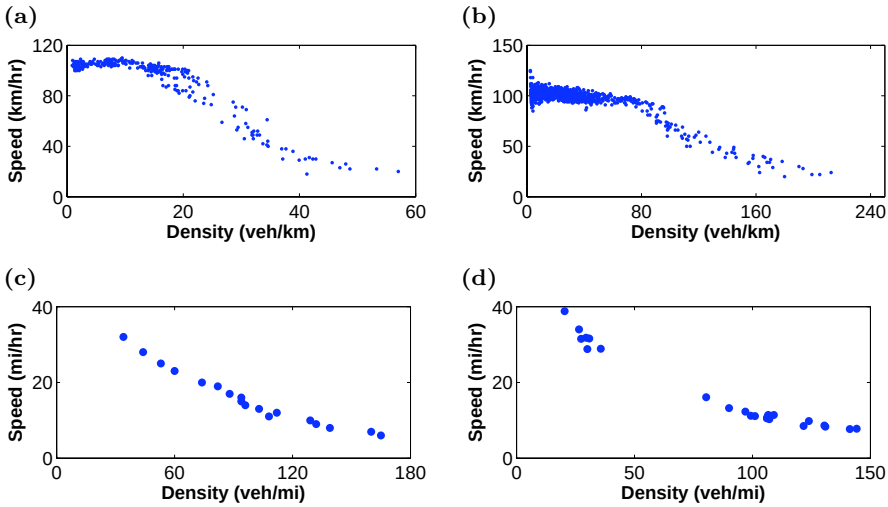


Figure 5.6 The velocity as a function of the density as measured for different roadways. Shown is (a) a highway near Toronto, (b) a freeway near Amsterdam, (c) the Lincoln Tunnel, and (d) the Merritt Parkway. Data for (a) and (b) are from Aerde and Rakha [1995], and (c) and (d) are from Greenberg [1959].

which have been invaluable for modern scientific research. However, what is interesting is the rather tight pattern in the earlier data as compared to the scatter in the more recent results. This begs the question of whether these earlier experimentalists were more careful, or did they force the results. It makes one wonder. A second observation concerns the difference in the densities between Toronto and Amsterdam, which differ by almost a factor of four. Any theory why this happens?

5.4.1 Constant Velocity

The simplest assumption is that v is constant in terms of its dependence on ρ, in other words, $v = a$. In this case the balance law (5.7) reduces to

$$\frac{\partial \rho}{\partial t} + a\frac{\partial \rho}{\partial x} = 0. \tag{5.9}$$

This is known as the advection equation. In looking at the data in Figure 5.6 one might conclude that assuming v is constant borders on delusional. The value of this assumption is not its realistic portrayal of traffic but, rather, what it provides in terms of insights into the type of mathematical problem that arises in traffic flow. The analysis of this problem will provide the foun-

dation needed for solving the more difficult nonlinear problems arising from more realistic velocity functions.

5.4.2 Linear Velocity

The most widely used, and most well known, constitutive laws are linear. For the traffic problem this means we assume $v = a - b\rho$, where a, b are constants. Those working in traffic flow refer to this as the Greenshields model, and the usual way this is written is

$$v = v_M \left(1 - \frac{\rho}{\rho_M}\right), \tag{5.10}$$

where the constants v_M, ρ_M are the maximum velocity and density, respectively. The values of these constants can almost be read off the plot in Figure 5.6. However, a more systematic way to find them is to use a least squares fit. Using the data for the Lincoln Tunnel and Merritt Parkway one finds that $v_M = 36.821$ mph, $\rho_M = 166.4226$ cars/mi and the resulting function is plotted in Figure 5.7 along with the original data. It is seen that even though this function misses the values at the extreme ends, where $\rho = 0$ or $\rho = 180$, it does show the correct monotonic dependence of the velocity on density. This would seem an acceptable approximation, and the traffic flow equation (5.7) reduces to

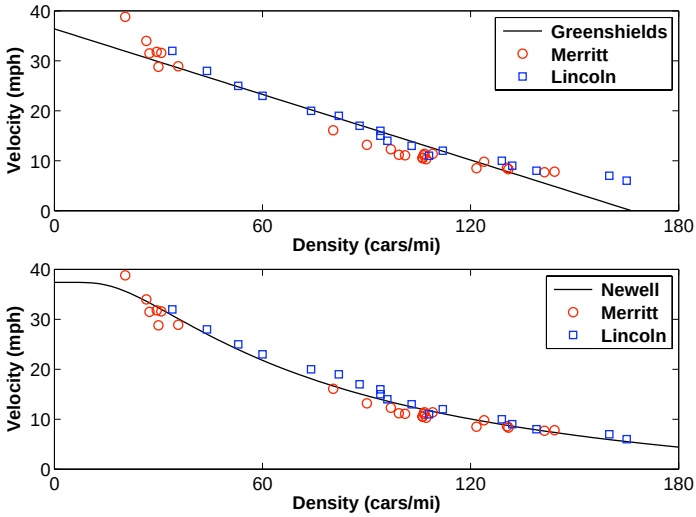


Figure 5.7 Curve fit of the Greenshields law (5.10) and the Newell law (5.17) to traffic data for the Merritt Parkway and the Lincoln Tunnel.

$$\frac{\partial \rho}{\partial t} + c(\rho)\frac{\partial \rho}{\partial x} = 0, \tag{5.11}$$

where

$$c = v_M \left(1 - \frac{2\rho}{\rho_M}\right). \tag{5.12}$$

This is a nonlinear conservation equation for ρ. It can be solved analytically, but it is certainly more challenging than the linear equation in (5.9). We will return to this problem once we have worked out the constant velocity case later in this chapter.

5.4.3 General Velocity Formulation

It is clear from the data in Figure 5.6 that the relationship between the velocity and density is not linear. In certain applications these differences are considered significant, and a more accurate function is needed. The general version of the constitutive law in this case has the form

$$v = F(\rho). \tag{5.13}$$

With this, the general formula for the flux is $J = \rho F(\rho)$. Assuming that F is a smooth function of ρ then, using the chain rule, it follows that $\frac{\partial}{\partial x} J = J'(\rho)\frac{\partial}{\partial x}\rho$. The general form of the balance law (5.5) now takes the form

$$\frac{\partial \rho}{\partial t} + c(\rho)\frac{\partial \rho}{\partial x} = 0, \tag{5.14}$$

where

$$c(\rho) = J'(\rho), \tag{5.15}$$

or equivalently

$$c(\rho) = F(\rho) + \rho F'(\rho). \tag{5.16}$$

The function $c(\rho)$ is known as the wave velocity, and it will play a critical role in the solution of the equation. A particular example of this function is given in (5.12), which is the wave velocity associated with the Greenshields constitutive law in (5.10).

It is not possible to use just any function for the constitutive law in (5.13). In particular, there are requirements that are needed to guarantee that (5.14) has a solution. These will become evident once we attempt to solve the problem. For the moment, we will take a more physical viewpoint, and impose conditions on the function $F(\rho)$ that are based on what is known about traffic flow. Interestingly, we will find that these physically based assumptions will overlap with the mathematical requirements needed to guarantee that the problem has a solution.

It has already been assumed that F is a smooth function of ρ. In addition to this, based on the data in Figure 5.6, the following assumptions are made.

NV1. $F'(\rho) \leq 0$ for $0 \leq \rho \leq \rho_M$.
This assumption comes from Figure 5.6 which shows v is a monotonically decreasing function of density. This requirement is consistent with the observation that (most) drivers leave a larger bumper-to-bumper spacing between cars as the speed increases. A consequence of this assumption is that $F(0) = v_M$ is the maximum velocity. This corresponds to the observation that on an uncongested highway, drivers tend to travel at the maximum allowable speed.

NV2. $F(\rho_M) = 0$.
This is based on the assumption that the closer the traffic gets to being bumper-to-bumper the closer the velocity gets to zero.

The list of functions that are capable of satisfying these rather general requirements is endless. It is for this reason that in selecting a particular function one should also consider simplicity. Given the uncertainty in the experimental data, and the approximate nature of the model, it is a waste of time to construct a function that hits every data point exactly. The problem is that the condition of simplicity, like beauty, is difficult to quantify. However, the above conditions require a function containing at least two parameters, namely v_M and ρ_M. The linear relationship in (5.10) is an example of a simple function with two parameters. Another possibility is the function proposed by Newell [1961], given as

$$v = v_M \left(1 - e^{-\lambda(1/\rho - 1/\rho_M)}\right). \tag{5.17}$$

Assuming that $\lambda \geq 0$, this is an example of a three parameter constitutive law that satisfies both NV1 and NV2. Fitting this to the data for the Lincoln Tunnel and Merritt Parkway one finds that $v_M = 37.4$ mph, $\rho_M = 271$ cars/mi, and $\lambda = 67.4$ mi/cars. The resulting function is plotted in Figure 5.7 along with the original data. It is evident that it is better than Greenshields at reproducing the data and, unlike the linear law, this function contains a plateau region near $\rho = 0$ that is seen in the Toronto and Amsterdam data in Figure 5.6. The penalty for this improvement is that the wave velocity, given in (5.16), is

$$c = v_M \left[1 - \left(1 + \frac{\lambda}{\rho}\right) e^{-\lambda(1/\rho - 1/\rho_M)}\right].$$

One therefore has to decide if the resulting complexity in the traffic flow equation (5.14) is worth the improvement in the data fit.

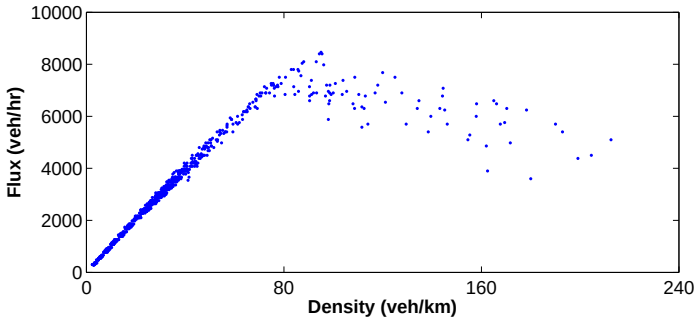


Figure 5.8 The flux as a function of the density measured on a freeway in Amsterdam (Aerde and Rakha [1995]).

5.4.4 Flux and Velocity

Our model has three dependent variables, flux, density, and velocity. Given that the equation of motion is written in terms of density and velocity the conventional approach is to propose a constitutive law that relates these two functions. However, it is worthwhile to consider other possibilities. One alternative is to relate the flux with the density using a constitutive law, and then use the equation $J = \rho v$ to determine the velocity. With this in mind the data in Figure 5.6 for the freeway in Amsterdam is plotted in Figure 5.8, giving the flux as a function of density. This is known as a fundamental diagram, and it is used extensively in developing traffic models. What is striking about this graph is that J has a well-defined dependence on ρ up to about $\rho = 80$ after which there is considerable scatter in the data. This spread is very typical of traffic flow, and it makes formulating a constitutive law for the flux problematic. In contrast, the v, ρ plots in Figure 5.6 show a more well-defined relationship over the entire density range, and for this reason it is more amenable to constitutive modeling.

One conclusion that can be made from Figure 5.8 is that the flux is concave down. From this we obtain an additional general rule for the general constitutive law $v = F(\rho)$, which is

NV3. $J''(\rho) \leq 0$, or equivalently, $2F'(\rho) + \rho F''(\rho) \leq 0$ for $0 \leq \rho \leq \rho_M$.

Recall that a smooth function is concave down if its derivative is monotone decreasing. Consequently, if the function $c(\rho) = J'(\rho)$ is monotone decreasing then the above condition is satisfied.

5.4.5 Reality Check

It is important to understand that even the most complex nonlinear expression relating the velocity and density is still, in the end, an approximation. Inevitably certain aspects of the problem are not accounted for, and many times this is intentional because the goal of the model is to capture the essential mechanisms responsible for the phenomena being studied. This has certainly been the case with the traffic flow problem. We have not included effects of intersections, inclement weather, adverse road conditions, or myriad other things that can influence traffic flow. There is also the problem that the cars are driven by people, who make individual decisions that can have dramatic effects on the traffic pattern. As a simple example, some drivers will speed up if there is lighter traffic ahead. This implies that the velocity depends on the density gradient. This is not accounted for in our model because we are assuming that the law has the form $v = F(\rho)$ and not $v = F(\rho, \rho_x)$. Some of the consequences of this extension are explored in Exercise 5.25. Generally, this sort of application is outside the scope of this textbook. However, a very humorous account of the role of human behavior, and how it affects traffic flow, can be found in Vanderbilt [2008].

A second comment that needs to be made is that the equation of motion (5.7) is general, and in terms of traffic flow can be applied to a multilane freeway or a small farm road. However, once a specific constitutive law for the velocity is introduced then the model becomes more limited in its applicability. For example, the traffic data given in Figure 5.6 measures the velocity on one side of the roadway (e.g., the velocity of the vehicles going east to west). This is reasonable because if both sides are counted, so the measured velocities can be either positive or negative, one could end up concluding that on average the velocity is zero at all density levels. In fact, it is not uncommon in traffic applications to have the constitutive law limited to a particular lane of traffic. For example, some roadways limit trucks to certain lanes of the roadway and this has a significant consequence for the velocity function. The point here is that the equation of motion is general but in applying it to particular problems, which requires the specification of a constitutive law, the model becomes more limited.

All of the above comments are evidence that we are studying a rich problem that has multiple research directions, and our model addresses one of them. Our objective is to understand how traffic flow behaves under the assumed conditions, and our next step is to figure out how to solve the mathematical problem we have produced.

5.5 Constant Velocity

To investigate the properties of the traffic flow problem we will begin with the assumption that the velocity is constant. The problem takes the form

$$\frac{\partial \rho}{\partial t} + a\frac{\partial \rho}{\partial x} = 0, \quad \text{for} \quad \begin{cases} -\infty < x < \infty \\ 0 < t, \end{cases} \tag{5.18}$$

where

$$\rho(x, 0) = f(x). \tag{5.19}$$

The partial differential equation (5.18) is known as the advection equation. The solution can be found if one notes that the equation can be written as

$$\left(\frac{\partial}{\partial t} + a\frac{\partial}{\partial x}\right)\rho = 0. \tag{5.20}$$

The idea is to transform x, t to new variables r, s in such a way that the derivatives transform as

$$\frac{\partial}{\partial r} = \frac{\partial}{\partial t} + a\frac{\partial}{\partial x}\,. \tag{5.21}$$

If this is possible then (5.20) becomes $\frac{\partial \rho}{\partial r} = 0$ and this equation is very easy to solve. With this goal in mind let $x = x(r, s), t = t(r, s)$, in which case using the chain rule the r-derivative transforms as

$$\frac{\partial}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial t}{\partial r}\frac{\partial}{\partial t}\,. \tag{5.22}$$

Comparing this with (5.20), we require $\frac{\partial x}{\partial r} = a$ and $\frac{\partial t}{\partial r} = 1$. Integrating these equations yields $x = ar + q(s)$ and $t = r + p(s)$. To determine the s dependence recall that the initial condition specifies the solution along the x-axis. To make it easy to apply the initial condition we will ask that the x-axis ($t = 0$) maps onto the s-axis ($r = 0$). In other words, $r = 0$ implies that $t = 0$ and $x = s$. Setting $r = 0$ and $t = 0$ we conclude $q(s) = s$ and $p(s) = 0$, and so, the change of variable we are looking for is

$$x = ar + s, t = r. \tag{5.23}$$

Inverting this transformation one finds that $r = t$ and $s = x - at$. We are now able to write (5.18) as $\frac{\partial \rho}{\partial r} = 0$, which means $\rho = \rho(s) = \rho(x - at)$. With the initial condition we therefore conclude that the solution of the problem is

$$\rho(x, t) = f(x - at). \tag{5.24}$$

Before making general conclusions about this solution we consider an example. This is worked out twice, first as a mathematical problem, and then as a problem in traffic flow.

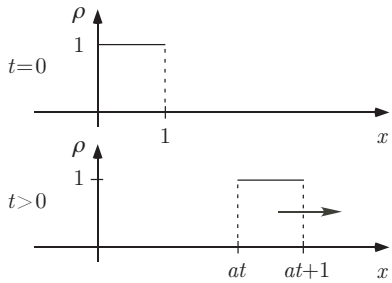


Figure 5.9 Solution of the advection equation (5.18). The top figure is the initial condition, as given in (5.27). The bottom figure is the solution at a later time, as given in (5.24).

Example: Mathematical Version

Suppose the initial condition is the square bump shown in Figure 5.9. In mathematical terms,

$$f(x) = \begin{cases} 1 & \text{if} \quad 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.25}$$

From (5.24) the solution is

$$\rho(x,t) = \begin{cases} 1 & \text{if} \quad 0 < x - at < 1, \\ 0 & \text{otherwise,} \end{cases}$$

or equivalently,

$$\rho(x,t) = \begin{cases} 1 & \text{if} \quad at < x < 1 + at, \\ 0 & \text{otherwise.} \end{cases} \tag{5.26}$$

A typical solution profile is also shown in Figure 5.9, and it is apparent that at any given time t, the solution is simply the original square bump that has moved over to occupy the interval $at \leq x \leq 1 + at$. ■

Example: Traffic Version

The previous example can be restated in physical terms. Suppose, at $t = 0$, that cars are uniformly spaced over the interval $0 < x < 1$, as shown in Figure 5.10. In this case the density has a constant, positive value for $0 < x < 1$, while the density outside this interval is zero. Also, assuming that each car travels with the same constant velocity a, then they will move as a unit. So, at any given time t, the group of cars will occupy the interval $at < x < at+1$. Because they are traveling at the same velocity, the spacing of the cars has not changed, and therefore the density in this interval is the same as it was at $t = 0$. This is the same result as obtained in the solution (5.26). ■

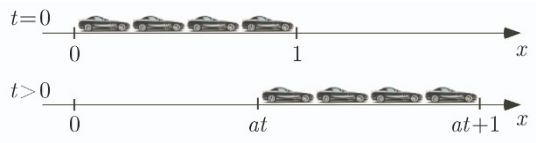


Figure 5.10 A uniformly spaced group of cars moves with constant velocity a along the x-axis.

In the above example, expressing the problem in terms of the motion of the individual cars is analogous to taking a microscopic point of view. In contrast, the macroscopic, or continuum, viewpoint is expressed in the solution given in (5.26). The attractive aspect of the microscopic point of view is that the solution is easy to understand, and it is obtained without having to solve a partial differential equation. Unfortunately, for more realistic problems, where the velocity depends on the density, the micro-scale version loses this advantage and the continuum problem becomes the easier one to solve.

From the above examples, and from the general formula in (5.24), we conclude that the solution is a traveling wave. The wave travels in only one direction, and for this reason (5.18) is sometimes called a one-way wave equation. In the case of when $a > 0$ the wave moves to the right with speed a. What is significant is that it moves at the same velocity as the vehicles, which, if you recall, is $v = a$. It might seem obvious that the wave moves with the vehicle velocity because, after all, the vehicles are responsible for the wave in the first place. However, the answer is not so simple. For example, the waves generated at sporting events by the fans in the audience are obtained not by the fans running around the stadium but, rather, by them periodically standing up and sitting down. Similarly, in heavy traffic if a car's taillights come on you will likely see a wave of taillights come on in the cars that follow. Not only is the wave of taillights not moving with the car's velocity, it is actually moving in the opposite direction. So, the connection between the motion of the constituents and the velocity of the wave requires some consideration. We will return to this point later when solving the problem of nonconstant velocity.

Another observation coming from the above example is that the shape and amplitude do not change as the wave travels along the x-axis. This is in marked contrast to the diffusion equation, where the corners or jumps in the initial condition are immediately smoothed out (see Figure 4.14). Because of this, one might question whether (5.26) is actually a solution since ρ_x is not defined at the jumps located at $x = at, 1 + at$. The short answer is that because there are only a finite number of jumps, everything is fine. What is necessary is to introduce the concept of a weak solution, and the interested reader is referred to Evans [1998] for an extended discussion of this subject. A slightly different approach to justifying the jumps, and understanding some of the difficulties of defining a continuum variable at a jump, are explored in Exercise 5.17.

5.5.1 Characteristics

There is another way to look at this solution that will prove to be particularly worthwhile. It is based on the observation that, from the formula $\rho(x,t) = f(x - at)$, if we hold $x - at$ fixed then the solution is constant. In other words, if $x - at = x_0$ then $\rho = f(x_0)$ along this line (see Figure 5.11). These lines are called characteristics for the equation, and the method we used to find the solution is called the method of characteristics. The observation that the solution is constant along the characteristics can be used to evaluate the solution anywhere in the x, t-plane. The next example illustrates how this is done.

Example

Suppose we want to determine $\rho(0,1)$. To use characteristics to find this value, we need to determine the line $x - at = x_0$ that passes through $(x,t) = (0,1)$ (see Figure 5.12). Plugging $x = 0$ and $t = 0$ into the equation $x - at = x_0$ we obtain $x_0 = -a$. Therefore, $\rho(0,1) = f(x_0) = f(-a)$. As it should, this result agrees with what is obtained from the formula given in (5.24). ■

In general, to determine $\rho(x_1, t_1)$ using characteristics, one first finds the characteristic that passes through (x_1, t_1). The equation for this line is $x - at = x_1 - at_1$. The solution is constant along this line, and because the x-intercept is $x_0 = x_1 - at_1$, it follows that $\rho(x_1, t_1) = f(x_0)$.

Example: Red Light - Green Light

As a second example of how the characteristics can be used to construct the solution, consider the situation of cars waiting at a stoplight. It is assumed that at $t = 0$ the light turns from red to green. We will locate the light at $x = 0$, and assume that at the start the cars have a constant density to

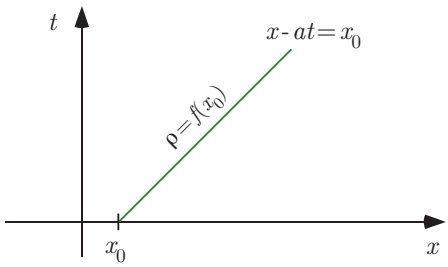


Figure 5.11 The characteristics for (5.18) are the straight lines $x - at = x_0$. Along each line the solution is constant.

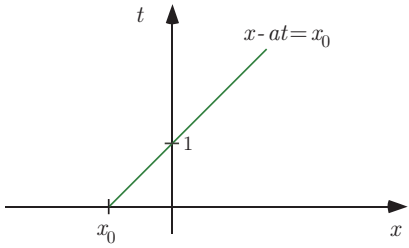


Figure 5.12 The characteristics used in the example to determine the value of $\rho(0,1)$.

the left of the light. The initial condition that will be used to describe this situation is

$$\rho(x,0) = \begin{cases} 1 & \text{if} \quad x \leq 0 \\ 0 & \text{if} \quad x > 0. \end{cases} \tag{5.27}$$

It is also assumed that $a > 0$. The characteristics for this problem are shown in Figure 5.13(a). Because of where the characteristics intersect the x-axis, the solution in the region covered by the solid lines is $\rho = 1$, while along the dashed lines the solution is $\rho = 0$. The characteristic that separates these two regions is the one that starts at the jump in the initial condition (5.27). Namely, it is the line $x = at$, and it is shown in Figure 5.13(a) using a line with small dots. The resulting solution is shown in Figure 5.13(b), and the corresponding formula is

$$\rho(x,t) = \begin{cases} 1 & \text{if} \quad x \leq at \\ 0 & \text{if} \quad x > at. \end{cases} \tag{5.28}$$

A somewhat more traditional view of the solution is given in Figure 5.13(c), where it is apparent that the solution consists of a wave that moves with speed a. ■

The two previous examples were used to introduce how characteristics can be used to find the solution, but in both cases the solution can be determined directly for the formula in (5.24). This is not true for the next, and final, example.

Example: Finite Length Highways

Up to this point our highways have been infinitely long. In the real world this is rather rare, and in this example we consider what happens when the road occupies the interval $0 \leq x \leq \ell$. This gives rise to the question as to what can, or should, be specified for boundary conditions at $x = 0, \ell$. A mathematically correct choice is to specify a boundary condition at $x = 0$ and not specify one at $x = \ell$. The reason is due to the fact that information in this problem goes

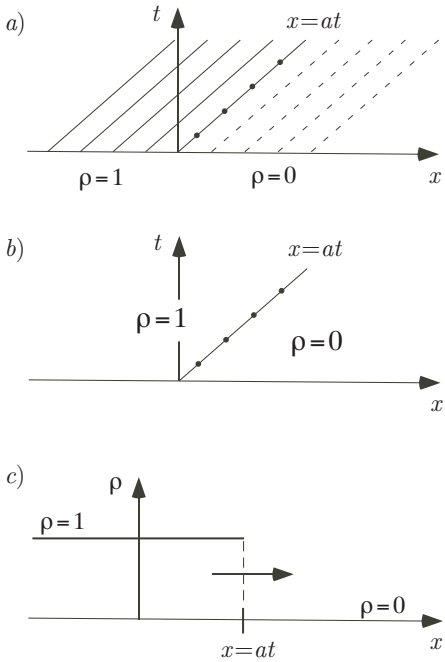


Figure 5.13 The solution of (5.18) when given the initial condition (5.27).

in only one direction, from left to right. Why this is important will become evident once we study the solution in more detail. To this end, we consider solving the equation

$$\frac{\partial \rho}{\partial t} + a\frac{\partial \rho}{\partial x} = 0, \quad \text{for} \quad \begin{cases} 0 < x < \ell \\ 0 < t, \end{cases} \tag{5.29}$$

along with the initial condition

$$\rho(x, 0) = f(x),$$

and the boundary condition

$$\rho(0, t) = g(t).$$

Using characteristics this is not hard to solve. We know that the solution of (5.29) is constant along any line of the form $x - at = const$ and these lines are shown in Figure 5.14. The analysis naturally separates into two components.

Solid Lines: In the region containing the characteristics that are solid lines, the solution is determined by the initial condition. Because the lines in

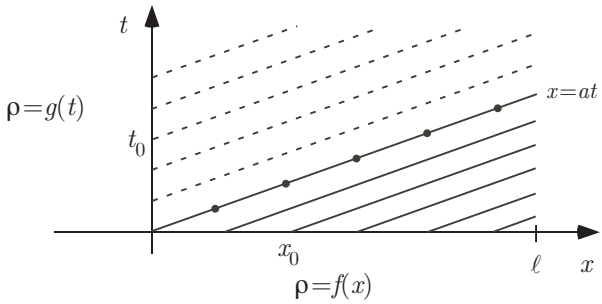


Figure 5.14 Characteristics used in solving the traffic flow problem over a finite interval.

this region have the form $x - at = x_0$, where x_0 is the x-intercept, then in this region the solution is $\rho(x,t) = f(x_0) = f(x - at)$.

Dashed Lines: To find the solution in the region where the characteristics are dashed lines, consider the characteristic shown in Figure 5.14 that has t-intercept t_0. The general form for the equation of this line is $x - at = constant$. Because it must pass through the point $(x,t) = (0,t_0)$, it follows that the equation is $x - at = -at_0$. Because the solution is constant along this line, and we are told that $\rho(0,t_0) = g(t_0)$, then it follows that along this characteristic $\rho(x,t) = g(t_0) = g(t - x/a)$.

Putting this information together, the solution is

$$\rho(x,t) = \begin{cases} f(x - at) & \text{if} \quad 0 \le t < x/a, \\ g(t - x/a) & \text{if} \quad x/a < t. \end{cases}$$

The value at $x = at$ depends on what value the function has at $(x,t) = (0,0)$. If $\rho(0,0) = f(0)$ then $\rho = f(0)$ for $x = at$, while if $\rho(0,0) = g(0)$ then $\rho = g(0)$ for $x = at$. ■

Returning to the question of whether it is possible to impose a boundary condition at $x = \ell$, suppose that $f(x) = 1$. In Figure 5.14, in the region covered with the solid lines the solution is $\rho = 1$. Any boundary condition imposed at $x = \ell$, other than $\rho = 1$, would be in contradiction to the known solution. That is why, in the case of when $a > 0$, it is more natural to impose a boundary condition at the left end of the interval. If one is insistent on specifying a boundary condition at $x = \ell$, it would then be necessary not to include either an initial condition or a boundary condition at $x = 0$. This idea is explored further in Exercise 5.9.

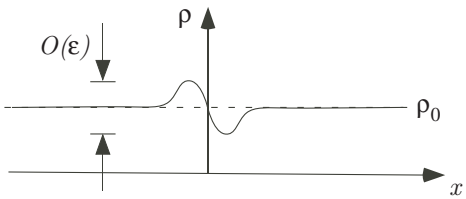


Figure 5.15 Small disturbance imposed onto constant density solution at $t = 0$. The resulting initial condition is given in (5.32)

5.6 Nonconstant Velocity

The linear wave equation studied in the previous section is a valuable source of information about some of the more basic properties of the solution. The fact is, however, the assumption that the velocity is independent of the density is not correct for traffic flow. This is evident in the data given in Figure 5.6. Precisely what constitutive law is used will be left unspecified for the moment other than to assume $v = F(\rho)$, where F is smooth. As shown in Section 5.4.3, the traffic flow equation takes the form

$$\frac{\partial \rho}{\partial t} + c(\rho)\frac{\partial \rho}{\partial x} = 0, \tag{5.30}$$

where the wave velocity is

$$c(\rho) = F + \rho F'. \tag{5.31}$$

Written this way the equation resembles the constant velocity version in (5.18) we studied earlier. One significant difference is that the wave velocity c can depend on the unknown ρ, and if this happens then (5.30) is nonlinear. Generally nonlinear partial differential equations are very difficult to solve. One option, which works on a wide variety of problems, is to introduce a small disturbance approximation, and this is discussed below. However, for this problem it is possible to solve the fully nonlinear equation using the method of characteristics and this will be considered in Section 5.6.2.

Although the nonlinear traffic flow equation is very general, a couple of restrictions are needed to help guarantee that there is a solution. One is that whatever function $c(\rho)$ is used in this equation, it is a smooth function of ρ. A second condition is related to the observation made in Section 5.4.4 that the flux is concave down. This is equivalent to $c(\rho)$ being a monotonically decreasing function of ρ. Mathematically, what is needed is that $c(\rho)$ is monotonic, either decreasing or increasing, and this is assumed in what follows.

5.6.1 Small Disturbance Approximation

One method for studying nonlinear wave problems is based on a small disturbance approximation. The basic idea is that a particular solution has been determined. This is usually an equilibrium solution, and it is very common that it is a constant. What is investigated is how small perturbations of this particular solution behave. To explain what this entails note that a constant function $\rho = \rho_0$ is a solution of the traffic flow equation (5.30). So, suppose that the traffic is flowing along smoothly with a uniform density $\rho = \rho_0$ and then one or more of the cars change speed slightly and cause a small perturbation in the density. For example if someone applies their brakes then the immediate affect will be to reduce the density in front of their car and to increase the density right behind them. A function that mimics this change in the density is shown in Figure 5.15.

To analyze this situation we will assume the disturbance occurs at $t = 0$. The initial condition that corresponds to this is

$$\rho(x, 0) = \rho_0 + \epsilon g(x). \tag{5.32}$$

The specific form of the function $g(x)$ is not important but we will illustrate the analysis using the example in Figure 5.15. Due to the initial condition the appropriate expansion for the solution is $\rho \sim \rho_0 + \epsilon\rho_1(x, t) + \cdots$. In this case, using Taylor's theorem,

$$\begin{aligned} c(\rho) &\sim c(\rho_0 + \epsilon\rho_1 + \cdots) \\ &\sim c(\rho_0) + (\epsilon\rho_1 + \cdots)c'(\rho_0) + \frac{1}{2}(\epsilon\rho_1 + \cdots)^2 c''(\rho_0) + \cdots \\ &\sim c(\rho_0) + \epsilon\rho_1 c'(\rho_0) + \cdots . \end{aligned}$$

The equation of motion (5.30) takes the form

$$\epsilon\frac{\partial\rho_1}{\partial t} + \cdots + [c(\rho_0) + \epsilon\rho_1 c'(\rho_0) + \cdots]\left(\epsilon\frac{\partial\rho_1}{\partial x} + \cdots\right) = 0, \tag{5.33}$$

where, from (5.32),

$$\rho_0 + \epsilon\rho_1(x, 0) + \cdots = \rho_0 + \epsilon g(x). \tag{5.34}$$

Setting $c_0 = c(\rho_0)$ then the $O(\epsilon)$ problem is

$$\frac{\partial\rho_1}{\partial t} + c_0\frac{\partial\rho_1}{\partial x} = 0, \tag{5.35}$$

where $\rho_1(x, 0) = g(x)$. This is known as the small disturbance equation for the problem and in this case it is a linear wave equation. Using (5.24), the solution is $\rho_1(x, t) = g(x - c_0 t)$. Therefore, the two term small disturbance

approximation of the solution is

$$\rho(x,t) \sim \rho_0 + \epsilon g(x - c_0 t). \tag{5.36}$$

It is clear from this that the initial disturbance propagates as a traveling wave, with velocity c_0. We will explore some of the consequences of this in the next example, but it is first necessary to comment on the accuracy of this approximation. If you compare (5.36) with, say, the numerical solution it is found that as time passes the approximation becomes less accurate. This is due to a slow change in the solution that is not accounted for in (5.36), and which over time starts to affect its accuracy. It is possible to use multiple scales, as described in Section 2.6, to improve the approximation. However, later in the chapter, after the nonlinear problem is solved, we will derive an exact solution of the problem.

Example: Phantom Traffic Jams

To investigate the properties of (5.36) we will use the Greenshields constitutive law and assume

$$v = v_M \left(1 - \frac{\rho}{\rho_M}\right). \tag{5.37}$$

In this case, from (5.31),

$$c = v_M \left(1 - \frac{2\rho}{\rho_M}\right), \tag{5.38}$$

and the flux is

$$J = v_M \left(1 - \frac{\rho}{\rho_M}\right) \rho. \tag{5.39}$$

These functions are sketched in Figure 5.16. Note that for a given value of the flux that there are two possible densities. Those satisfying $0 < \rho < \frac{1}{2}\rho_M$ are commonly referred to as light traffic while those satisfying $\frac{1}{2}\rho_M < \rho < \rho_M$ are heavy traffic. Also note that $c = J'$, in other words it equals the slope of the flux function. This means c is negative for lighter traffic and it is positive in heavier traffic.

Based on the above discussion, our conclusion is that in light traffic, where $c > 0$, the disturbance moves forward, and in heavy traffic, where $c < 0$, the disturbance moves backward. Given that $c \leq v$, the disturbance does not move faster than the flow of traffic. In other words, whoever was responsible for generating this disturbance would see it move backward relative to their position, but someone watching from an overpass would see it move forward in light traffic and move backward in heavy traffic. The one exception to this last statement is if the traffic density is $\rho_M/2$, in which case the disturbance would stay in the region where it was generated. Another point to make here is that, unlike the constant velocity example, the wave propagates at

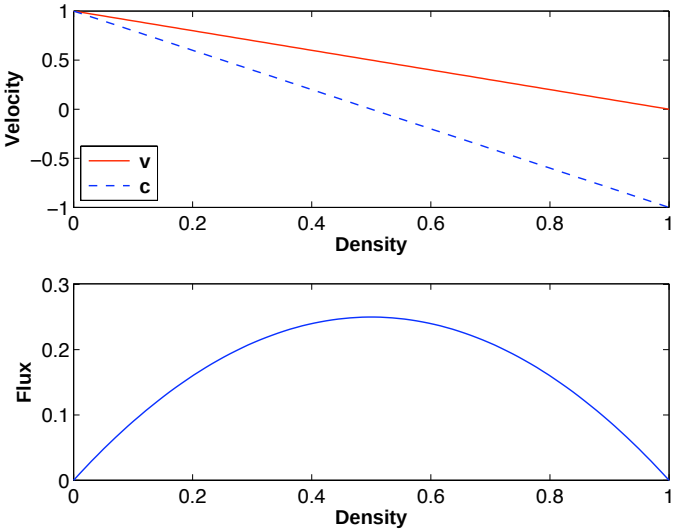


Figure 5.16 Velocities (5.37), (5.38), and the flux (5.39) when using the Greenshields law. In these plots $v_M = 1$ and $\rho_M = 1$.

a velocity that is different from the velocity of the vehicles that form the system.

The solution obtained using a small disturbance approximation provides an explanation of one of the mysteries of driving called the phantom traffic jam. This is the situation when there is no visible reason for a traffic slowdown, as there is no accident, construction, etc. As shown in Figure 5.17 some earlier perturbation in the traffic can result in a density wave propagating backwards along the highway. A driver who enters this region will see no apparent reason for its existence and once through the disturbance will return to the uniform flow they had earlier. One cause of such situations is weaving.

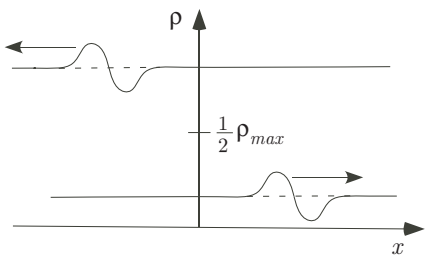


Figure 5.17 Disturbances move to the right if $\rho_0 < \frac{1}{2}\rho_M$ and move toward the left if $\frac{1}{2}\rho_M < \rho_0$. The signal velocity in both cases is $c_0 = c(\rho_0)$.

In heavier traffic drivers who change lanes frequently cause the drivers behind them to slow down or brake to leave room between them and the lane changer. This produces a small disturbance and this propagates along the highway behind the originators of this situation. ■

5.6.2 Method of Characteristics

As it turns out, the method of characteristics we developed to solve the constant velocity problem can be adapted so it also works on the nonlinear equation (5.30). In the constant velocity case, we found that the solution is constant along curves of the form $x = x_0 + at$. So, in a similar manner we will investigate if it is possible to find curves $x = X(t)$ on which the solution of (5.30) is constant. What we are looking for are curves with the property that $\frac{d}{dt}\rho(X(t), t) = 0$. Expanding this using the chain rule it follows that we need to select $X(t)$ in such a way that

$$\rho_t + X'(t)\rho_x = 0. \tag{5.40}$$

To find a function $X(t)$ that works in this equation, recall that ρ satisfies the traffic flow equation

$$\rho_t + c(\rho)\rho_x = 0. \tag{5.41}$$

Comparing this with (5.41) it is evident that $X(t)$ should be selected so that

$$X'(t) = c(\rho). \tag{5.42}$$

Before integrating to find the function $X(t)$, remember that ρ is constant along the curve. Consequently, if the curve begins at $x = x_0$ then at any point along the curve we have $\rho = \rho_0$ where $\rho_0 = f(x_0)$ (see Figure 5.18). Introducing this into (5.42), and integrating, we obtain $X = x_0 + c(\rho_0)t$. Therefore, the characteristic that begins at $x = x_0$ is

$$x = x_0 + c(\rho_0)t, \tag{5.43}$$

and along this characteristic the solution is

$$\rho = \rho_0, \tag{5.44}$$

where $\rho_0 = f(x_0)$. It might seem odd that the characteristics for a nonlinear equation turn out to be linear. However, the nonlinearity does have an affect as it determines the slope of the characteristics and, as we will see, this has major consequences on the solution.

The two expressions (5.43) and (5.44) form the solution of the problem. To explain how they are used, suppose one wants to calculate the value of ρ at a particular point, say at (x_1, t_1). In some cases, the value of $\rho(x_1, t_1)$ is

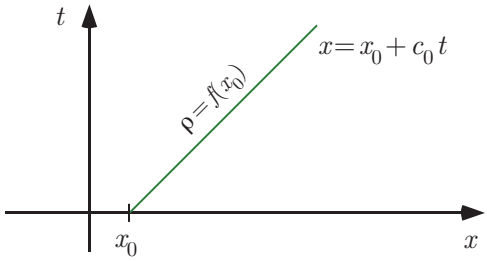


Figure 5.18 The method of characteristics involves finding the curves $x = X(t)$ along which the solution of (5.30) is constant.

easy to determine, and this happens in the next example when $\rho = \rho_L$ and when $\rho = \rho_R$. If the value is not obvious, then it is necessary to calculate the result, and this involves the following steps.

Step 1. Find the characteristic that passes through (x_1, t_1).
Given that the general form of the characteristic is $x - c_0 t = x_0$, then we require that $x_1 - c_0 t_1 = x_0$.

Step 2. Find c_0 in terms of x_0.
From the initial condition, we have that $c_0 = c(f(x_0))$. As an example, using the Greenshields law,

$$c(\rho_0) = v_M \left(1 - 2\frac{f(x_0)}{\rho_M}\right).$$

Step 3. Solve $x_1 = x_0 + c_0 t_1$ for x_0.
In the case of when the Greenshields law is used then the equation to solve is

$$x_1 = x_0 + v_M t_1 \left(1 - 2\frac{f(x_0)}{\rho_M}\right).$$

How difficult this equation is to solve for x_0 depends on $f(x_0)$. We will be using piecewise linear functions, so it is possible to solve the above equation relatively easily.

Once x_0 is known then the solution is $\rho(x_1, t_1) = f(x_0)$. This procedure is not particularly difficult, but it comes with caveats. In particular, it assumes that there is a characteristic passing through (x_1, t_1). As we will see shortly, this might not happen. We will postpone analyzing such difficulties until after we have more experience using the method when all goes according to plan.

Example: Modified Red Light - Green Light

To use the above solution for traffic flow we consider a modified version of the red light - green light problem. It is assumed that the traffic is initially

constant to the left of $x = -\epsilon$ and to the right of $x = \epsilon$. Also, there is a transition region, of width 2ϵ, where the density changes linearly between the left and right values. This situation is shown in Figure 5.19. It is assumed the faster cars are in the front, and so, $\rho_L > \rho_R$. The specific function used for the initial condition is

$$\rho(x,0) = \begin{cases} \rho_L & \text{if} \quad x \le -\epsilon \\ \rho_L + \frac{\rho_R - \rho_L}{2\epsilon}(x+\epsilon) & \text{if} \ -\epsilon < x < \epsilon \\ \rho_R & \text{if} \quad \epsilon \le x. \end{cases} \tag{5.45}$$

We also need to be specific about what constitutive law is being used for the velocity, and in what follows we use the Greenshields law. Consequently, $v = v_M(1 - \rho/\rho_M)$ and the wave velocity is

$$c(\rho) = v_M \left(1 - \frac{2\rho}{\rho_M}\right). \tag{5.46}$$

To sketch the characteristics, we consider what happens for different starting positions x_0.

- If x_0 is on the left, so $x_0 < -\epsilon$, then ρ_0 has the constant value ρ_L. This means that the characteristics in this region all have the same slope, and this is shown in Figure 5.20(a). Given that the solution is constant along each of these lines it follows that $\rho = \rho_L$ in the region of the x,t-plane to the left of the characteristic $x = -\epsilon + c_L t$, where $c_L = c(\rho_L)$. This is shown in Figure 5.20(b).
- Using a similar argument, the characteristics that start on the right, where $x_0 > \epsilon$, all have the same slope. Because $\rho_L > \rho_R$ then the characteristics on the left have a steeper slope than those on the right, and this is shown in Figure 5.20(a). The solution is constant along each of these lines, and so it follows that $\rho = \rho_R$ in the region of the x,t-plane to the right of the characteristic $x = \epsilon + c_R t$, where $c_R = c(\rho_R)$. This is shown in Figure 5.20(b).
- To determine what happens when $-\epsilon < x_0 < \epsilon$, it is seen in Figure 5.19 that the initial density is continuous over this interval. This means that $c(\rho_0)$ varies continuously from c_L at $x_0 = -\epsilon$, to c_R at $x_0 = \epsilon$. The resulting

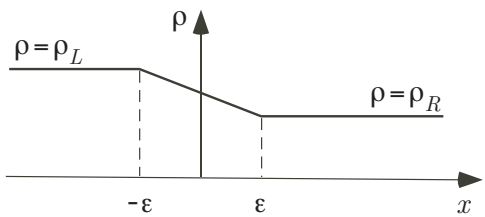


Figure 5.19 Initial density $\rho(x,0)$ for the modified red light - green light problem.

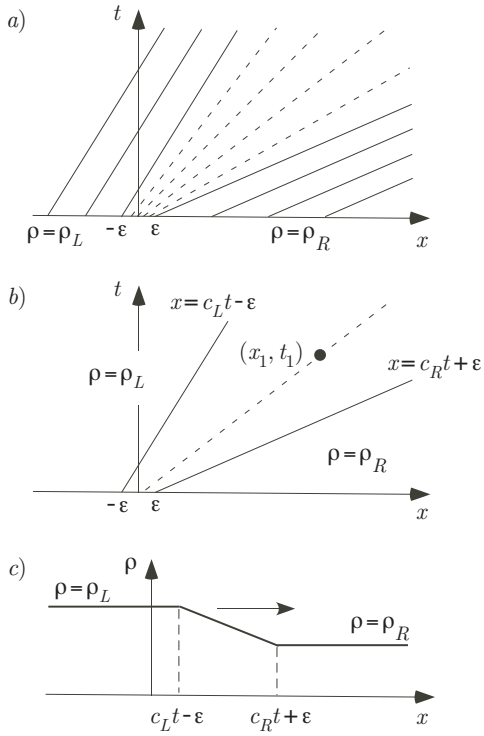


Figure 5.20 The solution of the modified red light - green light problem. The width of the linear transition region between the left and right constant states increases with time because $c_L < c_R$.

characteristics are shown in Figure 5.20(a) using dashed lines. To find the solution at a point (x_1, t_1) in this region, as illustrated in 5.20(b), we need to find the characteristic that passes through this point. This requires finding x_0. Because the density is constant on the characteristic, once x_0 is known then $\rho(x_1, t_1) = \rho(x_0, 0)$. Now, the general formula for the characteristics is $x = x_0 + c_0 t$, and so it is required that $x_1 = x_0 + c_0 t_1$. Given (5.46) and (5.45) we have that

$$\begin{aligned} c_0 &= v_M \left(1 - \frac{2\rho_0}{\rho_M}\right) \\ &= v_M \left[1 - \frac{2}{\rho_M}\left(\rho_L + \frac{\rho_R - \rho_L}{2\epsilon}(x_0 + \epsilon)\right)\right]. \end{aligned}$$

Substituting this into the equation $x_1 = x_0 + c_0 t_1$ and then solving for x_0 one finds that

$$x_0 = \frac{x_1 - t_1(c_L + c_R)/2}{1 + t_1(c_R - c_L)/(2\epsilon)}.$$

With this, and the initial condition in (5.45), the density is

$$\begin{aligned}\rho(x_1, t_1) &= \rho(x_0, 0) \\ &= \rho_L + (\rho_R - \rho_L)\frac{x_1 + \epsilon - c_L t_1}{2\epsilon + (c_R - c_L)t_1}.\end{aligned} \tag{5.47}$$

The formula for the solution is therefore

$$\rho(x,t) = \begin{cases} \rho_L & \text{if} \quad x \le c_L t - \epsilon \\ \rho_L + (\rho_R - \rho_L)\frac{x+\epsilon-c_L t}{2\epsilon+(c_R-c_L)t} & \text{if} \quad c_L t - \epsilon < x < c_R t + \epsilon \\ \rho_R & \text{if} \quad c_R t + \epsilon \le x. \end{cases} \tag{5.48}$$

According to this, between the two constant states the density varies linearly, just as it did in the initial condition. There is nothing unusual in this solution as it shows the expected result that the slower group on the left gradually separates from the faster group on the right. This is illustrated in Figure 5.20(c). ■

5.6.3 Rankine-Hugoniot Condition

As will become evident as we study the nonlinear traffic flow equation in Section 5.6.5, the solution has a propensity to evolve into a function with one or more jump discontinuities that move along the x-axis. We studied such a solution with the red light-green light problem for the linear equation, and the result is shown in Figure 5.13. The nonlinear equation is a different animal, and we are going to have to be a bit more careful any time a jump is present. To investigate what happens, suppose we have a situation as shown in Figure 5.21, which consists of a jump that is located at $x = s(t)$. Given that x-derivatives are not defined at such points we will reformulate the problem by integrating over a small spatial interval, $s - \epsilon \le x \le s + \epsilon$, around the jump. So, integrating $\rho_t + J_x = 0$ and remembering that the density is constant on either side of the jump we obtain

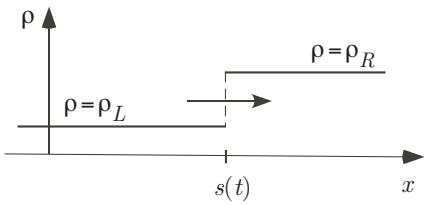


Figure 5.21 A jump discontinuity in the solution, located at $x = s(t)$.

$$\int_{s-\epsilon}^{s+\epsilon} \rho_t dx + J(\rho_R) - J(\rho_L) = 0. \tag{5.49}$$

From the Fundamental Theorem of Calculus recall that

$$\frac{d}{dt}\int_{s-\epsilon}^{s+\epsilon} \rho dx = \int_{s-\epsilon}^{s+\epsilon} \rho_t dx + s'(t)\rho|_{x=s+\epsilon} - s'(t)\rho|_{x=s-\epsilon}.$$

From this and (5.49) it follows that

$$\frac{d}{dt}\int_{s-\epsilon}^{s+\epsilon} \rho dx - \rho_R s'(t) + \rho_L s'(t) + J(\rho_R) - J(\rho_L) = 0. \tag{5.50}$$

Now, using the piecewise constant nature of the density

$$\begin{aligned}\int_{s-\epsilon}^{s+\epsilon} \rho dx &= \int_{s-\epsilon}^{s} \rho dx + \int_{s}^{s+\epsilon} \rho dx \\ &= \epsilon(\rho_L + \rho_R),\end{aligned}$$

and so

$$\frac{d}{dt}\int_{s-\epsilon}^{s+\epsilon} \rho dx = 0.$$

It follows from (5.50) that

$$s'(t) = \frac{J(\rho_R) - J(\rho_L)}{\rho_R - \rho_L}. \tag{5.51}$$

This equation is known as the *Rankine-Hugoniot condition* and it determines the velocity of a jump discontinuity in the solution.

It is useful to express (5.51) in terms of the wave velocity function c. Recalling that $c = J'(\rho)$, and $J(0) = 0$, then

$$J(\rho) = \int_0^{\rho} c(\bar{\rho}) d\bar{\rho}. \tag{5.52}$$

With this, the Rankine-Hugoniot condition takes the form

$$s'(t) = \frac{1}{\rho_R - \rho_L}\int_{\rho_L}^{\rho_R} c(\rho) d\rho. \tag{5.53}$$

This is an interesting result as it shows that any jump in the solution travels at the wave velocity averaged over the given density interval.

There are two types of jumps, and they are determined by what happens to the velocity v at the jump. If ρ has a jump discontinuity at $x = s(t)$, but v is continuous at $x = s(t)$, then the jump is called a contact discontinuity. An example is the red light-green light solution shown in Figure 5.13. The velocity is constant, hence it is continuous no matter where the jumps occur.

Note that because $v = a$ and $J = \rho v$ then the Rankine-Hugoniot condition (5.51) reduces to $s' = a$. In other words, the jumps move with the given constant velocity, and this is what was determined in Figure 5.13.

If v is not continuous at $x = s(t)$ then the jump is called a shock, and the motion of this jump produces a *shock wave*. As shown in the next examples, the velocity of the shock is strongly dependent on the constitutive law.

Examples

1. Greenshields Law. Using the linear law in (5.10), and the fact that $J = \rho v$, then the Rankine-Hugoniot condition (5.51) simplifies to the following

$$\begin{aligned} s'(t) &= \frac{1}{\rho_R - \rho_L}\left[\rho_R v_M\left(1 - \frac{\rho_R}{\rho_M}\right) - \rho_L v_M\left(1 - \frac{\rho_L}{\rho_M}\right)\right] \\ &= v_M\left(1 - \frac{1}{\rho_M}(\rho_R + \rho_L)\right) \\ &= \frac{1}{2}\left(c_R + c_L\right). \end{aligned} \tag{5.54}$$

In other words, when using the Greenshields law, the shock moves at a speed determined by the average of the jump in the wave velocity across the shock. ■

2. Newell Law. Using (5.17) then the Rankine-Hugoniot condition (5.51) is

$$\begin{aligned} s'(t) &= \frac{1}{\rho_R - \rho_L}\left[\rho_R v_M\left(1 - e_R\right) - \rho_L v_M\left(1 - e_L\right)\right] \\ &= v_M\left(1 - \frac{\rho_R e_R - \rho_L e_L}{\rho_R - \rho_L}\right), \end{aligned} \tag{5.55}$$

where

$$\begin{aligned} e_L &= e^{-\lambda(1/\rho_L - 1/\rho_M)}, \\ e_R &= e^{-\lambda(1/\rho_R - 1/\rho_M)}. \quad ■ \end{aligned}$$

When we first started out studying traffic flow, we had only one variable with the dimension of velocity. Now, we have three variables with this dimension. They are:

1. $v(x,t)$. This is the velocity of the car located at x at time t.

2. $c(\rho)$. This is the wave velocity, and it is defined in (5.31). It determines the slopes of the characteristic curves.

3. $s'(t)$. This is the velocity of the jumps in the solution, and it is defined in (5.51).

These velocities all play a critical role in the evolution of the solution and are distinct in the sense that, in most nonlinear problems, they are not simple multiples of each other. This is evident in the definitions of c and s', as well as from the expressions derived in Exercise 5.12. What we conclude from this is that this interesting problem is rich enough that a single velocity is not enough to describe the solution.

5.6.4 Expansion Fan

Now that we have some idea what happens when jumps occur in the solution, we will investigate a problem that starts out with a jump. The initial condition is shown in Figure 5.22, and it is given as

$$\rho(x,0) = \begin{cases} \rho_L & \text{if } x \leq 0, \\ \rho_R & \text{if } 0 < x, \end{cases} \tag{5.56}$$

where $0 < \rho_R < \rho_L$. This piecewise constant function gives rise to what is known as a Riemann problem. This problem is interesting because the solution is not obvious. In fact, it is so unclear that it is possible to produce a plausible argument for at least three different solutions. Before describing what these are we first state what we are certain of about the solution. This comes from the characteristics, and these are shown in Figure 5.23(a). As illustrated in Figure 5.23(b),(c), we conclude that $\rho = \rho_L$ for $x < c_L t$ and $\rho = \rho_R$ for $x > c_R t$. This leaves unresolved what the solution is for $c_L t < x < c_R t$ because there are no characteristics in this region. It is what happens in this sector that produces the three possible solutions.

1. The cars starting on the left, where $x < 0$, travel with velocity v_L while those on the right have velocity v_R. Because $v_L < v_R$ then one might argue, based on physical grounds, that the sector in question is nothing

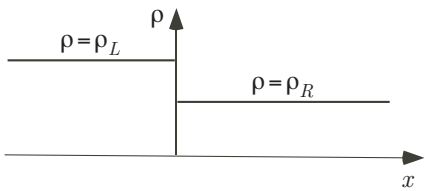


Figure 5.22 Initial density $\rho(x,0)$, where the slow cars start out behind the faster cars (i.e., $\rho_L > \rho_R$).

more than the gap between the slow cars on the left and the fast cars on the right. In other words, for points in this sector the density is just zero and the apparent solution is

$$\rho(x,t) = \begin{cases} \rho_L & \text{if} \quad x \le v_L t, \\ 0 & \text{if } v_L t < x < v_R t, \\ \rho_R & \text{if} \quad v_R t \le x. \end{cases} \tag{5.57}$$

The first indication that there is something wrong with this expression is that the sector is determined by the velocity of the cars, and not the wave velocity. This is a problem because $c(\rho_R) < v(\rho_R)$ and $c(\rho_L) < v(\rho_L)$, so the sector in (5.57) is different from the one shown in Figure 5.23. In other words, the above expression contradicts what we certain of, and that is the solution shown in Figure 5.23(c). Therefore, (5.57) is not the solution. Those who rely on physically motivated arguments to explain what is happening mathematically will almost certainly complain about this result. The reason is that the solution does not agree with what is expected

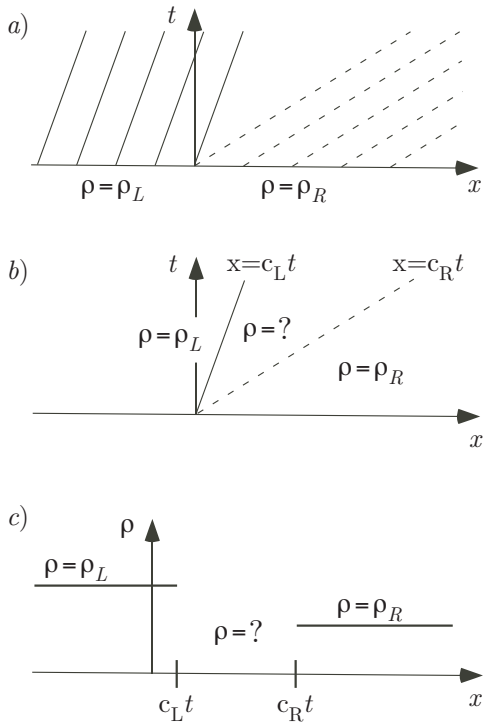


Figure 5.23 The solution obtained using the method of characteristics when the initial density is given in Figure 5.22. As shown in (a) and (b), there are no characteristics in the sector $c_L t < x < c_R t$, and so the solution in that region is unclear.

in the physical problem. More precisely, it does not agree with what might be expected based on a cursory analysis of the situation.

2. As another attempt at finding out what happens in the sector one might argue that the solution of the linear traffic flow equation (5.18), using the initial condition in (5.56), is a traveling wave with a single jump that moves with velocity a. Assuming the nonlinear equation also produces a single jump then the apparent solution is

$$\rho(x,t) = \begin{cases} \rho_L & \text{if } x \le s(t), \\ \rho_R & \text{if } s(t) < x. \end{cases} \tag{5.58}$$

The function $s(t)$ is determined from the Rankine-Hugoniot condition (5.53). Although it is not clear whether (5.58) is the solution, it has promise. For example, it is not hard to show that the line $x = s(t)$ is between $x = c_L t$ and $x = c_R t$. This means that (5.58) agrees with what we already know using characteristics, unlike what happened with (5.57). Moreover, in the special case of when c is constant, (5.58) reduces to the correct solution of the linear problem. These two observations are encouraging, but they do not guarantee that (5.58) is the solution of the Riemann problem we are studying.

3. A third attempt at finding the solution makes use of the modified red-light green-light problem shown in Figure 5.20. The solution of this modified problem should converge to the solution of our Riemann problem when $\epsilon \to 0$. This, in effect, takes the dashed characteristics in Figure 5.20 and pinches them together at the origin with the result shown in Figure 5.23. The radial characteristics form what is known as an *expansion fan*, or rarefaction wave, and it connects the constant states on the left and right. The formula for the solution, which is obtained from (5.48), is

$$\rho(x,t) = \begin{cases} \rho_L & \text{if} \quad x \le c_L t, \\ \rho_L + (\rho_R - \rho_L)\frac{x - c_L t}{(c_R - c_L)t} & \text{if } c_L t < x < c_R t, \\ \rho_R & \text{if} \quad c_R t \le x. \end{cases} \tag{5.59}$$

The resulting solution looks much like the one in Figure 5.20(c) in the sense that the expansion fan is responsible for a linear transition between the constant solutions on the left and right.

From the above discussion we have two contenders for the solution, namely (5.58) and (5.59). The fact that we have multiple possible solutions is because the nonlinear traffic flow problem is ill-posed, which in this case means that the problem is incomplete. What is required is an additional piece of information that will enable us to uniquely determine the solution. Moreover, it must be consistent with the physics of the problem. As an example, equations like the one we are dealing with arise in gas dynamics, and the approach used

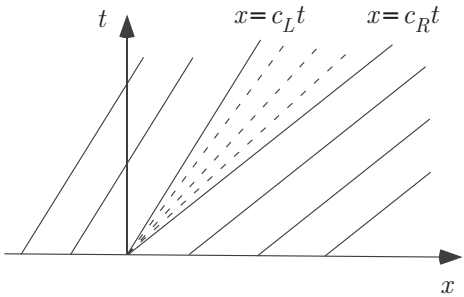


Figure 5.24 By letting $\epsilon \to 0$ the dashed characteristics in Figure 5.20 form an expansion fan between $x = c_L t$ and $x = c_R t$.

there is to introduce entropy and then employ the second law of thermodynamics to derive the needed condition. An effort has been made to define a concept similar to entropy for traffic flow, what is known as "driver's ride impulse," and then use a second law type of argument (Ansorge [1990]). We will take a different tack, and use a more mathematical argument.

The assumption is one of continuity. Namely, the jump appearing in the initial condition is almost impossible to produce physically, and in most experiments there is not a jump, but a small interval where the density changes in a rapid and continuous fashion from ρ_L to ρ_R. In this sense the initial condition containing a jump is simply a mathematical idealization of the true situation. Given that the solution with a continuous, but rapid, transition is known and given in Figure 5.20, the condition we are searching for must be consistent with this result. In other words, the condition must be able to tell us that (5.59) is the solution to this problem.

There are various ways to write the needed condition, and we will use the one introduced by Lax [1973]. The statement is that if the solution contains a jump, at $x = s(t)$, then the wave speed behind the jump is larger than the wave speed in front of it. In other words, the requirement is

$$c(\rho_R) < s' < c(\rho_L). \tag{5.60}$$

This is an example of what is known as an *admissibility condition*, because it provides the necessary information to determine the physically or mathematically admissible solution. In traffic flow it is often called the entropy condition, even though its connection to entropy is not at all clear for traffic problems.

One immediate consequence of the admissibility condition (5.60) is that the solution will only contain a jump if $c_L > c_R$. For our initial condition, given in (5.56), the assumption is that $c_L < c_R$. Therefore, a solution with a jump is not possible, and the solution in the region in question is an expansion

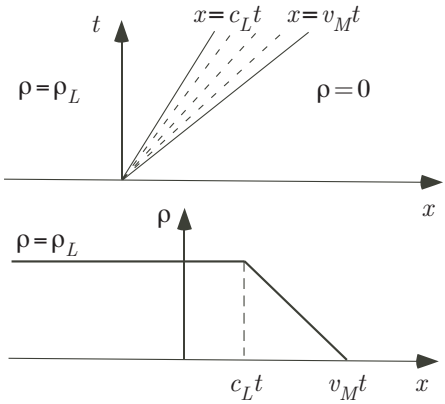


Figure 5.25 The upper plot shows the solution on the left and right, and the characteristics for the expansion fan. The lower plot shows the solution after the light turns green.

fan. In other words, (5.59) is the solution of the stated Riemann problem. The proof of this statement can be found in Lax [1973].

Example: Red Light - Green Light

Suppose a stoplight is located at $x = 0$, and it turns from red to green at $t = 0$. Also, assume that the light was red for so long that there are no cars on the right. In other words, the initial condition is

$$\rho(x,0) = \begin{cases} \rho_L & \text{if } x \le 0, \\ 0 & \text{if } 0 < x. \end{cases} \tag{5.61}$$

From (5.59), the solution of the traffic flow equation is

$$\rho(x,t) = \begin{cases} \rho_L & \text{if} \quad x \le c_L t, \\ \rho_L \frac{v_M t - x}{(v_M - c_L)t} & \text{if } c_L t < x < v_M t, \\ 0 & \text{if} \quad v_M t \le x. \end{cases} \tag{5.62}$$

The solution is shown in Figure 5.25, along with the associated characteristic curves. This shows that once the light turns green the cars move to the right, with the front moving at the maximum allowable velocity v_M. ■

The exact form of the expansion fan solution (5.59) relies on the specific formula for the wave velocity $c(\rho)$. In general, a fan appears when there is a gap between characteristics as shown in Figure 5.24. This occurs when $f(x)$ has a jump at a point $x = x_0$, with $c_R > c_L$ (see Figure 5.26). The equation for each of the dashed lines making up the fan has the form $x = x_0 + c(\rho)t$, where $c(\rho)$ satisfies $c_L < c < c_R$. There are a couple of methods that can

be used to prove this, other than taking a limit as we did earlier, and one is explored in Exercise 5.20. To determine the density at a point (x, t) in the fan, it is necessary to solve the equation $c(\rho) = (x - x_0)/t$ for ρ. This is where the specific form of c affects the solution, and (5.59) is what is obtained when using the Greenshields law. Also, in formulating the nonlinear traffic flow equation in Section 5.6, we made the assumption that $c(\rho)$ is monotonic. This is one of the places where we need that assumption because it guarantees that $c(\rho) = (x - x_0)/t$ has a unique solution.

After reading the above paragraphs one might decide that the best thing to do is avoid using an initial condition with a jump. After all, when using the continuous function in (5.45) the characteristics worked without complication and there was no doubt about the solution. However, as we will see in the next section, this nonlinear equation can take a continuous initial condition and cause it to form jumps. So, even if we do not feed it jumps at the beginning it can easily grow its own and this means there is no avoiding having to consider an admissibility condition.

As a final comment, the admissibility condition that should be used in traffic flow is a topic that continues to receive attention in the research literature. One question is whether the entropy based conditions that are used in gas dynamics are applicable in traffic problems, particularly those that involve unusual flux functions. An example of an unusual function is one that is not convex. Those who are interested in investigating this topic should consult Ansorge [1990], Velan and Florian [2002], Gasser [2003], and Knowles [2008].

5.6.5 Shock Waves

As stated earlier, at a shock wave both the density and velocity are discontinuous. Calling the solution shown in Figure 5.21 a shock wave gives the impression the cars are running into each other. They are not, and what happens when the shock passes over a car is that it immediately undergoes

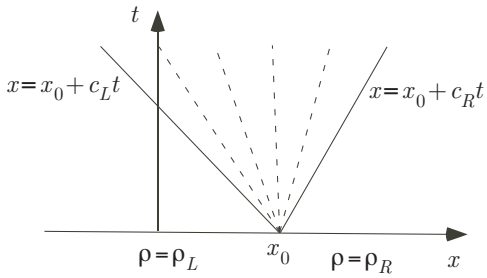


Figure 5.26 An expansion fan is generated at a point x_0 where the initial function $f(x)$ has a jump discontinuity, with $c_R > c_L$.

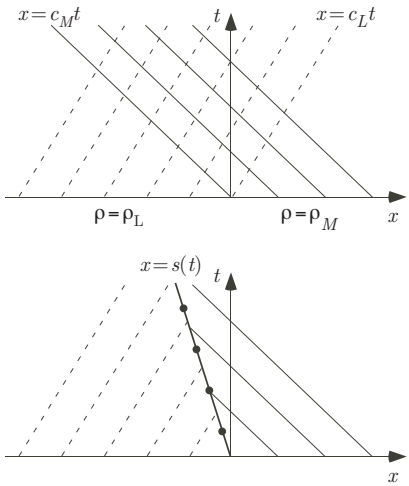


Figure 5.27 The traffic jam problem. The upper plot shows the characteristics associated with the initial condition. The lower plot shows the resulting shock location.

a jump in velocity. This is a bit unrealistic, and we will return to this point later.

Characteristics are used to determine when a shock wave is present in the solution. In contrast to an expansion fan, a shock appears when characteristics overlap, and the values on the characteristics are not equal. The easiest way to explain this is to work through a couple of examples.

Example: Traffic Jam

The first shock solution we will consider involves a traffic jam. Suppose that at the start, for $x \geq 0$, the density is ρ_M. This is the maximum density and means the cars can not move. For the interval $x < 0$ we will assume that the cars have density ρ_L, where $0 < \rho_L < \rho_M$. This means that the cars on the left move right with a constant velocity, in the direction of the traffic jam. Once they reach the jam the cars stop, and the result is that the traffic jam spreads leftward along the negative x-axis. To quantify these statements, the characteristics are shown in Figure 5.27. In the upper graph, along the solid lines the density is $\rho = \rho_M$ while along the dashed lines $\rho = \rho_L$. Clearly, there is a problem in the region where the characteristics overlap. The conclusion is that there is a curve $x = s(t)$ in this overlap region where the solution jumps from ρ_L to ρ_M. The resulting characteristics, and shock curve, are shown in the lower graph in Figure 5.27. The location of the shock, according to (5.53), moves with a velocity determined by an averaged value of the wave speed. Using the Greenshields law, the formula for the velocity is given in (5.55). Given that $c_L = v_M(1 - 2\rho_L/\rho_M)$ and $c_R = -v_M$ then

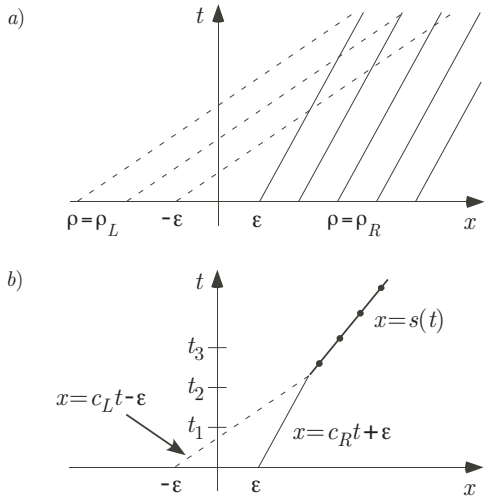


Figure 5.28 Overlapping characteristics are shown in (a), which indicates the existence of a shock wave in this region. The position of the shock is shown in (b), along with the two characteristics that intersect to initiate the formation of the shock at $t = t_s$.

$$s'(t) = -v_M \frac{\rho_L}{\rho_M} \,. \tag{5.63}$$

Integrating this, and using the fact that the shock starts at $(x, t) = (0, 0)$, we have that the position of the shock is given as

$$s(t) = -v_M \frac{\rho_L}{\rho_M} t. \tag{5.64}$$

With this the solution is

$$\rho(x, t) = \begin{cases} \rho_L & \text{if } x < s(t), \\ \rho_M & \text{if } s(t) \leq x. \end{cases} \tag{5.65}$$

As one final comment, it is important to point out that this solution satisfies the admissibility condition (5.60). This is because $c_L = v_M(1 - 2\rho_L/\rho_M)$, $c_R = -v_M$, and $\rho_L < \rho_M$. ■

Example: No Initial Jumps

As a second example suppose the density does not begin with a jump, but is continuous and has the form in (5.48). Now, however, we place the faster cars on the left so $\rho_L < \rho_R$. As usual, we will use the Greenshields law. The characteristics that are produced by these two constant values are shown in Figure 5.28(a). In the region covered by the dashed lines the solution is

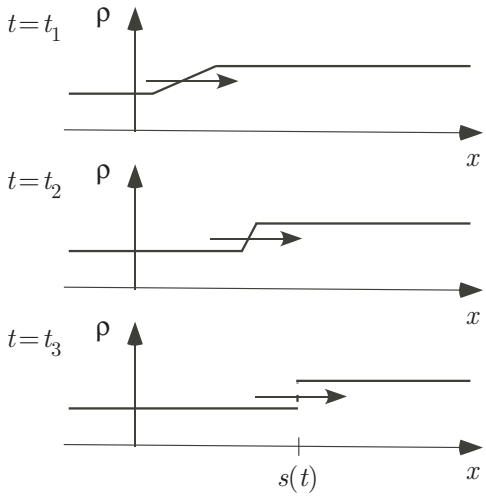


Figure 5.29 The solution of the traffic flow problem at the times shown in Figure 5.20(b). The width of the linear transition region between the left and right groups decreases with time until the left group catches the right group, and that time a shock wave appears.

$\rho = \rho_L$, while in the region covered by the solid lines the solution is $\rho = \rho_R$. The exception to this statement is where the dashed and solid lines overlap. In this region there is a shock wave, that begins where the characteristic $x = -\epsilon + c_L t$ intersects the characteristic $x = \epsilon + c_R t$. This intersection point is (x_s, t_s), where $t_s = 2\epsilon/(c_L - c_R)$ and $x_s = c_R t_s + \epsilon$, and the shock is shown in Figure 5.28(b). To determine the equation of this curve, we have from (5.55) that $s' = \frac{1}{2}(c_L + c_R)$. Integrating this equation yields

$$s(t) = c_s(t - t_s) + x_s, \tag{5.66}$$

where $c_s = \frac{1}{2}(c_L + c_R)$. It remains to determine the solution in the triangular region shown in Figure 5.28(b), which is bounded by the characteristics $x = -\epsilon + c_L t$ and $x = \epsilon + c_R t$. This is the same problem as finding the solution at (x_1, t_1) in Figure 5.20(b), and the solution is given in (5.47). Assembling all of this information, we therefore have that the solution for $t < t_s$ is

$$\rho(x,t) = \begin{cases} \rho_L & \text{if} \quad x \le c_L t - \epsilon, \\ \rho_L + \frac{\rho_R - \rho_L}{2\epsilon}(x + \epsilon) & \text{if } c_L t - \epsilon < x < c_R t + \epsilon, \\ \rho_R & \text{if} \quad c_R t + \epsilon \le x, \end{cases} \tag{5.67}$$

and for $t \ge t_s$ the solution is

$$\rho(x,t) = \begin{cases} \rho_L & \text{if } x < s(t), \\ \rho_R & \text{if } s(t) < x. \end{cases} \tag{5.68}$$

This solution is shown in Figure 5.29 for the three time values shown in Figure 5.20(b). At $t = t_1$ the solution consists of the two constant densities that are connected by a linear function. Because the cars on the left are faster than those on the right, at the later time $t = t_2$ the linear connection between the two densities has been reduced considerably. The effect of this transition region shrinking is to steepen the wave as it moves. The faster cars catch the slower ones in front, at $t = t_s$, at which point a shock forms. This is seen at time $t = t_3$, which shows the solution after the shock has formed. ■

The properties of the solution at a shock wave brings out one of the flaws in the traffic model. Specifically, as a shock passes over a car it immediately undergoes a jump in velocity. This is unrealistic and the reason it happens is that the model does not account for the momentum of the cars. Related to this is the assumption implicit in the constitutive law $v = F(\rho)$. For this to hold, the velocity must instantly adjust to the value of the density. This means that it is impossible to have the cars start from rest unless the density is at its maximum value of ρ_M. There are traffic models that account for the acceleration of the cars, and one is the cellular automata model studied later in the chapter. Also, in the next chapter we will significantly extend the continuum model in such a way that momentum is a central component of the model.

5.6.6 Return of Phantom Traffic Jams

The last example we will work out is the problem that introduced the phenomenon of a phantom traffic jam. The initial condition used here is

$$\rho(x,0) = \begin{cases} \rho_0 & \text{if} \quad x < a \\ \rho_a + \frac{\rho_b - \rho_a}{b-a}(x-a) & \text{if } a \le x \le b \\ \rho_0 & \text{if} \quad b < x, \end{cases} \tag{5.69}$$

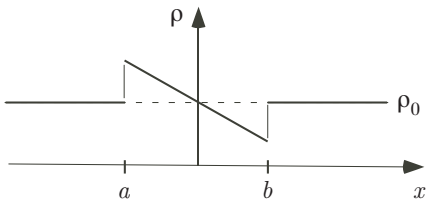


Figure 5.30 Disturbance imposed onto constant density solution at $t = 0$. The resulting initial condition is given in (5.69)

where $\rho_b < \rho_0 < \rho_a$. This function is shown in Figure 5.30. Although it is not continuous, this function is similar to the one in Figure 5.15, and is chosen to make the problem easier to solve. However, unlike what was done in Figure 5.15, we do not assume that the disturbance is necessarily small. Also, as usual, the Greenshields constitutive law is used.

To determine the solution, it is seen in Figure 5.30 that at each jump the faster cars are on the left. This means that a shock wave is going to be generated at each of these points. This is evident if one looks at the characteristics, which are shown in Figure 5.31(a). The characteristics that start to the left of $x = a$ have the form $x = x_0 + c_0 t$, where $c_0 = c(\rho_0)$. Similarly, the one that starts at $x = a$ is $x = a + c_a t$, where $c_a = c(\rho_a)$. Because $\rho_0 < \rho_a$ then $c_0 > c_a$. This means that the characteristic $x = a + c_a t$ is going to overlap with those on the left, as shown in Figure 5.31(a). A similar conclusion applies to the characteristics on the other end, where $x = b$.

The resulting shock waves are shown in Figure 5.31(b). The one on the left end is, from (5.55),

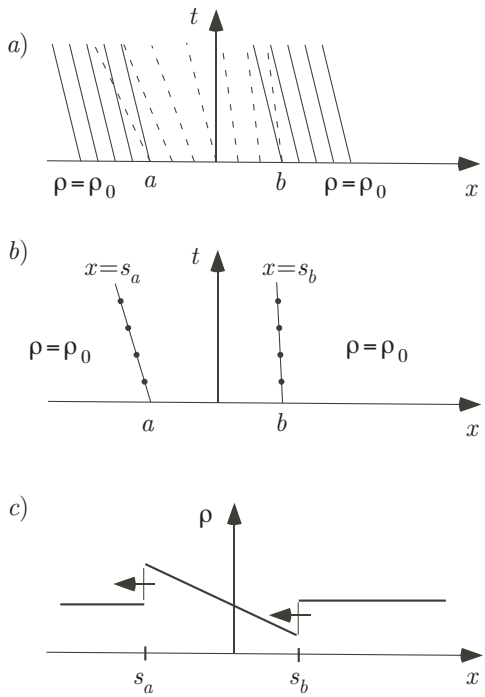


Figure 5.31 The solution of the phantom traffic jam problem, which uses (5.69) as the initial condition. The characteristics, and shock wave, have been drawn for the case of when $c < 0$.

$$s_a(t) = a + \frac{1}{2}(c_0 + c_a)t, \tag{5.70}$$

and the one on the right is

$$s_b(t) = b + \frac{1}{2}(c_0 + c_b)t. \tag{5.71}$$

Carrying out an analysis very similar to the one used for the modified red light - green light example of Section 5.6.2, one finds that the solution is linear in the interval $s_a \leq x \leq s_b$. The resulting solution is, therefore,

$$\rho(x,t) = \begin{cases} \rho_0 & \text{if} \quad x < s_a \\ \rho_a + (\rho_b - \rho_a)\frac{x - s_a}{s_b - s_a} & \text{if } s_a < x < s_b \\ \rho_0 & \text{if} \quad s_b < x. \end{cases} \tag{5.72}$$

This is shown in Figure 5.31(c), and because of its shape it is known as an N-wave. It has the properties mentioned earlier for a phantom traffic jam. Namely, a driver who comes in from the left will be happily driving on a road with a uniform density. When they reach the jam at $x = s_a$ they will have to immediately reduce their speed to adjust for the unexpected increase in the density. They will be able to gradually increase their speed, due to the decrease in the density over the interval $s_a < x < s_b$. However when they reach $x = s_b$ they will have gotten through the disturbance and will need to adjust their speed to match the uniform flow. This is basically the same conclusion reached with the small disturbance approximation in (5.36). What the approximation misses is the change in the width of the disturbance, which it states is constant. The solution in (5.72) shows that the width is $s_b - s_a = b - a + \frac{1}{2}(c_b - c_a)t$, which increases with time.

5.6.7 Summary

The results we have derived for the traffic flow problem are scattered through the preceeding pages, and it is worth finishing up this material by collecting them together. The problem consists of the first-order partial differential equation

$$\frac{\partial \rho}{\partial t} + c(\rho)\frac{\partial \rho}{\partial x} = 0, \tag{5.73}$$

with the initial condition

$$\rho(x,0) = f(x). \tag{5.74}$$

Assuming $c(\rho)$ is a smooth function, with $c'(\rho) \neq 0$, for $0 \leq \rho \leq \rho_M$, then the solution is constructed using the following information.

a) The solution is constant along the characteristic curves $x = x_0 + c_0 t$ (see Figure 5.18).

b) Characteristics Overlap. In a region containing overlapping characteristic curves the solution contains a shock wave at $x = s(t)$. The velocity of this wave is

$$s'(t) = \frac{\rho_R v_R - \rho_L v_L}{\rho_R - \rho_L} \,. \tag{5.75}$$

On either side of the shock, the respective characteristics determine the solution (as illustrated in Figure 5.27). As an example, if $f(x)$ is piecewise constant with a jump discontinuity at x_0, with $c_R < c_L$, then the solution starts out with a shock wave of the form $x = x_0 + s_0' t$, where s_0' is determined from (5.75).

c) Characteristics Separate. In a region with no characteristics, the solution is an expansion fan. An example is shown in Figure 5.26, where $f(x)$ has a jump discontinuity at x_0, with $c_R > c_L$. In this case, in the region $c_L t < x < c_R t$ the solution is found by solving $c(\rho) = (x - x_0)/t$.

The above conclusions are general in the sense that they apply to traffic flow, where $c' < 0$, but also to the case of when $c' > 0$. The latter occurs, for example, for gas flow, and is the subject of Exercises 5.13 and 5.14.

5.7 Cellular Automata Modeling

The viewpoint of the continuum model derived in Section 5.3 is that the motion of the individual cars can be approximated using an averaging process, giving rise to the density and flux functions. It is interesting to explore how to retain the individuality of the cars, and one approach incorporates ideas from cellular automata. The first step in constructing the model is to divide the road into equal segments, each with length ℓ as shown in Figure 5.32. Conventionally, this distance is taken to be the length of an average car, or vehicle, on the road. Time is also divided into equal segments, producing a time step Δt. The objective of the model is, given the positions of the cars at time t_{old}, to determine their positions at $t_{new} = t_{old} + \Delta t$. With this in mind we introduce an integer variable m that equals the number of road segments the car moves in a time step. For example, $m = 1$ means the car moves one segment, $m = 2$ means it moves two segments, etc. It is assumed that there is a maximum number of segments M that a car is allowed to move in a time step. This is equivalent to assuming there is a maximum velocity v_M on the highway. Given that a car's velocity in this formulation is $v = m\ell/\Delta t$, then $v_M = M\ell/\Delta t$.

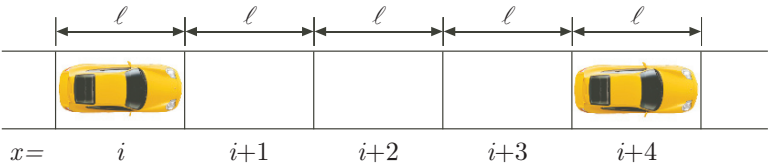


Figure 5.32 In traffic cellular automaton models the roadway is divided into equal segments, and the segments are numbered. For the car on the left, $x = i$ and its gap, which is the number of empty segments in front of it, is $g = 3$.

Example

Taking $\ell = 16$ ft (4.9 m) and $\Delta t = 1$ sec then $m = 1$ corresponds to a velocity of 10.9 mph (17.5 kph), while $m = 6$ corresponds to a velocity of 65.4 mph (105.2 kph). ■

In the model, each car has three integers associated with it, and they are (x, m, g). Here x is the position of the car and its value is determined by the road segment currently occupied by the car. The integer m was defined earlier, and g is called the gap and it is the number of spaces between the car and the one in front of it. For example, for the car on the left in Figure 5.32, $x = i$ and $g = 3$, while for the car on the right $x = i + 4$.

The basic idea in the model is that at time t_{old} we know the values of $(x_{old}, m_{old}, g_{old})$ for each car, and what the model does is to determine their values $(x_{new}, m_{new}, g_{new})$ at time $t_{new} = t_{old} + \Delta t$. This is done by applying the following four rules to each car on the road:

1. *Speedup.*
   If $m_{old} \neq M$, then $m_{new} = m_{old} + 1$.
2. *Do Not Overrun.*
   If $m_{new} > g_{old}$ then $m_{new} = g_{old}$.
3. *Randomization.*
   If $m_{new} \neq 0$ then, with probability p, take $m_{new} = m_{new} - 1$.
4. *Move the Car.*
   Take $x_{new} = x_{old} + m_{new}$.

These four steps constitute what is known as the stochastic traffic cellular automaton (SCTA) model.

The first three steps of the SCTA model contain assumptions, and potential modifications, that need to be discussed.

Speedup. It is assumed that a driver will attempt to drive at the maximum allowed velocity. Assuming the car is not moving at the maximum velocity then in this step the number of segments a car moves is increased by one. It is

certainly possible to consider what happens if there are larger accelerations, and increase the movement by two or more segments, but this will not be investigated here.

Do Not Overrun. The idea here is that if there are, say, three empty spaces in front of the car then it cannot move any more than three spaces in the time step. One can argue that the car in front will likely move in the time step and therefore there will be more than three available spaces. This is correct but it is not accounted for in the model.

Randomization. The previous two steps are intuitive, but not so with the randomization. Numerous reasons have been given to justify this assumption, and this includes the statements that it mimics delayed acceleration or that it accounts for an overreaction in braking. These explanations are rather vague, so instead we will concentrate on what effect the probability p has on the motion. If p is close to zero then it is unlikely the velocity is reduced and there will be little noticeable affect on the car's movement. This is not the case for larger values of p. Any time a car slows down there is a potential to effect the motion of those who are following, and the more often this happens the greater the affect on the flow. The extreme case of when $p = 1$ is examined in Exercise 5.28.

Given the recursive nature of how the four rules are used, it is difficult to determine exactly what happens using analytical methods. The approach, therefore, is to use computer simulations and this brings us to the next example.

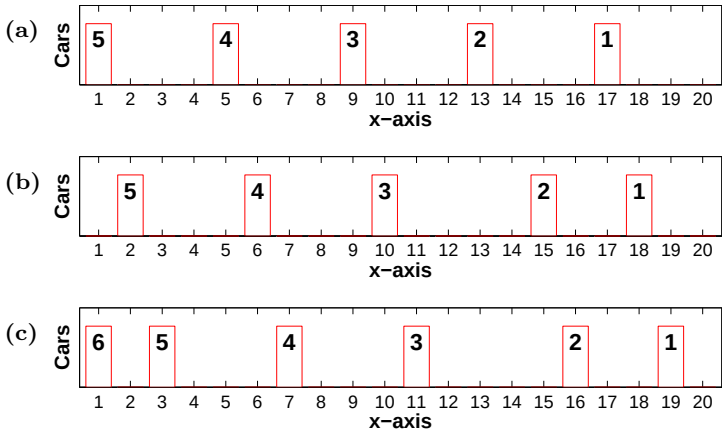


Figure 5.33 In (a) five cars are placed uniformly along a roadway at $t = 0$. Their positions calculated using the SCTA model are shown in (b) at $t = \Delta t$, and in (c) at $t = 2\Delta t$.

Example

Suppose the cars start out uniformly distributed along the highway, where the gap g is the same for all cars. This is shown in Figure 5.33(a), with $n = 20$ and $g = 3$. The lower two graphs are the positions of the cars at the first two time steps, assuming $p = 9/10$, $m = 1$, and $M = 2$. It is seen that after the first time step all cars have moved one space to the right, except for the second that has moved two spaces. To explain this, the speedup step sets $m = 2$ and this means the cars move forward two spaces. However, because of the large value of p it is likely that the velocity of each car in the randomization step is reduced to $m = 1$, and this is borne out in the plot. This one space movement is also seen at the next time step, given in the lower plot. What is new at this time step is the appearance of a sixth car on the left. This is not from the model but, rather, something that is included in the computer code. Specifically, if the first space is empty then the computer adds in a car at this location with a probability equal to the original uniform distribution, namely with probability $1/(g + 1)$. The computer code also removes cars on the right if they pass $n = 20$, although none of the cars in Figure 5.33 have traveled far enough for this to happen. ■

It is possible to take the computer code that produced Figure 5.33 and calculate car positions using a large number of road segments and time steps. Although this generates interesting pictures, little is learned in the process. It is better to study specific situations and compare the results with what is expected on a real roadway. For this we turn to the red light-green light and the green light-red light examples introduced earlier for the continuum model.

Example: Red Light - Green Light

For this the road is divided into 1000 segments, and the stoplight is located at $x = 333$. It is assumed the cars are bumper-to-bumper to the left of the light. This is shown in the lower plot of Figure 5.34(a), where the solid black block on the left are the 333 cars waiting for the light to turn green. Also shown in this plot is the corresponding density, using the continuum model. The solution after a few time steps, for both the SCTA and the continuum models, is shown in Figure 5.34(b). In the calculation $M = 2$ and $p = 1/10$. Both are behaving as expected. In the SCTA model the cars on the right have pulled ahead while those in the block on the left are waiting for room to open up so they can move. For the continuum model we have a expansion fan. From (5.59), in the case of when the light is located at $x = \bar{x}$, the solution is

$$\rho(x,t) = \begin{cases} \rho_L & \text{if} \quad x \le \bar{x} - v_M t, \\ \rho_L \frac{v_M t + \bar{x} - x}{2 v_M t} & \text{if} \quad \bar{x} - v_M t < x < \bar{x} + v_M t, \\ 0 & \text{if} \quad \bar{x} + v_M t \le x, \end{cases} \tag{5.76}$$

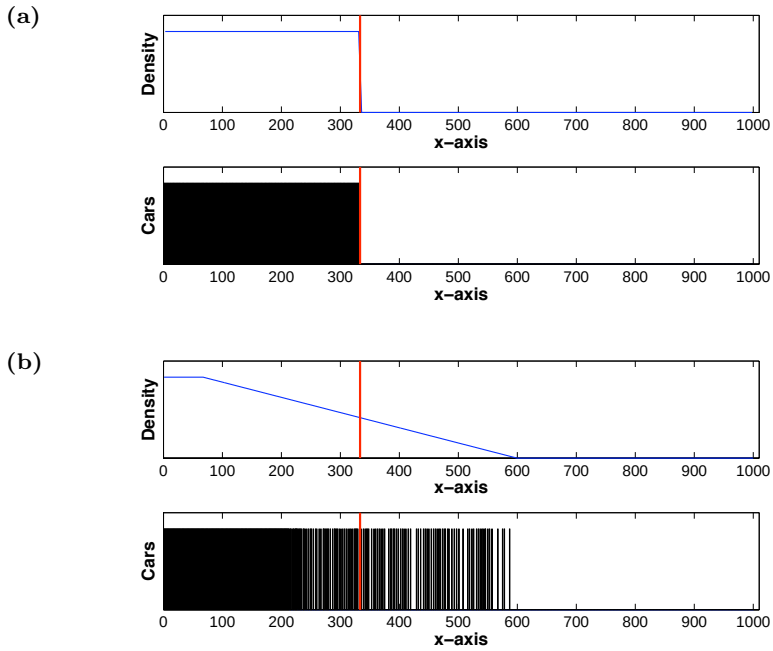


Figure 5.34 Solution of the red light-green light problem. Show are (a) the density and positions of the cars at $t = 0$, and (b) the density and positions after several time steps. The density is computed using (5.76) and the car positions are determined using the SCTA model.

where $\rho_L = 1/\ell$. The linear transition between $\rho = \rho_L$ and $\rho = 0$ is seen in the upper plot of Figure 5.34(b). We are now able to ask the big question, namely, how do the two models compare? To investigate this note that for the SCTA model the car in the front, after N time steps, can move no farther than MN spaces. The actual number will be smaller, depending on how many time steps it takes to accelerate to velocity M and the always present reduction of the velocity due to the randomization step. If p is close to zero, then the first car moves approximately $N_M = M[N - \frac{1}{2}(M-1)]$ spaces, where the $\frac{1}{2}M(M-1)$ term is due to the number of steps it takes the car to reach speed M. In Figure 5.34(b) this car is located at about $x = 600$. The leftmost car that is able to move, for p close to zero, is located approximately N spaces to the left of the light, which in Figure 5.34(b) is near $x = 200$. At the other extreme, the closer p gets to one the closer the number of spaces on either the left or right gets to zero. In this discussion we will assume a linear approximation in p. Therefore, the approximate spatial interval involving the cars that are in motion after N time steps is

$$\bar{x}^* - N\ell(1-p) \le x^* \le \bar{x}^* + N_M\ell(1-p). \tag{5.77}$$

In comparison, for the continuum model the interval is

$$\bar{x}^* - NM\Delta t \leq x^* \leq \bar{x}^* + NM\Delta t. \tag{5.78}$$

This result brings out one clear difference between the two models. Specifically, the continuum model predicts the interval is symmetric about the light's position, while for the SCTA model the interval is generally nonsymmetric. This is seen in Figure 5.34(b). The interval is symmetric in the continuum model because of the constitutive law. The speed the interval expands is determined from the slope of the flux function and, as seen in Figure 5.16, this function is symmetric. In real life, as indicated in Figure 5.8, the flux is not symmetric and this means if we were to use a more realistic constitutive law for v then we would obtain a nonsymmetric interval. Can we find a constitutive law that produces the same result as the SCTA model? Well, this rather interesting question will be left for you to think about. ■

Example: Green Light - Red Light

For this example the road is again divided into 1000 segments, with a stoplight located at $x = 1000$. It is assumed that the cars are uniformly spaced with three spaces between them, and each starts out at the maximum velocity $v_M = 2$. The randomization probability in this example is $p = \frac{1}{4}$. This is shown in Figure 5.35(a). Also shown in this plot is the corresponding density $\rho = 1/(4\ell)$ for the continuum model. The solution after a few time steps, for both the SCTA and the continuum models, is shown in Figure 5.35(b). Both are behaving as expected. In the SCTA model the cars coming in from the left stop when they arrive at the traffic jam, which in Figure 5.35(b) is located near $x = 800$. For the continuum model we have a shock wave that moves leftward. Because $\rho_L = 1/(4\ell)$ and $\rho_R = 1/\ell$ then, after N time steps, the shock is located at $s = \bar{x} - MN\Delta t/4$, where $\bar{x}$ is the location of the stoplight. In Figure 5.35(b), where $N = 450$, we have that $s = 780$.

One of the more obvious differences in the two models, when looking at Figure 5.35(b), is the lack of uniformity in the density to the left of the traffic jam in the SCTA description. This is not unexpected and is due to the randomization step. The second difference is the location of the traffic jam. It is difficult to predict where the jam is located using the SCTA model because there is no simple formula as there is in the continuum case. What is interesting in Figure 5.35(b) is that it appears that in the SCTA model the jam affects the motion of the cars before they reach the jam. Specifically, there is an increase in the density as the cars approach the jam. This can be explained by the fact that any time a car slows down this information is sent backwards along the road (Step 2). Therefore, when a car slows down as it arrives at the jam this affects the cars that follow. This is not in the continuum model and consequently represents a fundamental difference between the two descriptions. ■

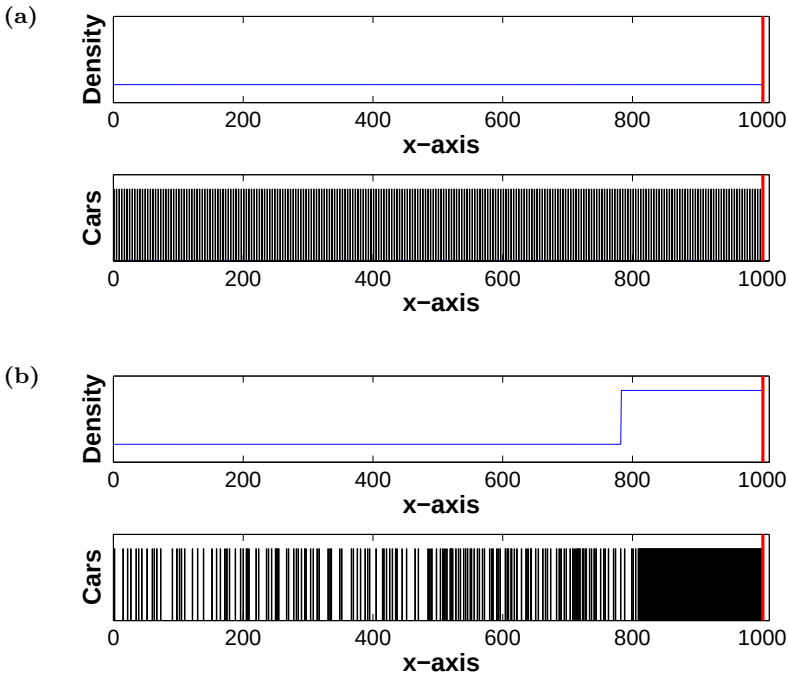


Figure 5.35 Solution of the green light-red light problem. Shown are (a) the density and positions of the cars at $t = 0$, and (b) the density and positions after several time steps. The density is computed using (5.65) and the car positions are determined using the SCTA model.

Exercises

5.1. This problem considers various consequences of the traffic flow equation.

(a) Show that given any two points a and b on the x-axis, with $a < b$,

$$\frac{d}{dt}\int_a^b \rho(x,t)dx = J(\rho_a) - J(\rho_b),$$

where $\rho_a = \rho(a,t)$ and $\rho_b = \rho(b,t)$. Interpret the above equations in physical terms.

(b) Show that

$$\rho(x,t) = f(x) - \frac{\partial}{\partial x}\int_0^t J(x,z)dz.$$

5.2. Consider the situation of when two lanes of traffic merge down to one lane, as shown in Figure 5.36. The densities and velocites at the far left and

right are known. Assume a steady flow, so the density and velocity do not depend on time, and assume all variables are non-negative.

(a) Using the result of Exercise 5.1(a), find an equation that relates the values on the right with those on the left.

(b) What does the equation in part (a) reduce to if the Greenshields law is used?

(c) Suppose $\rho_2 = \rho_1$, $v_2 = v_1$, and the Greenshields law is used. Find ρ_3 in terms of ρ_1. Your solution should give $\rho_3 = 0$ if $\rho_1 = 0$. With this, describe what happens to the flow of cars on the right as ρ_1 is increased, starting from $\rho_1 = 0$. Make sure to explain what happens as ρ_1 nears $\frac{1}{4}(2-\sqrt{2})\rho_M$.

5.3. There are various recommendations concerning safe following distances for cars. Below are a few of the more commonly sited rules. Find the resulting constitutive law relating density and velocity if you assume the cars are uniformly spaced according to the given rule. The function $F(\rho)$ must be continuous, and if you need to make additional assumptions to derive the requested constitutive law make sure to state what they are. Finally, determine which, if any, of the three requirements NV1-NV3, given in Sections 5.4.3, 5.4.4 the constitutive law satisfies.

(a) The National Safety Council recommends the 3-second rule. This means that you allow at least 3 seconds between you and the vehicle in front of you.

(b) In the early days of motoring, it was recommended that you keep one car length back (about 20 feet) for each ten miles per hour of speed.

(c) According to an insurance company, you should allow at least 4 seconds between you and the vehicle in front of you, but if traveling more than 50 mph that this time interval should be at least 6 seconds.

(d) According to a motoring society, the minimum safe distance to the car in front is made up of the sum of two terms. One accounts for the distance traveled due to reaction time, which is usually assumed to be 0.7 seconds. The second term is calculated assuming a constant deceleration, and it accounts for the distance the car will travel after the brakes are applied.

5.4. Apparently drivers do not follow the advice given out by insurance companies or motor clubs, and the claim is that they prefer to select their speed according to the rule

$$\rho = \frac{1}{\alpha + \beta v + \gamma v^2},$$

where α, β, γ are positive constants (Zhou and Peng [2005]).

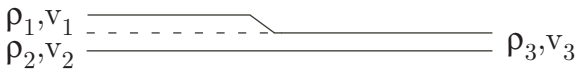


Figure 5.36 Configuration of roadway used for Problem 5.2.

(a) What are the units of α, β, γ?
(b) What is the resulting constitutive law for the velocity? Which of the three conditions on the constitutive law listed in Sections 5.4.3, 5.4.4 does this expression satisfy?
(c) It was found experimentally that in many cases γ is negative. This happens because γ is close to zero, and small variations in the data can cause the curve fitting program to produce a negative value. How does this affect your conclusions in part (b)?

5.5. Assume the flux J is given in Figure 5.37.
(a) Solve the traffic flow equation in the case of when $\rho(x, 0) = 4$.
(b) Solve the traffic flow equation when $\rho(x, 0) = (5 + x^2)/(1 + x^2)$.
(c) Using the information in the graph, find the velocity in terms of the density.

5.6. In the traffic flow problem suppose the velocity of cars, as a function of the density, is measured on a highway and the data shown in Figure 5.38 are obtained.
(a) Formulate a constitutive law for v as a function of ρ based on these data. Provide an explanation of how you reach your conclusion.
(b) In the traffic model it is assumed that $c'(\rho) \neq 0$ for $0 \leq \rho \leq \rho_M$. Does your constitutive law satisfy this condition?

5.7. This problem explores some of the consequences of the Greenshields model as identified in a typical traffic engineering manual.
(a) Sketch the flux as a function of density. At what density is the flux a maximum?
(b) The constant ρ_M is called the jam density, v_M is called the free-flow velocity, and $\frac{1}{4} v_M \rho_M$ is the capacity. Explain why they are given these names.
(c) The headway is defined as the time interval between a common point on the vehicles (e.g., the front bumper) passing a fixed point in space. How is this related to the flux or velocity?
(d) In the example of Section 5.2 for uniform cars the maximum merge density ρ_{merge} was calculated. Use this and the data in Figure 5.6 to find an approximate value for the maximum merge velocity v_{merge}, which is the velocity corresponding to the maximum merge density.

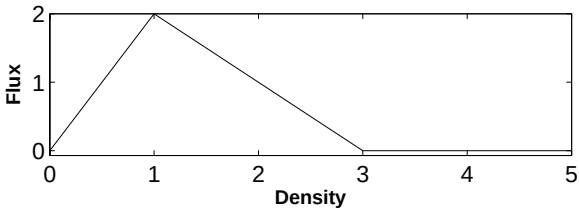


Figure 5.37 Flux-density data used in Exercise 5.5.

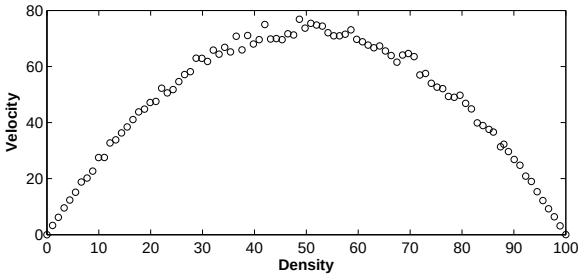


Figure 5.38 Data used for Problem 5.6.

5.8. Solve the following problems by extending the method that was used in Section 5.5 to solve the advection equation.

(a)
$$\frac{\partial \rho}{\partial t} + 2\frac{\partial \rho}{\partial x} = 1,$$
where $\rho(x, 0) = f(x)$.

(b)
$$\frac{\partial \rho}{\partial t} - 6\frac{\partial \rho}{\partial x} = \rho,$$
where $\rho(x, 0) = f(x)$.

(c)
$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} = \rho^2,$$
where $\rho(x, 0) = f(x)$.

(d)
$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} = x, \text{ for } x > 0, t > 0,$$
where $\rho(x, 0) = 0$ for $x > 0$, and $\rho(0, t) = 0$ for $t > 0$.

5.9. This problem explores the finite highway problem associated with Figure 5.14.

(a) Suppose it is found that $\rho(\ell, t) = 2$. What is $f(x)$ and what is $g(t)$?

(b) Suppose it is found that $\rho(\ell, t) = e^{-t}$. What is $f(x)$ and what is $g(t)$?

5.10. This problem considers possible solutions of the traffic flow equation when using the Greenshields law. Assume that $\rho(x, 0) = f(x)$, where $f(x)$ is piecewise constant. Also, make sure to justify your answers.

(a) Give an example of $f(x)$ that produces a solution with two expansion fans and no shock waves.

(b) Give an example of $f(x)$ that produces a solution that starts out with two shock waves and no expansion fans.

(c) Give an example of $f(x)$ that produces a solution that starts out with one shock wave and one expansion fan.

(d) In part (a), explain why the two expansion fans can not overlap.
(e) In part (b), the two shock waves eventually intersect. Explain why this is expected based on the way the cars are positioned at the very start. When the shocks intersect, a single shock is formed. Find the resulting solution.

5.11. This problem considers what values a solution of the traffic flow equation can have when using the Greenshields law, with $v_M = 1$ and $\rho_M = 10$. Assume that $\rho(x, 0) = f(x)$ is piecewise constant.
(a) Assuming that $3 \leq f(x) \leq 4$, explain why it is impossible for $\rho(x, t) = 5$ at any value of (x, t). Is it possible for $\rho(x, t) = 2$?
(b) Suppose that $f(x)$ is piecewise constant, and only takes on the values 1 and 3. Give an example to show that it is possible for $\rho(x, t) = 2$ for one or more points (x, t). For your example, what other values does the solution take on?
(c) Give an example for $f(x)$, so that the only values ρ takes on are 1, 2, and 3.

5.12. This problem explores some of the connections between the velocity functions that arise with nonlinear traffic flow.
(a) Show that

$$v = \frac{1}{\rho} \int_0^\rho c(\bar{\rho}) d\bar{\rho}.$$

(b) Show that

$$s'(t) = \frac{\rho_R v_R - \rho_L v_L}{\rho_R - \rho_L}.$$

(c) Show that if v is a monotonically decreasing function of ρ then $c \leq v$.
(d) Give an example to show that to have c monotonically decreasing, it is not enough to assume v is monotonically decreasing.
(e) Is it possible for a shock wave to stay in one place? You can assume the Greenshields law is used.
(f) Is it possible for the wave velocity c to be independent of ρ without assuming the car velocity v is independent of ρ?

5.13. In fluid dynamics one solves the nonlinear equation (5.30), but the wave velocity is $c(\rho) = \rho$. Using this function, assume that the initial condition is

$$\rho(x, 0) = \begin{cases} 1 & \text{if } x \leq -1 \\ \frac{1}{2}(1 - x) & \text{if } -1 \leq x \leq 1 \\ 0 & \text{if } 1 \leq x. \end{cases}$$

(a) Sketch the characteristics in the x, t-plane.
(b) Find the solution, and sketch it as a function of x, for $t > 0$.
(c) Show that $v = \frac{1}{2}\rho$.

5.14. As in the previous problem, suppose $c(\rho) = \rho$ but now the initial condition is

$$\rho(x,0) = \begin{cases} 0 & \text{if } x \le -1 \\ \frac{1}{2}(1+x) & \text{if } -1 \le x \le 1 \\ 1 & \text{if } 1 \le x. \end{cases}$$

(a) Sketch the characteristics in the x,t-plane.
(b) Find the solution, and sketch it as a function of x, for $t > 0$.
(c) Find the points in the x,t-plane where $\rho = \frac{1}{3}$.
(d) Show that $v = \frac{1}{2}\rho$. With this, determine the flux J.

5.15. This problem examines what happens on a finite length highway when the velocity is not constant. The equation is

$$\frac{\partial \rho}{\partial t} + c(\rho)\frac{\partial \rho}{\partial x} = 0, \quad \text{for} \quad \begin{cases} 0 < x < \ell \\ 0 < t, \end{cases}$$

where

$$\rho(x,0) = f(x), \quad \text{for } 0 \le x \le \ell,$$

and

$$\rho(0,t) = g(t), \quad \text{for } 0 < t.$$

Assume the Greenshields law is used, and so the function c is given in (5.38).
(a) Assuming that $2\rho_R < \rho_M$, find the solution when $f(x) = \rho_R(1 - x/\ell)$ and $g(t) = \rho_R$.
(b) For part (a), explain why there is no solution if $2\rho_R > \rho_M$. Which condition, the one at $t = 0$ or the one at $x = 0$, should be dropped so there is a solution?
(c) Find the solution when $f(x) = 0$ and $g(t) = \rho_L$.
(d) Find the solution when $f(x) = \rho_R$ and $g(t) = 0$.

5.16. To investigate just how much influence the constitutive law has on the solution, suppose it is assumed that $v = v_M((1 - (\rho/\rho_M)^2)$. This is a special case of what is known as Drew's constitutive law (Drew [1968]).
(a) What is the wave velocity $c(\rho)$? Is it a monotonic function of ρ?
(b) What is the solution of the modified red light - green light problem? As in Figure 5.20, sketch the solution as a function of x and comment on how the solution differs from the one in Figure 5.20. You can assume in this problem that $\rho_R = 0$.
(c) What is the solution of the red light - green light problem, where $\rho_R = 0$? As in Figure 5.25, sketch the solution as a function of x. Also, comment on how the solution differs from the one in Figure 5.25.
(d) What is the solution of the traffic jam problem? As in Figure 5.27, sketch the solution as a function of x. Also, compare the velocity of the shock with the value obtained using the Greenshields law.

5.17. One way to explain weak solutions is to consider a smooth version of the jump initial condition. Specifically, let $\rho(x,0) = 1/(1 + \alpha e^{x/\epsilon})$, where ϵ and α are positive. Also, this problem considers the linear equation, so $v = a$ is constant.

(a) Find the solution using the above initial condition. Explain why the resulting solution is smooth and sketch it assuming that ϵ is small.
(b) Explain what happens both to the initial condition and solution when $\epsilon \to 0$. Make sure to explain what happens if $x = at$.
(c) Part (b) helps explain why using a jump initial condition is consistent with what is known for smooth solutions, with the exception of what happens at the jump itself. This raises the question of what value the density can have at a jump. Given the definition of the density in (5.1), what should the value of the density be at $x = 0$ and at $x = 1$ in Figure 5.9 when $t = 0$?
(d) Show that the limiting value you found in part (b), when $x = at$, can be obtained for the solution in Figure 5.9 by modifying the averaging interval in (5.1). What this shows is that in a continuum theory, the value at a discontinuity is dependent on the averaging method.

5.18. Suppose one is interested in knowing the position of a particular car when using the continuum model. Assume that at $t = 0$ the car is located at $x = A$, and its position at later times is given by $x = \chi(t)$. This problem is concerned with how to find the function $\chi(t)$. In doing this it is assumed that the traffic flow equation has been solved, so the density $\rho(x, t)$ and the velocity $v(x, t)$ functions are known.
(a) Explain why, to find $\chi(t)$, one solves the differential equation $\chi' = v(\chi, t)$, where $\chi(0) = A$.
(b) For the red light-green light solution given in (5.62), assume $\rho_L = \rho_M$ and $A < 0$. What is the resulting velocity function? With this show that

$$\chi(t) = \begin{cases} A & \text{if } 0 \le t \le -A/v_M \\ v_M t - 2\sqrt{-A v_M t} & \text{if } -A/v_M < t. \end{cases}$$

On the same axes, sketch $\chi(t)$ for $A = -v_M$, $A = -2v_M$, and $A = -3v_M$.
(c) For the red light-green light problem, which cars are able to get through the light if it is green for $0 \le t < t_R$ and turns red at $t = t_R$?
(d) For the traffic jam example studied in Section 5.6.5, find $\chi(t)$ for $A < 0$. On the same axes, sketch $\chi(t)$ for $A = -v_M$, $A = -2v_M$, and $A = -3v_M$.

5.19. It is observed that when a stoplight turns green, the density of traffic passing through the light increases in time up to a constant value ρ_0. Assuming the light is located at $x = 0$, a boundary condition that mimics this observed behavior is

$$\rho(0, t) = \begin{cases} \rho_0 t/t_s & \text{if } 0 \le t \le t_s \\ \rho_0 & \text{if } t > t_s. \end{cases}$$

The domain over which the traffic flow problem is solved is $0 < x$ and $0 < t$. Assume here that $\rho(x, 0) = 0$ and the Greenshields constitutive law is used.
(a) In the case of when $\rho_0 = \frac{1}{3}\rho_M$ find, and then sketch, the solution.
(b) Suppose that $\rho_0 = \frac{2}{3}\rho_M$. Sketch the characteristics, and use this to explain why there is no solution. In fact, explain why there is no solution for any density that satisfies $\rho_0 > \frac{1}{2}\rho_M$.

5.20. This problem investigates how to use similarity variables to find an expansion fan.

(a) Assume $\rho(x,t) = R(\eta)$, where $\eta = x/t$. Show that the traffic flow equation reduces, in the case of when ρ is not constant, to the equation $c(\rho) = x/t$.

(b) Using the Greenshields law, solve $c(\rho) = x/t$ for ρ.

(c) Show that your solution in part (b) is the same as the one given in (5.59).

5.21. This problem examines what happens to the traffic flow problem when cars are allowed to enter or exit the highway. It is assumed this occurs not at discrete locations but continuously along the highway.

(a) Assume that over an interval $x_0 - \Delta x < x < x_0 + \Delta x$ the number that enter (or exit) from $t = t_0 - \Delta t$ to $t = t_0 + \Delta t$ is $4\Delta x \Delta t Q$, where $Q(x,t)$ is the net rate per unit length at which cars are entering or leaving the highway. Show that the resulting balance law for traffic flow is

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x} + Q.$$

(b) One possible constitutive law for this new variable is $Q = \alpha(\rho - \beta)$ where α, β are constants. Can you explain how this assumption could arise for traffic flow? Is there any reason you should assume α is either positive or negative? Any suggestion on how to choose β?

(c) Use the procedure to solve the $\alpha = 0$ case to solve the equation derived in part (a) along with the constitutive assumption in part (b). Assume a constant velocity.

(d) Based on your solution from part (c), what is the effect of Q on the density? Is the solution still a traveling wave? Demonstrate your conclusion using the initial distribution $\rho(x,0) = e^{-x^2}$ by sketching the solution at later times.

5.22. Suppose you had an experimental apparatus that enabled you to measure the velocity of a shock wave. Explain how you could use this to determine a constitutive law for the velocity.

5.23. A variable related to density is the volume fraction $\phi(x,t)$, which is used to determine how much of the highway is taken up by cars (versus empty road). In reference to Figure 5.4,

$$\phi(x_0,t_0) \approx \frac{\text{total length of cars from } x_0 - \Delta x \text{ to } x_0 + \Delta x \text{ at } t = t_0}{2\Delta x}.$$

The value of $\phi(x_0,t_0)$ is the limit of the right-hand side as $\Delta x \to 0$.

(a) For evenly spaced cars as in Example 5.2, show that $\phi(x,t) = \ell/(\ell + d)$, and therefore $\phi = \ell\rho$.

(b) If the cars are not necessarily evenly spaced but still are all of length ℓ show that it is still true that $\phi = \ell\rho$.

(c) Assuming $\phi = \ell\rho$, where ℓ is constant, rewrite the traffic flow equation (5.30) in terms of $\phi(x,t)$.

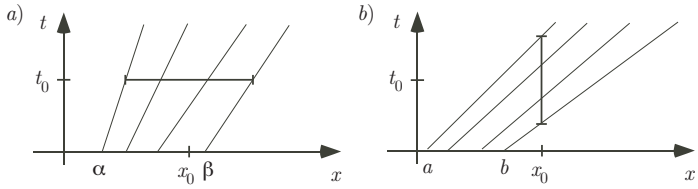


Figure 5.39 Averaging intervals used to define (a) the density, and (b) the flux. The horizontal bar in (a) has length $2\Delta x$, and the vertical bar in (b) has length $2\Delta t$. The four slanted lines shown in each figure are the paths of individual cars. These figures are used in Problem 5.26.

5.24. Suppose the density is given in terms of the velocity, and so, assume $\rho = H(v)$.
(a) Show how the traffic flow equation can be written as $v_t + d(v)v_x = 0$.
(b) Find H for the Greenshields (5.10) and Newell (5.17) functions.
(c) The initial condition used for the small disturbance approximation is $v(x, 0) = v_0 + \epsilon h(x)$. Find the resulting two term expansion for the velocity.

5.25. One might argue that if a driver is in a relatively high-density region and sees lower density traffic up ahead that they will speed up with the objective of traveling in the lower-density region.
(a) Explain why an assumption that accounts for this behavior is a constitutive law of the form $v = F(\rho, \rho_x)$.
(b) Write down a simple, three parameter, constitutive law for v that involves ρ and ρ_x.
(c) With the constitutive law from part (b) what is the resulting traffic flow equation?
(d) What is the resulting small disturbance equation and how does it differ from (5.35)?

5.26. This problem examines the averaging used to define the flux and density, and how they relate with the velocity. It is assumed that a car with initial location x_0 has velocity $f(x_0)$. Consequently, the position of this car at any later time t is $x = x_0 + f(x_0)t$. Example paths for the cars are shown in Figure 5.39. Therefore, in this problem, each car has a constant velocity, but different cars can have different velocities. For simplicity, it is assumed that $f(x) = v_0 + w_0 x$, where v_0 and w_0 are constants.
(a) The averaging interval used to define the density in (5.1) is shown in Figure 5.39(a), and it is the same as the one shown in Figure 5.4. Explain why $x_0 - \Delta x = \alpha + f(\alpha)t_0$ and $x_0 + \Delta x = \beta + f(\beta)t_0$. Use these equations to find α and β in terms of x_0 and t_0.
(b) The averaging interval used to define the flux in (5.3) is shown in Figure 5.39(b). Explain why $t_0 - \Delta t = b + f(b)t_0$ and $t_0 + \Delta t = a + f(a)t_0$. Use these equations to find a and b in terms of x_0 and t_0.

(c) Assuming that the cars are continuously distributed, show that the average velocity for the cars in the horizontal bar in Figure 5.39(a) is $v_0 + w_0(x_0 - v_0 t_0)/(1 + w_0 t_0)$.
(d) Assuming that the cars are continuously distributed, find the average velocity for the cars in the vertical bar in Figure 5.39(b). Assuming that Δt is small, show that the average velocity is, approximately, $v_0 + w_0(x_0 - v_0 t_0)/(1 + w_0 t_0)$.
(e) Use the fact that the average velocities in parts (c) and (d) are the same to explain why this provides additional evidence of the validity of the equation $J = \rho v$.

5.27. This problem considers various formulas for the SCTA model.
(a) Show that the first two steps in the SCTA model can be combined into the formula $m_{new} = \min\{m_{old} + 1, M, g_{old}\}$.
(b) Show that the first three steps in the SCTA model can be combined into the formula $m_{new} = \max\{0, \min\{m_{old}+1, M, g_{old}\} - \chi\}$, where $\chi = 1$ with probability p (otherwise, $\chi = 0$).

5.28. Suppose that in the SCTA model the cars start out uniformly spaced with $g = M$. Assume the randomization probability is $p = 1$.
(a) Let $M = 1$. What happens if the cars start out with velocity M? What happens if the cars start out with zero velocity?
(b) Let $M = 3$. What happens if the cars start out with velocity M? What happens if the cars start out with velocity $m = 2$? What happens if the cars start out with velocity $m = 1$? What happens if the cars start out with zero velocity?
(c) Generalize your conclusions from part (b) to describe what happens if $M \geq 2$.

5.29. Suppose there is a stoplight located at $x = 0$. When it turns red assume the cars are uniformly spaced in the region $x < 0$, with three spaces between the cars, and each car has $m = 1$. The maximum movement is $M = 2$. A space here is one car length.
(a) Assuming the Greenshields constitutive law is used, what is the resulting solution of the traffic flow problem? What is the velocity of the shock wave?
(b) In the SCTA model suppose the randomization is turned off (i.e., $p = 0$). Show that the approximate velocity of the shock wave is $-2\Delta x/(3\Delta t)$. How does this compare with the continuum result from part (a)?
(c) In the SCTA model suppose that in the randomization step $p = 1$. Explain why there is a shock-like solution but the jam density is half of what is obtained from the continuum solution. Also show that the shock moves with approximate velocity $-\Delta x/\Delta t$.
(d) Given the solution in part (c) describe, in general terms, what happens when p is close to one.

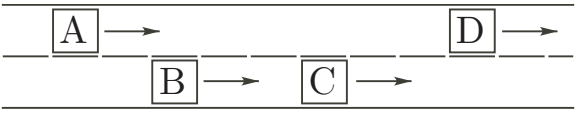


Figure 5.40 Possible car positions when deciding to make a lane change, as considered in Problem 5.30.

(e) Using your results from parts (b) and (d), explain why $-2(1-p)\Delta x/(3\Delta t)$ provides an approximation of the shock velocity for the SCTA model. Given this, what should the randomization probability be so that the SCTA velocity agrees with the continuum model?

5.30. To extend the SCTA model to multilane roads, where individual cars are able to change lanes, consider Figure 5.40. Assume a driver will switch lanes whenever they are able to travel farther in a time step in the other lane. Safe lane changing requires consideration of the backward gap in the other lane, so the driver in car B must consider the position and velocity of car A when deciding to switch. Write down a set of rules for moving the cars along the highway that includes lane changes. Assume in this problem that the randomization step is omitted.

Chapter 6
Continuum Mechanics: One Spatial Dimension

6.1 Introduction

In the previous chapter we investigated how to model the spatial motion of objects (cars, molecules, etc.) but omitted the possibility that the objects exert forces on each other. The objective now is to introduce this into the modeling. The situations where this is needed are quite varied and include the deformation of an elastic bar, the stretching of a string, or the flow of air or water. Each of these has an internal material structure that resists either stretching (the string and bar) or compression (air, water, and bar). Illustrations of particular applications are shown in Figure 6.1. In this chapter the mathematical model for these systems is derived, and in dong this we will limit our attention to one-dimensional motion. Also, the only problems solved in this chapter are to find the steady-state solution. The question of how to solve the time-dependent mathematical problem coming from the model will be taken up in the next chapter.

6.2 Coordinate Systems

The first step in continuum modeling is to define the coordinates that will be used. We will find two systems invaluable, and we will continually switch back and forth between them. One of them follows the material as it moves, and not surprisingly, this is called the material system. The other is fixed in space, and this is the spatial system.

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_6,
© Springer Science+Business Media, LLC 2009

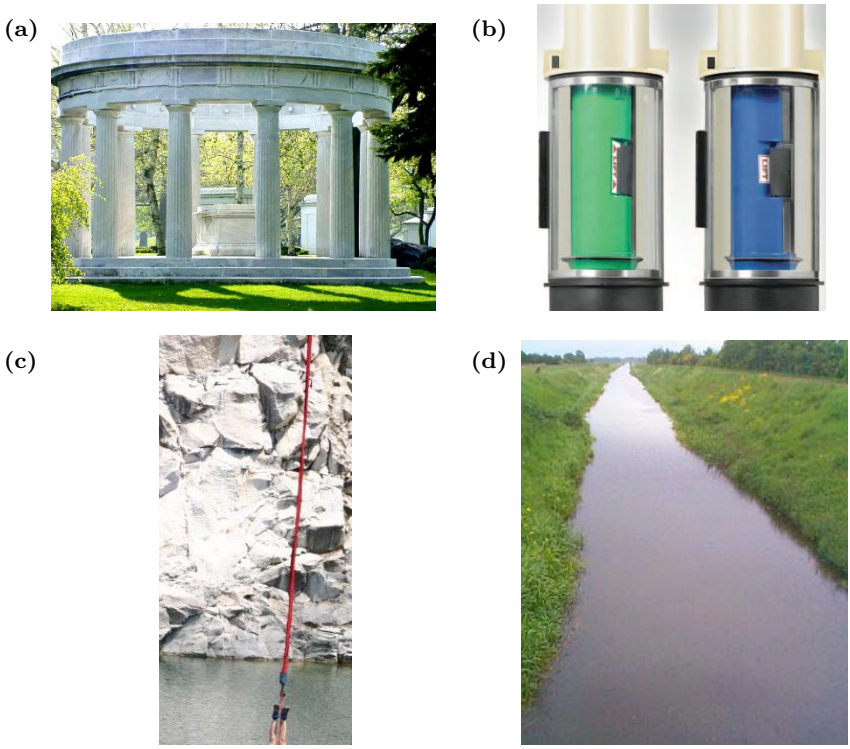


Figure 6.1 Examples of uniaxial motion or deformation. (a) Axial compression of a bar. These Doric-style columns are under compression due to the annular entablature they support. (b) Air flow through a pneumatic tube. (c) Longitudinal stretching of bungie cord. (d) Flow of water along a straight pathway.

6.2.1 Material Coordinates

To define this system, consider a cylindrical bar as shown in Figure 6.2. The top figure shows the bar at $t = 0$ and identifies a particular cross-section located at $x = A$. The lower figure shows the bar at a later time and the cross-section has moved to $x = X(A, t)$. To be consistent it is required that $X(A, 0) = A$.

Given the position function X, we can track each cross-section as it moves back and forth along the x-axis. In this way, the coordinates A, t can be used to locate the cross-sections, and together they constitute what is known as the material coordinate system. In physics this is usually referred to as Lagrangian coordinates.

We are particularly interested in how far each section moves from its original position, and for this reason we introduce the displacement function, defined as

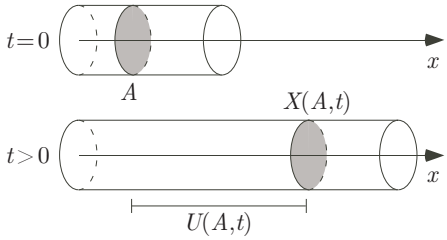


Figure 6.2 Axial motion of a cross-section that begins at $x = A$ and moves to $x = X(A,t)$. The resulting displacement of the cross-section is $U(A,t)$.

$$U(A,t) = X(A,t) - A. \tag{6.1}$$

Because $X(A,0) = A$ then it is required that

$$U(A,0) = 0. \tag{6.2}$$

Connected with the displacement is the velocity of the cross-section, defined as

$$V(A,t) = \frac{\partial U}{\partial t}\,. \tag{6.3}$$

The acceleration and other higher time derivatives can be calculated similarly.

One last comment to make is that in the material coordinate system the positions of the cross-sections at $t = 0$ comprise what is known as the reference configuration.

6.2.2 Spatial Coordinates

For an external observer watching the motion, the material description is not particularly convenient. For example, if you were to observe fluid flow in a river it is likely you would simply stand on the river bank and take your measurements at that fixed position. In doing this, you would be determining the properties of the motion without knowing where the individual water molecules were located at $t = 0$. For this reason the spatial, or Eulerian, coordinate system is introduced. The idea here is that you select the spatial location x and let the cross-sections come to you. In spatial coordinates the displacement function is denoted as $u(x,t)$ and the velocity is $v(x,t)$. In this context, $u(x,t)$ is the displacement of the cross-section that is located at x and time t, and $v(x,t)$ is the velocity of that cross-section. So, at a fixed location x, as long as the velocity is nonzero, the cross-sections at x keep changing. In contrast, in the material system, a fixed A means you are following a particular cross-section, but your spatial location is changing.

The values obtained from the spatial coordinate system must be exactly the same as obtained when they are computed using material coordinates. To determine how this happens, the transformation between material and spatial coordinates is through the formula $x = X(A,t)$. Because the two coordinate systems must give the same value, we have that

$$U(A,t) = u(X(A,t),t), \tag{6.4}$$

and

$$V(A,t) = v(X(A,t),t). \tag{6.5}$$

If, on the other hand, we know the material variables U and V and want to calculate the spatial versions it is first necessary to solve $x = X(A,t)$ for A. Physically this corresponds to determining the original position A of the cross-section that is currently located at x. Writing this as $A = a(x,t)$ then

$$u(x,t) = U(a(x,t),t), \tag{6.6}$$

and

$$v(x,t) = V(a(x,t),t). \tag{6.7}$$

The above expressions will prove to be invaluable when transforming between spatial to material coordinates.

Example

Suppose the bar deforms in such a way that the cross-sections move according to the rule that $x = A(1+2t)/(1+t)$. In this case,

$$\begin{aligned} U(A,t) &= X - A \\ &= \frac{tA}{1+t}, \end{aligned}$$

and

$$\begin{aligned} V(A,t) &= \frac{\partial U}{\partial t} \\ &= \frac{A}{(1+t)^2}. \end{aligned}$$

This means, for example, that the cross-section that starts at $A = 1$ moves with velocity $V = 1/(1+t)^2$. To express these formulas in spatial coordinates, we solve the rule to obtain $A = x(1+t)/(1+2t)$. In this case, from (6.6),

$$\begin{aligned} u(x,t) &= U|_{A=\frac{1+t}{1+2t}x} \\ &= \frac{tx}{1+2t}, \end{aligned}$$

and, from (6.7),

$$\begin{aligned} v(x,t) &= V|_{A=\frac{1+t}{1+2t}x} \\ &= \frac{x}{(1+2t)(1+t)}\,. \end{aligned} \tag{6.8}$$

One significant observation from the above calculation is that $v \neq \frac{\partial u}{\partial t}$. Also, the change of variables is reversible, and so

$$\begin{aligned} U(A,t) &= u(x,t) \\ &= u(A\frac{1+2t}{1+t},t) \\ &= \frac{tA}{1+t}\,. \quad \blacksquare \end{aligned}$$

One of the distinctive differences in the two coordinate systems is the domain over which they apply. To demonstrate, suppose that the bar starts off, at $t = 0$, with length ℓ_0 and occupies the interval $0 \leq A \leq \ell_0$. The range of A values for any material variable, such as $U(A,t)$ or $V(A,t)$, is determined by the original position of the bar. In other words, they are defined for $0 \leq A \leq \ell_0$. In contrast, the range of x values for a spatial variable, such as $u(x,t)$ or $v(x,t)$, depends on the current position of the bar. The left end of the bar is at $x = X(0,t) = U(0,t)$ and the right end is at $x = X(\ell_0,t) = \ell_0 + U(\ell_0,t)$. Consequently, the spatial variables are defined for $U(0,t) \leq x \leq \ell_0 + U(\ell_0,t)$.

Example (cont'd)

In the previous example, suppose the bar initially occupies the interval $0 \leq A \leq 3$. In this case, $U(A,t)$ and $V(A,t)$ are defined, for all values of t, in the interval $0 \leq A \leq 3$. To determine the interval for the spatial coordinates, note that the position of the left end is at

$$x = U(0,t) = 0,$$

and the position of the right end is at

$$x = 3 + U(3,t) = 3 + \frac{3t}{1+t}.$$

Therefore, the spatial variables are defined for $0 \leq x \leq 3(1+2t)/(1+t)$. $\blacksquare$

A comment is needed about partial derivatives and the two coordinate systems. The independent variables in the material system are A and t, while the independent variables in the spatial system are x and t. Partial derivatives with respect to these variables will be written using one of several notations.

In particular, the following are three different ways to designate the first partial derivative

$$\frac{\partial U}{\partial A}\,, \quad U_A\,, \quad \partial_A U\,.$$

The second partial derivative are written in either of the following three ways:

$$\frac{\partial^2 U}{\partial A^2}\,, \quad U_{AA}\,, \quad \partial_A^2 U\,.$$

Correspondingly, mixed second partial derivatives are written in one of the following ways:

$$\frac{\partial^2 U}{\partial A \partial t}\,, \quad U_{At}\,, \quad \partial_A \partial_t U\,.$$

There is nothing particularly special about any one of the these forms, and the choice of what to use is mostly determined by the format of the mathematical expression under consideration.

6.2.3 Material Derivative

The above example shows that even though $V = \frac{\partial U}{\partial t}$ it turns out that in spatial coordinates $v \neq \frac{\partial u}{\partial t}$. This begs the question as how it might be possible to determine v if we know u. To answer this suppose $F(A,t)$ is a function in material coordinates, and assume that its spatial version is $f(x,t)$. In this case $\frac{\partial F}{\partial t}$ represents the time rate of change of the variable for the cross-section that began at A. To determine what this is in spatial coordinates note that F and f must produce the same value. For example, if the cross-section that started at A is currently located at $x = X(A,t)$ then $F(A,t) = f(x,t)$. In other words,

$$F(A,t) = f(X(A,t),t). \tag{6.9}$$

Taking the time derivative, and using the chain rule, we have that

$$\begin{aligned}\frac{\partial F}{\partial t} &= \frac{\partial f}{\partial x}\frac{\partial X}{\partial t} + \frac{\partial f}{\partial t} \\ &= \frac{\partial f}{\partial x} V(A,t) + \frac{\partial f}{\partial t} \\ &= \frac{\partial f}{\partial x} v(x,t) + \frac{\partial f}{\partial t}\,.\end{aligned} \tag{6.10}$$

This derivative plays such an important role in what follows that it gets its own name and notation. It is called the *material derivative* of f and it is defined as

$$\frac{Df}{Dt} \equiv \frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x}\,. \tag{6.11}$$

The reason this is important is because $\frac{Df}{Dt}$ is the time rate of change of the function following a material cross-section, but expressed in spatial coordinates.

A particularly important application of the above formula concerns the spatial coordinate description of the displacement and velocity functions. Taking $F(A,t) = U(A,t)$, then $f(x,t) = u(x,t)$ and from (6.10) we have that

$$\frac{\partial U}{\partial t} = \frac{Du}{Dt}\,. \tag{6.12}$$

Because $V = \frac{\partial U}{\partial t}$, and $V(A,t) = v(x,t)$, it follows that

$$v = \frac{Du}{Dt}\,. \tag{6.13}$$

Using the definition of the material derivative in (6.11) the above equation reduces to

$$v = \frac{u_t}{1 - u_x}\,. \tag{6.14}$$

In other words, the velocity $v(x,t)$ of the cross-section located at x, at time t, is $\frac{Du}{Dt}$ and not $\frac{\partial u}{\partial t}$. This result explains why in spatial coordinates we usually end up with $v \neq u_t$.

Example (cont'd)

Using the functions from the previous example,

$$\frac{\partial u}{\partial x} = \frac{t}{1+2t}\,,$$
$$\frac{\partial u}{\partial t} = \frac{x}{(1+2t)^2}\,,$$

and so, from (6.14)

$$v = \frac{x}{(1+t)(1+2t)}\,.$$

This agrees with the formula for the spatial velocity calculated by converting the material velocity V to v in (6.8). ■

In modeling the deformation of a bar we will also need to consider the spatial changes in the material. The variable of interest in such situations will be the material gradient $\frac{\partial F}{\partial A}$. To express this in spatial coordinates we again start by differentiating (6.9) but this time with respect to A. In a similar manner as before, using the chain rule one finds that

Material Expression		Spatial Expression
$\frac{\partial F}{\partial t}$	$=$	$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x}$
$\frac{\partial F}{\partial A}$	$=$	$\frac{1}{1-u_x}\frac{\partial f}{\partial x}$
$\frac{1}{1+U_A}\frac{\partial F}{\partial A}$	$=$	$\frac{\partial f}{\partial x}$
$\frac{\partial F}{\partial t} - \frac{V}{1+U_A}\frac{\partial F}{\partial A}$	$=$	$\frac{\partial f}{\partial t}$

Table 6.1 Relationship between various spatial and material derivatives. Here $F(A,t)$ is a function in material coordinates and $f(x,t)$ is the function in spatial coordinates.

$$\begin{aligned}\frac{\partial F}{\partial A} &= \frac{\partial f}{\partial x}\frac{\partial X}{\partial A} \\ &= \frac{\partial f}{\partial x}\left(1 + \frac{\partial U}{\partial A}\right). \end{aligned} \tag{6.15}$$

In the special case of when $F = U$ and $f = u$, this reduces to $U_A = (1+U_A)u_x$. Solving for U_A we obtain

$$\frac{\partial U}{\partial A} = \frac{u_x}{1-u_x}. \tag{6.16}$$

Substituting this into (6.15), yields

$$\frac{\partial F}{\partial A} = \frac{1}{1-u_x}\frac{\partial f}{\partial x}. \tag{6.17}$$

This result is listed in Table 6.1 along with other useful equalities between various derivatives.

It would appear from the formulas in Table 6.1 that we might have a problem if $u_x = 1$, or if $U_A = -1$. Later in the chapter we will find that for the spatial coordinate system to be defined it is necessary to require $u_x < 1$. Similarly, the requirement for the material coordinate system to be defined is that $U_A > -1$.

6.2.4 End Notes

Before moving on to the derivation of the equations of motion there are a few aspects of the development that need to be stated explicitly. First, it is important to remember that the points making up a material cross-section, such as shown in Figure 6.2, move as a unit and the motion is only along the x-axis. One consequence of this assumption is that the cross-sectional area

is constant. A second point is the implicit assumption that cross-sections do not pass each other. For example, if a cross-section starts out to the left of another cross-section then it is always to the left of it. This is reasonable from a physical viewpoint, and it is known as the impenetrability of matter assumption. It is also important for mathematical reasons because it guarantees that we can uniquely convert back and forth between spatial and material coordinates, and this is related to the inequalities mentioned at the end of the previous paragraph. A third implicit assumption being made concerns the smoothness of the motion. We will differentiate variables any number of times based on the assumption that the motion is smooth enough to permit this.

There is more than a passing connection between this chapter and the previous one on traffic flow. In a mathematical sense the cars of the last chapter are cross-sections in this chapter. Both are objects moving along the x-axis. For this reason it should not be a surprise that the cross-sections satisfy a conservation law related to their density, just as the cars did in the last chapter as expressed in (5.7). A significant difference is that in this chapter the objects (i.e., the cross-sections) exert forces on their neighbors, and we will derive a force balance equation from this. Another difference is that only a spatial coordinate system was used in traffic flow. It was not necessary to introduce the material coordinate system, which would describe the motion from the driver's point of view, but in this chapter the material system is important. To explain why, using the car analogy, we will assume that the car directly in front of the driver, and the car directly behind, exert a force on the driver's car. This will happen, for example, if adjacent cars were connected by springs. These forces change as the distance between the cars change, and to keep track of this we will need to follow the cars. Hence, the need for material coordinates. At the same time, there will be situations where the spatial coordinate system is preferable, and this means we will switch back and forth between the two systems. If you are interested how material coordinates can be used in traffic flow, Exercise 5.18 should be consulted.

6.3 Mathematical Tools

To derive the traffic flow equation in the last chapter we used a control volume approach. With the objective of trying different ideas, we will derive the equations in this chapter using what is known as the integral method. This requires a bit more mathematical background but the benefit is that the derivation is easier. It is the purpose of this section to present the needed mathematical tools. As with seemingly everything in mathematical modeling, these results are personalized and given names.

The first result is straight out of mathematical analysis, and it tells us when we are able to conclude a function is identically zero.

Theorem 6.1. *If $f(x)$ is continuous and $\int_a^b f(x)dx = 0$, $\forall a, b$ with $a < b$, then $f(x) = 0$, $\forall x$.*

This is known as the du Bois-Reymond lemma, and the usual proof of this result involves contradiction. Assuming there is a point where $f(x) \neq 0$ then continuity requires that there is a small interval where the function is either positive or negative. The existence of such an interval contradicts the zero integral assumption and therefore such a point cannot exist. Just as a note in passing, the fellow this result is named for, Paul du Bois-Reymond, was the brother of Emil, the noted physiologist.

The second result we need involves the rate of change of an integral in which the interval of integration depends on time, and it is known as the Reynolds Transport Theorem.

Theorem 6.2. *Assuming $\alpha(t), \beta(t), f(x,t)$ are smooth functions then*

$$\frac{d}{dt}\int_{\alpha(t)}^{\beta(t)} f(x,t)dx = \int_{\alpha(t)}^{\beta(t)} \frac{\partial f}{\partial t}dx + f(\beta,t)\beta'(t) - f(\alpha,t)\alpha'(t). \tag{6.18}$$

In looking at this result it might appear we have simply restated a version of the second fundamental theorem of calculus. This observation is correct. The Reynolds Transport Theorem, as usually stated, is for the time rate of change of a volume integral when the integration domain is time dependent. Our version is what is obtained when reducing the motion to the x-axis. In calculus this identity is sometimes identified as Leibniz's rule for differentiation of integrals.

Given that we are establishing the mathematical tools for the derivation of the equations of motion it is appropriate to mention the inverse function theorem. As stated earlier, the impenetrability of matter requires the invertibility of $x = X(A,t)$. What this means is that given t, if x_0 is in the range of X then there is only one A that maps onto x_0. This brings us to the following result, the Inverse Function Theorem.

Theorem 6.3. *Assume $\frac{\partial X}{\partial A}$ is defined and continuous. Also, for t fixed, assume $x_0 = X(A_0,t)$. If $\frac{\partial X}{\partial A}(A_0,t) \neq 0$ then $x = X(A,t)$ can be solved uniquely for A for x near x_0. Moreover, $\frac{\partial A}{\partial x}$ is defined and continuous in this interval.*

What this result shows is that in the interval occupied by the material, the impenetrability of matter requirement will be satisfied if we assume

$$\frac{\partial U}{\partial A} \neq -1.$$

This result follows directly from the definition of the displacement (6.2) and the requirement that $X_A \neq 0$. We also know, from (6.2), that $U_A(A,0) = 0$.

Assuming the motion is smooth then the impenetrability of matter requirement takes the form

$$\frac{\partial U}{\partial A} > -1. \tag{6.19}$$

In spatial coordinates, the requirement is that

$$\frac{\partial u}{\partial x} < 1. \tag{6.20}$$

Like the smoothness assumptions made earlier, the above inequalities will always be assumed to hold.

6.4 Continuity Equation

We will assume that mass is neither created nor destroyed. To understand the mathematical consequences of this assumption suppose at $t = 0$ we identify a segment of the bar, and assume this is an interval of the form $A_L \leq x \leq A_R$. At any later time this segment occupies an interval $\alpha(t) \leq x \leq \beta(t)$, where the endpoints are determined from the formulas $\alpha(t) = X(A_L, t)$ and $\beta(t) = X(A_R, t)$. Our assumption means that the total mass of the material in this interval does not change. If we let $\rho(x, t)$ designate the mass density of the material (i.e., mass per unit volume) then our assumption states that

$$\int_{\alpha(t)}^{\beta(t)} \sigma\rho(x, t)dx = \int_{A_L}^{A_R} \sigma\rho(x, 0)dx, \tag{6.21}$$

where σ is the (constant) cross-sectional area of the bar. Differentiating both sides with respect to t gives

$$\frac{d}{dt}\int_{\alpha(t)}^{\beta(t)} \sigma\rho(x, t)dx = 0. \tag{6.22}$$

Recalling that $\frac{\partial X(A,t)}{\partial t} = V(A, t)$ and $V(A, t) = v(X(A, t), t)$ then $\beta' = v(\beta, t)$ and $\alpha' = v(\alpha, t)$. With this, and the Reynolds Transport Theorem, we have the following

$$\begin{aligned}
&\frac{d}{dt}\int_{\alpha(t)}^{\beta(t)} \sigma\rho(x,t)dx \\
&\qquad = \int_{\alpha(t)}^{\beta(t)} \sigma\frac{\partial\rho}{\partial t}dx + \sigma\rho(\beta,t)\beta' - \sigma\rho(\alpha,t)\alpha' \qquad (6.23)\\
&\qquad = \int_{\alpha(t)}^{\beta(t)} \sigma\frac{\partial\rho}{\partial t}dx + \sigma\rho v|_{x=\beta} - \sigma\rho v|_{x=\alpha} \\
&\qquad = \int_{\alpha(t)}^{\beta(t)} \sigma\frac{\partial\rho}{\partial t}dx + \int_{\alpha(t)}^{\beta(t)} \sigma\frac{\partial}{\partial x}(v\rho)dx \qquad (6.24)\\
&\qquad = \int_{\alpha(t)}^{\beta(t)} \sigma\left(\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(v\rho)\right)dx.
\end{aligned}$$

Substituting this into (6.22) we have that

$$\int_{\alpha(t)}^{\beta(t)} \left(\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(v\rho)\right)dx = 0, \qquad (6.25)$$

and this is zero no matter what A_L and A_R we choose. Therefore, from the du Bois-Reymond lemma we conclude

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) = 0. \qquad (6.26)$$

This is the continuity, or mass conservation, equation in spatial coordinates. It is also the traffic flow equation. As with traffic flow, the mathematical formulation is not complete because v is unknown. The difference now is that there are forces within the material, and by accounting for them we will be able to derive an equation for the velocity. This does not mean, however, that we are out of the constitutive law business. As will be seen shortly, that step has just been postponed until later in the development.

As a final comment, the Reynolds Transport Theorem was used to obtain (6.23), and this resulted in the evaluation of the integrand at the endpoints. In (6.24) these were then combined into a single integral using the Fundamental Theorem of Calculus. These two steps are the core of the integral method for deriving an equation of motion. It should be expected that any time the method is used that these steps will be present in the derivation.

6.4.1 Material Coordinates

There are situations when it is easier if the equations are expressed in material coordinates. To do this for the continuity equation let $R(A,t)$ be the density in material coordinates. This means that $R(A,t) = \rho(X(A,t),t)$, and from

(6.10) we have that $\frac{\partial R}{\partial t} = \rho_t + v\rho_x$. Also, from Exercise 6.5, we have that $v_x = \frac{\partial}{\partial t}\ln(1+U_A)$. With these two formulas, the continuity equation (6.26) transforms into

$$\frac{\partial R}{\partial t} + R\frac{\partial}{\partial t}\ln\left(1+\frac{\partial U}{\partial A}\right) = 0.$$

Solving this first-order equation for R yields

$$R(A,t) = \frac{R_0(A)}{1+U_A}\,, \tag{6.27}$$

where $R_0(A) = R(A,0)$. It would appear that by using material coordinates we have solved the continuity equation. This conclusion is correct although the solution contains the displacement gradient and this is unknown until the momentum equation is solved.

6.5 Momentum Equation

We will now introduce Newton's Second Law, $F = ma$, into the formulation. Actually, for our problem it is convenient to write this in momentum form as $\frac{d}{dt}(mv) = F$. To begin, we itemize the forces that are involved and how they usually enter the mathematical problem. They include:

External Surface Forces. These are, as the name implies, forces that affect the motion across the outer surface of the bar. For example, if you were to pick the bar up and stretch it you would be applying a surface force. Generally, for us these will only appear in the problem through boundary conditions.

External Body Forces. These affect all material points in the bar and will appear in the equation as a known forcing function. In the formulation below $f(x,t)$ will represent the external body forces per unit mass, and so, ρf is the resultant body force. The standard example of an external body force is gravity, in which case $f = -g$ and the resultant force is $-g\rho$.

Internal Forces. These are the forces that the constituents making up the bar exert on each other. For example, if the bar is stretched then the material points in the bar pull on each other in an effort to restore the bar to its original length. To get a handle on these forces, note that given any cross-section, except those at the ends, you can separate the bar into a left and right side as shown in Figure 6.3. In this context the left side can be thought of as being pulled (or pushed) by the material on the right. Although we do not yet know exactly what this force is, let $\tau(x,t)$ denote its value per unit area. Because τ has the dimensions of force/area it is a stress function. At the moment it is unknown and in the next section we will discuss at some length how to remedy this situation. Nevertheless, a

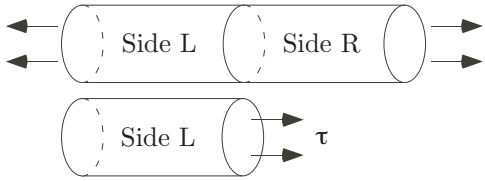


Figure 6.3 Any internal material cross-section can be thought of as separating the bar into a left (L) and right (R) side. As shown, the bar is being stretched and this results in a stress τ on side L due to the material on side R.

few things can be said about the stress. Newton's Third Law states that for every action, there is an equal and opposite reaction. What this means, in regard to Figure 6.3, is that if the stress on L due to R is τ, then $-\tau$ is the stress of L and R. The convention is that a force in the positive x-direction is positive. Consequently, if $\tau > 0$ then R is pulling on L, while if $\tau < 0$ then R is pushing on L.

The assumption made here is that the time rate of change of the momentum equals the sum of forces on the material. To understand the mathematical consequences of this assumption suppose that at $t = 0$ we identify a segment of the bar, and assume this is an interval of the form $A_L \leq x \leq A_R$. At any later time this segment occupies an interval $\alpha(t) \leq x \leq \beta(t)$, where the endpoints are determined from the formulas $\alpha(t) = X(A_L, t)$ and $\beta(t) = X(A_R, t)$. The total momentum of this segment is

$$\int_{\alpha(t)}^{\beta(t)} \sigma \rho v dx, \tag{6.28}$$

and the total body force on the segment is

$$\int_{\alpha(t)}^{\beta(t)} \sigma \rho f dx. \tag{6.29}$$

The force on the left end of the segment is $-\sigma\tau(\alpha, t)$, and on the right end it is $\sigma\tau(\beta, t)$. Therefore, from Newton's Second Law we obtain the equation

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma \rho v dx = \int_{\alpha(t)}^{\beta(t)} \sigma \rho f dx + \sigma\tau(\beta, t) - \sigma\tau(\alpha, t). \tag{6.30}$$

Using the same steps as in (6.24), differentiation of the integral on the left-hand side of the above equation yields

$$\frac{d}{dt} \int_{\alpha(t)}^{\beta(t)} \sigma \rho v dx = \int_{\alpha(t)}^{\beta(t)} \sigma \left(\frac{\partial(\rho v)}{\partial t} + \frac{\partial}{\partial x}(v^2 \rho) \right) dx. \tag{6.31}$$

As for the right-hand side of (6.30), we can write it as

$$\int_{\alpha(t)}^{\beta(t)} \sigma\rho f dx + \sigma\tau(\beta,t) - \sigma\tau(\alpha,t) = \int_{\alpha(t)}^{\beta(t)} \sigma\left(\rho f + \frac{\partial\tau}{\partial x}\right) dx. \tag{6.32}$$

Combining (6.31) and (6.32) we have that

$$\int_{\alpha(t)}^{\beta(t)} \sigma\left(\frac{\partial(\rho v)}{\partial t} + \frac{\partial}{\partial x}(v^2\rho) - \rho f - \frac{\partial\tau}{\partial x}\right) dx = 0. \tag{6.33}$$

This holds for any segment, and so from the du Bois-Reymond lemma it follows that

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial}{\partial x}(v^2\rho) - \rho f - \frac{\partial\tau}{\partial x} = 0. \tag{6.34}$$

Using the continuity equation (6.26), this can be written as

$$\rho\frac{Dv}{Dt} = \rho f + \frac{\partial\tau}{\partial x}\,, \tag{6.35}$$

where the material derivative $\frac{Dv}{Dt}$ is given in (6.11). This is the momentum equation for the bar expressed in spatial coordinates.

6.5.1 Material Coordinates

It is straightforward to rewrite the momentum equation in material coordinates. We know that $\rho(x,t) = R(A,t)$ and $\frac{Dv}{Dt} = \frac{\partial V}{\partial t}$. Also, letting $T(A,t)$ be the stress in material coordinate then $T(A,t) = \tau(x,t)$ for $x = X(A,t)$. With this, and Table 6.1,

$$\frac{\partial\tau}{\partial x} = \frac{1}{1+U_A}\frac{\partial T}{\partial A}\,. \tag{6.36}$$

Similarly, letting $F(A,t)$ be the body force in material coordinates then $f(x,t) = F(A,t)$. Substituting all of this information into (6.35), and making use of the continuity equation (6.37), one obtains

$$R_0\frac{\partial^2 U}{\partial t^2} = R_0 F + \frac{\partial T}{\partial A}\,, \tag{6.37}$$

where $R_0(A)$ is the initial density.

6.6 Summary of the Equations of Motion

To summarize the formulation of the equations of motion up to this point, we have found that in spatial coordinates the continuity and momentum equations are, respectively,

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) = 0, \tag{6.38}$$

$$\rho\left(\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x}\right) = \rho f + \frac{\partial \tau}{\partial x}\,. \tag{6.39}$$

In the above equations, ρ is the density, v is the velocity, τ is the stress, and f in the force per unit mass.

In material coordinates the equations take the form

$$R(A,t) = \frac{R_0}{1+U_A}\,, \tag{6.40}$$

$$R_0\frac{\partial^2 U}{\partial t^2} = R_0 F + \frac{\partial T}{\partial A}\,, \tag{6.41}$$

where $R_0 = R(A,0)$. In the above equations, R is the density, V is the velocity, T is the stress, and F in the force per unit mass.

Comparing the spatial and material versions it is easy to come to the conclusion that the material form is easier to use. For one reason the continuity equation does not need to be solved. Also, the material version of the momentum equation does not explicitly contain nonlinear terms, such as vv_x appearing in the spatial momentum equation. This does not mean, however, that the material version is linear as we have yet to determine how the stress is related to the density and displacement. Even so, the evidence appears to support the conclusion that the material version is easier to use. The fact is, however, that there are situations where the spatial version is preferred. Examples are easy to find in fluid dynamics because it is common when studying fluid motion to observe the flow from a fixed spatial position. In such cases the spatial version is more natural. At this point we will keep an open mind on the subject and use whichever seems to produce the easiest problem to solve.

As another observation, the mathematical problem consists of two equations involving several variables. The body force term in the momentum equation is assumed known. This leaves us with what looks to be three dependent variables to solve for, the density, the velocity, and the stress. So, we have either one too many unknowns or we are short one equation. The approach taken in continuum mechanics is to introduce a constitutive law for the stress, which relates it with the other two dependent variables. This is not a new situation for us as we had to do something similar in the traffic flow problem. Before doing this, we examine the steady-state solution.

6.7 Steady-State Solution

A simple yet informative problem involves the steady-state. This is the situation where the bar has come to rest, so the variables are independent of

time. Assuming there are no body forces then the momentum equation (6.41) reduces to

$$\frac{\partial T}{\partial A} = 0. \tag{6.42}$$

Therefore, at a steady-state with no body forces the stress T is a constant. To determine the value of T we need to know what was done to the bar to cause it to deform. In other words, we need to know the boundary conditions. Experimentally there are two commonly used testing methods, and they are considered in the examples below.

Example: Stress Relaxation Test

Consider the situation of when a bar of length ℓ_0 is stretched to length ℓ and held in this position. This is illustrated in Table 6.2. As shown, the bar initially occupies the interval $0 \leq x \leq \ell_0$, and it is then stretched so it occupies $0 \leq x \leq \ell$ for $t \geq t_0$. Because the left end of the bar is held fixed, it is relatively easy to write down the corresponding boundary condition. In material coordinates it is

$$U|_{A=0} = 0. \tag{6.43}$$

The bar is stretched to length ℓ, and for the steady-state problem this translates into the following boundary condition,

$$U|_{A=\ell_0} = \ell - \ell_0. \tag{6.44}$$

Although we know the displacement at the two ends we do not know how to relate this to the stress or displacement within the bar. This will have to wait until we specify the constitutive law for the stress. One last comment to make is that it is possible to express this steady-state problem in terms

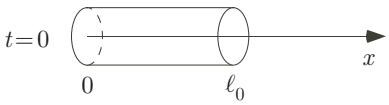
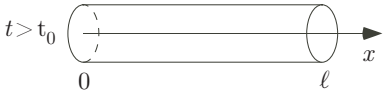
Configuration	Time	Material	Spatial
	$t = 0$	$0 \leq A \leq \ell_0$	$0 \leq x \leq \ell_0$
		$U(0,0) = 0$	$u(0,0) = 0$
		$U(\ell_0, 0) = 0$	$u(\ell_0, 0) = 0$
	$t > t_0$	$0 \leq A \leq \ell_0$	$0 \leq x \leq \ell$
		$U(0,t) = 0$	$u(0,t) = 0$
		$U(\ell_0, t) = \ell - \ell_0$	$u(\ell, t) = \ell - \ell_0$

Table 6.2 The differences between how a stretched, or compressed, bar is described using material and spatial coordinates for the stress relaxation example.

Figure 6.4 Typical testing system in which a material sample (on right) is tested to determine its deformational properties.

of the spatial variables, and the associated boundary conditions are given in Table 6.2. ■

Example: Creep Test

Another common method of stretching, or compressing, a bar is to apply a force on one of its ends. A situation where this arises is when the bar is vertical and a weight is attached to the lower end, which causes the bar to stretch. To put this into mathematical terms, assume the bar is held at $x = 0$, so the condition at this end is the same as before; namely, $U(0, t) = 0$. At the other end assume a constant force F_0 is applied. The boundary condition in this case is $T(\ell_0, t) = F_0/\sigma$. At steady-state the stress is constant throughout the bar, and this means $T(A, t) = F_0/\sigma$. Because we have been able to determine the stress in the bar we have gotten a bit further in solving the problem than in the stress relaxation example. However, we still do not know the displacement of the bar except at $x = 0$. Again, the issue is how to relate the stress to the displacement, and this is the reason for needing a constitutive law. ■

6.8 Constitutive Law for an Elastic Material

To complete the formulation we need to determine a constitutive law for the stress. This step requires more thought than is usually realized. It is not uncommon in modeling textbooks simply to state a law and then get on with the mathematics. Such an approach ignores some of the more interesting, and

important, questions that arise in applied mathematics. The reason is that determining a constitutive law requires close interaction of the mathematics with experiments, and even after decades of research the principles that underpin constitutive laws are still not completely understood.

The first question to address is what properties of the solution we can determine based on what is known about the problem so far. The objective is to compare the mathematical model with what is determined experimentally. An obvious choice is the steady-state solution obtained from stretching, or compressing, a bar from length ℓ_0 to length ℓ. As shown in the preceding section, at steady-state the stress T is constant throughout the bar. This is useful information because one of the most common material testing experiments involves measuring the steady-state stress as a function of ℓ. A typical experimental setup is shown in Figure 6.4. For testing in tension, so $\ell > \ell_0$, samples are cut into small strips and inserted into a computer-controlled testing machine. The range of such experimental systems is enormous, from huge machines that are capable of testing samples the size of a car, down to microscopic systems that are used to test single molecules. Given this, it is not surprising that the types of materials tested in this way are quite varied, and four examples are given in Figure 6.5. In this figure the measured stress is given as a function of the extension ratio

$$\lambda = \frac{\ell}{\ell_0}\,. \tag{6.45}$$

With this, the material is in tension if $\lambda > 1$, and it is in compression if $\lambda < 1$. The range of extension in the figure varies significantly with the material. For example, the range for steel is much smaller than it is for rubber. This difference is not surprising. Also, steel has the odd feature that the stress

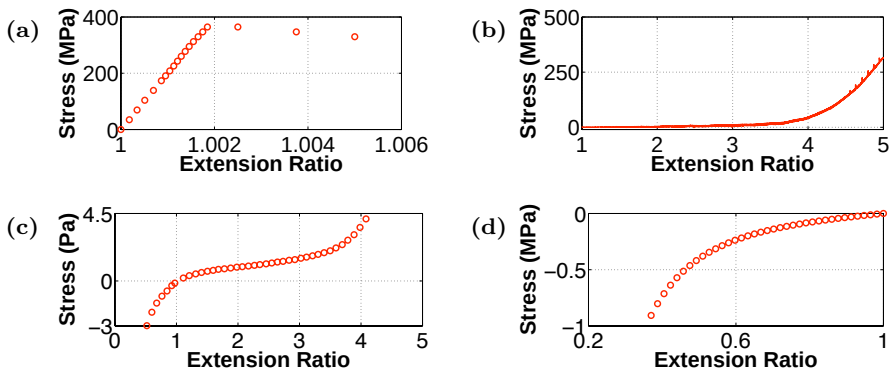


Figure 6.5 Stress measured as a function of the extension ratio (6.45), for (a) steel, (b) capture silk from a spider web (Blackledge and Hayashi [2006]), (c) rubber (Raos [1993]), and (d) articular cartilage (Kwan [1985]).

starts decreasing at larger extension ratios. This is due to the metal being pulled apart, and it is characteristic of what are called ductile materials. In contrast, brittle materials, such as glass, simply break. We will assume the displacements are not so extreme as to cause this irreversible behavior to occur, and when the force is removed that the material will return to its original shape.

In looking at the data in Figure 6.5 it is apparent that, for the materials shown, the stress increases with the imposed displacement $U = \ell - \ell_0$. This is consistent with the everyday observation that the more you stretch something the greater it resists. Based on this it might seem reasonable that for our constitutive law we should assume $T = T(U)$. The fact is, however, that this is not possible. We found earlier that at a steady-state the stress is constant. If $T = T(U)$ then this would require that the displacement is also constant. The difficulty with this is that we require $U = 0$ at the left end of the bar and $U = \ell - \ell_0$ at the right end. It is therefore impossible for the displacement to be constant and, consequently, it is not possible to assume the stress is a function only of the displacement.

6.8.1 Derivation of Strain

A way to correct the difficulty discussed in the previous paragraph is to assume the stress depends on the relative displacement. There are various ways to measure relative displacement and an example is $(\ell - \ell_0)/\ell_0$, which compares the displacement $\ell - \ell_0$ to the original length ℓ_0. There are other ratios for measuring relative displacement and some of the more commonly used are listed in Table 6.3. At this point there is no clear reason why you would want to pick one over another and we will use the Lagrangian strain, leaving the others for the exercises. For cultural reasons it is worth saying something about the names given the different strain measures. The ratio used to derive the Lagrangian strain is known in the literature as the engineering or nominal strain. You will also see the Hencky strain referred to as the natural, or true, strain. In this text whenever referring to strain it is understood we are using the Lagrangian strain as defined in Table 6.3.

The basic assumption for our constitutive law is that the stress depends on the relative displacement $(\ell - \ell_0)/\ell_0$. To be more precise, we will assume that the stress at a material point depends on the relative displacement in the immediate vicinity of this point. To translate this into mathematical terms, given a cross-section located initially at A, consider cross-sections at $A \pm \Delta A$ that are just to the left and right (see Figure 6.6). After time t, the cross-section on the right moves to $X(A + \Delta A, t)$, and the one on the left moves to $X(A - \Delta A, t)$. The length of this small segment of the bar is $X(A + \Delta A, t) - X(A - \Delta A, t)$, while the original length was $2\Delta A$. Recalling that $X = A + U$, then the ratio for the relative displacement is

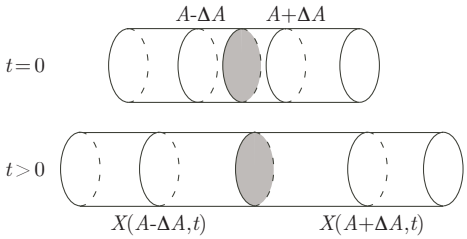


Figure 6.6 A segment starts off centered at A with length $2\Delta A$. At time t the segment has length $X(A+\Delta A,t)-X(A-\Delta A,t)$.

$$\frac{\text{new length} - \text{original length}}{\text{original length}} = \frac{X(A+\Delta A,t)-X(A-\Delta A,t)-2\Delta A}{2\Delta A} = \frac{U(A+\Delta A,t)-U(A-\Delta A,t)}{2\Delta A}. \tag{6.46}$$

Assuming ΔA is small, then using Taylor's theorem,

$$U(A\pm\Delta A,t) = U(A,t)\pm\Delta A\frac{\partial U}{\partial A}(A,t)+\frac{1}{2}\Delta A^2\frac{\partial^2 U}{\partial A^2}(A,t)\pm\frac{1}{6}\Delta A^3\frac{\partial^3 U}{\partial A^3}(A,t)+\cdots.$$

Introducing these into (6.46) we obtain

$$\frac{\text{new length} - \text{original length}}{\text{original length}} = \frac{\partial U}{\partial A}(A,t)+O(\Delta A^2). \tag{6.47}$$

Therefore, a local measure of the relative distortion in the vicinity of a material point is

Name	Ratio	Definition
Lagrangian Strain	$(\ell-\ell_0)/\ell_0$	$\epsilon = U_A$
Eulerian Strain	$(\ell-\ell_0)/\ell$	$\epsilon_e = u_x$
Green Strain	$(\ell^2-\ell_0^2)/(2\ell_0^2)$	$\epsilon_g = U_A+\frac{1}{2}U_A^2$
Almansi Strain	$(\ell^2-\ell_0^2)/(2\ell^2)$	$\epsilon_a = u_x-\frac{1}{2}u_x^2$
Midpoint Strain	$2(\ell-\ell_0)/(\ell+\ell_0)$	$\epsilon_m = U_A/(1+\frac{1}{2}U_A)$
Hencky Strain	$\ln(\ell/\ell_0)$	$\epsilon_h = \ln(1+U_A)$

Table 6.3 Various strain measures used in continuum mechanics.

$$\epsilon = \frac{\partial U}{\partial A}\,, \tag{6.48}$$

which is known as the Lagrangian strain, or in this textbook, simply the strain.

With the definition of strain given in (6.48), the assumed constitutive law for the stress is $T = T(\epsilon)$. A material for which this holds is said to be *elastic*. The momentum equation, in material coordinates, in this case is

$$R_0 \frac{\partial^2 U}{\partial t^2} = R_0 F + T'\left(\frac{\partial U}{\partial A}\right) \frac{\partial^2 U}{\partial A^2}\,. \tag{6.49}$$

This is a wave equation for U, and it is nonlinear if the stress is a nonlinear function of the strain. Not just any function can be used for the stress, and later in the chapter we will investigate some of the restrictions that must be imposed on how it depends on strain.

One last useful piece of information concerns the extension ratio (6.45). Given the result in (6.47), when deriving the continuum formulation by letting $\Delta A \to 0$, the extension ratio λ turns into $1 + \epsilon$. The reason for pointing this out is that in the simplification of the constitutive law for the stress that is given below, we will investigate how the measured stress behaves in the neighborhood of $\lambda = 1$. In the continuum formulation this is equivalent to looking at how the stress behaves around $\epsilon = 0$.

6.8.2 Material Linearity

With the assumption that $T = T(\epsilon)$, we return to the stress curves in Figure 6.5. The dependence of T on ϵ clearly depends on the material. This is reasonable as the morphological and mechanical characteristics of these materials are markedly different. Even so, there is a region for each material, containing $\epsilon = 0$, where the stress is approximately a linear function of strain. The constitutive law in this case reduces to

$$T = E\,\frac{\partial U}{\partial A}\,, \tag{6.50}$$

where E is known as Young's, or the elastic, modulus. A material that follows this law is said to be linearly elastic. The momentum equation (6.49) in this case reduces to

$$R_0 \frac{\partial^2 U}{\partial t^2} = R_0 F + E\,\frac{\partial^2 U}{\partial A^2}\,. \tag{6.51}$$

This is a linear wave equation for the displacement U.

In the parlance of continuum mechanics, (6.50) is an assumption of material linearity. It should be understood that this constitutive law is not based on a requirement of a small strain. The strains for which (6.50) is valid need

Material	Young's Modulus (GPa)	Density (kg/m^3)
Multiwall carbon nanotube	1200	2600
Diamond	1000	3500
Stainless Steel	200	8030
Glass	65	2600
White Oak	12	680
Beeswax	0.2	960
Rubber	0.007	1200
Silica Aerogel	0.001	100

Table 6.4 Young's modulus and density of various materials. These represent the current range of values for these parameters.

not be small, and it is only required that (6.50) furnishes an accurate approximation of the stress over the given range of strains. For example, in Figure 6.5, steel is linearly elastic for strains up to about 0.002 while the capture fibers from a spider web are linear up to strains of approximately 2.0.

In terms of dimensional units, because strain is dimensionless, the elastic modulus has the same dimensions as stress. The basic unit for stress is the Pascal (Pa) and $1\,\mathrm{Pa} = 1\ \mathrm{N/m^2}$. Relative to the strength of most materials, however, a Newton (N) is relatively small. To illustrate this, 1 N is the force on an object with the mass of approximately an apple, and it takes a lot of apples to deform steel or glass a noticeable amount. For this reason, the elastic moduli of most materials run in the MPa or GPa range, where $\mathrm{M} = 10^6$ and $\mathrm{G} = 10^9$. A few examples are given in Table 6.4. An observation coming from this table is that one should not assume that denser materials have larger elastic moduli. However, the density and modulus are connected through the molecular structure of the material, and this will be discussed later in the chapter.

Example: Stress Relaxation

In Section 6.7 we were unable to solve the steady-state problem. The situation has improved with the introduction of the linear constitutive law in (6.50). The reduced momentum equation (6.46) now takes the form

$$\frac{\partial^2 U}{\partial A^2} = 0.$$

The solution of this that satisfies the boundary conditions (6.43) and (6.47) is $U = (\ell - \ell_0)A/\ell_0$. With this the stress is $T = E(\ell - \ell_0)/\ell_0$. Also, from (6.40), the density is $R = ER_0/(T + E)$. Therefore, as advertised, with the inclusion of the constitutive law, we have been able to determine the stress, displacement, and density in the bar. ■

Example: Bungie Cord

Apparently, for bungie jumpers it is more entertaining if the cord is long enough that the jumper comes very close to hitting the ground. Getting the right length cord is not a simple matter of just knowing the weight of the jumper and the height of the jump. The reason is that the weight of the cord will cause it to extend. To determine this suppose the cord is hung with the upper end attached and the lower end free (see Figure 6.7). Assume gravity is the only force present, and the cord starts off with a constant density R_0 and length ℓ_0. Also, assume that the cord is at rest and its length after being hung is ℓ. Given that the cord is at a steady-state then the material momentum equation (6.41) takes the form

$$\frac{dT}{dA} = -R_0 g.$$

As for the boundary condition, the lower end is free and this means the stress is zero there. The boundary condition in this case is $T = 0$ at $A = \ell_0$. From the above momentum equation, and the given boundary condition, we obtain

$$T = R_0 g(\ell_0 - A). \tag{6.52}$$

To determine the displacement of the cord we need to specify the constitutive law for the stress, and we will use (6.50). With this, and (6.52), we have that

$$\frac{dU}{dA} = \frac{R_0 g}{E}(\ell_0 - A).$$

Integrating this equation, and using that fact that the upper end is fixed, so $U = 0$ at $A = 0$, then

$$U = \frac{R_0 g}{2E} A(2\ell_0 - A). \tag{6.53}$$

The bungie cord problem is now solved, and with the solution it is possible to determine just how far the cord will stretch. The displacement of the free end is obtained from (6.53) by setting $A = \ell_0$. The total length ℓ of the stretched cord is obtained by adding this displacement to the original length ℓ_0, and

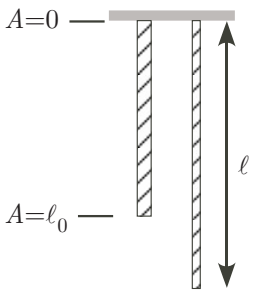


Figure 6.7 A bungie cord, originally with length ℓ_0, stretches to length ℓ after having been hung.

the result is

$$\ell = \ell_0 \left(1 + \frac{R_0 g}{2E} \ell_0\right). \tag{6.54}$$

This shows that a less stiff cord (E small) stretches longer. This is expected, but the above result shows that the stretched length is not a simple multiple of the modulus. For example, it does not happen that reducing the modulus by a factor of two causes the length to double. Also, because of the nonlinear dependence of ℓ on ℓ_0, we have found that longer cords stretch proportionally longer than shorter cords. ■

6.8.3 End Notes

The basic equations for elasticity were developed by Robert Hooke, and (6.50) is sometimes referred to as Hooke's law. Given this, it might seem odd that the one parameter that appears in the equations is named after a physician named Thomas Young. The reason for this is that Hooke's original statement that "as is the extension, so is the force" implies that the force is proportional to displacement. For springs this might be acceptable but as we saw earlier this assumption is inapplicable to elastic bars. It was Young who interpreted it correctly using strain.

The statement that the stress is a linear function of strain depends on the strain and coordinate used in the formulation. For example, using (6.16), the constitutive law (6.50) expressed in spatial coordinates is

$$\tau = E \frac{u_x}{1 - u_x}. \tag{6.55}$$

Consequently, an assumption of material linearity using one strain measure does not necessarily mean it is linear using another strain measure.

In the experiments used to produce the data in Figure 6.5, the experimenter waits until the motion stops before measuring the stress. This means that the constitutive law for the stress is determined using the steady-state response. Even so, the linear constitutive law (6.50) is assumed to apply even when the bar is in motion. If the stress also depends on rate variables, such as v or v_x, our approach of using the steady-state to determine the constitutive law would miss this completely. There are materials that depend strongly on rate variables and examples are water, jello, and silly putty. To determine the correct rate dependence requires dynamic tests and several are commonly used in material testing. Exactly how this is done will be explained when we study viscoelasticity in the next chapter.

6.9 Morphological Basis for Deformation

The constitutive law used for the stress is a mathematical expression for how the material reacts to deformation. The materially linear assumption in (6.50) is routinely used to describe such diverse materials as steel, rubber, and skin. Given the differences in the atomic, or molecular, structure of these materials it is of interest to be able to understand how the substructural changes that take place during deformation give rise to the constitutive law for the stress.

6.9.1 Metals

Undoubtedly, the most studied metal is steel. A typical stress-strain curve for steel is given in Figure 6.5(a), but a more complete version is shown in Figure 6.8. Because of the larger range of the extension ratio, the linear portion of the curve is not as evident as it is in Figure 6.5(a). However, what is apparent is that at larger strains the material is far from a simple linear function. It is also evident that the curve is not monotonic, and as explained in the next section this generates rather serious mathematical difficulties.

To understand how the microstructure of metal accounts for the observed deformation, the atoms in most metals are arranged in a periodic array, forming a lattice pattern. In this description the atoms are modeled as spheres. As examples, the radius of iron is 1.24 nm, the radius for copper is 1.28 nm, and for silver it is 1.44 nm. The dominant attractive force on the atoms is due to metallic bonding, which arises from the positively charged metal atoms sharing electrons. The resulting force has the form $F_a = \alpha/r^m$, where r is the separation distance between atoms, and α is a constant determined by the electronic characteristics of the material. For many metals $m = 4$. There is also a repulsive force that comes into play if the electron shells of the atoms overlap, and it is based on the Pauli exclusion principle. The associated form

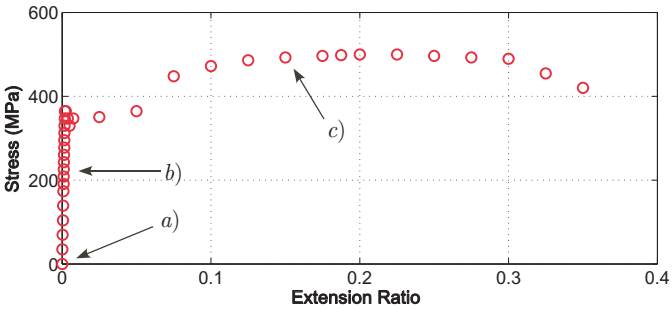


Figure 6.8 The complete stress-strain curve for the steel sample shown in Figure 6.5(a). The atomic configuration at (a), (b), and (c) is shown in Figure 6.9.

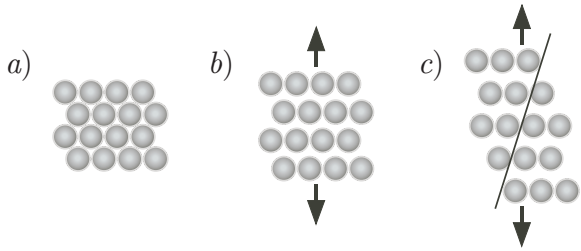


Figure 6.9 The atomic configurations in a metal during deformation: (a) the atoms when no load is applied; (b) their position in the elastic region; and (c) the appearance of a slip plane for larger strain values.

of this force is $F_r = -\beta/r^n$, where β is a constant. The value of n depends on the material, and typical values are $n = 11$ for copper and $n = 12$ for silver (Kimura et al. [2000]). The resulting force is

$$F = \frac{\alpha}{r^m} - \frac{\beta}{r^n}\,, \tag{6.56}$$

where $1 < m < n$, and α, β are positive constants.

In material science the properties of metals are often characterized using energy, and for this reason the force is written in terms of a potential function V. This is done by writing $F = \frac{dV}{dr}$, where

$$V = \frac{-\alpha}{(m-1)r^{m-1}} + \frac{\beta}{(n-1)r^{n-1}}\,. \tag{6.57}$$

This function, along with F, is sketched in Figure 6.10.

When no load is applied, so the atoms are in their equilibrium configuration, the two forces balance. Setting $F = 0$ determines the equilibrium interatomic spacing r_0, and one finds that

$$r_0 = \left(\frac{\beta}{\alpha}\right)^{\frac{1}{n-m}}\,. \tag{6.58}$$

This configuration is shown in Figure 6.9(a). As examples, for copper $r_0 = 1.25$ nm, and for silver $r_0 = 1.29$ nm. This means that the distance between the atoms, at equilibrium, is slightly larger than twice the atomic radius.

As the metal bar is stretched, the distance between the atomic layers increases, and the bonds between the atoms resist this change as described in (6.56). This is illustrated in Figure 6.9(b). If the load is not too large the bonds do not break, and when the load is removed the atoms return to their original positions in the lattice, shown in Figure 6.9(a). To relate the stress with the interatomic force, Figure 6.9(b) shows four cross-sections that are made up of atoms. To calculate the force between any two such cross-sections,

note there are approximately $\sigma/(4r_0^2)$ atoms in a square cross-section of area σ. So, the stress is approximately

$$\begin{aligned} T &\approx \frac{\sigma}{4r_0^2}\frac{F}{\sigma} \\ &= \frac{F}{4r_0^2}\,. \end{aligned} \tag{6.59}$$

In a similar manner, for r near r_0, the linear elastic law (6.50) can be approximated as

$$T(r_0) + (r - r_0)T'(r_0) \approx E\,\frac{r - r_0}{r_0}\,. \tag{6.60}$$

Combining (6.59) and (6.60), it follows that

$$\begin{aligned} E &\approx \frac{1}{4r_0}\frac{dF}{dr}\Big|_{r=r_0} \\ &= \frac{(n-m)\alpha}{4r_0^{m+2}}\,. \end{aligned} \tag{6.61}$$

Consequently, the elastic modulus has a strong dependence on the interatomic spacing. Another observation is that the atomic mechanisms involved with tension, where $r > r_0$, are fundamentally different than those involved with compression, where $r < r_0$. This is why knowing the stress-strain function for a nonlinearly elastic material for tension provides little insight into what the stress function is for compression.

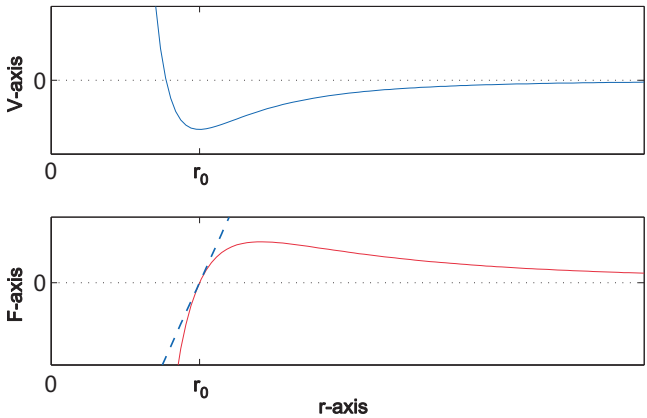


Figure 6.10 The force (6.56) and the potential (6.57) on the atoms in a metal. The dashed line is the tangent to the force curve at the point where $F = 0$. The slope of this line is used to obtain an approximation of the Young's modulus in (6.61).

The largest load that, when removed, results in the atoms returning to their original configuration is known as the elastic limit. From Figure 6.5(a) the elastic limit corresponds to an extension ratio of approximately 0.002. If a larger load is used slip planes will start to appear, and one is illustrated in Figure 6.5(c). As the name suggests, at a slip plane the atoms slide over each other along a plane. This is a permanent modification and if the load is removed the slip planes remain. In this situation the material is said to be plastic. Stretching the bar any farther produces more slip planes. Other defects in the atomic structure appear, including dislocations, and the specific events depend on the metal being tested. Eventually the metal is not capable of withstanding the stretching and breaks, a point material scientists call fracture. In Figure 6.8 this happens when the extension ratio reaches about 0.35.

Given the rather complicated atomic interactions taking place in the steel bar, it is natural to ask whether our continuum model can be used once slip planes start to appear. The answer is yes and no. It is no because the shearing motion generated by the slip planes violates our assumption that the cross sections move as a unit. However, if we were to use a continuum model with three spatial dimensions then this would not be an issue. In this case the answer is yes, although it would require some effort to determine the appropriate constitutive law for the stress function.

The function in (6.57) is a special case of what is known as a Sutton-Chen potential. We only considered what is effectively a nearest neighbor approximation using this potential, which means that we only considered the forces between a molecule and it nearest neighbor in the adjoining cross-section. A more realistic description would account for the other molecules, in which case F would consist of a sum of attractive and repulsive forces. It is also interesting that several other functions have been proposed, each accounting for different interatomic forces. One of the better known is the Lennard-Jones potential, where the attractive force is due to van der Waals bonding. Mathematically, there is little difference in the model as (6.57) still applies but $m = 7$, which reflects that fact that van der Waals bonding is weaker than metallic bonding. There has been considerable research in the last few years into what is called atomistic-based continuum theory, where the material's continuum properties are derived using interatomic potential functions. An introduction to this can be found in Finnis [2004].

6.9.2 Elastomers

An elastomer is rubber made with a loosely cross-linked molecular structure. To explain what this means, natural rubber is made up of long individual molecules, or more specifically, from long polymer chains. In effect, it is the molecular version of spaghetti. As with spaghetti, the molecules of natural

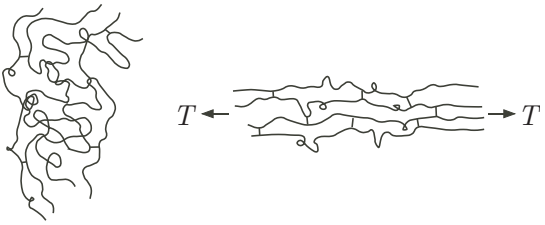


Figure 6.11 Elastomer network, before and after the application of an axial load. The cross-links between the rubber molecules increase its ability to resist the load, and enable the network to return to its original configuration once the load is removed.

rubber are not connected and this means it is more like a liquid than a solid. This changes if sulfur is added because this produces atomic bridges between the polymer chains. The consequence of this is a material that consists of long entangled molecules that are cross-linked, and a schematic of this is given in Figure 6.11 Assuming that the number of cross-links is not too large, one produces what is known as an elastomer. Such materials are formed from a three-dimensional molecular network in which highly flexible molecules are connected at points provided by cross-links between the molecules.

In stretching such an elastomer, the entangled polymer chains start to straighten. They are very extendable but at large extensions the cross-links mean that the movement of chains relative to one another is minimal. Consequently, upon the application and release of a stress, the molecules quickly revert to their normal crumpled form in the unstressed configuration, and this is the basis of the reversible high extensibility of elastomeric solids. This scenario applies to Figure 6.5(b),(d). Both the capture silk and rubber offer relatively little resistance for extension ratios up to about 3. This is the interval over which the polymer molecules are uncoiled. Once that happens, and the cross-links become engaged, both materials show significant resistance and the stress increases almost exponentially. An example of a constitutive law incorporating this into the formulation is examined in Exercise 6.25. A more extensive investigation into the molecular contributions to the elastic behavior of an elastomer can be found in Mark and Erman [2007].

6.10 Restrictions on Constitutive Laws

One of the central problems in continuum modeling is finding the appropriate constitutive law for the stress. What is appropriate depends on what the model is describing. If the goal is to determine the deformation of a table due to the load of a computer then the strains are likely so small that a linear theory can be used. On the other hand, if you are interested in the deflection of a trampoline then the strains are likely so large that a nonlinear theory

would be required. One question that arises in such cases is, what function should be used to describe this nonlinear behavior? As an example, for the data for rubber in Figure 6.5(c), the curve resembles a cubic. Based on this observation, one might assume that

$$T = a\epsilon + b\epsilon^3, \tag{6.62}$$

where $\epsilon = \frac{\partial U}{\partial A}$. On the other hand, the data for capture silk in Figure 6.5(b) look to follow more of an exponential function, and a possible constitutive law that could be used in this case is

$$T = a\left(e^{b\epsilon} - 1\right). \tag{6.63}$$

One of the standard answers to the question of what function to use is that it is simple, and it does a reasonable job describing the stress-strain data. Although reproducing the experimental results is a worthy goal, you want the model to also describe the motion in situations for which you do not have data. As an example, we known that strains must satisfy $-1 < \epsilon < \infty$. So, suppose one of the above nonlinear functions is used to fit data in the range $-0.5 < \epsilon < 50$. It is questionable that either one would successfully describe what happens for $-1 < \epsilon < -0.5$ because both predict a finite stress when the material is compressed to zero (i.e., when $\epsilon \to -1$). Because of this, it is worth imposing a requirement on the constitutive law to guarantee the right behavior under the extreme condition of letting $\epsilon \to -1$. It is the objective of this section to develop some general requirements that can be used to help formulate the constitutive law.

6.10.1 Frame-Indifference

Considering what happens to the stress when $\epsilon \to -1$ falls into the extreme behavior category. Another category relates to consistency. An example of this came up earlier when we concluded that it is impossible for the constitutive law for the stress to depend on the displacement. A general principle, that includes this situation, is given below.

Principle of Material Frame-Indifference: The response of any material must be independent of the observer.

To illustrate how this affects the constitutive law for the stress, suppose two observers are moving in opposite directions. Because they will record different velocities of a cross-section, velocity is not a frame-indifferent function. Therefore, any constitutive law depending on velocity will be observer-dependent. The above principle excludes such functions from consideration in the formulation of the constitutive law for the stress.

The argument used to rule out velocity needs to be made mathematically precise, and this leads us to the following definition.

Definition 6.1. Suppose two spatial coordinate systems (x, t) and (x^*, t^*) are related through a change of coordinates given as $x^* = x + b(t)$ and $t^* = t - t_0$. Using a superscript $*$ to denote the value of a quantity in the (x^*, t^*) system then a function $f(x, t)$ is *frame-indifferent* if $f(x, t) = f^*(x^*, t^*)$ for all smooth b and t_0.

As expressed in this definition, observers are related through a translation, with the spatial motion allowed to be time dependent. This is known as an Euclidean transformation. It differs from a Galilean transformation, often used in Newtonian physics, because the latter assumes b to be a linear function of time.

The assumption made in continuum mechanics is that the density ρ and stress τ are frame-indifferent. This is important for us because it means that the constitutive law for the stress can only depend on frame-indifferent quantities. Examples of frame-indifferent, and non-frame-indifferent, functions will be given below. However, before doing that, the principle is stated in terms of spatial coordinates, and we need to understand how it can be used when using material coordinates. This can be explained using an example.

In the case of when the material is elastic, we know that the constitutive law for the stress, in material coordinates, can be written in the general form

$$T = T(U_A). \tag{6.64}$$

This is equivalent to the statement that the constitutive law for the stress, in spatial coordinates, for an elastic material can be written in the general form

$$\tau = \tau(u_x). \tag{6.65}$$

The reason this is equivalent is because, from (6.16),

$$\frac{\partial U}{\partial A} = \frac{u_x}{1 - u_x}\,. \tag{6.66}$$

This enables us to transform (6.64) into an expression of the form given in (6.65). For example, if $T = \alpha U_A^3$ then, from (6.66), $\tau = \alpha u_x^3/(1 - u_x)^3$. What the equivalence of (6.64) and (6.65) gives us is that if (6.65) satisfies the Principle of Material Frame-Indifference then so does (6.64). In other words, the material coordinate version of a constitutive law satisfies the Principle of Material Frame-Indifference if its spatial version does.

Examples

1. The displacement function is not frame-indifferent. To explain why, in each coordinate system we have a different position function, so $x = X(A, t)$ and

$x^* = X^*(A, t^*)$. Given that the change of coordinates is $x^* = x + b(t)$, then the position functions satisfy $X^* = X + b$. Expressing this equation in terms of the displacement function we have that

$$U^*(A, t^*) = U(A, t) + b(t). \tag{6.67}$$

Given (6.4), in terms of spatial coordinates, the above equation takes the form $u^*(x^*, t^*) = u(x, t) + b(t)$. To be frame-indifferent we must be able to conclude that $u^* = u$, no matter what b we select. Clearly this does not happen and the conclusion is that the displacement is not frame-indifferent. Therefore, from the Principle of Material Frame-Indifference, τ cannot be assumed to depend on u, and T can not be assumed to depend on U. So, we are able to eliminate this possibility without resorting to special solutions of the steady-state problem as was done in Section 6.8.

2. The velocity function is not frame-indifferent. This follows by taking the time derivative of (6.67) and concluding that $V^* = V + b'(t)$. From (6.5), this can be written as $v^* = v + b'(t)$. Given that $b'(t)$ is not necessarily zero, it follows that the velocity is not frame-indifferent.

3. The strain function $\frac{\partial U}{\partial A}$ is frame-indifferent. This follows by taking the A derivative of (6.67) and concluding that $U_A^* = U_A$. We also conclude, from (6.66), that the strain function $\frac{\partial u}{\partial x}$ is frame-indifferent. Therefore, the assumption underlying the constitutive law for an elastic material satisfies the Principle of Material Frame-Indifference.

4. The function $\frac{\partial \rho}{\partial t}$ is not frame-indifferent. To prove this, the density in the two coordinate systems must satisfy $\rho(x, t) = \rho^*(x^*, t^*) = \rho^*(x + b, t - t_0)$. Consequently,

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \frac{\partial \rho^*}{\partial t} + \frac{\partial \rho^*}{\partial x^*}\frac{\partial x^*}{\partial t} \\ &= \frac{\partial \rho^*}{\partial t^*} + b'(t)\frac{\partial \rho^*}{\partial x^*}\,. \end{aligned}$$

Any change of coordinates with $b' \neq 0$ means $\rho_t \neq \rho^*_{t^*}$, and so this function is not frame-indifferent. ■

Given this requirement on the constitutive law it is worth having a small list of functions that are frame-indifferent. Functions that are frame-indifferent include

$$\rho, \frac{D\rho}{Dt}, \frac{\partial U}{\partial A}, \frac{\partial u}{\partial x}, \frac{\partial V}{\partial A}, \frac{\partial v}{\partial x}\,. \tag{6.68}$$

Functions that are not frame-indifferent include

$$\frac{\partial \rho}{\partial t}, U, u, V, v\,. \tag{6.69}$$

Are there materials that use multiple frame-indifferent functions in the constitutive model? The answer is yes, and they are very common. A simple example is a viscoelastic material, where one assumes the stress depends on the strain U_A and the strain rate V_A. Examples such as this are explored in the next chapter.

6.10.2 Entropy Inequality

There are several other principles used to formulate constitutive laws. We will only consider one more, and it is related to the second law of thermodynamics. This requires the introduction of three new variables, and the first is connected with the energy. As with all mechanical systems, the energy involves both kinetic and potential components. It is relatively easy to identify the kinetic energy density, and it is $\frac{1}{2}\rho v^2$. The potential energy has multiple sources, and one comes from the external forcing. Another comes from the ability of the material to store energy, in the same way a spring stores energy when it is compressed. Because this component arises from the properties of the material, it is known as the internal energy. We want to determine this in our continuum theory, and with this in mind let $\chi(x,t)$ be the internal energy density per unit mass.

Like the density and momentum, the energy is assumed to satisfy a balance law, and it is

$$\frac{d}{dt}\int_{\alpha(t)}^{\beta(t)} \sigma\rho\left(\frac{1}{2}v^2+\chi\right)dx = \int_{\alpha(t)}^{\beta(t)} \sigma\rho v f dx + \sigma v\tau\Big|_{x=\alpha}^{x=\beta}. \tag{6.70}$$

In words, the above equation states that the rate of change of the total energy of a material segment equals the sum of the rate of work of the external forces and the rate of work done by the forcing on the ends of the segment. Using the same argument employed to derive the continuity and momentum equations, the above expression results in the following equation

$$\rho\frac{D\chi}{Dt} = \tau\,\frac{\partial v}{\partial x}\,. \tag{6.71}$$

This gives us an equation that can be solved to determine the function χ.

The second variable that needs to be introduced is $\eta(z,t)$, which is the entropy density per unit mass. As expressed in the second law of thermodynamics, it is assumed that the entropy does not decrease. In other words, it is assumed that

$$\frac{D\eta}{Dt} \geq 0. \tag{6.72}$$

In continuum mechanics this is known as the Clausius-Duhem inequality. It is assumed here that there is no supply or flux of entropy. This can occur, for example, when there is heat flow in the system. In our development, the thermal affects are omitted.

The third, and final, function that needs to be introduced is the *Helmholtz free energy density* ψ, defined as

$$\psi = \chi - \theta\eta, \tag{6.73}$$

where θ is the absolute temperature. Consistent with our earlier assumptions, θ is assumed to be constant. The reason for calling ψ the free energy is that it represents the energy remaining to do work after accounting for what is invested in the entropic state of the material.

Solving (6.73) for η, and then substituting the result into the Clausius-Duhem inequality (6.72), yields

$$-\rho\frac{D\psi}{Dt} + \tau\frac{\partial v}{\partial x} \geq 0. \tag{6.74}$$

This is known as the *reduced entropy inequality.* In material coordinates this inequality takes the form

$$-R_0\frac{\partial \Psi}{\partial t} + T\frac{\partial V}{\partial A} \geq 0, \tag{6.75}$$

where Ψ is the material form of the Helmholtz free energy. It is assumed here that the spatial and material forms of the free energy functions give the same value. Therefore, if a cross-section that starts at A is currently located at $x = X(A,t)$ then $\Psi(A,t) = \psi(x,t)$.

We are now in position to state the second requirement imposed on constitutive laws.

Principle of Dissipation. A constitutive law must satisfy the reduced entropy inequality (6.74), or equivalently (6.75), for all values of its arguments.

Now comes the question of exactly how we use this condition because it involves the yet to be determined Helmholtz free energy ψ. We will show that in certain cases the stress τ can be determined from ψ. This means that instead of formulating a constitutive law for the stress, we can specify one for ψ, and then use this to determine τ. In doing this it is assumed that the constitutive law for ψ depends on the same variables used for the stress.

Example

For an elastic material, the general form of the constitutive law in spatial coordinates is $\tau = \tau(u_x)$. The corresponding assumption for the constitutive

law for the free energy function is $\psi = \psi(u_x)$. To see what this gives us, note, using the chain rule and Exercise 6.5(f),

$$\begin{aligned}\frac{D\psi}{Dt} &= \psi' \frac{D}{Dt} u_x \\ &= \psi'(1 - u_x) v_x.\end{aligned}$$

With this, (6.74) reduces to

$$[\tau - \rho(1 - u_x)\psi'(u_x)]\, v_x \geq 0. \tag{6.76}$$

According to the Principle of Dissipation, this inequality must hold for all values of v_x. For example, it must hold when $v_x = 1$, and when $v_x = -1$. Because the quantity in the square brackets does not depend on v_x, it must be that $\tau - \rho(1 - u_x)\psi'(u_x) = 0$. Therefore, for an elastic material, the stress is determined from the free energy as follows,

$$\tau = \rho(1 - u_x)\psi'(u_x). \quad \blacksquare \tag{6.77}$$

In the above example spatial coordinates were used. If one uses material coordinates, and assumes that the material is elastic then the constitutive law has the form $T = T(\epsilon)$. The corresponding assumption for the free energy function is $\Psi = \Psi(\epsilon)$. Using (6.75), and an argument similar to the one used in the above example, one finds that

$$T = R_0 \Psi'(U_A). \tag{6.78}$$

Elastic materials for which the stress can be derived from the Helmholtz free energy function are called *hyperelastic*. With this we have changed the question of how the stress depends on strain to how the free energy depends on strain. In other words, if we specify the constitutive law for the Helmholtz free energy function then the above equations are used to determine the stress function. By doing this, and assuming that the free energy depends only on frame-indifferent variables, then the resulting stress will satisfy both the Principle of Material Frame-Indifference and the Principle of Dissipation.

The usual way the free energy function method is employed starts with using experimental observations to determine the functional form of the stress. With this one then shows there is a free energy function that will produce the given stress function. This is the approach used in the following examples.

Examples

1. For a linearly elastic material, $T = EU_A$. According to (6.78), the free energy function must satisfy $R_0 \Psi'(\epsilon) = E\epsilon$. Integrating this expression we obtain $R_0 \Psi(\epsilon) = \frac{1}{2} E\epsilon^2$. The constant of integration has not been included

here as it has no impact on the stress function. ■

2. If $T = a\left(e^{b\epsilon} - 1\right)$, where $b \neq 0$, then integrating $R_0\Psi'(\epsilon) = a\left(e^{b\epsilon} - 1\right)$ yields $R_0\Psi(\epsilon) = a\left(\frac{1}{b}e^{b\epsilon} - \epsilon\right)$. ■

3. For a viscous fluid it is assumed that $\tau = \tau(\rho, v_x)$. Assuming ψ depends on the same quantities, the Clausius-Duhem inequality (6.74) takes the form

$$\frac{\partial v}{\partial x}\left(\tau - \rho^2\frac{\partial \psi}{\partial \rho}\right) + \rho\frac{\partial \psi}{\partial v_x}\frac{Dv_x}{Dt} \geq 0. \tag{6.79}$$

To obtain this result, the continuity equation $\rho_t = -v\rho_x - \rho v_x$ has been used. The above inequality must hold when $v_x = 0$ and $D_t v_x = \pm 1$. The only way for this to happen is that

$$\frac{\partial \psi}{\partial v_x} = 0. \tag{6.80}$$

Consequently, even though the stress might depend on v_x, the Clausius-Duhem inequality shows that the free energy function does not. Now, for a linear viscous model it is assumed that $\tau - \rho^2\psi_\rho = \alpha + \beta v_x$, where α and β are constants. Substituting this into (6.79), and making use of (6.80), we obtain $(\alpha + \beta v_x)v_x \geq 0$. This must hold for all values of v_x, and from this we conclude that $\alpha = 0$, and $\beta \geq 0$. Setting $p = -\rho^2\psi_\rho$ then the resulting constitutive law for the stress is

$$\tau = -p + \beta\frac{\partial v}{\partial x}. \tag{6.81}$$

The function p is the pressure and the constant β is the viscosity. This means that for a viscous fluid there is an additional function to determine, and that is the pressure. This requires an additional equation, and for compressible fluids this is done by prescribing an equation of state. As an example, for an ideal gas it is assumed that $p = a\rho^\gamma$. ■

The thermodynamic foundations of continuum mechanics have been introduced only in the briefest terms, just enough to obtain the reduced entropy inequality. The fact is, this is a rich area, one that has generated more than its share of challenging mathematical and physical questions. For those who might want to learn more about this subject, the source for this material, and one that is oddly entertaining, is Truesdell [1984]. In fact, the review of this book by Aris [1987] is also recommended.

6.10.3 Hyperelasticity

As stated earlier, a hyperelastic material is one for which there is a Helmholtz free energy function ψ that is a function of the strain $\epsilon = \frac{\partial U}{\partial A}$. After working through the above example one might wonder if it is really necessary to introduce this idea. After all, the stress function can be deduced directly from the experimental data. As long as it is assumed that T depends on ϵ then the Principle of Dissipation is satisfied. The reason for this is that once $T(\epsilon)$ is known you just integrate to find ψ, and this automatically guarantees the Principle of Dissipation is satisfied. Although this observation has merit, there are several reasons why the energy method is worth considering. One, very significant, reason is that in three-dimensional problems the integration method does not work except if the stress depends on the strain in a particular way.

Another reason for introducing the energy formulation relates to the mathematical problem derived from the constitutive law. If T is a nonlinear function of strain then the momentum equation (6.49) is a nonlinear partial differential equation for the displacement. We saw in the last chapter how difficult it can be to determine whether a nonlinear equation has a solution, or whether it has just one solution. The energy formulation helps answer these questions, and this is illustrated in the next example.

Example: Bungie Cord Revisited

For the bungie cord example we solved the momentum equation to find the stress, given in (6.52). To determine the displacement of the cord the linear elastic constitutive law was used. Suppose, instead, the material is nonlinear. We will consider three different nonlinear stress-strain laws, along with their corresponding free energy functions:

$$T_1 = E\epsilon^2, \quad \psi_1 = \frac{1}{3}E\epsilon^3, \tag{6.82}$$

$$T_2 = E\epsilon^3, \quad \psi_2 = \frac{1}{4}E\epsilon^4, \tag{6.83}$$

$$T_3 = E\arctan(\epsilon), \quad \psi_3 = E\left(\epsilon\arctan(\epsilon) - \frac{1}{2}\ln(1+\epsilon^2)\right). \tag{6.84}$$

These functions are plotted in Figure 6.12. The strain interval in this figure is larger than what is possible physically, but is used to help make the points to follow. For the problem at hand the question is, given the stress, can we uniquely determine the displacement? For each stress function we have the following observations:

T_1: Given a value for the stress, other than zero, there are two possible values of the strain if $T_1 > 0$, and no strain values when $T_1 < 0$. In other words, except for zero, there is no solution or else the solution is not unique.

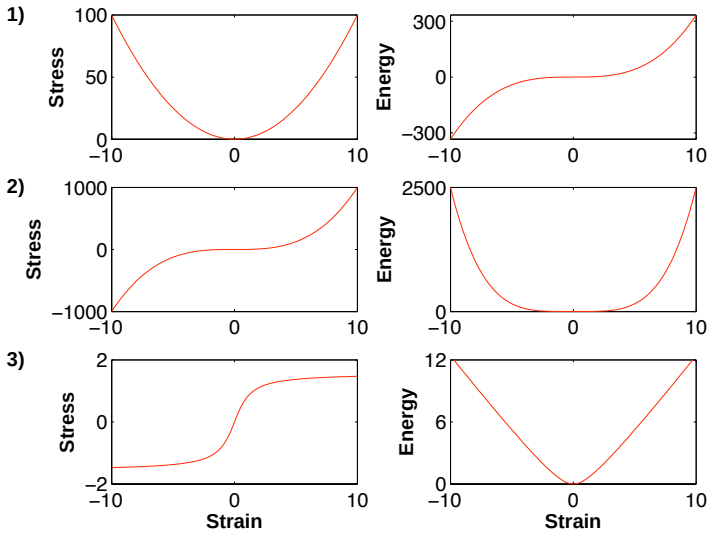


Figure 6.12 Nonlinear stress-strain functions and their corresponding Helmholtz free energy functions.

T_2: Given any value for the stress value there is a unique strain. In other words, there is a solution and it is unique. Note that the free energy for this stress function is concave up, or equivalently, convex.

T_3: For each stress value there is a unique strain. However, there are stress values, such as $T_3 = 2$, for which there is no corresponding strain. In other words, if there is a solution it is unique, but there are stress values for which there is no solution. Note that the free energy for this stress function is convex. ■

In general, to prevent multiple strain values as happened with T_1, but not with T_2 and T_3, the stress-strain law must be strictly monotonic increasing. This translates into the requirement that ψ is a strictly convex function of the strain, and this occurs if

$$\frac{d^2\psi}{d\epsilon^2} > 0. \tag{6.85}$$

However, this assumption is not enough. As shown with T_3, to guarantee that a solution exists, the stress values must have the right limiting behavior. This is the extreme value issue that was discussed earlier. Given that the strain interval is $-1 < \epsilon < \infty$, the specific requirement for this one-dimensional problem is that

$$\lim_{\epsilon \to -1} T = -\infty \quad \text{and} \quad \lim_{\epsilon \to \infty} T = \infty. \tag{6.86}$$

This means that it is assumed that it takes infinite energy to expand a finite bar to one of infinite length, and it also takes infinite energy to compress a bar down to one with zero length. None of the above three energy functions satisfies the $\epsilon \to -1$ condition, but examples of those that do can be found in Exercises 6.15 and 6.25.

The requirements for multidimensional problems are harder to determine. For example, when there is more than one spatial dimension, it has been shown that a free energy function that is convex will not be frame-indifferent. This was the motivation for introducing a milder form of convexity, something called polyconvexity. This is beyond the scope of this textbook, and the interested reader should consult Marsden and Hughes [1994] for further details.

Examples

1. For a linearly elastic material, $R_0\Psi(\epsilon) = \frac{1}{2}E\epsilon^2$. From the convexity condition, it is required that $E \geq 0$. ■

2. If $T = a\left(e^{b\epsilon} - 1\right)$, then $R_0\Psi(\epsilon) = a\left(\frac{1}{b}e^{b\epsilon} - \epsilon\right)$. From the convexity condition, it is required that $ab \geq 0$. ■

Exercises

6.1. Assume the motion is described by $X(A,t) = Ae^t$.
(a) Consider the cross-section that at $t = 5$ is located at $x = 1$. Where was it at $t = 0$?
(b) Find $u(x,t)$ and $U(A,t)$.
(c) Find $v(x,t)$ and $V(A,t)$.
(d) What is the velocity of the cross-section that is at $x = 2$ at time t? What is the velocity at time t of the cross-section that starts at $x = 2$?
(e) Suppose the temperature of the bar is $\theta(x,t) = x^5 + 4t$. What is the rate of change of θ following a material section?

6.2. Suppose that at $t = 0$ the bar occupies the interval $0 \leq A \leq 1$ and the motion of the bar is governed by the equation $X(A,t) = A + At^2$.
(a) What spatial interval does the bar occupy at $t = 2$?
(b) Find $V(A,t)$. What are the limits on A?
(c) Find $v(x,t)$. What are the limits on x?
(d) Suppose the temperature of the bar is $\theta(x,t) = xt^3$. What is the rate of change of θ following a material section?

6.3. This problem considers how the displacement can be determined from the velocity when using spatial coordinates. Therefore, in this problem, it is assumed that $v(x,t)$ is known.

(a) The direct approach to finding u uses (6.14). Show that this leads to a first-order partial differential equation for u. What is the initial condition for u?
(b) Another approach involves first converting to material coordinates. Explain why this results in having to solve

$$\frac{\partial U}{\partial t} = v(U + A, t),$$

where $U(A, 0) = 0$. Once $U(A, t)$ is known, explain how to determine $u(x, t)$.
(c) Using the approach in either part (a) or part (b), find u if $v = x/(\alpha + t)$, where α is a positive constant.

6.4. This problem explores the transformation between the material and spatial coordinate systems.
(a) Explain why $x = X(a(x, t), t)$ and $A = a(X(A, t), t)$.
(b) Show that $\frac{\partial X}{\partial A}\frac{\partial a}{\partial x} = 1$.

6.5. Prove the following.
(a) $v = \dfrac{u_t}{1 - u_x}$.
(b) $\dfrac{\partial u}{\partial x} = \dfrac{U_A}{1 + U_A}$.
(c) $\dfrac{\partial v}{\partial x} = \dfrac{\partial}{\partial t} \ln\left(1 + \dfrac{\partial U}{\partial A}\right)$.
(d) $\dfrac{\partial V}{\partial A} = \dfrac{v_x}{1 - v_x}$.
(e) $\dfrac{\partial^2 U}{\partial A^2} = \dfrac{u_{xx}}{(1 - u_x)^3}$.
(f) $\dfrac{D}{Dt} u_x = (1 - u_x) v_x$.

6.6. Assuming that $f(x, t)$ and $g(x, t)$ are smooth functions, show the following.
(a) $\dfrac{D(f + g)}{Dt} = \dfrac{Df}{Dt} + \dfrac{Dg}{Dt}$.
(b) $\dfrac{D(fg)}{Dt} = f\dfrac{Dg}{Dt} + g\dfrac{Df}{Dt}$.
(c) Explain why it is not necessarily true that $\frac{D}{Dt}\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\frac{Df}{Dt}$. What about the equation $\frac{D}{Dt}\frac{\partial f}{\partial t} = \frac{\partial}{\partial t}\frac{Df}{Dt}$?
(d) If $h = h(x)$, then show that $\frac{D}{Dt} h(f) = h'(f)\frac{Df}{Dt}$.

6.7. The deformation gradient, in material coordinates, is defined as $F(A, t) = \frac{\partial X}{\partial A}$. This function is used extensively in continuum mechanics when studying nonlinear elastic materials.
(a) Show that $F(A, t) = 1 + \frac{\partial U}{\partial A}$, and $F(A, 0) = 1$.

(b) Letting $f(x,t)$ denote the deformation gradient in spatial coordinates show that $\frac{Df}{Dt} = \frac{\partial v}{\partial x} f$.
(c) The function $C(A,t) = F^2$ is known as the Cauchy-Green deformation tensor in material coordinates. Letting $c(x,t)$ denote this function in spatial coordinates show that $\frac{Dc}{Dt} = 2\frac{\partial v}{\partial x} f^2$.

6.8. This problem considers some of the restrictions on the displacement function.
(a) Why is it not possible to have $X(A,t) = \frac{1}{2}A\cos(t)$?
(b) Why is it not possible to have, at any given value of t, $u(0,t) = -1$ and $u(1,t) = 1$?
(c) Why is it not possible to have $U(0,t) = 1$ and $U(1/2,t) = 0$?
(d) Prove that if $A_1 < A_2$ then $U(A_1,t) < A_2 - A_1 + U(A_2,t)$.
(e) Explain why it is not possible to have a displacement function of the form $U = \alpha\sin(A)$, where $\alpha > 1$.

6.9. Show that (6.19) in spatial coordinates is $\frac{\partial u}{\partial x} < 1$.

6.10. Suppose in the bungie cord example the initial density is not constant, and $R(A,0) = \alpha(1 + A/\ell_0)$. What is the steady-state length ℓ of the bungie cord?

6.11. In the steady-state bungie cord problem, if $U(A,t) = -\frac{g}{2E}A(2\ell_0 - A)e^{A/\ell_0}$ then find the density and stress in both material and spatial coordinates.

6.12. In three dimensions it is more common to use the Green strain, and this problem explores some of the differences between it and the Lagrangian strain.
(a) Rewrite the Lagrangian and Green ratios listed in Table 6.3 in terms of $\lambda = \ell/\ell_0$ and then on the same axes, sketch each ratio for $0 < \lambda < \infty$.
(b) Derive the formula for the Green strain ϵ_g.
(c) In the case of when U_A is close to zero, explain why the Green strain reduces to the Lagrangian strain.
(d) Under what circumstances would it be more appropriate to assume $T = E\epsilon_g$ rather than $T = E\epsilon$?
(e) What is the resulting equation of motion for the displacement U if one assumes $T = E\epsilon_g$?

6.13. This exercise examines the Hencky and midpoint strains listed in Table 6.3.
(a) Rewrite the Hencky and midpoint ratios in terms of $\lambda = \ell/\ell_0$ and then on the same axes sketch each ratio for $0 < \lambda < \infty$.
(b) Derive the formulas for the Hencky and midpoint strains from their corresponding ratios.

(c) Expand the Hencky and midpoint strain formulas in a Taylor series in the case of when U_A is small. It has been stated that the midpoint strain can be used as an approximation of the Hencky strain. Comment on this based on your two expansions as well as your results from part (a).
(d) What is the resulting equation of motion for the displacement U if one assumes $T = E\epsilon_h$? What is it if one assumes $T = E\epsilon_m$?

6.14. In modeling rubber as a chained polymer using what is known as a fixed junction model it is determined that two useful strain measures are $\lambda^2 - 1/\lambda$ and $\lambda - 1/\lambda^2$, where $\lambda = \ell/\ell_0$ in the extension ratio.
(a) On the same axes, sketch each strain measure for $0 < \lambda < \infty$.
(b) Derive the strain for each strain measure.
(c) Does either strain in part (b) reduce to U_A if U_A is small?

6.15. The Mooney-Rivlin model for rubber assumes $T = (\alpha + \frac{\beta}{\lambda})(\lambda^2 - \frac{1}{\lambda})$, where $\lambda = 1 + \epsilon$, and α, β are positive constants.
(a) Sketch the stress for $-1 < \epsilon < \infty$.
(b) By assuming ϵ is close to zero, determine how α, β are related to Young's modulus.
(c) Find the Helmholtz free energy function Ψ?

6.16. This problem considers if various constitutive laws satisfy the Principle of Material Frame-Indifference.
(a) Show that $\frac{\partial v}{\partial t}$ is not frame-indifferent. Explain why this shows that $\tau = \tau(v_t)$ does not satisfy the Principle of Material Frame-Indifference.
(b) Does $T = T(V_t)$ satisfy the Principle of Material Frame-Indifference?
(c) Show that $\tau = \tau(u_x, u_{xt})$ satisfies the Principle of Material Frame-Indifference.
(d) Does $T = T(U_A, U_{At})$ satisfy the Principle of Material Frame-Indifference?

6.17. Transform the initial conditions $U|_{t=0} = G(A)$ and $V|_{t=0} = H(A)$ into initial conditions for u and u_t.

6.18. A linearly elastic bar is made of two different materials and before being stretched it occupies the interval $0 \le A \le \ell_0$. Also, before being stretched, for $0 \le A < A_0$ the modulus and density are $E = E_L$ and $R = R_L$, while for $A_0 < A \le \ell_0$ they are $E = E_R$ and $R = R_R$. Both R_L and R_R are constants.
(a) The requirements at the interface, where $A = A_0$, are that the displacement and stress are continuous. Express these requirements mathematically using one-sided limits.
(b) Suppose the bar is stretched and the boundary conditions are $U(0,t) = 0$ and $U(\ell_0, t) = \ell - \ell_0$. Assume there are no body forces. Find the steady-state solution for the density, displacement and stress.

6.19. Suppose it is assumed that $\Psi = \Psi(\epsilon_g)$, where $\epsilon_g = U_A + \frac{1}{2}U_A^2$ is the Green strain.
(a) Given that $U_A > -1$, sketch ϵ_g as a function of U_A.

(b) Show that $T = R_0(1 + U_A)\Psi'(\epsilon_g)$.
(c) Suppose it is known that $T = E\epsilon_g$. Use this to show that the free energy function is

$$\Psi(\epsilon_g) = \frac{E}{3R_0}(\epsilon_g - 1)\sqrt{1 + 2\epsilon_g}.$$

6.20. Suppose it is assumed that $\psi = \psi(\epsilon_a)$, where $\epsilon_a = u_x - \frac{1}{2}u_x^2$ is the Almansi strain.
(a) Given that $u_x < 1$, sketch ϵ_a as a function of u_x.
(b) Show that $\frac{D}{Dt}\psi = (1 - u_x)^2 v_x \psi'(\epsilon_a)$.
(c) Show that $\tau = \rho(1 - u_x)^2\psi'(\epsilon_a)$.
(d) Suppose it is known that $\tau = E\epsilon_a$. What is the free energy function?

6.21. In the derivation of the equations of motion it was assumed that the cross-sectional area is constant. This problem examines what happens if this assumption is dropped and $\sigma = \sigma(x)$.
(a) Derive the resulting continuity equation in spatial coordinates, and then show that the material coordinates version is

$$R(A, t) = \frac{R_0}{1 + U_A}e^{-\kappa U},$$

where $\kappa = \sigma'(X)/\sigma(X)$.
(b) Derive the momentum equation in spatial coordinates, and then show that the material coordinates version is

$$R_0\frac{\partial^2 U}{\partial t^2} = R_0 F + e^{\kappa U}\frac{\partial T}{\partial A} + \kappa(1 + U_A)e^{\kappa U}T.$$

(c) Show that

$$\frac{\partial \sigma}{\partial A} = (1 + U_A)\sigma'.$$

(d) Assuming $F = 0$, show that the steady-state solution of the momentum equation is

$$T = \frac{f_0}{\sigma(X)},$$

where f_0 is a constant.
(e) The equations in part (a) and (b) are often used when designing loudspeakers, and an assumption often made is that the loudspeaker is an exponential horn. This means that $\sigma = \sigma_0 e^{\mu x}$, where σ_0 and μ are positive constants. How does this assumption simplify the momentum equation?

6.22. Instead of using a material volume to derive the equations of motion, it is possible to use a fixed spatial region. This is the control volume approach used to derive the traffic flow equation in the previous chapter.
(a) Given a spatial location x_0, consider the interval $x_0 - \Delta x \le x \le x_0 + \Delta x$. Explain where each term in the following equation comes from:

$$\int_{x_0-\Delta x}^{x_0+\Delta x} \sigma\rho(x,t)\bar{a}(x,t)dx$$
$$= \tau(x_0+\Delta x,t)\sigma - \tau(x_0+\Delta x,t)\sigma + \int_{x_0-\Delta x}^{x_0+\Delta x} \sigma\rho(x,t)f(x,t)dx,$$

where $\bar{a}(x,t)$ is the acceleration.

(b) Assuming small Δx, show that the equation in part (a) reduces to

$$2\sigma\Delta x\rho(x_0,t)\bar{a}(x_0,t) = 2\sigma\Delta x\tau_x(x_0,t) + 2\sigma\Delta x\rho(x_0,t)f(x_0,t) + O(\Delta x^2).$$

(c) Using the result from part (a), derive the momentum equation.

6.23. This problem derives general forms of the balance law, using the same notations used in (6.21), (6.30), and (6.70). With this in mind, let $f(x,t)$ be a quantity that is measured per unit volume.

(a) Explain where each term in the following balance equation comes from:

$$\frac{d}{dt}\int_{\alpha(t)}^{\beta(t)} \sigma f dx = -\sigma J\Big|_{x=\alpha}^{x=\beta} + \int_{\alpha(t)}^{\beta(t)} \sigma Q dx.$$

(b) Identify the functions f, J, and Q for the equations (6.21), (6.30), and (6.70).

(c) Show that the balance law in part (a) reduces to

$$\frac{\partial f}{\partial t} + \frac{\partial(vf)}{\partial x} = -\frac{\partial J}{\partial x} + Q.$$

6.24. The mechanical energy equation is

$$\rho\frac{D}{Dt}\left(\frac{1}{2}v^2\right) + \tau\frac{\partial v}{\partial x} = \frac{\partial}{\partial x}(v\tau) + \rho v f.$$

(a) Derive this directly from the momentum equation.

(b) Combine the result from part (a) with (6.71) to show that

$$\rho\frac{D}{Dt}\left(\frac{1}{2}v^2 + \chi\right) = \frac{\partial}{\partial x}(v\tau) + \rho v f.$$

The above equation represents the time rate of change of the total energy balanced with the energy flux associated with the stress and the rate of work of the body forces.

6.25. The Holmes energy function is

$$\Psi = \frac{\alpha}{\lambda^{2\beta}}e^{\beta\lambda^2}.$$

where $\lambda = 1+\epsilon$, and α, β are positive constants.

(a) Show that

$$T = \frac{1}{2} E \frac{\lambda^2 - 1}{\lambda^{2\beta+1}} e^{\beta(\lambda^2-1)},$$

where E is a positive constant.

(b) Show that if ϵ is small then the formula in part (a) reduces to the linearly elastic constitutive law given in (6.50).

(c) Show that T is a strictly monotonic increasing function of λ. Explain why this means that T is a strictly monotonic increasing function of ϵ.

(d) Show that T satisfies the limit conditions in (6.86).

6.26. This problem explores some additional ideas related to the morphological basis for deformation of a metal.

(a) The binding energy is $V_0 = V(r_0)$, where V is given in (6.57), and it is the minimum energy needed to break the atomic bonds. Show that

$$V = \frac{V_0}{m-n} \left[-(n-1) \left(\frac{r_0}{r} \right)^{m-1} + (m-1) \left(\frac{r_0}{r} \right)^{n-1} \right].$$

(b) Show that

$$E \approx -\frac{(n-1)(m-1)V_0}{4r_0^3}.$$

One conclusion that comes from this result is that materials with a high binding energy, and a small interatomic spacing, have a relatively large Young's modulus.

6.27. This problem examines the elastic modulus when the interatomic forces are described using the Morse potential function, which is

$$V = \beta \left(e^{-2\alpha(r-r_0)} - 2e^{-\alpha(r-r_0)} \right),$$

where α and β are positive constants.

(a) Show that $V'(r_0) = 0$.

(b) What is the resulting force function F? Identify the term accounting for the repulsive component of the force, and the term responsible for the attractive component.

(c) Sketch V and F. Because of the pronounced differences in the repulsive and attractive components of (6.56), it was stated that knowing the stress-strain function for a nonlinearly elastic material for tension provides little insight into what the stress function is for compression. Is this true when using the Morse potential? Are there any significant qualitative differences between (6.56) and the function you derive in part (b)?

(d) What is the resulting approximation for the elastic modulus?

(e) It has been found that for carbon nanotubes, $\beta = 3.77$ eV, $\alpha = 26.25$ nm^{-1}, and $r_0 = 0.14$ nm (Liew et al. [2005]). Use these values to estimate the Young's modulus. Note that 1 eV $= 1.6 \times 10^{-19}$ J.

Chapter 7
Elastic and Viscoelastic Materials

7.1 Linear Elasticity

A particularly successful application of continuum mechanics is linear elasticity. For a linearly elastic material, the constitutive law for the stress is

$$T = E\,\frac{\partial U}{\partial A}\,, \tag{7.1}$$

where E is Young's modulus. The momentum equation (6.41) in this case reduces to

$$\frac{\partial^2 U}{\partial t^2} = c^2\frac{\partial^2 U}{\partial A^2} + F, \tag{7.2}$$

where $c^2 = E/R_0$. It is assumed that both E and R_0 are constants. Therefore, the equation of motion for a linearly elastic material is a wave equation for the displacement. One of the objectives of this chapter is to solve this equation, and then use the solution to understand how an elastic material responds.

It is important to point out that the linear elastic model we are considering comes from assuming that the stress is a linear function of the Lagrangian strain (6.48). As is evident from Figure 6.5, exactly what strains this is valid for depends on the specific material under study. Also, if one of the other strains listed in Table 6.3 is used, a linear constitutive law for the stress does not lead to a linear momentum equation as happens in (7.2). This observation will be reconsidered later when discussing what is known as the assumption of geometric linearity.

There is a long list of methods that can be used to solve problems in linear elasticity, and this includes separation of variables, Green's functions, Fourier transforms, Laplace transforms, and the method of characteristics. The latter two will be used in this chapter, and the reasons for this will be explained as the methods are developed. Before doing this we consider a more basic issue, and this has to do with the form of the mathematical solution and its connection to the physical problem.

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_7,
© Springer Science+Business Media, LLC 2009

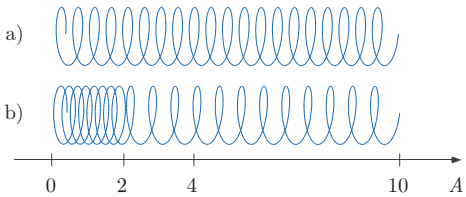


Figure 7.1 (a) A slightly extended slinky is held at $A = 0$ and at $A = 10$. (b) The loop that was at $A = 4$ is moved over to $A = 2$, producing a compression in the region $0 \leq A < 2$, and an expansion in $2 < A \leq 10$.

Example: Rubber Band at Rest

Suppose a rubber band is stretched a small amount with one end held at $A = 0$ and the other end held at $A = 10$. One then moves the cross section at $A = 4$ to $A = 2$. This situation is illustrated in Figure 7.1 for a slinky, which is not exactly a rubber band but behaves in a similar manner. For the spring, the distance between the loops is a measure of the strain. As an example, in Figure 7.1(b), the loops in $0 \leq A < 2$ and in $4 < A \leq 10$ are both uniformly placed, indicating a uniform strain in these two regions. The fact that the loops in $0 \leq A < 2$ are closer together than they are in the upper figure indicates a constant compressive strain. For a similar reason there is a constant tensile strain in $4 < A \leq 10$. Returning to the rubber band, we will assume that at rest it can be modeled as a linearly elastic material. To satisfy the given boundary conditions, it is required that the displacement satisfy $U = 0$ at $A = 0, 10$. Also, given that the cross-section that was at $A = 4$ is moved over to $A = 2$, then it is required that $U = -2$ at $A = 4$. From (7.2), at steady-state we know $U_{AA} = 0$, and this means U is a linear function of A. More precisely, it is linear for $0 < A < 4$, and it is another linear function for $4 < A < 10$. For $0 < A < 4$, the linear function that satisfies $U(0) = 0$ and $U(4) = -2$ is $U = -A/2$. For $4 < A < 10$, the linear function that satisfies $U(10) = 0$ and $U(4) = -2$ is $U = (A-10)/3$. We therefore have the piecewise linear solution

$$U = \begin{cases} -A/2 & \text{if } 0 \leq A \leq 4, \\ (A-10)/3 & \text{if } 4 \leq A \leq 10. \end{cases} \tag{7.3}$$

The conventional method for plotting such a function is given in Figure 7.2(b). It shows, for example, that the point that started at $A = 4$ moves in the negative direction to $A = 2$. Although there is nothing wrong with this plot, it obfuscates what is happening in the rubber band and seems to have no connection with what is illustrated in Figure 7.1. Another method for plotting the solution is given in Figure 7.2(a). The upper bar shows cross-sections equally spaced along the rubber band, before the rubber band is pulled. In the lower bar in Figure 7.2(a) the positions of the same cross-sections are shown after the rubber band has been pulled. The position of any given cross-section is $X = A + U$, where U is given in (7.3). What is seen is that

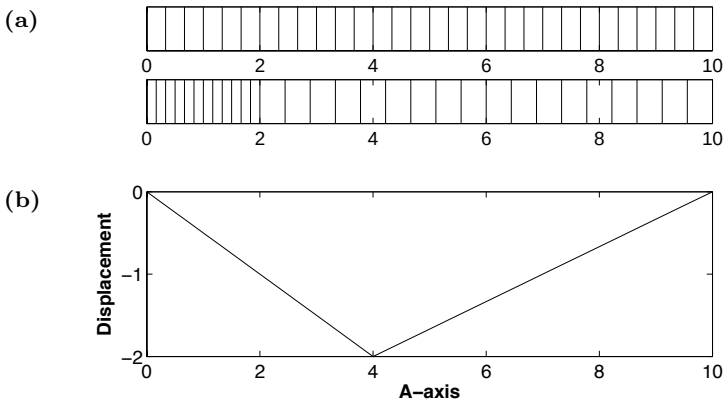


Figure 7.2 The rubber band at rest example. In (a) the upper bar shows evenly spaced cross-sections in the rubber band before it is pulled, and the lower bar shows where they are located after it is pulled. In (b) the displacement (7.3) is plotted in a more traditional method.

the cross-sections that started out uniformly spaced in $0 \leq A \leq 4$ end up uniformly spaced in the interval $0 \leq A \leq 2$. The difference is that they are closer together due to the fact that the rubber band is being compressed in this region. In contrast, the cross-sections that started out in $4 \leq A \leq 10$ get farther apart after pulling, and this is due to the stretching of the rubber band in this region. ■

The solution in the rubber band example illustrates some general characteristics that arise in elasticity. Whenever the strain is negative, so $U_A < 0$, the cross section is said to be in compression. This means that the cross-sections in this vicinity are closer together than they were before the load was applied. In contrast, whenever the strain is positive the cross-section is in tension. In Figure 7.2(a), the cross-sections that start out in $4 < A < 10$ end up in $2 < A < 10$, and are therefore in tension because $U_A = 1/3$. Similarly, those that start out in $0 < A < 4$ end up in $0 < A < 2$, and they are in compression because $U_A = -1/2$.

7.1.1 Method of Characteristics

Suppose the bar is very long, so it is reasonable to assume $-\infty < A < \infty$. Also, it is assumed that there are no body forces. The two initial conditions that will be used are

$$U(A,0) = f(A), \;\; U_t(A,0) = g(A). \tag{7.4}$$

With the given assumptions, the wave equation (7.2) can be written as

$$\left(\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial A^2}\right)U = 0.$$

Factoring the derivatives, the equation takes the form

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial A}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial A}\right)U = 0. \tag{7.5}$$

Our goal is to change coordinates, from (A, t) to (r, s), so the above equation can be written as

$$\frac{\partial}{\partial r}\left(\frac{\partial U}{\partial s}\right) = 0. \tag{7.6}$$

What we want, therefore, is the following

$$\frac{\partial}{\partial r} = \frac{\partial}{\partial t} - c\frac{\partial}{\partial A}, \tag{7.7}$$

$$\frac{\partial}{\partial s} = \frac{\partial}{\partial t} + c\frac{\partial}{\partial A}. \tag{7.8}$$

To determine how this can be done assume $A = A(r, s), t = t(r, s)$. In this case, using the chain rule

$$\frac{\partial}{\partial r} = \frac{\partial A}{\partial r}\frac{\partial}{\partial A} + \frac{\partial t}{\partial r}\frac{\partial}{\partial t}, \tag{7.9}$$

$$\frac{\partial}{\partial s} = \frac{\partial A}{\partial s}\frac{\partial}{\partial A} + \frac{\partial t}{\partial s}\frac{\partial}{\partial t}. \tag{7.10}$$

Comparing (7.7) and (7.9), we require $\frac{\partial A}{\partial r} = -c$ and $\frac{\partial t}{\partial r} = 1$. Similarly, comparing (7.8) and (7.10), we require $\frac{\partial A}{\partial s} = c$ and $\frac{\partial t}{\partial s} = 1$. Solving these equations gives us that $A = c(-r+s)$ and $t = r+s$. Inverting this transformation one finds,

$$r = -\frac{1}{2c}(A - ct), \quad s = \frac{1}{2c}(A + ct). \tag{7.11}$$

This change of variables reduces the wave equation to (7.6). The general solution of this is $U = F(r) + G(s)$ where F and G are arbitrary functions. Reverting back to A, t, and absorbing the $\frac{1}{2c}$ into the arbitrary functions, we obtain the solution

$$U(A, t) = F(A - ct) + G(A + ct), \tag{7.12}$$

where F, G are determined from the initial conditions. With this we have that the general solution of the problem consists of the sum of two traveling waves. One, with profile F, moves to the right with speed c, and the other, with profile G, moves to the left with speed c.

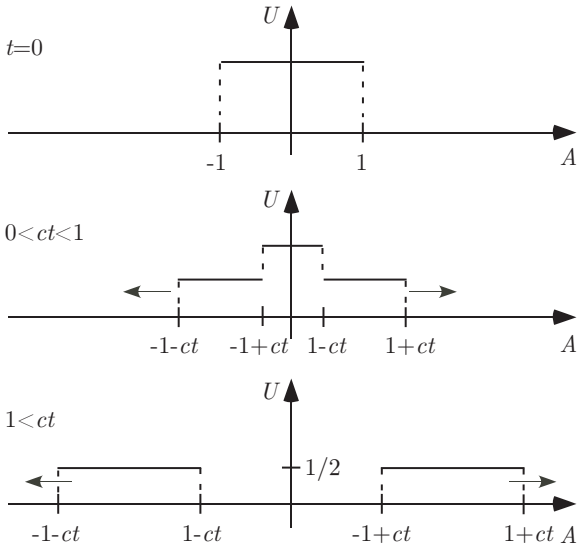


Figure 7.3 Solution of the wave equation obtained using the d'Alembert solution (7.15).

It remains to have (7.12) satisfy the initial conditions (7.4). Working out the details, one finds that the solution is

$$U(A,t) = \frac{1}{2}f(A-ct) + \frac{1}{2}f(A+ct) + \frac{1}{2c}\int_{A-ct}^{A+ct} g(z)dz. \tag{7.13}$$

This is known as the d'Alembert solution of the wave equation. It is crystal clear from this expression how the initial conditions contribute to the solution. Specifically, the initial displacement $f(A)$ is responsible for two traveling waves, both moving with speed c and traveling in opposite directions. The initial velocity $g(A)$ contributes over an ever-expanding interval, the endpoints of this interval moving with speed c.

Example

As an example, suppose the initial conditions are $U_t(A,0) = 0$ and $U(A,0) = f(A)$, where $f(A)$ is the rectangular bump

$$f(A) = \begin{cases} 1 & \text{if } -1 \le A \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7.14}$$

From (7.13) the solution is

$$U(A,t) = \frac{1}{2}f(A-ct) + \frac{1}{2}f(A+ct). \tag{7.15}$$

This is shown in Figure 7.3 and it is seen that the solution consists of two rectangular bumps, half the height of the original, traveling to the left and right with speed c. ■

The nice thing about the method of characteristics is that it produces a solution showing the wave-like nature of the response. Its flaw is that the derivation assumes that the interval is infinitely long. It is possible in some cases to use it on finite intervals, by accounting for the reflections of the waves at the boundaries. The mathematical representation of such a solution is obtained in the slinky example in the next section. For finite intervals other methods can be used. One is separation of variables, which is a subject often covered in elementary partial differential equation textbooks. Another in the Laplace transform, and this is the one pursued here.

7.1.2 Laplace Transform

Earlier, in Chapter 4, we used the Fourier transform to solve the diffusion equation. This could also be used on the wave equation, but the Laplace transform is used instead. One reason is that it is an opportunity to learn something new. Another reason is that the Laplace transform is particularly useful for cracking open some of the problems that will arise later in the chapter when studying viscoelasticity.

The Laplace transform of a function $U(t)$ is defined as

$$\widehat{U}(s) = \int_0^\infty U(t)e^{-st}dt. \tag{7.16}$$

We will need to be able to determine $U(t)$ given $\widehat{U}(s)$, and for this we need the inverse transform. It can be shown that if U is continuous at t then

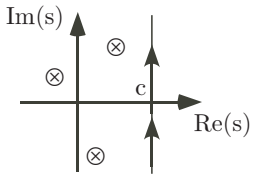


Figure 7.4 Contour used in the formula for the inverse Laplace transform (7.17). It must be to the right of any singularities of $\widehat{U}$, which are indicated using the symbol $\otimes$.

$$U(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \widehat{U}(s) e^{st} ds. \tag{7.17}$$

The integral here is a line integral in the complex plane, along the vertical line $\mathrm{Re}(s) = c$ (see Figure 7.4). It is evident from the above line integral that the variable s in (7.16) is complex valued. A second observation is that the inverse transform (7.17) is not as simple as might be expected from (7.16). Although some of the more entertaining mathematical problems arise when inverting the Laplace transform using contour integration in the complex plane, most people rely on tables. This will be the approach used here, and we will mostly determine the inverse using the relatively small collection of formulas listed in Table 7.1.

It is convenient to express the Laplace transform in operator form, and write (7.16) as $\widehat{U} = \mathcal{L}(U)$. Using this notation, the inverse transform (7.17) is $U = \mathcal{L}^{-1}(\widehat{U})$. It should be restated that the inverse formula assumes that U is continuous at t. If it is not, and U has a jump discontinuity at t, then the right-hand side of (7.17) equals the average of the jump. This means that, for $t > 0$,

$$\frac{1}{2}\left(U(t^+) + U(t^-)\right) = \mathcal{L}^{-1}(\widehat{U}),$$

and if $t = 0$ then

$$\frac{1}{2} U(0^+) = \mathcal{L}^{-1}(\widehat{U}).$$

This result, that one obtains the average of the function at a jump, is consistent with what was found for the inverse Fourier transform.

Examples

1. For the function $U(t) = e^{-t}\sin(3t)$ the Laplace transform is

$$\begin{aligned}
\widehat{U}(s) &= \int_0^\infty \sin(3t) e^{-(s+1)t} dt \\
&= \left[-\frac{s+1}{(s+1)^2+9} e^{-(s+1)t} \sin(3t) - \frac{3}{(s+1)^2+9} e^{-(s+1)t} \cos(3t) \right]_{t=0}^{\infty} \\
&= -\frac{3}{(s+1)^2+9}. \quad \blacksquare
\end{aligned}$$

2. For the piecewise constant function

$$U(t) = \begin{cases} 0 & \text{if} \quad t \le 1, \\ 2 & \text{if} \quad 1 < t \le 3, \\ -1 & \text{if} \quad 3 < t, \end{cases}$$

the Laplace transform is

	$\widehat{U}(s)$	$u(t)$
1.	$a\widehat{U}(s) + b\widehat{V}(s)$	$aU(t) + bV(t)$
2.	$\widehat{V}(s)\widehat{U}(s)$	$\int_0^t V(t-r)U(r)dr$
3.	$s\widehat{U}(s)$	$U'(t) + U(0)$
4.	$\frac{1}{s}\widehat{U}(s)$	$\int_0^t U(r)dr$
5.	$e^{-as}\widehat{U}(s)$	$U(t-a)H(t-a)$
6.	$\widehat{U}(s-a)$	$e^{at}U(t)$
7.	$\frac{1}{(s+a)^n}$	$\frac{1}{(n-1)!}t^{n-1}e^{-at}$ for $n = 1, 2, 3, \ldots$
8.	$\frac{bs+c}{(s+a)^2+\omega^2}$	$e^{-at}\left(b\cos(\omega t) + \frac{c-ab}{\omega}\sin(\omega t)\right)$ for $\omega > 0$
9.	$\frac{cs+d}{(s+a)(s+b)}$	$\frac{1}{b-a}\left((bc-d)e^{-bt} - (ac-d)e^{-at}\right)$ for $a \neq b$
10.	$\frac{1}{\sqrt{s+a}}$	$\frac{1}{\sqrt{\pi t}}e^{-at}$
11.	$\frac{1}{(s+a)\sqrt{s+b}}$	$\begin{cases} \frac{1}{\sqrt{b-a}}e^{-at}\mathrm{erf}(\sqrt{(b-a)t}) & \text{if} \quad b > a \\ \frac{2}{\sqrt{(a-b)\pi}}\int_0^{(a-b)t} e^{r^2}dr & \text{if} \quad b < a \end{cases}$
12.	$\frac{1}{\sqrt{s}(\sqrt{s}+a)}$	$e^{a^2t}\mathrm{erfc}(a\sqrt{t})$
13.	$\frac{1}{s}e^{-as}$	$H(t-a)$ for $a > 0$
14.	$e^{-a\sqrt{s}}$	$\frac{a}{2\sqrt{\pi}}t^{-3/2}e^{-a^2/(4t)}$ for $a > 0$
15.	$\frac{1}{\sqrt{s}}e^{-a\sqrt{s}}$	$\frac{1}{\sqrt{\pi t}}e^{-a^2/(4t)}$ for $a > 0$
16.	$\frac{1}{s}e^{-a\sqrt{s}}$	$\mathrm{erfc}(a/(2\sqrt{t}))$ for $a > 0$
17.	$\frac{1}{s^{\nu+1}}e^{-a^2/(4s)}$	$\left(\frac{2}{a}\right)^{\nu} t^{\nu/2}J_\nu(a\sqrt{t})$ for $\mathrm{Re}(\nu) > -1$
18.	$\frac{1}{q}e^{-cq}$ where $c \geq 0$ and $q = \sqrt{(s+a)(s+b)}$	$e^{-(a+b)t/2}I_0[(a-b)\sqrt{t^2-c^2}/2]H(t-c)$

Table 7.1 Inverse Laplace transforms. The Heaviside step function $H(x)$ is defined in (7.19), the complementary error function $\mathrm{erfc}(x)$ is given in (1.60), the error function $\mathrm{erf}(x) = 1 - \mathrm{erfc}(x)$, and J_ν, I_0 are Bessel functions.

$$\begin{aligned}\widehat{U}(s) &= \int_1^3 2e^{-st}dt - \int_3^\infty e^{-st}dt \\ &= -\frac{3}{s}e^{-3s} + \frac{2}{s}e^{-s}.\end{aligned}$$

It is interesting to see if we obtain the original function $U(t)$ by taking the inverse transform of $\widehat{U}(s)$. Using Property 13 from Table 7.1 it follows that

$$\begin{aligned}\mathcal{L}^{-1}(\widehat{U}) &= -3\mathcal{L}^{-1}(\frac{1}{s}e^{-3s}) + 2\mathcal{L}^{-1}(\frac{1}{s}e^{-3s}) \\ &= -3H(t-3) + 2H(t-1),\end{aligned} \tag{7.18}$$

where $H(x)$ is the Heaviside step function, and it is defined as

$$H(x) = \begin{cases} 0 & \text{if } x < 0, \\ \frac{1}{2} & \text{if } x = 0, \\ 1 & \text{if } 0 < x. \end{cases} \tag{7.19}$$

Writing out the definition of H in (7.18), the inverse transform is

$$\mathcal{L}^{-1}(\widehat{U}) = \begin{cases} 0 & \text{if} \quad t < 1, \\ 1 & \text{if} \quad t = 1, \\ 2 & \text{if} \quad 1 < t < 3, \\ \frac{3}{2} & \text{if} \quad t = 3, \\ -1 & \text{if} \quad 3 < t. \end{cases}$$

This result shows that $\mathcal{L}^{-1}(\widehat{U}) = U$ at values of t where U is continuous, but at the jump points the inverse equals the average of the jump in the function. ■

3. Suppose that

$$\widehat{U} = \frac{2}{s} - \frac{3}{s^2+4}\,.$$

According to Property 7, from Table 7.1, $\mathcal{L}^{-1}(\frac{1}{s}) = 1$, and from Property 8, $\mathcal{L}^{-1}((s^2+4)^{-1}) = \frac{1}{2}\sin(2t)$. Using Property 1 it therefore follows that

$$\begin{aligned}U(t) &= \mathcal{L}^{-1}\left(\frac{2}{s} - \frac{3}{s^2+4}\right) \\ &= 2\mathcal{L}^{-1}\left(\frac{1}{s}\right) - 3\mathcal{L}^{-1}\left(\frac{1}{s^2+4}\right) \\ &= 2 - \frac{3}{2}\sin(2t). \quad ■\end{aligned}$$

Given the improper integral in (7.16), it is necessary to impose certain restrictions on the function $U(t)$, although the requirements are much less severe than for the Fourier transforms studied in Chapter 4. It is assumed that $U(t)$ is piecewise continuous and has exponential order. This means that U grows no faster than a linear exponential function as $t \to \infty$. The specific requirement is that there is a constant α so that

$$\lim_{t\to\infty} Ue^{\alpha t} = 0. \tag{7.20}$$

As examples, any bounded function or any polynomial function has exponential order. On the other hand, e^{t^2} and e^{t^3} do not. With this, the Laplace transform (7.16) is defined for any s that satisfies $\mathrm{Re}(s) > \alpha$, and this gives rise to what is known as the half-plane of convergence for the Laplace transform. This comes into play when calculating the inverse transform (7.17), and the requirement is that c is in the half-plane of convergence. It is relatively easy to determine this half-plane from $\widehat{U}$. The requirement is that the half-plane of convergence is to the right of the singularities of $\widehat{U}$ (see Figure 7.4). As an example, if $\widehat{U} = 1/s$ then the half-plane of convergence must satisfy $\mathrm{Re}(s) > 0$, while if $\widehat{U} = 1/\sqrt{s(s-1)}$ then the half-plane of convergence must satisfy $\mathrm{Re}(s) > 1$.

One last comment to make before working out some of the properties of the Laplace transforms relates to the behavior of $\widehat{U}$ when $\mathrm{Re}(s) \to \infty$. Because of the negative exponential in the integral, it follows that

$$\lim_{\mathrm{Re}(s)\to\infty} \widehat{U} = 0. \tag{7.21}$$

This limit assumes that the original function U is piecewise continuous and has exponential order. The reason this result is useful is that it can be used to help check for errors in a calculation. For example, if you find that $\widehat{U} = s$, or $\widehat{U} = \sin(s)$, or $\widehat{U} = e^s$ then an error has been made. The reason is that none of these functions satisfies (7.21).

7.1.2.1 Transformation of Derivatives

One of the hallmarks of the Laplace transform, as with most integral transforms, is that it converts differentiation into multiplication. To explain what this means, we use integration by parts to obtain the following

$$\begin{aligned}\mathcal{L}(U') &= \int_0^\infty U'e^{-st}dt \\ &= Ue^{-st}\Big|_{t=0}^{\infty} + s\int_0^\infty Ue^{-st}dt \\ &= -U(0) + s\mathcal{L}(U).\end{aligned} \tag{7.22}$$

This formula can be used to find the transform of higher derivatives, and as an example

$$\begin{aligned}\mathcal{L}(U'') &= -U'(0) + s\mathcal{L}(U') \\ &= -U'(0) + s\left(-U(0) + s\mathcal{L}(U)\right) \\ &= s^2\mathcal{L}(U) - U'(0) - sU(0). \end{aligned} \tag{7.23}$$

Generalizing this to higher derivatives

$$\mathcal{L}(U^{(n)}) = s^n\mathcal{L}(U) - U^{(n-1)}(0) - sU^{(n-2)}(0) - \cdots - s^{n-1}U(0). \tag{7.24}$$

7.1.2.2 Convolution Theorem

A common integral arising in viscoelasticity is a convolution integral of the form

$$T = \int_0^t G(t-\tau)V(\tau)d\tau. \tag{7.25}$$

Taking the Laplace transform of this equation we obtain

$$\begin{aligned}\mathcal{L}(T) &= \int_0^\infty \int_0^t G(t-\tau)V(\tau)e^{-st}d\tau dt \\ &= \int_0^\infty \int_\tau^\infty G(t-\tau)V(\tau)e^{-st}dtd\tau \\ &= \int_0^\infty \int_0^\infty G(r)V(\tau)e^{-s(r+t)}drd\tau \\ &= \int_0^\infty V(\tau)e^{-s\tau}\int_0^\infty G(r)e^{-sr}drd\tau \\ &= \widehat{G}(s)\widehat{V}(s). \end{aligned}$$

Using the inverse transform this can be written as

$$\mathcal{L}^{-1}(\widehat{G}(s)\widehat{V}(s)) = \int_0^t G(t-\tau)V(\tau)d\tau. \tag{7.26}$$

This is Property 2, in Table 7.1, and it is known as the convolution theorem.

7.1.2.3 Solving the Problem for Linear Elasticity

The problem that will be solved using the Laplace transform consists of the wave equation

$$\frac{\partial^2 U}{\partial t^2} = c^2\frac{\partial^2 U}{\partial A^2} + F(A,t), \tag{7.27}$$

where the boundary conditions are

$$U(0,t) = p(t), \quad U(\ell,t) = q(t), \tag{7.28}$$

and the initial conditions are

$$U(A,0) = f(A), \quad U_t(A,0) = g(A). \tag{7.29}$$

It is understood that the only unknown is $U(A,t)$, and all the other functions in the above equations are given. The first step is to take the Laplace transform of both sides of the wave equation to obtain

$$\mathcal{L}(U_{tt}) = c^2\mathcal{L}(U_{AA}) + \mathcal{L}(F). \tag{7.30}$$

Using (7.23), and the given initial conditions,

$$\mathcal{L}(U_{tt}) = s^2\widehat{U} - g(A) - sf(A).$$

Also, because the transform is in the time variable, $\mathcal{L}(U_{AA}) = \widehat{U}_{AA}$. Introducing these observations into (7.30) we have that

$$c^2\widehat{U}_{AA} - s^2\widehat{U} = -\widehat{F}(A,s) - g(A) - sf(A). \tag{7.31}$$

The solution of this equation must satisfy the transform of the boundary conditions (7.28), and this means that

$$\widehat{U}(0,s) = \widehat{p}(s), \quad \widehat{U}(\ell,s) = \widehat{q}(s). \tag{7.32}$$

where $\widehat{p} = \mathcal{L}(U_L)$ and $\widehat{q} = \mathcal{L}(q)$.

Solving (7.31) for $\widehat{U}$ depends on what functions are used for the forcing, boundary, and initial conditions, and we consider two examples. Before doing this, note that by taking the Laplace transform that the initial conditions have become forcing functions in the differential equation (7.31). This limits the usefulness of this method. The reason is that even simple looking initial conditions can result in solutions of (7.31) that are complicated functions of the transform variable s. By complicated it is meant that the inverse transform is not evident, and even manipulating the contour integral in the definition of the inverse transform does not help. This observation should not be interpreted to mean that the method is a waste of time. Rather, it should be understood that the Laplace transform is an important tool for analyzing differential and integral equations, but like all other methods, it has limitations.

Example 1: Slinky

Suppose there is no forcing, so $F(A,t) = 0$, and $f(A) = g(A) = 0$. Also, the boundary conditions are $p = U_0$ and $q = 0$. Physically, this corresponds to

taking an elastic bar at rest and pushing on the left end a fixed amount U_0. A similar situation is shown in Figure 7.5 for a slinky. What happens is that the disturbance propagates along the slinky, reaches the right end, reflects, and then moves leftward. The result is a disturbance that moves back and forth along the spring. This is mentioned as it is worth having some expectation on what the mathematics will produce. Proceeding on to solving the problem, with the stated assumptions (7.31) takes the form

$$c^2\widehat{U}_{AA} - s^2\widehat{U} = 0, \tag{7.33}$$

and the boundary conditions (7.28) are

$$\widehat{U}(0,s) = \frac{U_0}{s}, \quad \widehat{U}(\ell,s) = 0. \tag{7.34}$$

The general solution of (7.33) is

$$\widehat{U} = \alpha e^{sA/c} + \beta e^{-sA/c},$$

where α and β are arbitrary constants. This function must satisfy the boundary conditions (7.34), and from this it follows that

$$\widehat{U}(A,s) = \frac{U_0}{s}\frac{\sinh(s(\ell - A)/c)}{\sinh(s\ell/c)}. \tag{7.35}$$

Now comes the big question, can we find the inverse transform of (7.35)? Some of the more extensive tables listing inverse Laplace transforms do include this particular function, but most do not. Given the propensity of second order differential equations to generate solutions involving the ratio of exponential functions, as in (7.35), it is worth deriving the inverse from scratch. The first step is to use the definition of the sinh function to write

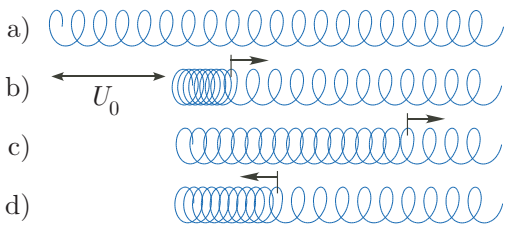


Figure 7.5 (a) A slightly extended slinky is held at $A = 0$ and at $A = \ell$. (b) The left end is then moved a distance U_0, producing a compressed region in the spring. (c) This region spreads down the spring towards the right end. (d) When the compression reaches $A = \ell$, it reflects and then starts moving in the opposite direction. In an elastic spring this back-and-forth motion will continue indefinitely.

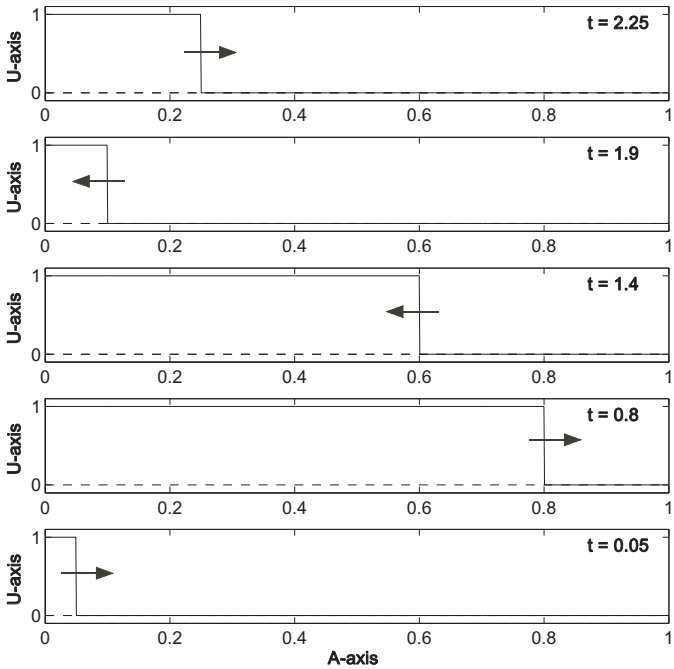


Figure 7.6 Solution of the elastic bar given in (7.36). The solution consists of a traveling wave that starts at $A = 0$ and then propagates back and forth along the A-axis.

$$\begin{aligned}\frac{\sinh(\alpha s)}{\sinh(\beta s)} &= \frac{e^{\alpha s} - e^{-\alpha s}}{e^{\beta s} - e^{-\beta s}} \\ &= e^{-\beta s}\frac{e^{\alpha s} - e^{-\alpha s}}{1 - e^{-2\beta s}}\,.\end{aligned}$$

It is assumed here that $0 < \beta$. Using the geometric series on the denominator we obtain

$$\begin{aligned}\frac{\sinh(\alpha s)}{\sinh(\beta s)} &= e^{-\beta s}\left(e^{\alpha s} - e^{-\alpha s}\right)\left(1 + e^{-2\beta s} + e^{-4\beta s} + \cdots\right) \\ &= e^{-\beta s}\left(e^{\alpha s} - e^{-\alpha s}\right) + e^{-3\beta s}\left(e^{\alpha s} - e^{-\alpha s}\right) + e^{-5\beta s}\left(e^{\alpha s} - e^{-\alpha s}\right) + \cdots \\ &= \sum_{n=1}^{\infty} e^{-(2n-1)\beta s}\left(e^{\alpha s} - e^{-\alpha s}\right).\end{aligned}$$

Now, using Property 13 from Table 7.1,

$$\mathcal{L}^{-1}\left(\frac{1}{s}e^{-bs}(e^{\alpha s}-e^{-\alpha s})\right) = \mathcal{L}^{-1}\left(\frac{1}{s}e^{(\alpha-b)s}\right) - \mathcal{L}^{-1}\left(\frac{1}{s}e^{-(\alpha+b)s}\right)$$
$$= H\left[t+(\alpha-b)\right] - H\left[t-(\alpha+b)\right].$$

With this,

$$\mathcal{L}^{-1}\left(\frac{1}{s}\frac{\sinh(\alpha s)}{\sinh(\beta s)}\right) = \sum_{n=1}^{\infty}\left[H\left(t+\alpha-(2n-1)\beta\right) - H\left(t-\alpha-(2n-1)\beta\right)\right].$$

The inverse of (7.35) is, therefore,

$$U(A,t) = U_0\sum_{n=1}^{\infty}\left[H\left(t+\kappa_{-n+1}\right) - H\left(t-\kappa_n\right)\right], \tag{7.36}$$

where

$$\kappa_n = \frac{1}{c}(-A+2n\ell).$$

The solution is shown in Figure 7.6, for $\ell = c = U_0 = 1$. As expected from the slinky analogy, the solution is a traveling wave that starts at $A = 0$ and then moves back and forth over the bar. The amplitude is $U_0 = 1$, and the speed of the wave can be determined from the arguments of the Heaviside functions in (7.36). Namely, its speed is equal to c. This is not surprising as this is the speed of the traveling waves found using the method of characteristics, given in (7.13). As a final comment, there are different ways of writing the solution to this problem, and some are derived in Exercise 7.3. ■

Example 2: Resonance

In this example we investigate what happens to the bar when it is forced periodically. Specifically, it is assumed that the forcing function in (7.27) is $F(A,t) = a(t)\cos(\kappa_n A)$, where $a(t) = \sin(\omega t)$, $\kappa_n = n\pi/\ell$, and n is a positive integer. The bar is assumed to be stress free at the ends, and so the boundary conditions are

$$\frac{\partial U}{\partial A} = 0 \text{ at } A = 0, \ell. \tag{7.37}$$

Taking the Laplace transform, the boundary conditions become

$$\frac{\partial \widehat{U}}{\partial A} = 0 \text{ at } A = 0, \ell. \tag{7.38}$$

The initial conditions are $f(A) = g(A) = 0$. In this case (7.31) takes the form

$$c^2\widehat{U}_{AA} - s^2\widehat{U} = -\widehat{a}(s)\cos(\kappa_n A),$$

where $\widehat{a} = \mathcal{L}(a)$. The general solution of this equation is

$$\widehat{U} = \frac{\widehat{a}(s)}{\kappa_n^2 c^2 + s^2} \cos(\kappa_n A) + \alpha e^{sA/c} + \beta e^{-sA/c},$$

where α and β are arbitrary constants. This solution must satisfy the boundary conditions (7.38), and from this is follows that

$$\widehat{U} = \frac{\widehat{a}(s)}{\kappa_n^2 c^2 + s^2} \cos(\kappa_n A). \tag{7.39}$$

Again, the big question, can we find the inverse transform of (7.39)? Using Property 8, with $a = b = 0$ and $\omega = \kappa_n c$,

$$\mathcal{L}^{-1}\left(\frac{1}{\kappa_n^2 c^2 + s^2}\right) = \frac{1}{\kappa_n c} \sin(\kappa_n ct).$$

Therefore, using the convolution property (7.26), it follows that

$$\begin{aligned} U(A,t) &= \mathcal{L}^{-1}\left(\frac{\widehat{a}(s)}{\kappa_n^2 c^2 + s^2} \cos(\kappa_n A)\right) \\ &= \cos(\kappa_n A)\, \mathcal{L}^{-1}\left(\widehat{a}(s)\, \frac{1}{\kappa_n^2 c^2 + s^2}\right) \\ &= \frac{1}{\kappa_n c} b(t) \cos(\kappa_n A), \end{aligned}$$

where

$$b(t) = \int_0^t a(t-r) \sin(\kappa_n cr) dr.$$

Given that $a(t) = \sin(\omega t)$ it follows that

$$b(t) = \begin{cases} \dfrac{1}{\omega^2 - \kappa_n^2 c^2} \left(\omega \sin(\kappa_n ct) - \kappa_n c \sin(\omega t)\right) & \text{if } \omega \neq \kappa_n c, \\ -\dfrac{1}{2} t \cos(\kappa_n ct) + \dfrac{1}{2\kappa_n c} \sin(\kappa_n ct) & \text{if } \omega = \kappa_n c. \end{cases} \tag{7.40}$$

This shows that when $\omega \neq \kappa_n c$, the displacement is a combination of periodic functions. In contrast, when $\omega = \kappa_n c$ the solution grows, becoming unbounded as $t \to \infty$. This is a phenomenon known as resonance, and it is a characteristic of linearly elastic systems. The resonant frequencies are easily measured experimentally, and this provides a means to test the accuracy of the model. In the experiments of Bayon et al. [1993], an aluminum bar was tested and the first three measured resonant frequencies f_1, f_2, and f_3 are given in Table 7.2. Recall that circular and angular frequencies are related through the equation $f = 2\pi\omega$. In this case, $\omega = \kappa_n c$ reduces to

$$f_n = \frac{n}{2\ell} \sqrt{\frac{E}{R}}\,. \tag{7.41}$$

n	f_n Experimental	f_n Computed	Relative Error
1	15,322 Hz	15,541 Hz	1.4%
2	30,644 Hz	31,082 Hz	1.4%
3	45,966 Hz	46,623 Hz	1.4%

Table 7.2 Natural frequencies of an aluminum bar measured experimentally (Bayon et al. [1993]), and computed using (7.41).

To compare with the model, the bar in the experiment was 0.1647 m long. Also, using the conventional values for pure aluminum, $E = 70758$ MPa and $R = 2700$ kg/m^3. The resulting values of the angular frequencies are also shown in Table 7.2. The rather small difference between the experimental and computed values is compelling evidence that the linear elastic model is appropriate here. ■

Given the need to be able to determine the material parameters in a model, the question comes up whether the measured values for f_n can be used to determine E and R. The best we can do with (7.41) is to determine the ratio E/R. How it might be possible to use the resonant frequencies to find the material and geometrical parameters is one of the core ideas in inverse problems. As an example, a classic paper in this area is, "Can you hear the shape of a drum," by Kac [1966]. Considerable work has been invested in solving inverse problems, and some of the more recent discoveries are discussed in Gladwell [2004].

In a physical problem, the growth in the amplitude that occurs at the resonant frequency means that eventually the linear elasticity approximation no longer applies, and other effects come into play. These will generally mollify the amplitude, although not always. An example is the Tacoma Narrows Bridge. Although classic resonance was not the culprit, the same principle of unstable linear oscillations that feed large nonlinear motion was in play, and this eventually caused the bridge to collapse.

7.1.3 Geometric Linearity

The assumption in (7.1) that the stress is a linear function of the Lagrangian strain results in a linear momentum equation (7.2). This does not happen if any of the other strains listed in Table 6.3 are used. For example, when working in three dimensions it is conventional to use the Green strain ϵ_g. From (6.49), the assumption that $T = E\epsilon_g$ results in the momentum equation

$$\frac{\partial^2 U}{\partial t^2} = c^2\left(1 + \frac{\partial U}{\partial A}\right)\frac{\partial^2 U}{\partial A^2} + F.$$

In contrast to (7.2), this is a nonlinear wave equation for the displacement.

It is possible to obtain a linear momentum equation using the other strains, but it is necessary to impose certain restrictions on the motion. What is needed is geometric linearity, and to contrast this with our earlier assumption we have the following:

- *Material Linearity.* The assumption is that the stress-strain function is linear. Examples are $T = E\epsilon$, $T = E\epsilon_g$, and $\tau = E\epsilon_e$. Linearity in this context is relative to a particular strain measure.

- *Geometric Linearity.* It is assumed that there are only small deformations, and so ϵ is assumed close to zero. This is often referred to as an assumption of infinitesimal deformations.

The idea underlying geometric linearity is that the displacement is small in comparison to the overall length of the bar. This means that the displacement gradients $\frac{\partial U}{\partial A}$ and $\frac{\partial u}{\partial x}$ are close to zero. One consequence of this assumption is that all of the strains listed in Table 6.3 are effectively equal. For example, $\epsilon_g = U_A + \frac{1}{2}U_A^2 \approx U_A = \epsilon$. Similarly, $\epsilon = U_A = u_x/(1 - u_x) \approx u_x = \epsilon_e$. As is apparent in these calculations, the assumption of geometric linearity knocks out the nonlinear terms in the equations. Therefore, one ends up with a linear momentum equation.

7.2 Viscoelasticity

The slinky example in the previous section is interesting but unrealistic from a physical point of view. The reason is that the traveling waves shown in Figure 7.6 continue indefinitely. In contrast, in a real system the motion eventually comes to rest. The reason is that energy is lost due to dissipation. This is similar to what occurs when dropping an object and letting it fall through the air. The faster the object moves the greater the air resistance on the object. The usual assumption in this case is that there is a resistance force that is proportional to the object's velocity. This same idea is used when formulating the equations for a damped oscillator, and in the next section we use this observation to develop the theory of viscoelasticity.

7.2.1 Mass, Spring, Dashpot Systems

It is informative to review the equation for a damped oscillator, as shown in Figure 7.7. From Newton's second law, the displacement $u(t)$ of the mass in the mass, spring, dashpot system satisfies

$$mu'' = F_s + F_d, \tag{7.42}$$

where m is the mass, F_s is the restoring force in the spring, and F_d is the damping force. Assuming the spring is linear, then from Hooke's law

$$F_s = -ku, \tag{7.43}$$

where k is a positive constant. The mechanism commonly used to produce damping involves a dashpot, where the resisting force is proportional to velocity. The associated constitutive assumption is

$$F_d = -cu', \tag{7.44}$$

where c is a positive constant. With this, the total force $F = F_s + F_d$.

This example contains several ideas that will be expanded on below. First, it shows that the force includes an elastic component, which depends on displacement, and a damping component that depends on the velocity. As we saw earlier, when using a spring, mass system to help formulate a constitutive law for the stress, displacement is replaced with strain. Therefore, instead of assuming the force depends on displacement and velocity, in the continuum formulation the stress is assumed to depend on the strain $\epsilon = U_A$ and strain rate $\epsilon_t = U_{At}$. The question is, as always, exactly what function should we select and how do we make this choice. To help answer this question, we will examine spring and dashpot systems.

It is possible to generalize the above example and introduce the basic laws of viscoelasticity. This is done by putting the spring and dashpot in various configurations, and three of the more well-studied are shown in Figure 7.8. We start with the series orientation shown in Figure 7.8(a). The point m moves due to a force $-F(t)$, and its displacement $u(t)$ equals the sum of the displacement $u_s(t)$ of the spring and the displacement $u_d(t)$ of the dashpot. Converting to velocities we have that $u' = u'_s + u'_d$. Now, according

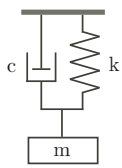


Figure 7.7 Mass, spring, dashpot system.

to Newton's third law, the force in the spring and dashpot equals F. From (7.43) we get $u_s = F/k$ and from (7.44) we have that $u_d' = F/c$. With this we obtain the following force, deflection relationship

$$u' = F'/k + F/c. \tag{7.45}$$

This is known as the Maxwell element in viscoelasticity.

In the next configuration, shown in Figure 7.8(b), the spring and dashpot are in parallel. For this we use the fact that forces add, and so $F_s + F_d = F$. Also, the displacement of the spring and dashpot are the same, and both are equal to $u(t)$. With this we obtain

$$F = ku + cu'. \tag{7.46}$$

This is known as the Kelvin-Voigt element in viscoelasticity.

The third configuration, shown in Figure 7.8(c), gives rise to what is known as the standard linear element. The force in the upper spring is $F_0 = k_0 u$, while for the lower spring, dashpot the force satisfies $u' = F_1'/k_1 + F_1/c_1$. The forces must balance, and this means that $F = F_0 + F_1$. In this case, $F_1 = F - F_0$, and so

$$\begin{aligned} u' &= (F - F_0)'/k_1 + (F - F_0)/c_1 \\ &= (F - k_0 u)'/k_1 + (F - k_0 u)/c_1. \end{aligned}$$

Rearranging things, it follows that

$$F + a_1 F' = a_2 u + a_3 u', \tag{7.47}$$

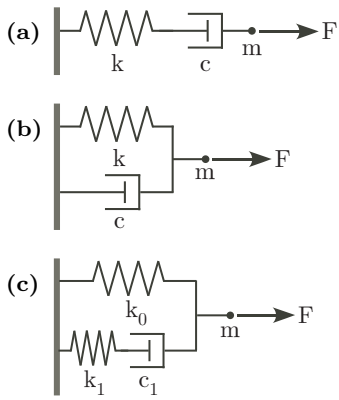


Figure 7.8 Spring, dashpot systems used to derive viscoelastic models: (a) Maxwell element; (b) Kelvin-Voigt element; and (c) standard linear element.

where $a_1 = c_1/k_1$, $a_2 = k_0$ and $a_3 = c_1(1 + k_0/k_1)$. The coefficients in this equation satisfy an inequality that is needed later. Because $c_1 = k_1 a_1$, then $a_3 = a_1(k_1 + a_2)$. With this we have that $a_3 > a_1 a_2$.

Each of the spring, dashpot examples can be generalized to a viscoelastic constitutive law that can be used in continuum mechanics. This is done by simply replacing u with the strain ϵ, u' with the strain rate ϵ_t, and F with the stress T. After rearranging the constants in the formulas, the resulting viscoelastic constitutive laws are

$$\text{Maxwell model:} \quad T + \tau_0 \frac{\partial T}{\partial t} = E\tau_1 \frac{\partial \epsilon}{\partial t}\,, \tag{7.48}$$

$$\text{Kelvin-Voigt model:} \quad T = E\left(\epsilon + \tau_1 \frac{\partial \epsilon}{\partial t}\right), \tag{7.49}$$

$$\text{standard linear model:} \quad T + \tau_0 \frac{\partial T}{\partial t} = E\left(\epsilon + \tau_1 \frac{\partial \epsilon}{\partial t}\right). \tag{7.50}$$

The strain in the above formulas, as usual, is

$$\epsilon = \frac{\partial U}{\partial A}\,.$$

In analogy with the linear elastic law (7.1), the constant E is the Young's modulus and it is assumed to be positive. The constants τ_0 and τ_1 have the dimensions of time, and are known as the dissipation time scales for the respective model. To be consistent with the expressions in (7.45) - (7.47), E and the τ_i's are assumed to be positive. In addition it is assumed in the standard linear model that $\tau_0 < \tau_1$. This condition comes from the same inequality that exists between the constants in (7.47).

7.2.2 Equations of Motion

The somewhat unusual forms of the viscoelastic constitutive laws generate several questions related to their mathematical and physical consequences. We begin with the mathematical questions, and with this in mind, remember that the reason for introducing a constitutive law is to complete the equations of motion. There are two functions that are solved for, which are the displacement and the stress. Using the standard linear model (7.50), and assuming there is no body force, the equations to solve are

$$R_0 \frac{\partial^2 U}{\partial t^2} = \frac{\partial T}{\partial A}\,, \tag{7.51}$$

$$T + \tau_0 \frac{\partial T}{\partial t} = E\left(\frac{\partial U}{\partial A} + \tau_1 \frac{\partial^2 U}{\partial A \partial t}\right). \tag{7.52}$$

To complete the problem, initial and boundary conditions must be specified and an example is presented below. Also, if one of the other viscoelastic models is used then (7.52) would change accordingly.

Example: Periodic Displacement

A common testing procedure involves applying a periodic displacement to one end of the material, while keeping the other end fixed. Assuming the bar occupies the interval $0 \leq A \leq \ell$, then the associated boundary conditions are

$$U(0,t) = a\sin(\omega t), \text{ and } U(\ell,t) = 0. \tag{7.53}$$

We are going to solve the system of equations (7.51), (7.52). In doing so, it is assumed that the elastic modulus E and density R_0 are known using one or more of the steady-state tests described in Section 6.7. Our goal here is to use the periodic displacement to determine the damping parameters τ_0 and τ_1. This will be accomplished by finding the periodic solution to the problem, which is the solution that appears long after the effects of the initial conditions have died out. To find this solution assume that

$$U(A,t) = \overline{U}(A)e^{i\omega t}, \tag{7.54}$$

and

$$T(A,t) = \overline{T}(A)e^{i\omega t}. \tag{7.55}$$

Using complex variables simplifies the calculations to follow, but it is necessary to rewrite the boundary condition at $A = 0$ in (7.53) to fit this formulation. This will be done by generalizing it to

$$U(0,t) = ae^{i\omega t}. \tag{7.56}$$

It is understood that we are interested in the imaginary component of whatever expression we obtain. Now, substituting (7.54), (7.55) into (7.52) we have that

$$\overline{T} = E\,\frac{1+i\omega\tau_1}{1+i\omega\tau_0}\,\frac{d\overline{U}}{dA}\,. \tag{7.57}$$

The momentum equation (7.51) in this case reduces to

$$\frac{d^2\overline{U}}{dA^2} = -\kappa^2\,\overline{U},$$

where

$$\kappa^2 = \frac{R_0\omega^2}{E}\,\frac{1+i\omega\tau_0}{1+i\omega\tau_1}\,.$$

The general solution of this is $\overline{U} = \alpha\exp(i\kappa A) + \beta\exp(-i\kappa A)$. Imposing the two boundary conditions gives us the following solution,

$$\overline{U}(A) = a\frac{e^{i\kappa A} - e^{-i\kappa A + 2i\kappa\ell}}{1 - e^{2i\kappa\ell}}. \tag{7.58}$$

To simplify the analysis we will assume the bar is very long and let $\ell \to \infty$. With this in mind, note

$$\kappa^2 = \frac{R_0\omega^2}{E}\frac{1 + \omega^2\tau_0\tau_1 + i\omega(\tau_0 - \tau_1)}{1 + \omega^2\tau_1^2}.$$

Given that $0 \le \tau_0 < \tau_1$, then $\mathrm{Re}(\kappa^2) > 0$ and $\mathrm{Im}(\kappa^2) < 0$. From this we have that $\mathrm{Im}(\kappa) < 0$, and so for large values of ℓ, (7.58) reduces to

$$\overline{U}(A) = ae^{-i\kappa A}.$$

With this, the displacement is

$$U(A,t) = ae^{i(\omega t - \kappa A)}. \tag{7.59}$$

One of the reasons that experimentalists use this test is to compare the stress measured at $A = 0$ with what is predicted from the model. With the solution in (7.59), and the formulas for the stress in (7.55) and (7.57), the stress at $A = 0$ is

$$T(0,t) = -i\kappa aE\,\frac{1 + i\omega\tau_1}{1 + i\omega\tau_0}e^{i\omega t}. \tag{7.60}$$

To determine the imaginary component of this expression set

$$\begin{aligned} r_0e^{i\delta} &= \kappa E\,\frac{1 + i\omega\tau_1}{1 + i\omega\tau_0} \\ &= \omega\sqrt{R_0E}\sqrt{\frac{1 + i\omega\tau_1}{1 + i\omega\tau_0}} \\ &= \omega\sqrt{R_0E}\sqrt{\frac{1 + \omega^2\tau_0\tau_1 + i\omega(\tau_1 - \tau_0)}{1 + \omega^2\tau_0^2}}. \end{aligned}$$

Taking the modulus of this we have that

$$r_0 = \omega\sqrt{R_0E}\left(\frac{1 + \omega^2\tau_1^2}{1 + \omega^2\tau_0^2}\right)^{1/4}, \tag{7.61}$$

and, taking the ratio of the imaginary and real components,

$$\tan(2\delta) = \frac{\omega(\tau_1 - \tau_0)}{1 + \omega^2\tau_0\tau_1}. \tag{7.62}$$

With this, (7.60) reduces to

$$T(0,t) = ar_0\sin(\omega t + \delta - \pi/2). \tag{7.63}$$

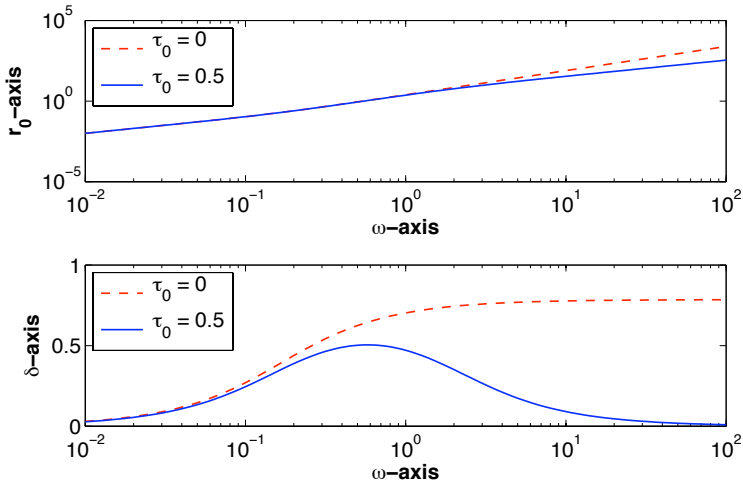


Figure 7.9 The amplitude r_0 and phase δ in response to a periodic forcing. Shown are the curves for a Kelvin-Voigt model, $\tau_0 = 0$, and for a standard linear model, where $\tau_0 = 1/2$.

We are now in position to determine some of the affects of viscoelasticity. First, because $1 + \omega^2\tau_1^2 > 1 + \omega^2\tau_0^2$, the amplitude ar_0 of the observed stress is increased due to the viscoelasticity. This conclusion is consistent with the understanding that damping increases the resistance to motion. However, as shown in Figure 7.9, the r_0 curves for the two viscoelastic models are rather similar, although they show some differences for very large values of ω. What this means is that the r_0 curve is not particularly useful in identifying which viscoelastic model to use. This is not the case with the phase δ. As shown in (7.63), for a viscoelastic material the phase difference between the stress and displacement is $\delta - \pi/2$. The characteristics of δ differ markedly between the two models. For the Kelvin-Voigt model, so $\tau_0 = 0$, the formula in (7.62) reduces to

$$\delta = \frac{1}{2}\arctan(\omega\tau_1). \tag{7.64}$$

In this case, δ is a monotonically increasing function of ω, and the larger the driving frequency the closer δ gets to $\pi/4$ (see Figure 7.9). In comparison, for the standard linear model with $0 < \tau_0 < \tau_1$, δ reaches a maximum value when $\omega = 1/\sqrt{\tau_0\tau_1}$, and approaches zero as $\omega \to \infty$. This difference provides a simple test to determine which of the two models should be used. It is also useful for determining the damping parameters from experiment. If one is able to measure the frequency ω_M, and phase δ_M, for the maximum phase, then from (7.62) one finds thats $\tau_1 = [\tan(2\delta_M) + \sec(2\delta_M)]/\omega_M$ and $\tau_0 = 1/(\tau_1\omega_M^2)$. The derivation of this result is the subject of Exercise 7.25.

To demonstrate that the frequency dependence shown in Figure 7.9 does indeed occur in applications, data for porcine cartilage is shown in Figure 7.10. The dependence appears to follow the standard linear model. Also, note that cartilage is strongly viscoelastic. The reason is that if an elastic model is assumed then $\delta = 0$, and this certainly does not happen in Figure 7.10. ■

The previous example demonstrates how a mathematical model can be used in conjunction with experimental measurements to help test that the model is applicable, and to also determine some of the parameters. The focus of the inquiry was on the resulting stress at the end of the bar. It is also interesting to study the response within the bar. For example, with (7.59), the displacement has the form

$$U(A,t) = ae^{-\kappa_i A}\sin(\omega t - \kappa_r A), \tag{7.65}$$

where $\kappa = \kappa_r - i\kappa_i$. This is a traveling wave which has an amplitude that decays with A. A similar conclusion holds for the stress. Exactly how the viscoelasticity affects the properties of the wave is important in many applications, such as in geophysics when studying earthquakes, and this is explored in Exercise 7.14.

7.2.3 Integral Formulation

One of the attractive features of the Kelvin-Voigt model is that it provides an explicit formula for the stress. This can be substituted directly into the momentum equation (7.51), to produce a single equation for U, which avoids a system formulation as in (7.51), (7.52). The other two viscoelastic models are implicit, and require a solution of a differential equation to determine T. There are reasons, particularly when solving the problem numerically, why one would want to keep the problem in system form. However, there are also reasons why it is worth expressing the problem as a single equation.

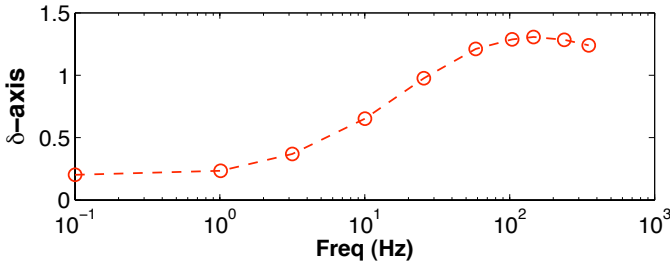


Figure 7.10 Measured values for δ for porcine cartilage (Morita et al. [2002]).

To solve the Maxwell model (7.48), note that it is a linear first-order equation for T. Solving this equation one finds that

$$T(A,t) = T_0(A)e^{-t/\tau_0} + \int_0^t Ee^{(\tau-t)/\tau_0}\frac{\partial \epsilon}{\partial \tau}d\tau,$$

where $T_0(A) = T(A,0)$. We will assume $T(A,0) = 0$, so the above solution reduces to

$$T = \int_0^t G(t-\tau)\frac{\partial \epsilon}{\partial \tau}d\tau, \tag{7.66}$$

where

$$G(t) = Ee^{-t/\tau_0}. \tag{7.67}$$

Substituting this into (7.51) we obtain

$$R_0\frac{\partial^2 U}{\partial t^2} = \int_0^t G(t-\tau)\frac{\partial^3 U}{\partial A^2 \partial \tau}d\tau, \tag{7.68}$$

which is an integro-differential equation for the displacement. The standard method for solving this equation is to use the Laplace transform. For the moment we will continue to concentrate on the formulation of the viscoelastic constitutive law and save the question of how to solve the problems until later.

The standard linear model is also a linear first-order equation for T, that can be solved using an integrating factor. One finds that

$$T = \int_0^t G(t-\tau)\frac{\partial \epsilon}{\partial \tau}d\tau, \tag{7.69}$$

where

$$G(t) = E\left(1 + \kappa e^{-t/\tau_0}\right), \tag{7.70}$$

and $\kappa = (\tau_1 - \tau_0)/\tau_0$ is a nonnegative constant. It has been assumed in deriving (7.69) that $\epsilon = 0$ and $T = 0$ at $t = 0$. With this we have obtained the same integral equation given in (7.66), except the function G is given in (7.70).

We now have two versions of the viscoelastic models. Those in (7.48)-(7.50) are differential equations and are examples of what are called rate-type laws. Expressing them in integral form we found that

$$T(A,t) = \int_0^t G(t-\tau)\frac{\partial \epsilon}{\partial \tau}(A,\tau)d\tau, \tag{7.71}$$

and this is known as a viscoelastic law of relaxation type. The function G is called the relaxation function. It is possible to rewrite the integral using integration by parts. The result is

$$T(A,t) = E\epsilon(A,t) + \int_0^t K(t-\tau)\epsilon(A,\tau)d\tau, \tag{7.72}$$

where $K(t) = G'(t)$ and $E = G(0)$. Written this way, the stress is expressed as the sum of an elastic component and an integral associated with the damping in the system. Either version, (7.71) or (7.72), shows that the stress depends on the values of the strain, or strain rate, over the entire time interval. For this reason, the Maxwell and standard linear models apply to materials with memory. It might seem unreasonable to expect that the stress at the current time depends on what was happening a long time ago. However, the decaying exponential in (7.70) reduces the contribution from earlier times, and the smaller the dissipation time scale τ_0 the less they contribute. In contrast, with the Kelvin-Voigt model the stress depends solely on the values of the strain and strain rate at the current time.

7.2.4 Generalized Relaxation Functions

The integral form of the stress law (7.71) is widely used in the engineering literature, and this is partly due to the information that is obtained from experiments. In many of the conventional tests used to determine material properties, the strain is imposed and the stress is measured. This information is then used to determine the parameters in the relaxation function. For this to work one must make a judicious choice for the functional form for G. The usual argument made in such situations is that real materials do not operate as a simple spring, dashpot system as in Figure 7.8, but involve many such elements. The consequence of this observation is that one does not end up with one exponential, as in (7.67) and (7.70), but a relaxation function of the form

$$G(t) = E\left(1 + \sum \kappa_i e^{-t/\tau_i}\right). \tag{7.73}$$

An example spring, dashpot system that produce a multi-exponential relaxation function is shown in Figure 7.14, and the specifics are worked out in Exercise 7.11. Although (7.73) is considered an improvement over the earlier simpler models, it still has flaws. Again, the argument is that because of the complexity of real materials, a finite number of elements is inadequate and one should use a continuous distribution. What happens in this case is that the sum in (7.73) is replaced with an integral. The resulting constitutive law for the relaxation function is

$$G(t) = E\left(1 + \int_0^\infty g(\tau)e^{-t/\tau}d\tau\right), \tag{7.74}$$

where $g(\tau)$ is a nonnegative function. This transfers the question of how to pick G to what to take for g, which is not much of an improvement in

terms of difficulty. The answer depends on the application. One approach is to attempt to formulate a general law that is still simple enough to allow analysis of the problem. For example, a commonly made choice used to model the viscoelastic properties of biological materials is

$$G(t) = E\left(1 + \kappa \int_{\tau_1}^{\tau_2} \frac{1}{\tau}\, e^{-t/\tau} d\tau\right), \tag{7.75}$$

where κ, τ_1, τ_2 are positive constants. This is known as the Neubert-Fung relaxation function. Exactly how to determine the three constants from experiment is discussed in Fung [1993].

We are in a quagmire that is common in viscoelasticity, which is having multiple constitutive laws to pick from but not knowing exactly which one to use. The answer, again, depends on the application. To illustrate, let's reconsider the slinky example of the previous section. As noted earlier, assuming the bar is linearly elastic means that the motion observed in Figure 7.6 never slows down, much less stops, and this was the motivation for introducing a viscoelastic model in the first place. We will assume that the damping, or dashpot, mechanism only acts when the bar is moving, and when at rest the bar can be modeled as a linearly elastic material. In terms of the differential forms in (7.48)-(7.50), this means that when $\epsilon_t = 0$ and $T_t = 0$ the formula reduces to $T = E\epsilon$. This eliminates the Maxwell model from consideration. This observation is why (7.49) and (7.50) are referred to as viscoelastic solids, while (7.48) is called a viscoelastic fluid. This still leaves open the question of whether to use a Kelvin-Voigt or a standard linear model. The answer bridges the mathematical and experimental worlds. It is not uncommon for an applied mathematician to ask an experimentalist to run a specific test that corresponds to a problem that the mathematician is able to solve. It is also not uncommon for the experimentalist to reply that the testing equipment does not have the particular capability that is requested. What is necessary in such cases is for the two to work out an experimental procedure that can provide useful information for building and testing the model. One that has found wide use in viscoelasticity involves periodic loading, and this was considered in an earlier example. There are certainly other methods, and a review of the possibilities can be found in Lakes [2004].

7.2.5 Solving Viscoelastic Problems

One of the standard tools for reducing viscoelastic models, both rate and integral type, is the Laplace transform. There are a couple of reasons for this. One is that it converts differentiation into multiplication. The second reason is it can handle the convolution integrals that arise with the integral type viscoelastic laws.

Example 1: Deriving the Relaxation Function

For the standard linear model the stress T is related to the strain ϵ through the equation

$$T + \tau_0 \frac{\partial T}{\partial t} = E\left(\epsilon + \tau_1 \frac{\partial \epsilon}{\partial t}\right).$$

To solve this for T take the Laplace transform of both sides to obtain

$$\mathcal{L}(T) + \tau_0 \mathcal{L}(T_t) = E\left(\mathcal{L}(\epsilon) + \tau_1 \mathcal{L}(\epsilon_t)\right).$$

It is assumed that $T = 0$ and $\epsilon = 0$ at $t = 0$. With this, and using Property 1 from Table 7.1, and (7.22), we obtain

$$\widehat{T} + \tau_0 s \widehat{T} = E\left(\widehat{\epsilon} + \tau_1 s\, \widehat{\epsilon}\right).$$

Solving for $\widehat{T}$ yields

$$\widehat{T} = E\, \frac{1 + \tau_1 s}{1 + \tau_0 s}\, \widehat{\epsilon}.$$

We are going to take the inverse transform to find T. At first glance it might appear that the convolution theorem, Property 2 from Table 7.1, can be used to find the inverse of the right hand side of the equation. However, the function multiplying $\widehat{\epsilon}$ does not satisfy (7.20), and therefore there is no inverse transform for this function. It is possible to modify the equation to get this to work, and the trick is to write the equation as

$$\widehat{T} = E\, \frac{1 + \tau_1 s}{s(1 + \tau_0 s)}\, s\, \widehat{\epsilon}. \tag{7.76}$$

From (7.22), $\mathcal{L}^{-1}(s\,\widehat{\epsilon}) = \epsilon_t$, and from Property 9 from Table 7.1

$$\mathcal{L}^{-1}\left(\frac{1 + \tau_1 s}{s(1 + \tau_0 s)}\right) = 1 + \left(1 - \frac{\tau_1}{\tau_0}\right) e^{-t/\tau_0}.$$

Applying the convolution theorem to (7.76), and then using integration by parts, the stress is

$$T = \int_0^t \left(1 + \left(\frac{\tau_1}{\tau_0} - 1\right) e^{-(t-r)/\tau_0}\right) \frac{\partial \epsilon}{\partial r} dr.$$

This result agrees with the solution given in (7.69) that was obtained using an integrating factor. For this particular problem the integrating factor method is easier to use, but its limitation is that it only works on first-order equations. The Laplace transform, however, also works on higher-order problems, and this is important for studying more complex viscoelastic models, such as those investigated in Exercises 7.11 and 7.12. ■

Example 2: Solving an Integro-Differential Equation

Suppose the bar is modeled as a Maxwell viscoelastic material, and the integral form of the stress law (7.66) is used. As shown in (7.68), the momentum equation in this case is

$$R_0 \frac{\partial^2 U}{\partial t^2} = \int_0^t G(t-\tau) \frac{\partial^3 U}{\partial A^2 \partial \tau} d\tau, \tag{7.77}$$

where $G(t) = Ee^{-t/\tau_0}$. It is assumed the bar occupies the interval $0 \leq A < \infty$, and the associated boundary conditions are

$$\frac{\partial U}{\partial A}(0,t) = F(t), \tag{7.78}$$

and

$$\lim_{A \to \infty} U(A,t) = 0. \tag{7.79}$$

The initial conditions are $U(A,0) = U_t(A,0) = 0$. Taking the Laplace transform of (7.77) we obtain

$$R_0 \mathcal{L}(U_{tt}) = \mathcal{L}\left(\int_0^t G(t-\tau) \frac{\partial^3 U}{\partial A^2 \partial \tau} d\tau \right). \tag{7.80}$$

Because of the initial conditions, from (7.23), we have that $\mathcal{L}(U_{tt}) = s^2 \mathcal{L}(U)$. Also, from the convolution theorem we know that

$$\mathcal{L}\left(\int_0^t G(t-\tau) V(\tau) d\tau \right) = \mathcal{L}(G)\mathcal{L}(V).$$

Consequently, (7.80) takes the form

$$R_0 s^2 \widehat{U} = \mathcal{L}(G)\, \mathcal{L}\left(\frac{\partial^3 U}{\partial A^2 \partial t} \right). \tag{7.81}$$

Now, basic integration gives us

$$\begin{aligned} \mathcal{L}(G) &= \int_0^\infty E e^{-t/\tau_0} e^{-st} dt \\ &= E \frac{\tau_0}{s + \tau_0}. \end{aligned}$$

Also,

$$\begin{aligned} \mathcal{L}\left(\frac{\partial^3 U}{\partial A^2 \partial t} \right) &= s\mathcal{L}\left(\frac{\partial^2 U}{\partial A^2} \right) \\ &= s \frac{\partial^2}{\partial A^2} \mathcal{L}(U). \end{aligned}$$

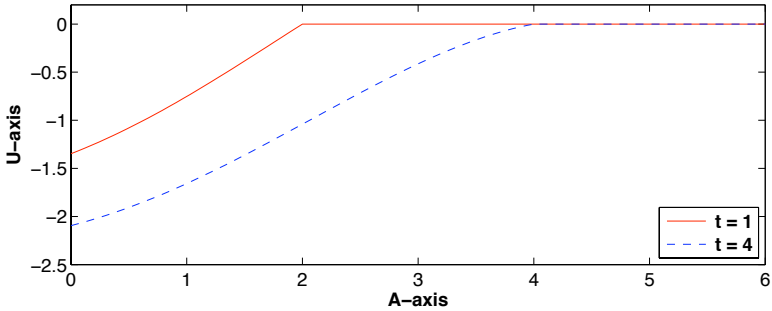


Figure 7.11 Solution of the Maxwell viscoelastic model as given in (7.86), in the case of when $F(t) = \sin(t)$.

Introducing these into (7.81) we obtain

$$R_0 s^2 \widehat{U} = E \, \frac{\tau_0 s}{s + \tau_0} \, \frac{\partial^2 \widehat{U}}{\partial A^2} \, .$$

The general solution of this second order differential equation is

$$\widehat{U} = \alpha e^{\omega A} + \beta e^{-\omega A}, \tag{7.82}$$

where

$$\omega = \sqrt{\frac{R_0 s(s + \tau_0)}{E \tau_0}} \, . \tag{7.83}$$

To find α and β we take the Laplace transform of the boundary conditions (7.78) and (7.79) to find that $\widehat{U}_A(0, s) = \widehat{F}$ and $\widehat{U}(A, s) \to 0$ as $A \to \infty$. In this case (7.82) reduces to

$$\widehat{U}(A, s) = -\frac{1}{\omega} \widehat{F}(s) e^{-\omega A}. \tag{7.84}$$

From Property 18, in Table 7.1 we have that

$$\mathcal{L}^{-1}\left(\frac{1}{\omega} e^{-\omega A}\right) = \kappa e^{-\tau_0 t/2} I_0 \left[\frac{1}{2} \tau_0 \sqrt{t^2 - \lambda^2}\right] H(t - \lambda), \tag{7.85}$$

where $\kappa = \sqrt{E\tau_0/R_0}$ and $\lambda = A/\kappa$. In the above expression, I_0 is the modified Bessel function of the first kind. With this, and the convolution theorem, it follows that

$$U(A, t) = H(t - \lambda) \int_\lambda^t Q(A, t - r) F(r) dr, \tag{7.86}$$

where

$$Q(A, t) = -\kappa e^{-\tau_0 t/2} I_0 \left[\frac{1}{2} \tau_0 \sqrt{t^2 - \lambda^2}\right] . \tag{7.87}$$

An interesting conclusion that can be made is that the effects of the boundary condition move through the material with finite velocity. According to (7.86), the solution starts to be nonzero when $t = \lambda$, and the corresponding velocity is $\sqrt{E\tau_0/R_0}$. On the other hand, the solution in (7.86) is not as satisfying a result as the previous example because the solution is in the form of a convolution integral involving a Bessel function. However, most math software programs, such as Maple and MATLAB, have the Bessel functions built in, so it is relatively easy to evaluate the integral. The result of such a calculation is shown in Figure 7.11, which gives the solution at two time points. The finite velocity of the wave is clearly seen in this figure. ■

Exercises

7.1. A linearly elastic bar is stretched by applying a constant stress T_0 to the right end. Assuming the original interval is $0 \le A \le \ell_0$ then the boundary conditions are $U(0,t) = 0$ and $T(\ell_0, t) = T_0$. Assume there are no body forces.

(a) Find the steady-state solution for the density, displacement and stress.
(b) What happens to the displacement and stress if Young's modulus is increased?

7.2. The equations for the linearly elastic bar are given in (7.27)-(7.29). This exercise shows that not just any smooth function can be used in the displacement initial condition.

(a) Based on the impenetrability of matter requirement, what condition must be imposed on $f(A)$ in (7.29)?
(b) Using the result from part (a), explain why it is not possible to take $f(A) = 3A(\ell - A)/\ell$, but it is possible to take $f(A) = A(\ell - A)/(2\ell)$.

7.3. This problem considers various ways to express the solution of the wave equation given in (7.36).

(a) Show that $-\kappa_{-n+1} \le \kappa_n$.
(b) Show that (7.36) can be written as

$$U(A,t) = \sum_{n=1}^{\infty} I_{(-\kappa_{-n+1},\kappa_n)}(t)\,.$$

(c) Show that (7.36) can be written as

$$U(A,t) = \begin{cases} U_0 & \text{if } 0 \le A < q(t), \\ U_0/2 & \text{if } A = q(t), \\ 0 & \text{if } q(t) < A \le \ell, \end{cases}$$

where $q(t)$ is a $2\ell/c$ periodic function.

7.4. Solve the following problems by extending the method that was used in Section 7.1.1 to solve the wave equation.

(a)
$$\frac{\partial^2 U}{\partial t^2} - 4\frac{\partial^2 U}{\partial A^2} = 1,$$
where $U(A,0) = f(A)$ and $U_t(A,0) = 0$.

(b)
$$\frac{\partial^2 U}{\partial t^2} + \frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial A^2} + \frac{\partial U}{\partial A},$$
where $U(A,0) = f(A)$ and $U_t(A,0) = 0$.

7.5. A steel bar is forced periodically, and it is found that the first three resonant frequencies are 7,861 Hz, 15,698 Hz, and 23,535 Hz (Bayon et al. [1994]).

(a) Explain why this result is consistent with the assumption that the bar is linearly elastic.
(b) If the bar is 0.32 mm long and has a density of 7893.16 kg/m^3, find Young's modulus for the bar.
(c) The same experimentalists found that the wave speed in the bar is 5037 m/sec. Use this to estimate the Young's modulus and compare the result from part (b).

7.6. The equations of motion when the cross-section is not constant were derived in Exercise 6.21. This problem explores how these equations can be linearized.

(a) Explain why assuming material linearity, so $T = E\epsilon$, does not result in the momentum equation being linear.
(b) Assuming geometric linearity, as well as material linearity, the momentum equation reduces to
$$\frac{\partial^2 U}{\partial t^2} = c^2\frac{\partial^2 U}{\partial A^2} + \frac{1}{\sigma}\frac{\partial \sigma}{\partial A}\frac{\partial U}{\partial A} + F.$$
This is known as Webster's equation, or Webster's horn equation, and it is linear. Explain how this result is obtained from the equations given in Exercise 6.21.

7.7. This problem explores the effect on the solution when using different materially linear theories. The two constitutive laws that are compared are: (i) $T = EU_A$, and (ii) $\tau = Eu_x$. As usual, let $\epsilon = U_A$ and $\epsilon_e = u_x$

(a) Transform constitutive law (ii) into material coordinates, that is, transform it into an expression involving T and ϵ. For labeling purposes, identify this stress as T_{ii} and label the one from (i) as T_i. On the same axes, sketch T_{ii} and T_i for $-1 < \epsilon < \infty$.
(b) Transform constitutive law (i) into spatial coordinates, that is, transform it into an expression involving τ and ϵ_a. For labeling purposes, identify

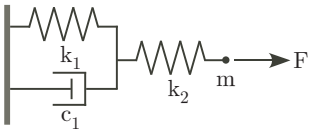


Figure 7.12 Three-parameter viscoelastic solid studied in Exercise 7.9.

this stress as τ_i and label the one from (ii) as τ_{ii}. On the same axes, sketch τ_i and τ_{ii} for $-\infty < \epsilon_e < 1$.

(c) Show that $T_{ii} < T_i$ if $\epsilon \neq 0$.

(d) Show that $\tau_{ii} > \tau_i$ if $\epsilon_a \neq 0$.

(e) Suppose the stress in the bar becomes unbounded for large tensile strains. Is T_{ii} or T_i the more appropriate constitutive law?

(f) Suppose the stress in the bar becomes unbounded for large compressive strains. Is T_{ii} or T_i the more appropriate constitutive law?

7.8. This problem explores what happens with a constitutive law when there is a jump in the solution. To do this, assume that at a given position A, the stress and strain are smooth except for a jump discontinuity when $t = t_s$.

(a) By integrating the constitutive law (7.50) over the time interval $t_s - \Delta t \leq t \leq t_s + \Delta t$, show that an expression of the following form is obtained,

$$\tau_0 \left[T(A, t_s + \Delta t) - T(A, t_s - \Delta t)\right] \\ = E\tau_1 \left[\epsilon(A, t_s + \Delta t) - \epsilon(A, t_s - \Delta t)\right] + \int_{t_s - \Delta t}^{t_s + \Delta t} q(A, t) dt.$$

(b) By letting $\Delta t \to 0$, show that

$$\tau_0 \left[T(A, t_s^+) - T(A, t_s^-)\right] = E\tau_1 \left[\epsilon(A, t_s^+) - \epsilon(A, t_s^-)\right].$$

This states how the stress and strain behave across a jump, similar to what is obtained from the Rankine-Hugoniot condition for traffic flow.

(c) A common experiment is to apply a constant stress at one end of the bar, which is assumed here to be at $A = 0$. This produces what is known as a creep response, and the associated boundary condition is $T(0, t) = T_0$ for $t > 0$. Assume that for $t < 0$ the bar is at rest with $T = 0$ and $\epsilon = 0$. Using the standard linear model show that an expression of the following form is obtained

$$E_i \frac{\partial U}{\partial A}(0, 0^+) = T_0,$$

where $E_i = E\tau_1/\tau_0$. In engineering E_i is called the instantaneous elastic modulus. Explain why it is larger than the elastic modulus.

(d) Find the instantaneous elastic modulus when using the Maxwell model.

7.9. This problem concerns the system shown in Figure 7.12, which is an example of what is known as a three- parameter viscoelastic solid.

(a) Show that the force F and the displacement u satisfy

$$F + \frac{1}{c_1(k_1+k_2)}F' = \frac{k_1k_2}{k_1+k_2}u + \frac{k_2}{c_1(k_1+k_2)}u'.$$

(b) Show that the continuum version of the result from part (a) has the form

$$T + \tau_0\frac{\partial T}{\partial t} = E\left(\epsilon + \tau_1\frac{\partial \epsilon}{\partial t}\right),$$

where $0 < \tau_0 < \tau_1$.

(c) Show that the viscoelastic constitutive law in part (b) can be expressed in integral form (7.66), where

$$G(t) = E(1 + \kappa e^{-t/\lambda}).$$

Assume that $\epsilon = 0$ and $T = 0$ at $t = 0$. Also, show that κ and λ are positive.

(d) Discuss the similarities, and differences, between the results from (a) and (b) with the formulas for the standard linear model.

7.10. This problem concerns the system shown in Figure 7.13, which is an example of what is known as a three-parameter viscoelastic fluid.

(a) Show that the force F and the displacement u satisfy

$$F + \frac{c_1+c_2}{k_1}F' = c_2u' + \frac{c_1c_2}{k_1}u''.$$

(b) Show that the continuum version of the result from part (a) has the form

$$T + \tau_0\frac{\partial T}{\partial t} = \tau_1\frac{\partial \epsilon}{\partial t} + \tau_2\frac{\partial^2 \epsilon}{\partial t^2},$$

where the τ_i's are positive with $\tau_0\tau_1 < \tau_2$.

(c) Show that the viscoelastic constitutive law in part (b) can be expressed in integral form as

$$T = \kappa_1\epsilon' + \int_0^t G(t-s)\epsilon'(s)ds,$$

where

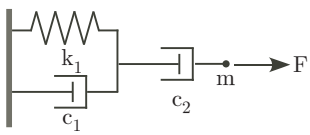


Figure 7.13 Three-parameter viscoelastic fluid studied in Exercise 7.10.

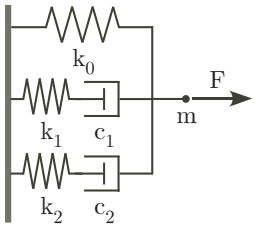


Figure 7.14 Five-parameter model for Exercise 7.11.

$$G(t) = \kappa_2 e^{-t/\lambda}.$$

Assume that $\epsilon = \epsilon' = 0$ and $T = 0$ at $t = 0$. Also, show that κ_i's and λ are positive.

7.11. This problem concerns the five-parameter model shown in Figure 7.14.

(a) Derive the differential equation that relates the force F with the displacement u.

(b) Show that the continuum version of the result from part (a) has the form

$$T + \tau_0 \frac{\partial T}{\partial t} + \tau_1 \frac{\partial^2 T}{\partial t^2} = E\left(\epsilon + \tau_2 \frac{\partial \epsilon}{\partial t} + \tau_3 \frac{\partial^2 \epsilon}{\partial t^2}\right),$$

where the τ_i's are positive, with $\tau_0 < \tau_2$ and $\tau_1 < \tau_3$.

(c) Show that the viscoelastic constitutive law in part (b) can be expressed in integral form (7.66), where

$$G(t) = E(1 + \kappa_1 e^{-t/\lambda_1} + \kappa_2 e^{-t/\lambda_2}).$$

Assume that $\epsilon = \epsilon' = 0$ and $T = T' = 0$ at $t = 0$. Also, show that the λ_i's are positive.

7.12. This problem concerns the four-parameter model shown in Figure 7.15, what is known as the Burger model.

(a) Derive the differential equation that relates the force F with the displacement u.

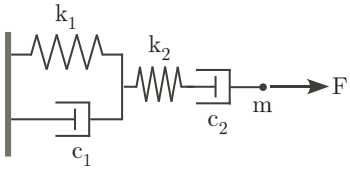


Figure 7.15 The Burger viscoelastic model used in Exercise 7.12.

(b) Show that the continuum version of the result from part (a) has the form

$$T + \tau_0 \frac{\partial T}{\partial t} + \tau_1 \frac{\partial^2 T}{\partial t^2} = \tau_2 \frac{\partial \epsilon}{\partial t} + \tau_3 \frac{\partial^2 \epsilon}{\partial t^2},$$

where the τ_i's are positive, with $4\tau_1 < \tau_0^2$.

(c) Show that the viscoelastic constitutive law in part (b) can be expressed in integral form (7.66), where

$$G(t) = \kappa_1 e^{-t/\lambda_1} + \kappa_2 e^{-t/\lambda_2}.$$

Assume that $\epsilon = \epsilon' = 0$ and $T = T' = 0$ at $t = 0$. Also, show that the λ_i's are positive.

7.13. In some applications it is easier to work with the stress rather than the displacement. This problem investigates this for the standard linear model.

(a) Derive (7.47).

(b) Starting from (7.50), show that

$$\epsilon = \int_0^t J(t-\tau) \frac{\partial T}{\partial \tau} d\tau.$$

Assume here that $\epsilon = 0$ and $T = 0$ at $t = 0$. The function J is called the creep function.

(c) Show that

$$\epsilon = J(0)T + \int_0^t J'(t-\tau) T d\tau.$$

(d) Use the result in part (b) to transform (7.52) into an equation for the stress T.

(e) By taking the Laplace transform of the creep and relaxation forms of the constitutive laws show that

$$\int_0^t G(s) J(t-s) ds = t.$$

7.14. This problem investigates the traveling waves that are obtained in the periodic displacement example using the standard linear model.

(a) Assuming ℓ is large, and letting $\kappa = \kappa_r - i\kappa_i$, show that (7.54) and (7.58) reduce to

$$U(A,t) = a e^{-\kappa_i A} \sin(\omega t - A\kappa_r).$$

Find the corresponding expression for the stress $T(A,t)$.

(b) Show that for high frequencies

$$\kappa \sim \sqrt{\frac{R_0 \tau_0}{E \tau_1}}\, \omega \left(1 - i \frac{\tau_1 - \tau_0}{2\tau_0 \tau_1 \omega} + O\left(\frac{1}{\omega^2}\right)\right).$$

(c) Show that for low frequencies

$$\kappa \sim \sqrt{\frac{R_0}{E}}\,\omega\left(1 - i\frac{1}{2}(\tau_1 - \tau_0)\omega + O(\omega^2)\right).$$

(d) Suppose the elastic modulus E and density R_0 are known. Can the phase velocity $v_p = \omega/\kappa_r$ of the wave, measured at both low and high frequencies, be used to determine the two viscoelastic constants τ_0 and τ_1? If the amplitude $ae^{-\kappa_i A}$ of the wave is also measured at both low and high frequencies, does this help in determining the two viscoelastic constants τ_0 and τ_1?

7.15. One of the consequences of damping is that it can mollify the effects of resonance. As an example, suppose that in (7.58) the viscoelasticity is turned off by letting $\tau_0 = \tau_1 = 0$. In this case there are frequencies for which (7.58) is undefined. Relate these to the resonance frequencies found in (7.40). Also, explain why (7.58) does not have this particular problem when the viscoelasticity is turned on (remember that $\tau_0 < \tau_1$).

7.16. This problem investigates the differences in the viscoelastic models when a periodic forcing is used. The boundary conditions in this case are

$$T(0,t) = b\sin(\omega t),\ U(\ell,t) = 0.$$

Assume the standard linear viscoelastic model is used.
(a) Assuming a periodic solution, find $\overline{U}(A)$ and $\overline{T}(A)$.
(b) Find $\overline{U}(0)$ assuming $\ell \to \infty$. Your answer should be in terms of one trig function, similar to what was done for the stress in (7.63).
(c) In the experiments one measures the displacement at the end and compares the data with the predictions from the model. One objective is to determine the viscoelastic parameters in the model. Does the periodic stress boundary condition provide any information not learned from the periodic displacement boundary condition?

7.17. Find the Laplace transform of the following functions. Make sure to state if there are conditions on s.
(a) $f(t) = te^{\alpha t}$
(b) $f(t) = \cosh^2 t$
(c) $Si(t) = \int_0^t \frac{\sin(r)}{r}dr$
(d) $f(t) = \frac{1}{t}\sin(t)$
(e) $f(t) = \sqrt{t}$

7.18. Using the Laplace transform, solve $y'' + 4y = f(t)$, where $y(0) = 0$, $y'(0) = -1$, and

$$f(t) = \begin{cases} \cos(2t) & \text{if } 0 \le t < \pi, \\ 0 & \text{otherwise.} \end{cases}$$

7.19. Using the Laplace transform, solve the system of equations

$$x' = 3x - 4y$$
$$y' = 2x + 3y$$

where $x(0) = 1$ and $y(0) = 0$.

7.20. This problem concerns solving the diffusion equation

$$Du_{xx} = u_t, \quad \text{for} \quad \begin{cases} 0 < x < \infty, \\ 0 < t, \end{cases}$$

where $u(x, 0) = 0$, $u \to 0$ as $x \to \infty$, and

$$u(0, t) = \begin{cases} T & \text{if} \quad 0 < t \leq b, \\ 0 & \text{if} \quad b < t. \end{cases}$$

Using the Laplace transform, find the solution of this problem.

7.21. Using the Laplace transform solve the integral equation

$$u(t) - \int_0^t e^{t-r} u(r) dr = f(t).$$

Assume this holds for $0 \leq t$ and that $f(t)$ is continuous.

7.22. Find the solution of the problem for a linearly elastic bar with zero initial conditions, zero external forcing, and boundary conditions $U(0, t) = 0$ and $U(\ell, t) = U_0$.

7.23. In solving the tautochrone problem one finds that it is necessary to solve the integral equation

$$\int_0^t \frac{u(r)}{\sqrt{t - r}} dr = \alpha, \text{ for } 0 < t,$$

where α is a positive constant. Find the solution using the Laplace transform.

7.24. Using the Laplace transform solve the integral equation

$$u(t) + \int_0^t \frac{u(r)}{\sqrt{t - r}} dr = f(t),$$

where $f(t)$ is smooth and satisfies $f(0) = 0$.

7.25. This problem explores some of consequences of the periodic displacement example of Section 7.2.2. Assume that $0 < \tau_0 < \tau_1$.

(a) Assume that the maximum value of δ, as determined from (7.62), is δ_M and occurs at frequency ω_M. Show that $\tau_1 = [\tan(2\delta_M) + \sec(2\delta_M)]/\omega_M$ and $\tau_0 = 1/(\tau_1 \omega_M^2)$.

(b) Use your results from part (a) to estimate τ_1 that was used in Figure 7.9.

Chapter 8
Continuum Mechanics: Three Spatial Dimensions

8.1 Introduction

The water in the ocean, the air in the room, and a rubber ball have a common characteristic, they appear to completely occupy their respective domains. What this means is that the material occupies every point in the domain. This observation is the basis of the continuum approximation, and it was used in Section 5.2 to define continuum variables such as density and flux. These variables can be defined as long as the individual nature of the constituent particles are not apparent. So, for example, the continuum approximation cannot be used on the nanometer scale, because atomic radii range from 0.2 to 3.0 nm. It can, however, be used down to the micron level. As an example, at 15° C, and one atmosphere, there are approximately 3×10^7 molecules in a cubic micron of air. Similarly, for water at room temperature there are approximately 3×10^{10} molecules in a cubic micron, and for a metal such as copper there are approximately 10^{11} atoms in a cubic micron. Consequently, the averaging on which the continuum approximation is based is applicable down to the micron scale. This is why continuum models are commonly used for microdevices, which involve both electrical and mechanical components. At the other extreme, continuum models are used to study the motion of disk galaxies and, more recently, to investigate the existence and properties of the "dark fluid" proposed to be responsible for the expansion of the universe. This range of applicability is why the continuum approximation, and the subsequent equations of motion, play a fundamental role in most branches of science and engineering. From a mathematical standpoint, the problems that come from continuum models have been almost single-handedly responsible for the development of an area central to applied mathematics, and this is the theory of nonlinear partial differential equations. In this chapter the fundamental concepts of continuum mechanics are introduced, and they are then used to derive equations of motion for viscous fluids and elastic solids.

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_8,
© Springer Science+Business Media, LLC 2009

8.2 Material and Spatial Coordinates

To define the material coordinate system, assume that at $t = 0$ a particular point in the material is located at $\mathbf{x} = \mathbf{A}$. It is assumed that as the material moves, the position of the point is given as $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$. To be consistent, the position function must satisfy $\mathbf{X}(\mathbf{A}, 0) = \mathbf{A}$. The resulting displacement and velocity functions are defined as

$$\mathbf{U}(\mathbf{A}, t) = \mathbf{X}(\mathbf{A}, t) - \mathbf{A}, \tag{8.1}$$

and

$$\mathbf{V}(\mathbf{A}, t) = \frac{\partial \mathbf{U}}{\partial t}\,. \tag{8.2}$$

Because $\mathbf{X}(\mathbf{A}, 0) = \mathbf{A}$, it follows that $\mathbf{U}(\mathbf{A}, 0) = \mathbf{0}$.

Instead of following particles as they move, one can select a spatial location and then let them come to you. This is the viewpoint taken for spatial coordinates. In this system, the displacement function is denoted as $\mathbf{u}(\mathbf{x}, t)$, and the velocity is $\mathbf{v}(\mathbf{x}, t)$. As is usual for displacement functions, it is required that $\mathbf{u}(\mathbf{x}, 0) = \mathbf{0}$.

Example

Suppose a particle that started at location $(1, -1, 1)$ is, at $t = 2$, located at $(3, 0, -1)$.

Material Coordinates: For this particle, $\mathbf{A} = (1, -1, 1)$, and its displacement at $t = 2$ is $(3, 0, -1) - (1, -1, 1) = (2, 1, -2)$. In other words,

$$\mathbf{U}(\mathbf{A}, 2) = (2, 1, -2), \quad \text{for } \mathbf{A} = (1, -1, 1). \tag{8.3}$$

We also have that

$$\mathbf{X}(\mathbf{A}, 0) = (1, -1, 1), \quad \text{for } \mathbf{A} = (1, -1, 1),$$

and

$$\mathbf{X}(\mathbf{A}, 2) = (3, 0, -1), \quad \text{for } \mathbf{A} = (1, -1, 1).$$

Spatial Coordinates: At $t = 2$, the displacement of the particle located at $(3, 0, -1)$ is $(2, 1, -2)$. In other words,

$$\mathbf{u}(\mathbf{x}, 2) = (2, 1, -2), \quad \text{for } \mathbf{x} = (3, 0, -1). \quad \blacksquare \tag{8.4}$$

The spatial system is the one usually used for fluids, which includes both gases and liquids. As an example, when measuring the properties of the atmosphere, observers are often fixed, and not moving with the air. This is the viewpoint taken by the spatial coordinate system, and hence the reason why

it is the default system in fluid dynamics. In contrast, the material system is associated with solid mechanics, because the configuration at $t = 0$, what is known as the reference state, is usually known for a solid. The fact is, however, that some of the more interesting contemporary applications of continuum mechanics involve both fluid and solid components. A particularly rich area for this is biology, which includes the study of how birds fly and the study of the internal mechanisms of cell function. For this reason, both coordinate systems need to be understood, and both are studied in this chapter.

The material and spatial descriptions for the displacement and velocity functions must be consistent. This was demonstrated in the last example, as given in (8.3) and (8.4). To express this in a general form, for a particle with position function $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$ it is required that $\mathbf{U}(\mathbf{A}, t) = \mathbf{u}(\mathbf{x}, t)$, and $\mathbf{V}(\mathbf{A}, t) = \mathbf{v}(\mathbf{x}, t)$. More expansively, these equations can be written as

$$\mathbf{U}(\mathbf{A}, t) = \mathbf{u}(\mathbf{X}(\mathbf{A}, t), t), \tag{8.5}$$

and

$$\mathbf{V}(\mathbf{A}, t) = \mathbf{v}(\mathbf{X}(\mathbf{A}, t), t). \tag{8.6}$$

The transformation between the two coordinate systems is assumed to be invertible, and so it is possible to solve $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$ uniquely for $\mathbf{A}$. Writing the solution as $\mathbf{A} = \mathbf{a}(\mathbf{x}, t)$ then

$$\mathbf{u}(\mathbf{x}, t) = \mathbf{U}(\mathbf{a}(\mathbf{x}, t), t), \tag{8.7}$$

and

$$\mathbf{v}(\mathbf{x}, t) = \mathbf{V}(\mathbf{a}(\mathbf{x}, t), t). \tag{8.8}$$

The above formulas will be invaluable when converting the equations of motion between the two coordinate systems

8.2.1 Deformation Gradient

The assumption that the transformation between the two coordinate systems is invertible is one of the fundamental hypotheses in continuum mechanics. To explore this a bit more, suppose that given a material point $\mathbf{A}_0$ that its spatial counterpart is $\mathbf{x}_0 = \mathbf{X}(\mathbf{A}_0, t)$. For material points $\mathbf{A} = \mathbf{A}_0 + \Delta\mathbf{A}$ near $\mathbf{A}_0$, we have from Taylor's theorem

$$\begin{aligned} \mathbf{x} &= \mathbf{X}(\mathbf{A}_0 + \Delta\mathbf{A}, t) \\ &\approx \mathbf{X}(\mathbf{A}_0, t) + \mathbf{F}\Delta\mathbf{A} \\ &= \mathbf{x}_0 + \mathbf{F}\Delta\mathbf{A}, \end{aligned} \tag{8.9}$$

where $\mathbf{F}$ is the Jacobian matrix for $\mathbf{X}$, evaluated at $\mathbf{A}_0$. Letting $\mathbf{X} = (X, Y, Z)$ and $\mathbf{A} = (A_1, A_2, A_3)$, then

$$\mathbf{F} = \begin{pmatrix} \dfrac{\partial X}{\partial A_1} & \dfrac{\partial X}{\partial A_2} & \dfrac{\partial X}{\partial A_3} \\ \dfrac{\partial Y}{\partial A_1} & \dfrac{\partial Y}{\partial A_2} & \dfrac{\partial Y}{\partial A_3} \\ \dfrac{\partial Z}{\partial A_1} & \dfrac{\partial Z}{\partial A_2} & \dfrac{\partial Z}{\partial A_3} \end{pmatrix}. \tag{8.10}$$

This matrix plays an important role in continuum mechanics, and is known as the *deformation gradient*. The reason it is important can be seen in (8.9), which shows that as a local approximation, $\mathbf{x} - \mathbf{x}_0 = \mathbf{F}(\mathbf{A} - \mathbf{A}_0)$. Consequently, $\mathbf{F}$ is a measure of how much the motion is distorting the material, where $\mathbf{F} = \mathbf{I}$ means that there is no distortion. We will return to this idea later in the chapter, once the equations of motion are derived.

If the transformation between the spatial and material coordinate systems is invertible, then it must be possible to find $\mathbf{A}$ given $\mathbf{x}$. We must, therefore, be able to solve $\mathbf{x} - \mathbf{x}_0 = \mathbf{F}(\mathbf{A} - \mathbf{A}_0)$ for $\mathbf{A}$. The result is that $\mathbf{A} = \mathbf{A}_0 + \mathbf{F}^{-1}(\mathbf{x} - \mathbf{x}_0)$. The requirement for this to hold is that $\mathbf{F}$ is invertible. This is known as the impenetrability of matter assumption, and its mathematical statement is that $\det(\mathbf{F}) \neq 0$. This means that $\det(\mathbf{F})$ is either always positive or it is always negative. Given that $\mathbf{X}(\mathbf{A}, 0) = \mathbf{A}$, so $\mathbf{F} = \mathbf{I}$ at $t = 0$, then the mathematical form for the impenetrability of matter assumption is

$$\det(\mathbf{F}) > 0. \tag{8.11}$$

The one-dimensional version of this condition is given in (6.19). As in Chapter 6, this inequality is assumed to hold whenever discussing the continuum theory.

Example: Uniform Dilatation

A motion given by

$$\mathbf{x} = \alpha(t)\mathbf{A}, \tag{8.12}$$

where $\alpha(0) = 1$, is called a uniform dilatation. As an example, suppose we start out, at $t = 0$, with a sphere of radius r that is centered at the origin (Figure 8.1). Under a uniform dilatation we still have a sphere, centered at the origin, but with a radius αr. If $\alpha > 1$ there is a uniform expansion while for $0 < \alpha < 1$ there is uniform contraction. This is illustrated in Figure 8.1 for a circular region in the plane. Calculating the displacement and velocity, in material coordinates, using (8.1) and (8.2), we have that $\mathbf{U}(\mathbf{A}, t) = (\alpha - 1)\mathbf{A}$ and $\mathbf{V}(\mathbf{A}, t) = \alpha'\mathbf{A}$. To find the spatial version, we solve (8.12) to obtain $\mathbf{A} = \mathbf{x}/\alpha$. From (8.7) and (8.8) it follows that $\mathbf{u}(\mathbf{x}, t) =$

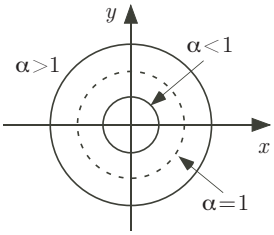


Figure 8.1 Uniform dilatation of a circular region.

$(\alpha - 1)\mathbf{x}/\alpha$ and $\mathbf{v}(\mathbf{x}, t) = \alpha'\mathbf{x}/\alpha$. Therefore, as in the one-dimensional case, $\mathbf{v} \neq \frac{\partial \mathbf{u}}{\partial t}$. Finally, from (8.10), the deformation gradient is $\mathbf{F} = \alpha\mathbf{I}$. To satisfy the impenetrability of matter condition (8.11), it is required that $\alpha > 0$. ■

Example: Simple Shear

A motion given by $x = A_1 + \alpha(t)A_2$, $y = A_2$, $z = A_3$, where $\alpha(0) = 1$, is an example of simple shear. An illustration of what happens in simple shear is shown in Figure 8.2, where a rectangle is transformed into a parallelogram with the same height. This is the type of motion one gets when pushing on the side of a deck of cards. To determine the kinematic variables, because $\mathbf{X} = (A_1 + \alpha(t)A_2, A_2, A_3)$ then, from (8.1), we have that $\mathbf{U}(\mathbf{A}, t) = (\alpha(t)A_2, 0, 0)$. To find the displacement in spatial coordinates, we solve $\mathbf{x} = (A_1 + \alpha(t)A_2, A_2, A_3)$ to obtain $(A_1, A_2, A_3) = (x - \alpha y, y, z)$. With this,

$$\begin{aligned}\mathbf{u}(\mathbf{x}, t) &= \mathbf{U}(x - \alpha y, y, z, t)\\ &= (\alpha(t)y, 0, 0).\end{aligned}$$

Finally, from (8.10)

$$\mathbf{F} = \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Given that $\det(\mathbf{F}) = 1$, then this motion satisfies the impenetrability of matter condition for any value of α. ■

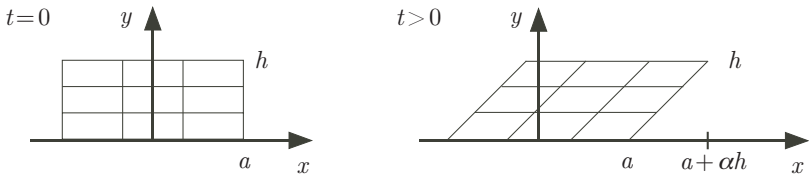


Figure 8.2 Simple shear of a rectangular region.

Example: Rigid Body Motion

A rigid motion is one given by

$$\mathbf{x} = \mathbf{Q}(t)\mathbf{A} + \mathbf{b}(t), \tag{8.13}$$

where $\mathbf{Q}(t)$ is a rotation matrix with $\mathbf{Q}(0) = \mathbf{I}$, and $\mathbf{b}(0) = \mathbf{0}$. Therefore, it consists of a rotation, determined by $\mathbf{Q}$, followed by a translation given by $\mathbf{b}$. To qualify for a rotation, the matrix $\mathbf{Q}$ must satisfy $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ and $\det(\mathbf{Q}) = 1$. As an example, consider a merry-go-round motion, where the points in the x, y-plane rotate around the z-axis. This happens if $\mathbf{b}(t) = \mathbf{0}$ and

$$\mathbf{Q}(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) & 0 \\ \sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{8.14}$$

In this case, the points rotate around the z-axis with an angular velocity ω. ■

8.3 Material Derivative

To derive the formula for the material derivative, suppose $F(\mathbf{A}, t)$ is a variable or function in material coordinates and its spatial version is $f(\mathbf{x}, t)$. In this case, $\frac{\partial F}{\partial t}$ is the time rate of change of the variable for the material point that began at $\mathbf{A}$. To determine what this is in spatial coordinates note that F and f must produce the same value. Therefore, if the material point that started at $\mathbf{A}$ is currently located at $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$ then it must be that

$$F(\mathbf{A}, t) = f(\mathbf{X}(\mathbf{A}, t), t). \tag{8.15}$$

Letting $\mathbf{x} = (x, y, z)$ and $\mathbf{X} = (X, Y, Z)$, we have that

$$\begin{aligned} \frac{\partial F}{\partial t} &= \frac{\partial f}{\partial x}\frac{\partial X}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial Y}{\partial t} + \frac{\partial f}{\partial z}\frac{\partial Z}{\partial t} + \frac{\partial f}{\partial t} \\ &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) \cdot \frac{\partial \mathbf{U}}{\partial t} + \frac{\partial f}{\partial t} \\ &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) \cdot \mathbf{V} + \frac{\partial f}{\partial t} \\ &= \nabla f \cdot \mathbf{v} + \frac{\partial f}{\partial t} \\ &= \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\right) f, \end{aligned} \tag{8.16}$$

where

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right). \tag{8.17}$$

Therefore, the *material derivative* in three spatial dimensions is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla. \tag{8.18}$$

This derivative is important because it is the time rate of change of a function following a material point, but expressed in spatial coordinates.

It is not hard to show that the material derivative satisfies some, but not all, of the elementary properties of differentiation. For example, for constants α, β and functions f, g,

$$\frac{D}{Dt}(\alpha f + \beta g) = \alpha \frac{Df}{Dt} + \beta \frac{Dg}{Dt}\,,$$
$$\frac{D}{Dt}(fg) = g\frac{Df}{Dt} + f\frac{Dg}{Dt}\,.$$

However, because of the $\mathbf{v}$ in the formula for the material derivative, it is generally true that

$$\frac{D}{Dt}\frac{\partial}{\partial x} \neq \frac{\partial}{\partial x}\frac{D}{Dt}\,,$$
$$\frac{D}{Dt}\frac{\partial}{\partial t} \neq \frac{\partial}{\partial t}\frac{D}{Dt}\,.$$

In other words, interchanging the order of differentiation requires some care when using the material derivative.

A particularly important example is the material derivative of the displacement function. Recalling that $\mathbf{V} = \frac{\partial}{\partial t}\mathbf{U}$, it follows from (8.16) that

$$\mathbf{v} = \frac{D\mathbf{u}}{Dt}\,. \tag{8.19}$$

This is the vector version of (6.13), and some of the complications that arise from this innocent-looking formula are explored in Exercise 8.10.

Example: Uniform Dilatation (cont'd)

For uniform dilatation we found that $\mathbf{u}(\mathbf{x}, t) = (\alpha - 1)\mathbf{x}/\alpha$ and $\mathbf{v}(\mathbf{x}, t) = \alpha'\mathbf{x}/\alpha$. To check on (8.19),

$$\begin{aligned}\frac{D\mathbf{u}}{Dt} &= \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{u} \\ &= \frac{\alpha'}{\alpha^2}\,\mathbf{x} + \frac{\alpha'}{\alpha}\,\mathbf{x}\left(1-\frac{1}{\alpha}\right) \\ &= \frac{\alpha'}{\alpha}\,\mathbf{x}.\end{aligned}$$

So, as expected, (8.19) holds. ■

The above derivation of the material derivative closely follows what was done for one dimension. In fact, this is true for much of what is done in the chapter. There are some notable exceptions to this statement, and this will become evident when we introduce the stress tensor in Section 8.6.1.

8.4 Mathematical Tools

The key tool in deriving the equations of motion is the Reynolds Transport Theorem. To state this result, consider a collection of material points that at $t = 0$ occupy a volume $R(0)$, as shown in Figure 8.3. Due to the motion, at later times these same points occupy the volume $R(t)$. The surface of this volume is denoted as $\partial R(t)$. For example, if R is the ball $||\mathbf{x}|| \leq 2t+1$, then $\partial R(t)$ is the sphere $||\mathbf{x}|| = 2t+1$. With this we have the following result, known as the Reynolds Transport Theorem.

Theorem 8.1. *Assuming $R(t)$ is regular, and f is a smooth function then*

$$\frac{d}{dt}\iiint\limits_{R(t)} f(\mathbf{x},t)dV = \iiint\limits_{R(t)} \frac{\partial f}{\partial t}dV + \iint\limits_{\partial R(t)} f\mathbf{v}\cdot\mathbf{n}\,dS, \tag{8.20}$$

where $\mathbf{n}$ is the unit outward normal to $R(t)$, $\mathbf{v}$ is the velocity of the points on the surface, dV is the volume element, and dS is the surface element.

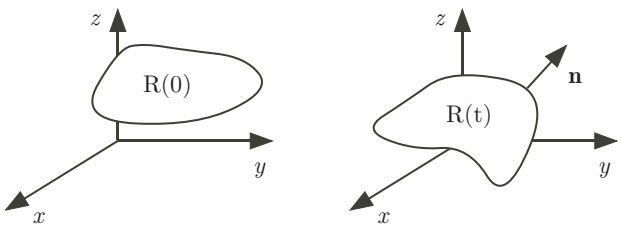


Figure 8.3 The material points that occupy the region $R(0)$ at $t = 0$ move and at later times occupy $R(t)$.

Stating that $R(t)$ is regular includes several requirements, all imposed on the original region $R(0)$ and the motion $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$. First, $R(0)$ is assumed to satisfy the conditions stated in the integral theorems of multivariable calculus. Namely, $R(0)$ is bounded with a boundary $\partial R(0)$ that consists of finitely many smooth, closed orientable surfaces. The second assumption is that the motion is smooth and satisfies (8.11). The reason for this is that in the proof of (8.20) a change of coordinates is made in the volume integral to transform it into an integral over the time-independent domain $R(0)$. To use the change of variables theorem from multivariable calculus the Jacobian for the transformation must be nonzero, and that is guaranteed if (8.20) holds.

To outline the proof of (8.20), the first step is to change variables in the integrals so the limits are not dependent on time. The natural choice is to use material coordinates, and let $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$. The Jacobian matrix for this change of variables is $\mathbf{F}$, given in (8.10). From the change of variables formula for multiple integrals,

$$\iiint\limits_{R(t)} f(\mathbf{x}, t) dxdydz = \iiint\limits_{R(0)} f(\mathbf{X}(\mathbf{A}, t), t) \det(\mathbf{F}) dA_1 dA_2 dA_3. \tag{8.21}$$

In the calculations to follow, we need a result from analysis for the derivative of a determinant (Bourbaki [2004]). Namely, given a smooth invertible matrix $\mathbf{M}(t)$,

$$\frac{d}{dt} \det(\mathbf{M}) = \det(\mathbf{M}) \ \mathrm{tr}\left(\mathbf{M}^{-1} \frac{d}{dt} \mathbf{M}\right), \tag{8.22}$$

where det() is the determinant and tr() is the trace. The trace is the sum of the diagonal entries of the matrix, and so $\mathrm{tr}(\mathbf{M}) = M_{11} + M_{22} + M_{33}$. Its basic properties, as well as those for the determinant, are given in Appendix D. Setting $J = \det(\mathbf{F})$ then from (8.22), and the results from Exercise 8.9,

$$\begin{aligned} \frac{\partial J}{\partial t} &= J \ \mathrm{tr}\left(\mathbf{F}^{-1} \frac{d}{dt} \mathbf{F}\right) \\ &= J \ \mathrm{tr}\left(\mathbf{F}^{-1} \nabla_A \mathbf{V}\right) \\ &= J \ \mathrm{tr}\left(\nabla \mathbf{v}\right). \end{aligned} \tag{8.23}$$

In the above expressions, letting $\mathbf{V} = (V_1, V_2, V_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$,

$$\nabla_A \mathbf{V} = \begin{pmatrix} \dfrac{\partial V_1}{\partial A_1} & \dfrac{\partial V_1}{\partial A_2} & \dfrac{\partial V_1}{\partial A_3} \\ \dfrac{\partial V_2}{\partial A_1} & \dfrac{\partial V_2}{\partial A_2} & \dfrac{\partial V_2}{\partial A_3} \\ \dfrac{\partial V_3}{\partial A_1} & \dfrac{\partial V_3}{\partial A_2} & \dfrac{\partial V_3}{\partial A_3} \end{pmatrix}, \tag{8.24}$$

and

$$\nabla \mathbf{v} = \begin{pmatrix} \dfrac{\partial v_1}{\partial x} & \dfrac{\partial v_1}{\partial y} & \dfrac{\partial v_1}{\partial z} \\ \dfrac{\partial v_2}{\partial x} & \dfrac{\partial v_2}{\partial y} & \dfrac{\partial v_2}{\partial z} \\ \dfrac{\partial v_3}{\partial x} & \dfrac{\partial v_3}{\partial y} & \dfrac{\partial v_3}{\partial z} \end{pmatrix}. \tag{8.25}$$

The above two tensors play an important role in continuum mechanics, and they are called velocity gradients. Specifically, $\nabla_A \mathbf{V}$ is the material velocity gradient tensor, and $\nabla \mathbf{v}$ is the spatial velocity gradient tensor. One important property that we need here is that

$$\operatorname{tr}(\nabla \mathbf{v}) = \nabla \cdot \mathbf{v}. \tag{8.26}$$

With this, (8.23) reduces to

$$\frac{\partial J}{\partial t} = J(\nabla \cdot \mathbf{v}). \tag{8.27}$$

It should be remembered that in the above expression J is a function of $\mathbf{A}$ and t, and $\nabla \cdot \mathbf{v}$ is evaluated at $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$.

We are now in position to differentiate (8.21). Letting $\mathbf{X} = (X, Y, Z)$,

$$\begin{aligned} &\frac{d}{dt} \iiint\limits_{R(t)} f(\mathbf{x}, t) dx dy dz \\ &= \iiint\limits_{R(0)} \left[\frac{\partial f}{\partial t} J + \frac{\partial f}{\partial x}\frac{\partial X}{\partial t} J + \frac{\partial f}{\partial y}\frac{\partial Y}{\partial t} J + \frac{\partial f}{\partial z}\frac{\partial Z}{\partial t} J + f \frac{\partial J}{\partial t} \right] dA_1 dA_2 dA_3 \\ &= \iiint\limits_{R(0)} \left[\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} V_1 + \frac{\partial f}{\partial y} V_2 + \frac{\partial f}{\partial z} V_3 + f \nabla \cdot \mathbf{v} \right] J dA_1 dA_2 dA_3 \\ &= \iiint\limits_{R(0)} \left[\frac{\partial f}{\partial t} + \nabla \mathbf{f} \cdot \mathbf{v} + f \nabla \cdot \mathbf{v} \right] J dA_1 dA_2 dA_3 \\ &= \iiint\limits_{R(t)} \left[\frac{\partial f}{\partial t} + \nabla \cdot (f \mathbf{v}) \right] dx dy dz. \end{aligned} \tag{8.28}$$

The next step requires the Divergence Theorem, which states that for a smooth function $\mathbf{w}$,

$$\iiint\limits_{R} \nabla \cdot \mathbf{w} \, dV = \iint\limits_{\partial R} \mathbf{w} \cdot \mathbf{n} \, dS.$$

Taking $\mathbf{w} = f\mathbf{v}$, then (8.28) reduces to (8.20), and the theorem is proved.

A useful form of the Reynolds Transport formula comes out of the proof, and it is worth restating the result. From (8.28), and the definition of the material derivative in (8.18), it follows that

$$\frac{d}{dt}\iiint\limits_{R(t)} f(\mathbf{x},t)dV = \iiint\limits_{R(t)} \left(\frac{Df}{Dt} + f\nabla\cdot\mathbf{v}\right) dV. \tag{8.29}$$

8.4.1 General Balance Law

The above integral theorems will be used to take balance laws that are formulated as integrals and express them as differential equations. The steps involved in this derivation are always the same, so it is worth deriving a general formula that can be used when needed. With this in mind, the general form of the balance laws can be written as

$$\frac{d}{dt}\iiint\limits_{R(t)} f(\mathbf{x},t)dV = -\iint\limits_{\partial R(t)} \mathbf{J}\cdot\mathbf{n}\,dS + \iiint\limits_{R(t)} Q(\mathbf{x},t)dV. \tag{8.30}$$

To state the above equation in physical terms, f can be thought of a density of a quantity, and examples are mass density, momentum density, and energy density. The above balance law states that the rate of change of the total amount of this quantity in a region $R(t)$ is due to the flux across the boundary and the creation or loss through the volume. The flux in this case is $\mathbf{J}$, and Q is the creation or loss density.

The integral balance law (8.30) is the three dimensional version of (4.47). Differentiating the integral using (8.29), and using the Divergence Theorem to convert the surface integral into a volume integral, (8.30) can be written as

$$\iiint\limits_{R(t)} \left(\frac{Df}{Dt} + f\nabla\cdot\mathbf{v}\right) dV = \iiint\limits_{R(t)} [-\nabla\cdot\mathbf{J} + Q(\mathbf{x},t)]dV.$$

The balance law is assumed to hold for all volumes R, and therefore from the du Bois-Reymond Lemma we have that

$$\frac{Df}{Dt} + f\nabla\cdot\mathbf{v} = -\nabla\cdot\mathbf{J} + Q, \tag{8.31}$$

or equivalently

$$\frac{\partial f}{\partial t} + \nabla\cdot(\mathbf{v}f) = -\nabla\cdot\mathbf{J} + Q. \tag{8.32}$$

This equation, in one form or another, has been used repeatedly in this textbook. The one-dimensional version (4.48) was used to derive the diffusion equation in Chapter 3, it was used in the derivation of the traffic flow equation

in Chapter 4, and it was used multiple times in Chapter 6. We are now going to use it to derive the equations of continuum mechanics.

8.5 Continuity Equation

The assumption is that mass is neither created nor destroyed. To express this mathematically, assume that at $t = 0$ a collection of material points occupies the volume $R(0)$. At any later time these same points occupy a spatial volume $R(t)$. Our assumption means that the total mass of the material in this region does not change. If we let $\rho(\mathbf{x}, t)$ designate the mass density of the material (i.e., mass per unit volume) then our assumption states that

$$\frac{d}{dt} \iiint\limits_{R(t)} \rho(\mathbf{x}, t) dV = 0. \tag{8.33}$$

In terms of the general law (8.30), we have that $f = \rho$, $\mathbf{J} = \mathbf{0}$, and $Q = 0$. Therefore, from (8.31) we have that the continuity equation is

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0. \tag{8.34}$$

8.5.1 Incompressibility

When studying the motions of liquids, such as water, it is very often assumed the liquid is incompressible. The idea is that even though a volume of material points moves, and changes shape, the total volume is constant. This assumption provides an addition balance law, and it is that

$$\frac{d}{dt} \iiint\limits_{R(t)} dV = 0. \tag{8.35}$$

In this case, in (8.30), $f = 1$, $\mathbf{J} = \mathbf{0}$, and $Q = 0$, and so from (8.31) the resulting differential equation is

$$\nabla \cdot \mathbf{v} = 0. \tag{8.36}$$

This is the continuity equation for an incompressible material, fluid or solid. You might be wondering what happens to the more general version given in (8.34). Well, in this case it reduces to $\frac{D\rho}{Dt} = 0$. This states that the density following a material point does not change in time. Therefore, if the density is initially constant, then it is constant for all time. In this textbook, whenever

discussing an incompressible material it will always be assumed that the initial density is constant, so ρ is constant.

Examples

1. For uniform dilatation, $\mathbf{v}(\mathbf{x}, t) = \alpha' \mathbf{x}/\alpha$. In this case, $\nabla \cdot \mathbf{v} = 3\alpha'/\alpha$. Assuming that α is not constant then $\nabla \cdot \mathbf{v} \neq 0$. Therefore, uniform dilatation is not possible for an incompressible material. ■

2. For translational motion the velocity $\mathbf{v}$ is independent of $\mathbf{x}$, but can depend on t. Given that $\mathbf{v} = \mathbf{v}(t)$ then $\nabla \cdot \mathbf{v} = 0$. This means that translational motion is possible for an incompressible material. This conclusion makes sense from a physical point-of-view because the volume of an object is unaffected by translation. By the same reasoning, it is expected that rigid body motion is possible for an incompressible material, and this is proved in Exercise 8.4. ■

8.6 Linear Momentum Equation

The derivation of the momentum equation requires more effort than for mass conservation. One complication is that we now need to distinguish between linear and angular momentum. We start with linear momentum, which is what was derived in the one-dimensional formulation in Chapter 6. To determine the forces within the material we consider what happens as it moves. With this in mind, we assume that at $t = 0$ a collection of material points occupies a volume $R(0)$, and at any later time these same points occupy a spatial volume $R(t)$, as shown in Figure 8.3. As in the one-dimensional formulation, it is assumed that R is subject to external body forces, measured with respect to unit mass, and these are denoted as $\mathbf{f}$. There are also forces that act on the surface of R due to the relative deformation of the material. The one-dimensional version of this is shown in Figure 6.3. The idea is that the material points external to R act on R across the surface ∂R. To incorporate this into the balance law, given a point $\mathbf{x}$ on the surface, let $\mathbf{t}$ be the force, per unit area, on R due the material exterior to R. Because of its units, $\mathbf{t}$ is referred to as a stress vector. Also, from Newton's Third Law, $-\mathbf{t}$ is the stress of the material in R on the exterior region.

With this the balance of linear momentum gives us that

$$\frac{d}{dt} \iiint_{R(t)} \rho \mathbf{v} \, dV = \iint_{\partial R(t)} \mathbf{t} \, dS + \iiint_{R(t)} \rho \mathbf{f} \, dV. \tag{8.37}$$

It is important to understand that the above equation is an assumption, and is one of the balance laws of continuum mechanics. We will reduce it to a

differential equation, but before doing that it is necessary to consider the stress vector $\mathbf{t}$ in more detail.

8.6.1 Stress Tensor

The equation in (8.37) is almost too general. To explain this statement, all four quantities in this equation, ρ, $\mathbf{v}$, $\mathbf{t}$, and $\mathbf{f}$, can depend on $\mathbf{x}$ and t. The new complication is that $\mathbf{t}$ also depends on direction. As an example, if you pull on a sheet of rubber, the stress $\mathbf{t}$ in the direction of the pull is different than the stress $\mathbf{t}$ perpendicular to the pull. Given that we are dealing with vectors in $\mathbb{R}^3$, you might think that if you know the stress vector for three different directions that you can use this to find the stress vector in any other direction. This is correct, and the next few paragraphs show how this is done. It is important to point out that this result is much more than simply constructing a basis in $\mathbb{R}^3$. It also requires the use of some of the fundamental laws of physics.

We are going to show that the stress at a point on the surface can be decomposed into stresses on three orthogonal coordinate surfaces. To explain how, for simplicity, assume that the point is the origin. A small square piece of the planar surface $y = 0$, centered at the origin, is shown in Figure 8.5(b). The material in $y > 0$ exerts a force $\mathbf{F}_y$ on this square on the material in $y < 0$. The resulting approximation for the stress is $\mathbf{F}_y/A$, where $A = h^2$ is the area of the square. Letting the area shrink to zero we obtain the stress vector $\mathbf{t}_y = (T_{21}, T_{22}, T_{23})^T$, and this vector is shown in Figure 8.5(b). It follows from Newton's Third Law, that the stress of $y < 0$ on $y > 0$ is $-\mathbf{t}_y$. In the same way, on the surface $x = 0$ there is a stress $\mathbf{t}_x = (T_{11}, T_{12}, T_{13})^T$, and on the surface $z = 0$ there is a stress $\mathbf{t}_z = (T_{31}, T_{32}, T_{33})^T$.

We are interested in the stress $\mathbf{t} = (t_1, t_2, t_3)^T$ on the surface that has unit outward normal $\mathbf{n} = (n_1, n_2, n_3)^T$, as illustrated in Figure 8.4. To relate this

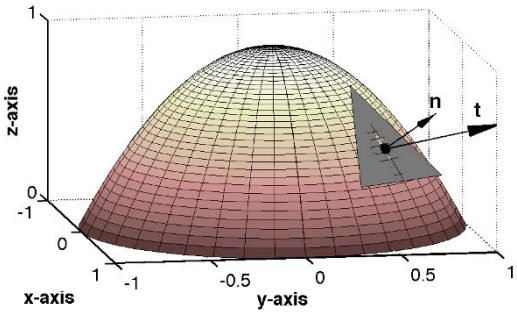


Figure 8.4 A triangular piece of the tangent plane to a surface, along with the unit outward normal $\mathbf{n}$ and stress vector $\mathbf{t}$.

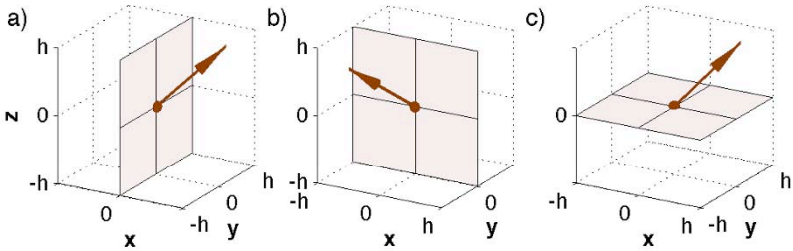


Figure 8.5 Stress vectors on three orthogonal coordinate surfaces. Shown are: (a) $\mathbf{t}_x$; (b) $\mathbf{t}_y$; and (c) $\mathbf{t}_z$.

with the stress vectors in Figure 8.5, consider a small tetrahedron as shown in Figure 8.6. The lateral face ABC of the tetrahedron is perpendicular to $\mathbf{n}$, and is the triangular piece of the tangent plane shown in Figure 8.4. The critical observation is that the forces on this small volume must balance. Letting ΔA_y be the area of triangular face ABO then the total force on the tetrahedron across this face is approximately $-\Delta A_y \mathbf{t}_y$. The forces on the other faces are determined in a similar manner, and from the requirement that they balance we have that

$$-\Delta A_x \mathbf{t}_x - \Delta A_y \mathbf{t}_y - \Delta A_z \mathbf{t}_z + \Delta A \mathbf{t} = \mathbf{0}, \tag{8.38}$$

where ΔA is the area of the lateral face. A particularly useful, if not well known, result from geometry is that the area of the four faces of the tetrahedron are related as follows

$$\Delta A_x = n_1 \Delta A, \quad \Delta A_x = n_2 \Delta A, \quad \Delta A_z = n_3 \Delta A.$$

Introducing these into (8.38), and cleaning things up a bit we have that

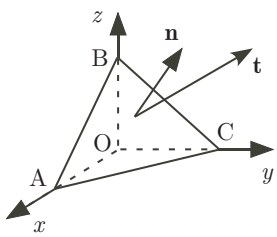


Figure 8.6 Cauchy stress tetrahedron used to derive the stress tensor in (8.40).

$$
\begin{aligned}
\mathbf{t} &= n_1\mathbf{t}_x + n_2\mathbf{t}_y + n_3\mathbf{t}_z \\
&= n_1\begin{pmatrix} T_{11} \\ T_{12} \\ T_{13} \end{pmatrix} + n_2\begin{pmatrix} T_{21} \\ T_{22} \\ T_{23} \end{pmatrix} + n_3\begin{pmatrix} T_{31} \\ T_{32} \\ T_{33} \end{pmatrix} \\
&= \begin{pmatrix} n_1T_{11} + n_2T_{21} + n_3T_{31} \\ n_1T_{12} + n_2T_{22} + n_3T_{32} \\ n_1T_{13} + n_2T_{23} + n_3T_{33} \end{pmatrix} \\
&= \mathbf{T}^T\mathbf{n},
\end{aligned} \tag{8.39}
$$

where

$$
\mathbf{T} = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} \tag{8.40}
$$

is known as the *Cauchy stress tensor*. This proves that three stress vectors can be used to determine the stress in any direction. One consequence of this is that we have nine unknown stress functions in (8.40). As in Chapter 6, we will use a constitutive law to determine these functions.

As defined, the rows of the stress tensor (8.40) correspond to the three stress vectors shown in Figure 8.5. The diagonal entries of $\mathbf{T}$ are referred to as the normal stresses, and the off-diagonal entries are the shear stresses. For example, the second row corresponds to $\mathbf{t}_y$ and the entry of this vector that is perpendicular, or normal, to the $y = 0$ plane is T_{22}. The other two entries T_{21} and T_{23} are the tangential, or shear, components.

Not everything was explained in the above derivation of the equation $\mathbf{t} = \mathbf{T}^T\mathbf{n}$, and this was done to simplify the presentation. In particular, it was stated that the force on ABO is approximately $-\Delta A_y\mathbf{t}_y$. The word "approximately" is used because the actual force on ABO is the integral of $-\mathbf{t}_y$ over ABO. We are using a first term Taylor series approximation of the stress, which is its value at $\mathbf{x} = \mathbf{0}$, and this gives rise to the stated approximation. A second point is that (8.38) comes directly from the force equation $\mathbf{F} = m\mathbf{a}$, but it does not contain the acceleration term. The reason is, again, Taylor's theorem. It is not difficult to show that the acceleration term has order $O(\Delta A^2)$. Therefore, it contributes to the second term in the expansion, and not to the first term approximation in (8.38). In the limit of letting the tetrahedron shrink to zero, the second terms in the approximation drop out, and for this reason they were not included in (8.38).

Before ending this section, a comment is necessary about the definition of the Cauchy stress tensor in (8.40). You would think that for a subject as old as continuum mechanics that there would be universal agreement about how to define such a fundamental concept. Well, apparently not because some define $\mathbf{T}$ to be the transpose of what is given in (8.40). It will be shown later that for most problems $\mathbf{T}$ is symmetric, so this difference has minimal consequences when using the spatial coordinate system.

8.6.2 Differential Form of Equation

Using (8.39), the balance law (8.37) for linear momentum becomes

$$\frac{d}{dt}\iiint\limits_{R(t)} \rho\mathbf{v}\,dV = \iint\limits_{\partial R(t)} \mathbf{T}^T\mathbf{n}\,dS + \iiint\limits_{R(t)} \rho\mathbf{f}\,dV. \tag{8.41}$$

This is now in a form that is the same as the general law in (8.30). Using (8.31) we therefore conclude that

$$\frac{D}{Dt}(\rho\mathbf{v}) + (\nabla\cdot\mathbf{v})(\rho\mathbf{v}) = \nabla\cdot\mathbf{T} + \rho\mathbf{f}. \tag{8.42}$$

The divergence of $\mathbf{T}$ in the above equation is defined as

$$\nabla\cdot\mathbf{T} = \begin{pmatrix} \dfrac{\partial T_{11}}{\partial x} + \dfrac{\partial T_{21}}{\partial y} + \dfrac{\partial T_{31}}{\partial z} \\ \dfrac{\partial T_{12}}{\partial x} + \dfrac{\partial T_{22}}{\partial y} + \dfrac{\partial T_{32}}{\partial z} \\ \dfrac{\partial T_{13}}{\partial x} + \dfrac{\partial T_{23}}{\partial y} + \dfrac{\partial T_{33}}{\partial z} \end{pmatrix}. \tag{8.43}$$

Expanding the material derivative, and using the continuity equation (8.34), we obtain

$$\rho\frac{D\mathbf{v}}{Dt} = \nabla\cdot\mathbf{T} + \rho\mathbf{f}. \tag{8.44}$$

This is the equation for linear momentum, or just the momentum equation for short. It is expressed in spatial coordinates, and the material coordinates version is given in Section 8.12.

8.7 Angular Momentum

Unlike continuity and linear momentum, the equation for angular momentum is simply the statement that the stress tensor is symmetric. To obtain this result, we consider the angular momentum of the volume $R(t)$. For a single point the angular momentum per unit volume is $\mathbf{x}\times(\rho\mathbf{v})$. Integrating this over the volume, and accounting for the same forces used for linear momentum, it follows that the balance law for angular momentum is

$$\frac{d}{dt}\iiint\limits_{R(t)} \mathbf{x}\times(\rho\mathbf{v})dV = \iint\limits_{\partial R(t)} \mathbf{x}\times(\mathbf{T}^T\mathbf{n})dS + \iiint\limits_{R(t)} \mathbf{x}\times(\rho\mathbf{f})dV. \tag{8.45}$$

Carrying out the cross products, writing the result in our standard balance law format, and then using the linear momentum equation to simplify the expression, the resulting equation is $(T_{32} - T_{23}, T_{13} - T_{31}, T_{21} - T_{12})^T = \mathbf{0}$. The conclusion is that $T_{32} = T_{23}$, $T_{13} = T_{31}$, $T_{21} = T_{12}$. Therefore, as stated earlier, to satisfy the balance law for angular momentum, $\mathbf{T}$ must be symmetric.

8.8 Summary of the Equations of Motion

To summarize the equations of motion up to this point, we have found that the continuity and momentum equations are, respectively,

$$\frac{D\rho}{Dt} + \rho\nabla \cdot \mathbf{v} = 0, \tag{8.46}$$

$$\rho\frac{D\mathbf{v}}{Dt} = \nabla \cdot \mathbf{T} + \rho\mathbf{f}, \tag{8.47}$$

where $\mathbf{T}$ is symmetric. If the material is assumed to incompressible, and the initial density is constant, then (8.46) is replaced with

$$\nabla \cdot \mathbf{v} = 0, \tag{8.48}$$

and ρ is a constant. Depending on the region the material occupies, boundary and initial conditions must be supplied to complete the problem.

Although the above equations are quite general, certain assumptions were made in the derivation. In particular, it was assumed that mass is not created or destroyed. If this occurs then both the continuity and the momentum equations are affected. We also assumed that there are no sources of angular momentum, other than what comes from the linear momentum forcing function $\mathbf{f}$. There are situations where this does not happen, and the most well known are micropolar materials. Those interested in investigating what this means in terms of the model and analysis should consult Eringen [2001].

One last comment worth making here is that the above equations are coordinate free. This means that if a particular orthogonal coordinate system is preferred, such as cylindrical coordinates, one only needs to determine the formulas for the divergence and gradient operators to be able to determine the equations of motion. An example of this will be given in Section 9.2.2.

8.9 Constitutive Laws

It remains to specify the constitutive law for the stress. As discussed in Chapter 6, there are certain requirements these laws are expected to satisfy. We

are interested here in only one, and it is the Principle of Material Frame-Indifference. The basic requirement is that the stress does not depend on the observer, assuming observers are connected by a rigid body motion.

To put this into a mathematical framework, suppose we change coordinates from $\mathbf{x}$ to $\mathbf{x}^*$ using rigid body motion. Specifically,

$$\mathbf{x}^* = \mathbf{Q}(t)\mathbf{x} + \mathbf{b}(t), \tag{8.49}$$

where $\mathbf{Q}(t)$ is a rotation matrix, and $\mathbf{Q}$ and $\mathbf{b}$ are smooth functions of time. To qualify for a rotation, the matrix $\mathbf{Q}$ must satisfy $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ and $\det(\mathbf{Q}) = 1$. An example of such a matrix is given in (8.14). One important property of rotations is that $\mathbf{Q}^{-1} = \mathbf{Q}^T$, and this is a direct consequence of the equation $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$. In the parlance of continuum mechanics, (8.49) is known as an Euclidean transformation. It differs from a Galilean transformation, often studied in Newtonian physics. This is because for a Galilean transformation $\mathbf{Q}$ is taken to be constant and $\mathbf{b}$ is linear in time.

The rigid body motion assumption has consequences for the material coordinate system. Given that the spatial system reduces to the material system when $t = 0$, then from (8.49) we have that

$$\mathbf{A}^* = \mathbf{Q}_0\mathbf{A} + \mathbf{b}_0, \tag{8.50}$$

where $\mathbf{Q}_0 = \mathbf{Q}(0)$ is a rotation, and both $\mathbf{Q}_0$ and $\mathbf{b}_0 = \mathbf{b}(0)$ are constants. This means that if our two observers want to revert to material coordinates, the above expression tells them how their material coordinate systems are related.

There are two tenets of the Principle of Material Frame-Indifference, and both are assumptions on the properties of the stress when measured by different observers.

Tenet 1: Objectivity

Given basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ in the $\mathbf{x}$ system, the corresponding basis vectors in the $\mathbf{x}^*$ system are $\mathbf{e}_1^* = \mathbf{Q}\mathbf{e}_1$, $\mathbf{e}_2^* = \mathbf{Q}\mathbf{e}_2$, $\mathbf{e}_3^* = \mathbf{Q}\mathbf{e}_3$. Using the respective basis vectors, the stress has a representation $\mathbf{T}$ in the $\mathbf{x}$ system, and it has a representation $\mathbf{T}^*$ in the $\mathbf{x}^*$ system. The Principle of Frame-Indifference requires that the stress obeys the usual change of basis formula from linear algebra, and so the requirement is that

$$\mathbf{T}^* = \mathbf{Q}\mathbf{T}\mathbf{Q}^T. \tag{8.51}$$

A tensor $\mathbf{T}$ that satisfies this equation is said to be *objective*, or Euclidean frame-indifferent.

Put into words, (8.51) states that the stress in the $\mathbf{x}^*$ system can be found by rotating back to the $\mathbf{x}$ system, calculating the stress there, and then rotating the result over to the $\mathbf{x}^*$ system. To show this mathematically, consider the stress vector shown in Figure 8.6 and given in (8.39). Letting

the unit outward normal in the $\mathbf{x}^*$ system be $\mathbf{n}^*$, then the stress vector in this system is $\mathbf{t}^* = \mathbf{T}^*\mathbf{n}^*$. Now, according to (8.51), $\mathbf{t}^* = \mathbf{Q}\mathbf{T}\mathbf{Q}^T\mathbf{n}^*$. Because $\mathbf{Q}^{-1} = \mathbf{Q}^T$, then $\mathbf{Q}^T\mathbf{n}^*$ is $\mathbf{n}^*$ rotated back into the $\mathbf{x}$ system, so $\mathbf{Q}^T\mathbf{n}^* = \mathbf{n}$. With this, $\mathbf{T}(\mathbf{Q}^T\mathbf{n}^*) = \mathbf{T}\mathbf{n} = \mathbf{t}$ is the stress vector in the $\mathbf{x}$ system. In this case, $\mathbf{Q}(\mathbf{T}\mathbf{Q}^T\mathbf{n}^*) = \mathbf{Q}\mathbf{t}$ is the rotation of the stress tensor back over to the $\mathbf{x}^*$ system.

Tenet 2: Form Invariance

The second tenet of Material Frame-Indifference concerns the functional form of the constitutive law. To explain, suppose that the proposed constitutive law for the stress states that it depends on a quantity $\mathbf{R}$. In other words, there is a function $\mathbf{G}$ so that

$$\mathbf{T} = \mathbf{G}(\mathbf{R}). \tag{8.52}$$

The assumption is that the form of the constitutive law does not depend on the observer. Therefore, if $\mathbf{T}^*$ and $\mathbf{R}^*$ are the $\mathbf{x}^*$ system versions of $\mathbf{T}$ and $\mathbf{R}$, then the requirement is that

$$\mathbf{T}^* = \mathbf{G}(\mathbf{R}^*). \tag{8.53}$$

In this case the constitutive law is said to be *form invariant*.

The conditions given in (8.51), (8.52), and (8.53) are combined to produce the requirement on the constitutive law coming from the Principle of Frame-Indifference. The statement is that the function $\mathbf{G}$ must satisfy

$$\mathbf{Q}\mathbf{G}(\mathbf{R})\mathbf{Q}^T = \mathbf{G}(\mathbf{R}^*), \tag{8.54}$$

for all possible values for $\mathbf{R}$, rotations $\mathbf{Q}$, and translations $\mathbf{b}$.

To use the requirement in (8.54), we need some basic information about how the variables transform under a rigid body motion. In the $\mathbf{x}$ coordinate system, the material points move according to the rule $\mathbf{x} = \mathbf{X}(\mathbf{A}, t)$, while in the $\mathbf{x}^*$ coordinate system the material points move according to the rule $\mathbf{x}^* = \mathbf{X}^*(\mathbf{A}^*, t)$. Given (8.49), it follows that $\mathbf{X}^* = \mathbf{Q}\mathbf{X} + \mathbf{b}$. Taking the time derivative of this equation it follows that

$$\mathbf{V}^* = \mathbf{Q}\mathbf{V} + \mathbf{Q}'\mathbf{X} + \mathbf{b}',$$

or equivalently

$$\mathbf{v}^* = \mathbf{Q}\mathbf{v} + \mathbf{Q}'\mathbf{x} + \mathbf{b}'. \tag{8.55}$$

In a similar manner it follows that the deformation gradient, given in (8.10), transforms as

$$\mathbf{F}^* = \mathbf{Q}\mathbf{F}. \tag{8.56}$$

This shows that $\mathbf{F}$ does not satisfy (8.51), and it is therefore not objective.

Examples

1. Suppose it is assumed that $\mathbf{T} = \mathbf{G}(\mathbf{v})$. In this case $\mathbf{R} = \mathbf{v}$, and from (8.55) we have that $\mathbf{R}^* = \mathbf{Q}\mathbf{v} + \mathbf{Q}'\mathbf{v} + \mathbf{b}'$. From this, (8.54) becomes

$$\mathbf{Q}\mathbf{G}(\mathbf{v})\mathbf{Q}^T = \mathbf{G}(\mathbf{Q}\mathbf{v} + \mathbf{Q}'\mathbf{x} + \mathbf{b}').$$

This must hold for every rotation $\mathbf{Q}$, and vectors $\mathbf{v}$ and $\mathbf{b}'$. In particular, taking $\mathbf{Q} = \mathbf{I}$ and $\mathbf{v} = \mathbf{0}$, then $\mathbf{G}(\mathbf{0}) = \mathbf{G}(\mathbf{b}')$ for all vectors $\mathbf{b}'$. The only function capable of this is the constant function, and so the stress must be independent of the velocity. The conclusion is that to be consistent with the Principle of Frame-Indifference, a constitutive law cannot depend on the spatial velocity. ■

2. Suppose it is assumed that $\mathbf{T} = \lambda\mathbf{F}\mathbf{F}^T$, where $\mathbf{F}$ is the deformation gradient and λ is a constant. In this case, the left-hand side of (8.54) is

$$\mathbf{Q}\mathbf{G}(\mathbf{R})\mathbf{Q}^T = \lambda\mathbf{Q}\mathbf{F}\mathbf{F}^T\mathbf{Q}^T.$$

Also, from (8.56) we have that the right hand side of (8.54) is

$$\begin{aligned}\mathbf{G}(\mathbf{R}^*) &= \lambda\mathbf{F}^*(\mathbf{F}^*)^T \\ &= \lambda(\mathbf{Q}\mathbf{F})(\mathbf{Q}\mathbf{F})^T \\ &= \lambda\mathbf{Q}\mathbf{F}\mathbf{F}^T\mathbf{Q}^T.\end{aligned}$$

This shows that this constitutive law satisfies (8.54), and is therefore consistent with the Principle of Frame-Indifference. ■

3. The equation for the velocity in (8.55) is an example of a relationship of the form $\mathbf{g}^*(\mathbf{x}^*, t) = \mathbf{g}(\mathbf{x}, t)$. To relate the derivatives of these two functions, from (8.49) we have that

$$\mathbf{x} = \mathbf{Q}^T(\mathbf{x}^* - \mathbf{b}).$$

Letting $\mathbf{g}^* = (g_1^*, g_2^*, g_3^*)$ and $\mathbf{g} = (g_1, g_2, g_3)$, then using the chain rule

$$\begin{aligned}\frac{\partial g_1^*}{\partial x^*} &= \frac{\partial g_1}{\partial x}\frac{\partial x}{\partial x^*} + \frac{\partial g_1}{\partial y}\frac{\partial y}{\partial x^*} + \frac{\partial g_1}{\partial z}\frac{\partial z}{\partial x^*} \\ &= \frac{\partial g_1}{\partial x}Q_{11} + \frac{\partial g_1}{\partial y}Q_{12} + \frac{\partial g_1}{\partial z}Q_{13}, \\ \frac{\partial g_1^*}{\partial y^*} &= \frac{\partial g_1}{\partial x}\frac{\partial x}{\partial y^*} + \frac{\partial g_1}{\partial y}\frac{\partial y}{\partial y^*} + \frac{\partial g_1}{\partial z}\frac{\partial z}{\partial y^*} \\ &= \frac{\partial g_1}{\partial x}Q_{21} + \frac{\partial g_1}{\partial y}Q_{22} + \frac{\partial g_1}{\partial z}Q_{23}.\end{aligned}$$

Carrying out the other derivatives the conclusion is that

$$(\nabla \mathbf{g})^* = (\nabla \mathbf{g})\mathbf{Q}^T,$$

where

$$(\nabla \mathbf{g})^* = \begin{pmatrix} \frac{\partial g_1^*}{\partial x^*} & \frac{\partial g_1^*}{\partial y^*} & \frac{\partial g_1^*}{\partial z^*} \\ \frac{\partial g_2^*}{\partial x^*} & \frac{\partial g_2^*}{\partial y^*} & \frac{\partial g_2^*}{\partial z^*} \\ \frac{\partial g_3^*}{\partial x^*} & \frac{\partial g_3^*}{\partial y^*} & \frac{\partial g_3^*}{\partial z^*} \end{pmatrix}, \qquad \nabla \mathbf{g} = \begin{pmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} & \frac{\partial g_1}{\partial z} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} & \frac{\partial g_2}{\partial z} \\ \frac{\partial g_3}{\partial x} & \frac{\partial g_3}{\partial y} & \frac{\partial g_3}{\partial z} \end{pmatrix}.$$

As an example, with the velocity given (8.55),

$$\begin{aligned} (\nabla \mathbf{v})^* &= (\mathbf{Q}\nabla \mathbf{v} + \mathbf{Q}')\mathbf{Q}^T \\ &= \mathbf{Q}(\nabla \mathbf{v})\mathbf{Q}^T + \mathbf{Q}'\mathbf{Q}^T. \quad \blacksquare \end{aligned} \tag{8.57}$$

It needs to be pointed out that Frame-Indifference is only imposed on the constitutive law for the stress, and not imposed on the equations of motion. As shown in Exercise 8.23, the momentum equation is not form invariant. More precisely, it is not invariant for Euclidean transformations, but it is for Galilean transformations. This fact is one of the reasons for the rather pointed controversies that surround the subject, and this will be discussed in more detail later.

8.9.1 Representation Theorem and Invariants

The assumptions that the stress is objective, and the constitutive law is form invariant, have some interesting consequences. One, which will play an important role in the constitutive modeling, is the next result, known as the Rivlin-Ericksen representation theorem.

Theorem 8.2. *Assume* $\mathbf{T}$ *and* $\mathbf{R}$ *are both symmetric and objective tensors. If* $\mathbf{T} = \mathbf{G}(\mathbf{R})$ *is form invariant then the constitutive law can be rewritten as*

$$\mathbf{T} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{R} + \alpha_2 \mathbf{R}^2, \tag{8.58}$$

where the coefficients α_0, α_1, *and* α_2 *are functions of*

$$I_R = tr(\mathbf{R}), \tag{8.59}$$

$$II_R = \frac{1}{2}\left(tr(\mathbf{R})^2 - tr(\mathbf{R}^2)\right), \tag{8.60}$$

$$III_R = det(\mathbf{R}). \tag{8.61}$$

In the above expressions, $tr()$ *is the trace, and* $det()$ *is the determinant.*

The proof of this theorem relies on picking different rotations that show how to simplify $\mathbf{G}$, and the steps are outlined in Exercise 8.26. A somewhat simplier proof, that requires an additonal hypothesis,is given in Exercise 8.25.

The three quantities I_R, II_R, and III_R are the principal invariants of $\mathbf{R}$. They are called invariants because their values do not change when changing coordinates using rigid body motion. The proof of this statement is based on the identities $\mathrm{tr}(\mathbf{RS}) = \mathrm{tr}(\mathbf{SR})$ and $\det(\mathbf{RS}) = \det(\mathbf{SR})$. For example, $\mathrm{III}_{R^*} = \det(\mathbf{R}^*) = \det(\mathbf{QRQ}^T) = \det(\mathbf{Q}^T\mathbf{QR}) = \det(\mathbf{R}) = \mathrm{III}_R$. To take advantage of this observation, recall that a symmetric matrix can be diagonalized, and the diagonal entries are the eigenvalues λ_1, λ_2, and λ_3. With this we have that

$$\mathrm{I}_R = \lambda_1 + \lambda_2 + \lambda_3, \tag{8.62}$$

$$\mathrm{II}_R = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3, \tag{8.63}$$

$$\mathrm{III}_R = \lambda_1\lambda_2\lambda_3. \tag{8.64}$$

This shows that in terms of their dependence on the eigenvalues, I_R is linear, II_R is quadratic, and III_R is cubic. It also shows that the three invariants are independent, in the sense that it is not possible to write one of them in terms of the other two. Some of the properties of the principal invariants are developed in Exercise 8.24.

A consequence of material frame-indifference is that the material must be isotropic, which means that the material properties are the same in all directions. To investigate what this means physically, assume that $\mathbf{v} = (u, v, w)$ and the constitutive law is

$$\mathbf{T} = \begin{pmatrix} a\dfrac{\partial u}{\partial x} & 0 & 0 \\ 0 & b\dfrac{\partial v}{\partial y} & 0 \\ 0 & 0 & c\dfrac{\partial w}{\partial z} \end{pmatrix}. \tag{8.65}$$

For an isotropic material it is required that $a = b = c$. To explain why, suppose an experimental device is aligned with the x-axis and it measures T_{11} for $\frac{\partial u}{\partial x} = 1$. If the device is picked up and rotated so it is aligned with the y-axis, and the experiment rerun, then one would measure T_{22} for $\frac{\partial v}{\partial y} = 1$. For an isotropic material, because the test in the y-direction is exactly the same as the one run earlier in the x-direction, the value of T_{22} must equal the stress T_{11} measured in the first experiment. For this to happen it is necessary that $a = b$, and by extension we have that $a = b = c$. The conclusion that the stress is directionally independent is a consequence of the assumption of form invariance (8.53). It is possible to generalize the formulation and include non-isotropic materials, and an introduction to this can be found in Batra [2002] and Soldatos [2008].

The ideas underlying the Principle of Frame-Indifference are almost universally accepted by those working in continuum mechanics, and the development used here was adapted from the work of Svendsen and Bertram [1999]. This does not mean that all issues related to this principle have been worked out. Most of the attention on this topic is not germane to this textbook, but it is worth providing a glimpse into some of the questions that have arisen. Constitutive laws are macroscopic functions representing the accumulated actions taking place on the atomic scale. This viewpoint was used in Chapter 6 to help explain how the atomic structure of a solid can give rise to the assumed linear law of elasticity. The idea is that it should be possible to derive the constitutive law from more fundamental principles, such as arise in statistical mechanics. This very attractive proposal brings with it a problem, which is that the Newtonian laws of microscopic physics generally do not satisfy the Principle of Frame-Indifference. Given this then how can one expect that the resulting macroscopic constitutive laws obey this principle? The resolution of this problem involves a closer look at the limit taken when moving from the microscopic to macroscopic scale, and this is discussed in the papers by Speziale [1987] and Murdoch [2006]. More general reviews on Frame-Indifference can be found in Speziale [1998] and Frewer [2009].

8.10 Newtonian Fluid

To apply the above theory to the study of fluid motion, the first question to consider is, what exactly is a fluid? It is certainly easy to list specific examples. This includes liquids, such as water and mercury at room temperature, as well as gases, such as air. Two of the central characteristics of fluids is their ability to flow, and their inability to retain a specific shape of their own. The important question for us is, how do we translate these observations into a mathematical formula for the stress?

8.10.1 Pressure

In developing the constitutive law for a solid, one of the first experiments we considered was what happens when the material is compressed. We found that the displacement in such situations was not constant, and this gave rise to the concept of a strain in an elastic solid. The reason a strain is possible within the solid is indicated in Figure 6.9. The forces holding the atoms in the lattice enable the solid to support a variable displacement. This is not possible in a fluid. The reason is that the fluid atoms are farther apart, and are able to move past each other with little difficulty. Under compression they will get closer together, and assuming the fluid has come to rest, they

are all approximately equidistant from each other. This is the situation, for example, that occurs after you have blown up a balloon. The compression does introduce a stress in the fluid, and it is the same in all directions. This is the concept underlying a pressure, and the resulting constitutive law is

$$\mathbf{T} = -p\mathbf{I}, \tag{8.66}$$

where $\mathbf{I}$ is the identity tensor and p is the pressure.

8.10.2 Viscous Stress

What about the stress when the fluid is moving? A hint on how to answer this is obtained from the usual explanation for how to account for air resistance. When modeling the motion of an object in a fluid, such as a ball falling through the air, it is usually assumed that the drag force is proportional to the velocity. The correct way to say this is that it is proportional to the relative velocity between the fluid and object. The idea is that when atoms of the fluid move together, in parallel, there is no relative velocity and therefore no resistance. It is when the atoms move past each other that the resistance force is generated. The resulting constitutive assumption is that the fluid stress depends on the spatial derivative of the fluid velocity. To get an idea of what this entails, all of the various spatial derivatives of $\mathbf{v}$ are collected together in the velocity gradient tensor $\nabla\mathbf{v}$, given in (8.25). The constitutive assumption is, therefore, that each of the six elements of the stress tensor are functions of the nine derivatives in the velocity gradient. This is not a very appealing result. For example, even if we try to make things easy and assume that the dependence is linear we end up with 54 parameters.

To derive a more manageable theory for fluids, we first use $\nabla\mathbf{v}$ to introduce two associated tensors. One is the *rate of deformation tensor*

$$\begin{aligned}\mathbf{D} &= \frac{1}{2}\left(\nabla\mathbf{v} + \nabla\mathbf{v}^T\right)\\ &= \begin{pmatrix} \dfrac{\partial v_1}{\partial x} & \dfrac{1}{2}\left(\dfrac{\partial v_1}{\partial y} + \dfrac{\partial v_2}{\partial x}\right) & \dfrac{1}{2}\left(\dfrac{\partial v_1}{\partial z} + \dfrac{\partial v_3}{\partial x}\right)\\ \dfrac{1}{2}\left(\dfrac{\partial v_1}{\partial y} + \dfrac{\partial v_2}{\partial x}\right) & \dfrac{\partial v_2}{\partial y} & \dfrac{1}{2}\left(\dfrac{\partial v_2}{\partial z} + \dfrac{\partial v_3}{\partial y}\right)\\ \dfrac{1}{2}\left(\dfrac{\partial v_1}{\partial z} + \dfrac{\partial v_3}{\partial x}\right) & \dfrac{1}{2}\left(\dfrac{\partial v_2}{\partial z} + \dfrac{\partial v_3}{\partial y}\right) & \dfrac{\partial v_3}{\partial z}\end{pmatrix},\end{aligned} \tag{8.67}$$

and the other is the vorticity or spin tensor

$$\mathbf{W} = \frac{1}{2}\left(\nabla\mathbf{v} - \nabla\mathbf{v}^T\right)$$
$$= \begin{pmatrix} 0 & \frac{1}{2}\left(\frac{\partial v_1}{\partial y} - \frac{\partial v_2}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right) & 0 & \frac{1}{2}\left(\frac{\partial v_2}{\partial z} - \frac{\partial v_3}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial v_3}{\partial x} - \frac{\partial v_1}{\partial z}\right) & \frac{1}{2}\left(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right) & 0 \end{pmatrix}. \tag{8.68}$$

The most obvious properties of these two tensors are that $\mathbf{D}$ is symmetric, $\mathbf{W}$ is skew-symmetric or antisymmetric, and $\nabla\mathbf{v} = \mathbf{D} + \mathbf{W}$.

8.10.2.1 Reduction of the Viscous Stress Function

In its general form, the fluid stress is assumed to not depend on $\nabla\mathbf{v}$ but, rather, on $\mathbf{D}$ and $\mathbf{W}$. Specifically, the assumption is that

$$\mathbf{T} = -p\mathbf{I} + \mathbf{G}(\mathbf{D}, \mathbf{W}), \tag{8.69}$$

where $\mathbf{G}(\mathbf{0}, \mathbf{0}) = \mathbf{0}$. In what follows, the specific form of the function $\mathbf{G}$ is reduced, using the properties of $\mathbf{T}$ and additional simplifying assumptions. Before doing this, note that we have assumed that $\mathbf{G}$ does not depend explicitly on $\mathbf{x}$. This means we are assuming that the fluid is homogeneous, so the constitutive law for the stress does not depend explicitly on position.

Simplification 1: $\mathbf{T} = -p\mathbf{I} + \mathbf{G}(\mathbf{D})$.

The conclusion that the stress does not depend on $\mathbf{W}$ is not too surprising because $\mathbf{T}$ is symmetric while $\mathbf{W}$ is skew-symmetric. The proof, however, comes from the Principle of Material Frame-Indifference. For the rigid body motion given in (8.49), it is shown in Exercise 8.23 that $\mathbf{D}^* = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$ and $\mathbf{W}^* = \mathbf{Q}\mathbf{W}\mathbf{Q}^T + \boldsymbol{\Omega}$, where $\boldsymbol{\Omega} = \mathbf{Q}'\mathbf{Q}^T$ is skew-symmetric. Therefore, from (8.54), it is required that

$$\mathbf{Q}\mathbf{G}(\mathbf{D}, \mathbf{W})\mathbf{Q}^T = \mathbf{G}(\mathbf{Q}\mathbf{D}\mathbf{Q}^T, \mathbf{Q}\mathbf{W}\mathbf{Q}^T + \boldsymbol{\Omega}). \tag{8.70}$$

To make our point we do not need to consider every rotation, and it is enough to consider those where $\mathbf{Q}(0) = \mathbf{I}$ and $\mathbf{Q}'(0) = \mathbf{M}$ is an arbitrary skew-symmetric matrix. With this, and setting $t = 0$ in (8.70), it follows that

$$\mathbf{G}(\mathbf{D}, \mathbf{W}) = \mathbf{G}(\mathbf{D}, \mathbf{W} + \mathbf{M}),$$

for every skew-symmetric matrix $\mathbf{M}$. The only function capable of this is one that does not depend on $\mathbf{W}$. The dependence of $\mathbf{G}$ on $\mathbf{D}$ is left open, other

than it must satisfy $\mathbf{Q}\mathbf{G}(\mathbf{D})\mathbf{Q}^T = \mathbf{G}(\mathbf{Q}\mathbf{D}\mathbf{Q}^T)$, for all rotations $\mathbf{Q}$.

Simplification 2: $\mathbf{G}(\mathbf{D}) = \alpha_0\mathbf{I} + \alpha_1\mathbf{D} + \alpha_2\mathbf{D}^2$

This result is an immediate consequence of the Rivlin-Ericksen theorem (8.58), because both $\mathbf{T}$ and $\mathbf{D}$ are symmetric and objective (see Exercise 8.23). In this case, the coefficients α_0, α_1, and α_2 have the dimensions of stress and, with one exception, are arbitrary functions of the three principal invariants

$$\begin{aligned} \mathrm{I}_D &= \mathrm{tr}(\mathbf{D}), \\ \mathrm{II}_D &= \frac{1}{2}\left(\mathrm{tr}(\mathbf{D})^2 - \mathrm{tr}(\mathbf{D}^2)\right), \\ \mathrm{III}_D &= \det(\mathbf{D}). \end{aligned}$$

The exception comes from the requirement that $\mathbf{G} = \mathbf{0}$ if $\mathbf{D} = \mathbf{0}$, which means that $\alpha_0 = 0$ if $\mathbf{D} = \mathbf{0}$.

Simplification 3: $\mathbf{G}(\mathbf{D}) = \lambda \mathrm{I}_D\,\mathbf{I} + 2\mu\mathbf{D}$.

This simplification is an assumption, and specifically it is assumed that the stress is a linear function of $\mathbf{D}$, with $\mathbf{G}(\mathbf{0}) = \mathbf{0}$. This means, using the result of Simplification 2, that $\alpha_2 = 0$ and $\alpha_1 = 2\mu$ is a constant. The coefficient α_0 can be a linear function of the elements of $\mathbf{D}$. As is evident in how the invariants depend on the eigenvalues, as given in (8.62)-(8.64), only I_D is linear in $\mathbf{D}$. Therefore, it follows that $\alpha_0 = \lambda \mathrm{I}_D$, where λ is a constant.

After working through the last few pages, the following question might arise. The conclusion of *Simplification 3* is the result of several assumptions, and some mathematical effort. Why not skip all this and simply make one assumption, which is the formula given below in (8.71)? This is, in fact, the approach of more elementary textbooks, and has the advantage of letting you get started solving fluid problems without a lot of preparatory work. The reason for using a sequence of simplification steps is to try to better understand the physical assumptions underlying the formulation of the constitutive law for the stress. Also, the assumptions made here are more fundamental in the sense that they can be used to obtain constitutive laws for other types of materials, and an example of this will come later in the chapter when we apply the theory to elastic solids.

A second comment that should be made concerns an alternative approach for the derivation of the fluid stress. Many textbooks on this subject start with the assumption that $\mathbf{T}$ is a linear function of the elements of $\nabla\mathbf{v}$. This idea was mentioned earlier, right before introducing $\mathbf{D}$ and $\mathbf{W}$. Using this approach, one employs isotropy and frame-indifference to reduce the general

formula for the stress down to the conclusion of *Simplification 3*. The reason for reversing the order here, and keeping the linearity assumption until the end, is to obtain a theory that can be generalized to nonlinear materials. For example, to study isotropic non-Newtonian fluids we have the general result given in *Simplification 2*. This result is, for example, used to derive the nonlinear theory for what are called Reiner-Rivlin fluids. It is also used in the next chapter to describe power-law fluids.

8.11 Equations of Motion for a Viscous Fluid

The conclusion from the previous section is that the constitutive law for a viscous fluid is

$$\mathbf{T} = -p\mathbf{I} + \lambda(\nabla \cdot \mathbf{v})\mathbf{I} + 2\mu\mathbf{D}, \tag{8.71}$$

where p is the pressure and $\mathbf{D}$ is the rate of deformation tensor given in (8.67). This is the constitutive law for what is called a *Newtonian fluid*, which means that the stress is a linear function of the rate of deformation tensor. The coefficients λ and μ are viscosity parameters. In the engineering literature μ is called the dynamic viscosity, or sometimes the shear viscosity.

With this, one finds that

$$\nabla \cdot \mathbf{T} = -\nabla p + \lambda\nabla(\nabla \cdot \mathbf{v}) + \mu\left(\nabla(\nabla \cdot \mathbf{v}) + \nabla^2\mathbf{v}\right), \tag{8.72}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{8.73}$$

is the Laplacian. The resulting equations of motion are

$$\frac{D\rho}{Dt} + \rho\nabla \cdot \mathbf{v} = 0, \tag{8.74}$$

$$\rho\frac{D\mathbf{v}}{Dt} = -\nabla p + (\lambda + \mu)\nabla(\nabla \cdot \mathbf{v}) + \mu\nabla^2\mathbf{v} + \rho\mathbf{f}. \tag{8.75}$$

The above momentum equation is known as the *Navier-Stokes equation*.

Looking at the equations in (8.74), (8.75) you soon realize that something is missing. Namely, there are four equations, but five unknowns (ρ, p, and $\mathbf{v}$). What is needed is a second constitutive law, an equation of state, that relates the pressure and density. Commonly used examples are the ideal gas law $p = \rho RT$, and the van der Waals equation

$$p = \rho RT\left(\frac{1}{1 - \beta\rho} - \frac{\alpha\rho}{RT}\right).$$

Both of these expressions contain the temperature T. For isothermal flows this is assumed constant, but if this is not the case then it is necessary to

derive a balance law for the energy. This was considered earlier, in Section 6.10.2, for one-dimensional motion, but will not be considered here.

A couple of comments are in order related to the derivation of the above equations of motion. First, it is important to note that a Newtonian fluid is an assumption of material linearity, and not an assumption of geometric linearity. The domains over which the Newtonian fluid equations are applicable can be, and are routinely are, highly variable. A second comment is that there are different ways to derive the constitutive law for a Newtonian fluid. As an example of a different approach, it is possible to obtain (8.71) using the Principle of Dissipation. This was the method used in Section 6.10.2, to derive the one-dimensional version of (8.71).

8.11.1 Incompressibility

If the fluid is assumed to be incompressible then the constitutive law for the stress is

$$\mathbf{T} = -p\mathbf{I} + 2\mu\mathbf{D}. \tag{8.76}$$

The equations of motion in this case reduce to the following

$$\rho\frac{D\mathbf{v}}{Dt} = -\nabla p + \mu\nabla^2\mathbf{v} + \rho\mathbf{f}, \tag{8.77}$$

$$\nabla\cdot\mathbf{v} = 0. \tag{8.78}$$

Assuming that the initial density is constant, then ρ is known, and it is constant. In this case the number of equations matches the number of unknowns (p and $\mathbf{v}$), and so an equation of state is not needed. Also, these equations appear to be somewhat simpler than the compressible versions in (8.74), (8.75). Although this may be true, both versions are formidable mathematical problems.

Given our interest in solving (8.77), (8.78), it is worth spending a moment considering what sort of mathematical problem we are facing. In terms of the velocity, (8.77) is a first-order equation in time and a second-order equation in space. In this sense it is the same as the diffusion equations studied in Chapter 4, and it should not be surprising to find that the kinematic viscosity $\nu = \mu/\rho$ has the same dimensions as a diffusion coefficient. One of the distinctive differences from a diffusion equation is the nonlinear term $(\mathbf{v}\cdot\nabla)\mathbf{v}$ hiding in the material derivative. This term is the type of nonlinearity we studied in traffic flow. In fact, in terms of its mathematical characteristics, you could say that (8.77), (8.78) is Chapter 4 meets Chapter 7. The nonlinearity, however, means that transform methods, both Fourier and Laplace, will not work. Also, the presence of the viscosity term means that the method of characteristics will not work. Aside from a numerical solution, this leaves similarity methods, perturbation methods, and guessing. We will make heavy use of guessing, and

this requires a well formulated mathematical problem, which means that we need boundary conditions.

Before moving on, it is worth commenting on the earlier statement about the formidability of solving the Navier-Stokes equation. Finding the solution is considered to be one of the greatest unsolved mathematical problems of our time, and this is the reason that it was included as one of the Millennium Prize Problems (Devlin [2002]). The person, or team, that first solves this problem will be awarded $1,000,000 (US).

8.11.2 Boundary and Initial Conditions

To be able to find the solution of the fluid equations it is necessary to know the boundary conditions. Of interest here are the conditions at a solid boundary, which for us will usually be the container holding the fluid. Although it is often the case that such boundaries are stationary, we will include the possibility that it moves. An example of such a situation arises with a water balloon, where the boundary, the balloon, does not just move, it also deforms.

In the following, let S be the boundary surface, and assume that the points on the boundary have a known velocity $\mathbf{v}_s(\mathbf{x}, t)$.

Impermeability Condition

The boundary is solid, and this means that the fluid can not flow through it. To translate this into a boundary condition, let $\mathbf{n}$ be the unit outward normal to the surface. With this, $\mathbf{v} \cdot \mathbf{n}$ is the velocity of the fluid in the normal direction, and $\mathbf{v}_s \cdot \mathbf{n}$ is the velocity of the surface in the normal direction. The boundary condition is one of continuity, namely that on S the normal velocity of the fluid is equal to the normal velocity of the surface. Therefore, the mathematical consequence of impermeability is that

$$\mathbf{v} \cdot \mathbf{n} = \mathbf{v}_s \cdot \mathbf{n} \quad \text{on } S. \tag{8.79}$$

If the fluid is incompressible, and S is the boundary of a bounded domain, then $\mathbf{v}_s$ must be consistent with the incompressibility assumption (see Exercise 8.14). Also, it should be pointed out that it is implicitly assumed here that the fluid does not separate from the boundary. There are situations where separation occurs, such as in cavitation, and this often results in a very challenging mathematical problem. For obvious reasons, we will avoid such situations in this introductory presentation.

No-Slip Condition

Because of the viscosity, it is assumed that the fluid sticks to the boundary. This means that the fluid velocity, on the boundary, equals the velocity of the boundary. The corresponding boundary condition is

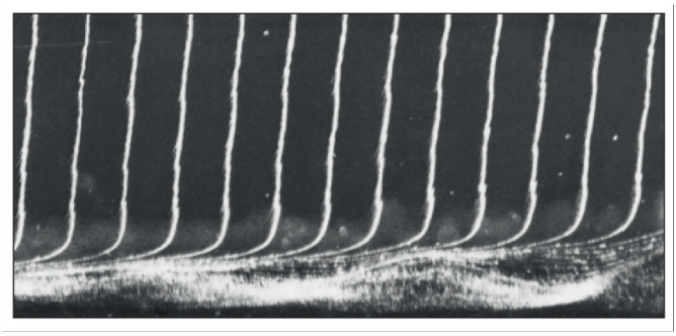


Figure 8.7 Fluid flow over a flat plate illustrating the no-slip boundary condition. The fluid is moving from left to right, and the plate is at the bottom and is not moving. The white curves are indicators of the fluid particles moving with the flow, which show a rapid transition from zero velocity on the plate, to the constant velocity in the upper region.

$$\mathbf{v} = \mathbf{v}_s \quad \text{on } S. \tag{8.80}$$

This is known as the no-slip condition. It means that not only the normal velocities are equal, as required by the impermeability condition (8.79), it also means that the tangential velocities are equal. In many fluid problems the boundary does not move, and in those cases the no-slip condition is simply $\mathbf{v} = \mathbf{0}$ on S. An illustration of this situation is shown in Figure 8.7.

It is common to have situations where the pressure is prescribed on the boundary. Rather than attempt to write down a general formulation of such situations, it is informative to consider an example.

Example: Flow in a Pipe

Most people, when attempting to drink using a straw, create a pressure difference between the two ends of the straw. This is the same idea used to have water flow through a hose or pipe. To formulate the boundary conditions for

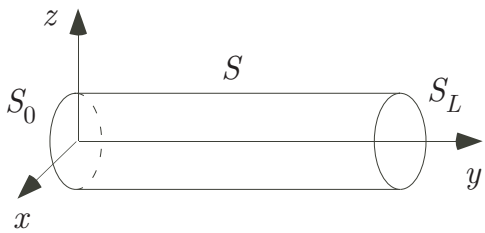


Figure 8.8 Geometry of pipe used in the boundary condition example.

such a situation, consider the straight pipe shown in Figure 8.8. The pipe is fixed, and water is flowing through the pipe due to a pressure difference between the two ends. From the no-slip condition we have that $\mathbf{v} = \mathbf{0}$ on S, which is the wall of the pipe. At the two ends, S_0 and S_L, inflow/outflow conditions are used. This means that instead of prescribing the three components of the velocity, we will prescribe its two tangential components and the pressure. Letting $\mathbf{v} = (u, v, w)$ then on S_0 we take $u = w = 0$ and $p = p_0$, while at S_L we take $u = w = 0$ and $p = p_L$. These boundary conditions, along with the equations of motion, form what is known as the Poiseuille flow problem, and the solution in the case of steady flow is derived in Section 9.2.2. ■

The usual initial condition used for incompressible fluid problems is simply

$$\mathbf{v}(\mathbf{x}, 0) = \mathbf{v}_0(\mathbf{x}), \tag{8.81}$$

where $\mathbf{v}_0(\mathbf{x})$ is given. Not just any function can be used here. In particular, it must be consistent with incompressibility, and therefore it is required that $\mathbf{v}_0(\mathbf{x})$ satisfies (8.78). A consistency requirement can also come from the boundary conditions. For example, at a solid surface the impermeability condition (8.79) must be satisfied. This means that $\mathbf{v}_0 \cdot \mathbf{n} = \mathbf{v}_s(\mathbf{x}, 0) \cdot \mathbf{n}$ on S. To obtain a well-posed mathematical problem it is not necessary that the tangential components of $\mathbf{v}_0(\mathbf{x})$ satisfy the no-slip condition. A more complete discussion of the various boundary and initial conditions that can be used to obtain a well-posed mathematical problem involving the Navier-Stokes equation can be found in Temam [2001].

Boundary conditions are of supreme importance in the formulation of any physical problem. This is brought up because the equations of motion have been derived from fundamental physical principles. The boundary conditions, in contrast, give the appearance of being tacked on using plausibility arguments why they should be used. The no-slip condition is an example of this. Make no mistake, it is almost universally used for viscous fluid problems. However, as a budding applied mathematician, you should be skeptical of this situation. One question you might ask is, can it be derived from more fundamental physical principles? Even more important, are there situations where it should not be used? These are difficult questions, and in the early development of fluid dynamics they were controversial topics. Eventually, based on the available experimental evidence, the no-slip condition became the accepted requirement. These questions, however, have started to be asked again. This is due to better experimental methods, and the application of the Navier-Stokes equations to small-scale systems where the no-slip condition is questionable. It should be pointed out that the questions apply to the tangential component of the no-slip condition. The normal component, which is the impermeability condition (8.80), can be derived from the continuity equation (Hutter and Johnk [2004]), and is not in question. Those interested

in the no-slip condition, and its limitations, should consult the review by Lauga et al. [2007].

8.12 Material Equations of Motion

Because elastic solids have the ability to hold their shape, they have a natural reference configuration. For this reason, the material coordinate system is more often used in elasticity. There are a couple of options here for how to determine the material version of the equations of motion. One is to use the chain rule, and convert the spatial derivatives in (8.46) and (8.47) into material derivatives. This is rather tedious, and not very enlightening. Another approach is to derive the equations from the material form of the balance laws, and this is the one used here.

The more interesting equation to derive is the one for linear momentum. To state this result, let B be a volume of material points with boundary surface ∂B. When using spatial coordinates this volume was designated as $R(0)$. Using the balance of linear momentum,

$$\frac{d}{dt}\iiint\limits_{B} R\mathbf{V}\,dV_A = \iint\limits_{\partial B} \bar{\mathbf{t}}\,dS_A + \iiint\limits_{B} R\bar{\mathbf{F}}\,dV_A. \tag{8.82}$$

This equation is simply the material version of (8.37). In this equation $R(\mathbf{A}, t)$ is the density in material coordinates, $\mathbf{V}(\mathbf{A}, t)$ is the velocity, $\bar{\mathbf{t}}(\mathbf{A}, t)$ is the force, per unit area, on B due the material exterior to B, and $\bar{\mathbf{F}}(\mathbf{A}, t)$ is the external body force in material coordinates. The subscript A on the volume and surface elements is to indicate that the integration is with respect to material coordinates. For example, if $dV = dxdydz$ then $dV_A = dA_1 dA_2 dA_3$.

Using exactly the same type of analysis that led to (8.39), one can show that

$$\bar{\mathbf{t}} = \mathbf{P}^T\mathbf{N}, \tag{8.83}$$

where

$$\mathbf{P} = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \tag{8.84}$$

is known as the first Piola-Kirchhoff stress tensor and $\mathbf{N}$ is the unit outward normal to ∂B. Substituting this into (8.82), and using the fact that this holds for all material volumes B we conclude that

$$R_0\frac{\partial^2\mathbf{U}}{\partial t^2} = \nabla_A \cdot \mathbf{P} + R_0\bar{\mathbf{F}}. \tag{8.85}$$

This is the momentum equation in material coordinates. The $\nabla_A \cdot \mathbf{P}$ term is similar to (8.43), except the derivatives are with respect to the elements of $\mathbf{A}$ instead of $\mathbf{x}$. In simplifying the above result the material version of the continuity equation has been used, which is

$$R = \frac{R_0}{\det(\mathbf{F})}\,, \tag{8.86}$$

where $\mathbf{F}$ is given in (8.10) and $R_0 = R(\mathbf{A}, 0)$ is the initial value for the density. Finally, the angular momentum equation in material coordinates is

$$\mathbf{P}^T\mathbf{F}^T = \mathbf{F}\mathbf{P}. \tag{8.87}$$

The derivation of this result, and the continuity equation, is left as an exercise.

The derivation of the linear momentum equation looks to be a replay of the spatial coordinate analysis, and this is correct. What is left is the more interesting step, and that is to relate the stress tensors $\mathbf{T}$ and $\mathbf{P}$. Because they consist of the stresses on three orthogonal coordinate surfaces, and the material is undergoing deformation, $\mathbf{T}$ and $\mathbf{P}$ are not necessarily equal. In fact, to jump ahead a bit, we will find that

$$\mathbf{P} = \det(\mathbf{F})\mathbf{F}^{-1}\mathbf{T}\,. \tag{8.88}$$

To derive this result, suppose that given a material point $\mathbf{A}_0$, its spatial counterpart is $\mathbf{x}_0 = \mathbf{X}(\mathbf{A}_0, t)$. Assuming $\mathbf{A}_0$ is on the surface ∂B then $\mathbf{x}_0$ is on the surface $\partial R(t)$. The definition of the stress vector $\bar{\mathbf{t}}$ uses the force on a small piece of the tangent plane at $\mathbf{A}_0$, and then lets this region shrink to zero. This is the same idea employed earlier when using the tetrahedron, shown in Figure 8.6, to define the stress vector in spatial coordinates. The difference here is that it is easier to use pieces of the tangent plane shaped as parallelograms instead of triangles. To construct the parallelogram let $\mathbf{A}_1$ and $\mathbf{A}_2$ be two points close to $\mathbf{A}_0$, and in the tangent plane. This means that $\mathbf{A}_i = \mathbf{A}_0 + \Delta\mathbf{A}_i$. The corresponding points in the spatial system are determined using (8.9), and the result is $\mathbf{x}_i \approx \mathbf{x}_0 + \mathbf{F}\Delta\mathbf{A}_i$, where $\mathbf{F}$ is the Jacobian matrix for $\mathbf{X}$, evaluated at $\mathbf{A}_0$. Now, the cross-product $(\mathbf{A}_2 - \mathbf{A}_0) \times (\mathbf{A}_1 - \mathbf{A}_0) = \Delta\mathbf{A}_2 \times \Delta\mathbf{A}_1$ determines the normal direction to the surface, and its length gives the area of the parallelogram. To determine the corresponding information for the spatial system, we use the vector $(\mathbf{x}_2 - \mathbf{x}_0) \times (\mathbf{x}_1 - \mathbf{x}_0) \approx (\mathbf{F}\Delta\mathbf{A}_2) \times (\mathbf{F}\Delta\mathbf{A}_1)$. We are going to compare the areas in these two coordinate systems, and for this we need the vector identity

$$\mathbf{B}^T(\mathbf{B}\mathbf{y} \times \mathbf{B}\mathbf{z}) = \det(\mathbf{B})(\mathbf{y} \times \mathbf{z}).$$

Setting $\Delta\mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_0$, then we have shown that

$$\begin{aligned}\Delta\mathbf{x}_2 \times \Delta\mathbf{x}_1 &\approx (\mathbf{F}\Delta\mathbf{A}_2) \times (\mathbf{F}\Delta\mathbf{A}_1) \\ &= \det(\mathbf{F})\mathbf{F}^{-T}(\Delta\mathbf{A}_2 \times \Delta\mathbf{A}_1).\end{aligned}$$

Letting the area of the parallelogram in the spatial system be ΔS, and letting ΔS_A be the area in the material system, then the above equation can be written as $\Delta S\,\mathbf{n} \approx (\Delta S_A)\det(\mathbf{F})\mathbf{F}^{-T}\mathbf{N}$. Taking the limit of $\Delta\mathbf{A}_i \to \mathbf{0}$ we obtain

$$\mathbf{n}dS = \det(\mathbf{F})\mathbf{F}^{-T}\mathbf{N}dS_A, \tag{8.89}$$

which is known as Nanson's formula. What this means is that, when changing from spatial to material coordinates in a surface integral, the following holds

$$\iint\limits_{\partial R(t)} \mathbf{T}^T\mathbf{n}\,dS = \iint\limits_{\partial B} \det(\mathbf{F})\mathbf{T}^T\mathbf{F}^{-T}\mathbf{N}\,dS_A.$$

The consequence of this is that expressing $\mathbf{T}^T\mathbf{n}$ in terms of material coordinates we obtain $\det(\mathbf{F})\mathbf{T}^T\mathbf{F}^{-T}\mathbf{N}$. Given that this equals $\mathbf{P}^T\mathbf{N}$, for all material regions B, it follows that $\mathbf{P}^T = \det(\mathbf{F})\mathbf{T}^T\mathbf{F}^{-T}$. Taking the transpose of this equation yields (8.88).

The equations of motion in material coordinates are given in (8.85)-(8.87). There is nothing particularly unusual about this system of equations. For example, as in the one-dimensional version given in Chapter 6, it is not necessary to solve a differential equation to find the density when using material coordinates. This is because of (8.86). The one new twist that arises in the material version is that the stress tensor $\mathbf{P}$ is not necessarily symmetric, but satisfies (8.87) instead. As will be seen shortly, this nonsymmetry does complicate the formulation of the constitutive law for the stress.

One last comment that needs to be made concerns naming conventions for the stress. Earlier it was mentioned that some authors define the Cauchy stress to be the transpose of the formula in (8.40). For the same reasons, they define the first Piola-Kirchhoff stress tensor to be the transpose of the formula given in (8.84), and then refer to (8.84) as the nominal stress tensor. This difference is of little consequence when using spatial coordinates, but the stress tensor in the material system is not symmetric so this difference is more of an issue.

8.12.1 Frame-Indifference

Before formulating constitutive laws in material coordinates, it is first necessary to understand how the Principle of Frame-Indifference applies. For one-dimensional motion, studied in Section 6.10, a constitutive law in material coordinates is frame-invariant if its spatial counterpart is frame-invariant. This statement also holds for the three-dimensional case we are now studying.

However, rather than doing this on a case-by-case basis, it is easier to simply determine the material form of the two tenets that make up the Principle of Frame-Indifference, as given in Section 8.9.

The rigid body change of coordinates used for the material version of frame-indifference is given in (8.50). With this, the form invariance assumption, Tenet 2, is unaffected. In particular, if the constitutive law is that $\mathbf{P} = \mathbf{G}(\mathbf{R})$, and if $\mathbf{P}^*$ and $\mathbf{R}^*$ are the $\mathbf{A}^*$ system versions of $\mathbf{P}$ and $\mathbf{R}$, then it must be that $\mathbf{P}^* = \mathbf{G}(\mathbf{R}^*)$.

The objectivity condition, Tenet 1, is affected by the change in coordinates. To determine how, recall that $\mathbf{T}^* = \mathbf{Q}\mathbf{T}\mathbf{Q}^T$ and $\mathbf{F}^* = \mathbf{Q}\mathbf{F}$. So, from (8.88),

$$\begin{aligned}
\mathbf{P}^* &= \det(\mathbf{F}^*)(\mathbf{F}^*)^{-1}\mathbf{T}^* \\
&= \det(\mathbf{Q}\mathbf{F})(\mathbf{Q}\mathbf{F})^{-1}\mathbf{Q}\mathbf{T}\mathbf{Q}^T \\
&= \det(\mathbf{Q})\det(\mathbf{F})\mathbf{F}^{-1}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{T}\mathbf{Q}^T \\
&= \det(\mathbf{F})\mathbf{F}^{-1}\mathbf{T}\mathbf{Q}^T \\
&= \mathbf{P}\mathbf{Q}^T. \qquad (8.90)
\end{aligned}$$

Therefore, $\mathbf{P}$ satisfies the material form of objectivity if $\mathbf{P}^* = \mathbf{P}\mathbf{Q}^T$, for all rotations $\mathbf{Q}$.

The result of the above discussion is that if the constitutive law is $\mathbf{P} = \mathbf{G}(\mathbf{R})$, then the Principle of Material Frame-Indifference, when using material coordinates, requires that

$$\mathbf{G}(\mathbf{R})\mathbf{Q}^T = \mathbf{G}(\mathbf{R}^*). \qquad (8.91)$$

This is the material coordinate version of (8.54).

Example

Suppose it is assumed that $\mathbf{P} = \mathbf{G}(\mathbf{U})$. It is not hard to show that $\mathbf{U}^* = \mathbf{Q}\mathbf{U} + (\mathbf{Q} - \mathbf{Q}_0)\mathbf{A} + \mathbf{b} - \mathbf{b}_0$, so from (8.91) it is required that

$$\mathbf{G}(\mathbf{U})\mathbf{Q}^T = \mathbf{G}(\mathbf{Q}\mathbf{U} + (\mathbf{Q} - \mathbf{Q}_0)\mathbf{A} + \mathbf{b} - \mathbf{b}_0).$$

This must hold for every rotation $\mathbf{Q}$, and vectors $\mathbf{U}$ and $\mathbf{b}$. In particular, taking $\mathbf{Q} = \mathbf{I}$ and $\mathbf{U} = \mathbf{0}$, then $\mathbf{G}(\mathbf{0}) = \mathbf{G}(\mathbf{b} - \mathbf{b}_0)$ for all vectors $\mathbf{b}$. The only function capable of this is the constant function, and so the stress must be independent of the displacement. This is the same conclusion reached in Section 6.7, and in Section 6.10, for one-dimensional motion. ■

8.12.2 Elastic Solid

The stress in an elastic solid is determined by the strain, and the reasons for this assumption were discussed in Chapter 6. The strain in this case being determined using $\frac{\partial U}{\partial A}$. The three-dimensional version of this is

$$\nabla_A \mathbf{U} = \begin{pmatrix} \dfrac{\partial U_1}{\partial A_1} & \dfrac{\partial U_1}{\partial A_2} & \dfrac{\partial U_1}{\partial A_3} \\ \dfrac{\partial U_2}{\partial A_1} & \dfrac{\partial U_2}{\partial A_2} & \dfrac{\partial U_2}{\partial A_3} \\ \dfrac{\partial U_3}{\partial A_1} & \dfrac{\partial U_3}{\partial A_2} & \dfrac{\partial U_3}{\partial A_3} \end{pmatrix}. \tag{8.92}$$

This is known as the displacement gradient. It is related to the deformation gradient, given in (8.10), through the identity $\nabla_A \mathbf{U} = \mathbf{F} - \mathbf{I}$.

An elastic solid is one for which the stress is a function of $\nabla_A \mathbf{U}$, or equivalently, of $\mathbf{F}$. It is more traditional to work with $\mathbf{F}$, and so, the constitutive assumption is that

$$\mathbf{P} = \mathbf{G}(\mathbf{F}). \tag{8.93}$$

To satisfy the Principle of Material Frame-Indifference (8.91), from (8.56), the function $\mathbf{G}$ must satisfy

$$\mathbf{G}(\mathbf{F})\mathbf{Q}^T = \mathbf{G}(\mathbf{QF}), \tag{8.94}$$

for all possible values for $\mathbf{F}$ and rotations $\mathbf{Q}$. One immediate conclusion is that it is not possible to have a linear constitutive law. For example, it is not possible to assume that $\mathbf{G} = \alpha\mathbf{F}$, because (8.94) requires

$$\mathbf{FQ}^T = \mathbf{QF}.$$

If this were to hold, then taking $\mathbf{F} = \mathbf{I}$ we would conclude that $\mathbf{Q}$ is symmetric. This clearly does not have to happen, as demonstrated by the rotation in (8.14).

To take stock of our situation, we have a stress that is not symmetric, a form of objectivity that does not fit the requirement of the Rivlin-Ericksen theorem, and a constitutive law that cannot be linear. One way to avoid these difficulties is to revert back to spatial coordinates, but this will not be done as material coordinates is a more natural system for elastic solids. There are a couple of ways to handle these difficulties, and the one we will use involves the introduction of the second Piola-Kirchhoff stress tensor, defined as

$$\mathbf{S} = \mathbf{PF}^{-T}, \tag{8.95}$$

or equivalently

$$\mathbf{S} = \det(\mathbf{F})\mathbf{F}^{-1}\mathbf{TF}^{-T}. \tag{8.96}$$

Because $\mathbf{T}$ is symmetric it follows that $\mathbf{S}$ is symmetric. Also, the requirement that $\mathbf{P}$ is materially objective is equivalent to the requirement that $\mathbf{S}^* = \mathbf{S}$, where the superscript $*$ indicates the value of the variable using the change of variables in (8.49). There is still a problem with linearity. For example, it not possible to assume $\mathbf{S} = \alpha\mathbf{F}$ or $\mathbf{S} = \alpha(\mathbf{F} + \mathbf{F}^T)$, because neither satisfies $\mathbf{S}^* = \mathbf{S}$. The stress $\mathbf{S}$ is assumed to depend on $\mathbf{F}$, but this dependence is through a quantity $\mathbf{C}$ that has the same properties as $\mathbf{S}$. Specifically, $\mathbf{C}$ is symmetric and $\mathbf{C}^* = \mathbf{C}$. The one used in elasticity is

$$\mathbf{C} = \mathbf{F}^T\mathbf{F}, \tag{8.97}$$

which is known as the right Cauchy-Green deformation tensor. A related quantity is

$$\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{I}), \tag{8.98}$$

which is known as the *Green strain tensor*. It is the three-dimensional version of the Green strain listed in Table 6.3.

A few quick comments about $\mathbf{E}$ are in order. Given (8.98), assuming $\mathbf{S}$ depends on $\mathbf{C}$ is equivalent to assuming it depends on $\mathbf{E}$. The explanation of why $\mathbf{E}$ is a measure of strain was given in Section 6.8. Instead of revisiting that discussion, we note two important properties that are required of all strain measures. One is that if there is no deformation, so $\mathbf{U} = \mathbf{0}$, then $\mathbf{E} = \mathbf{0}$. A second property comes from the general observation that for rigid body motion there is no relative deformation, and the strain is therefore zero. It is not hard to check that if $\mathbf{X}(\mathbf{A}, t) = \mathbf{Q}(t)\mathbf{A} + \mathbf{b}(t)$, where $\mathbf{Q}(0) = \mathbf{I}$ and $\mathbf{b}(0) = \mathbf{0}$, then $\mathbf{E} = \mathbf{0}$ (see Exercise 8.17).

The constitutive assumption is that $\mathbf{S} = \mathbf{G}(\mathbf{E})$. It is assumed that the material is stress free at $t = 0$. Given that $\mathbf{F} = \mathbf{I}$ at $t = 0$ then we require that $\mathbf{G}(\mathbf{0}) = \mathbf{0}$. We are not able to use the Rivlin-Ericksen theorem because $\mathbf{S}$ does not satisfy the form of objectivity required. Nevertheless, we can use what we learned using that theorem when modeling a viscous fluid. We are interested in a linear constitutive law, and in analogy with (8.71), it is assumed that

$$\mathbf{S} = \lambda_s \mathrm{I}_E \mathbf{I} + 2\mu_s \mathbf{E}, \tag{8.99}$$

where λ_s and μ_s are constants, and $\mathrm{I}_E = \mathrm{tr}(\mathbf{E})$. This expression satisfies the form invariance requirement, and it also produces a stress that satisfies $\mathbf{S}^* = \mathbf{S}$. The corresponding constitutive law for $\mathbf{P}$ is

$$\mathbf{P} = \lambda_s \mathrm{I}_E \mathbf{F}^T + 2\mu_s \mathbf{E}\mathbf{F}^T. \tag{8.100}$$

Using the terminology of Section 4.7, (8.99) is an assumption of material linearity. Geometric linearity has not been assumed, and this brings us to the next topic.

8.12.3 Linear Elasticity

The constitutive law for the stress $\mathbf{S}$ in (8.99) is based on the assumption of material linearity, or more specifically, on the assumption that the stress is a linear function of $\mathbf{E}$. We are going to assume that the motion is also geometrically linear. This means that the displacements are small enough that we are able to linearize the problem. For example, because $\mathbf{F} = \mathbf{I} + \nabla_A \mathbf{U}$, then

$$\begin{aligned}\mathbf{C} &= (\mathbf{I} + \nabla_A \mathbf{U})^T)(\mathbf{I} + \nabla_A \mathbf{U}) \\ &= \mathbf{I} + \nabla_A \mathbf{U} + (\nabla_A \mathbf{U})^T + (\nabla_A \mathbf{U})^T (\nabla_A \mathbf{U}) \\ &\approx \mathbf{I} + \nabla_A \mathbf{U} + (\nabla_A \mathbf{U})^T. \end{aligned} \tag{8.101}$$

With this we have that

$$\begin{aligned}\mathbf{E} &= \frac{1}{2}(\mathbf{C} - \mathbf{I}) \\ &\approx \frac{1}{2}(\nabla_A \mathbf{U} + (\nabla_A \mathbf{U})^T), \end{aligned} \tag{8.102}$$

and $\mathrm{I}_E \approx \nabla_A \cdot \mathbf{U}$. Also, from (8.100),

$$\begin{aligned}\mathbf{P} &\approx \left[\lambda_s (\nabla_A \cdot \mathbf{U})\mathbf{I} + \mu_s (\nabla_A \mathbf{U} + (\nabla_A \mathbf{U})^T)\right] (\mathbf{I} + \nabla_A \mathbf{U}^T) \\ &\approx \lambda_s (\nabla_A \cdot \mathbf{U})\mathbf{I} + \mu_s (\nabla_A \mathbf{U} + (\nabla_A \mathbf{U})^T). \end{aligned} \tag{8.103}$$

This can be rewritten as

$$\mathbf{P} = \lambda_s (\nabla_A \cdot \mathbf{U})\mathbf{I} + 2\mu_s \mathbf{E}_0, \tag{8.104}$$

where

$$\mathbf{E}_0 = \frac{1}{2}\left(\nabla_A \mathbf{U} + \nabla_A \mathbf{U}^T\right). \tag{8.105}$$

This is the constitutive law for linear elasticity, and it is the three-dimensional version of (6.50). The coefficients λ_s and μ_s are called the Lamè constants, and in the engineering literature μ_s is referred to as the shear modulus. Also, $\mathbf{E}_0$ is the linearized Green strain tensor, or what is identified as the Lagrangian strain in Table 6.3.

With (8.104) the equation of motion given in (8.85) reduces to

$$R_0 \frac{\partial^2 \mathbf{U}}{\partial t^2} = (\lambda_s + \mu_s)\nabla(\nabla \cdot \mathbf{U}) + \mu_s \nabla^2 \mathbf{U} + R_0 \bar{\mathbf{F}}. \tag{8.106}$$

This is known as the Navier equation.

There are striking similarities between the constitutive laws, and equations of motion, for fluids and elastic solids. For example, the constitutive law for a linearly elastic solid given in (8.104) is very similar to the fluid law in (8.71),

and the Navier equations (8.106) are similar to the Navier-Stokes equations (8.75). There are also differences, and one of the more obvious ones is that elasticity uses the displacement gradient while fluids use the velocity gradient in the formulation of the constitutive law for the stress. However, a perhaps more subtle difference is that the Navier-Stokes equations are obtained using an assumption of material linearity, while the Navier equations require both material and geometric linearity.

8.13 Energy Equation

One of the central components in characterizing a mechanical system is the energy. This was discussed in Chapter 6, and in the process a variety of new variables were introduced. A different tack is taken here, and we will derive the energy formulation from what we already know. The key player is the momentum equation (8.47). Taking the dot product with the velocity, and rearranging the terms one obtains the equation

$$\frac{1}{2}\rho\frac{D}{Dt}(\mathbf{v}\cdot\mathbf{v}) = \nabla\cdot(\mathbf{T}\mathbf{v}) - \mathrm{tr}(\mathbf{T}\mathbf{D}) + \rho\mathbf{v}\cdot\mathbf{f}. \tag{8.107}$$

This is known as the mechanical energy equation. Using the continuity equation (8.46), (8.107) can be rewritten as

$$\frac{D}{Dt}\left(\frac{1}{2}\rho\mathbf{v}\cdot\mathbf{v}\right) + \left(\frac{1}{2}\rho\mathbf{v}\cdot\mathbf{v}\right)(\nabla\cdot\mathbf{v}) = \nabla\cdot(\mathbf{T}\mathbf{v}) - \mathrm{tr}(\mathbf{T}\mathbf{D}) + \rho\mathbf{v}\cdot\mathbf{f}. \tag{8.108}$$

This has the form of the general balance law given in (8.31), where $f = \frac{1}{2}\rho\mathbf{v}\cdot\mathbf{v}$ is the kinetic energy density. To express this in integral form, from (8.30) we have that

$$\frac{d}{dt}\iiint\limits_{R(t)}\frac{1}{2}\rho\mathbf{v}\cdot\mathbf{v}\,dV = \iint\limits_{\partial R(t)}(\mathbf{T}\mathbf{v})\cdot\mathbf{n}\,dS + \iiint\limits_{R(t)}\left[-\mathrm{tr}(\mathbf{T}\mathbf{D}) + \rho\mathbf{v}\cdot\mathbf{f}\right]dV. \tag{8.109}$$

This shows that the rate of change of the kinetic energy is due to three contributions. The first term on the right, the surface integral, is the rate of work done by surface forces. Given the form of the general balance law in (8.30), this can be interpreted as an energy flux term, with $\mathbf{J} = -\mathbf{T}\mathbf{v}$. In a similar manner, the integral of $\rho\mathbf{v}\cdot\mathbf{f}$ is the rate of work done by the body forces. This leaves the integral involving $-\mathrm{tr}(\mathbf{T}\mathbf{D})$. Without the constitutive law for the stress it is not obvious how to interpret this term, and usually in continuum mechanics it is given the rather vague name of the stress power. In this regard it is stated to be rate of work by the stress per unit volume.

What we have in (8.108) and (8.109) are energy balance equations. Usually when considering energy there is a term for the kinetic energy, and another

for the potential energy. The kinetic term we have. If there is a contribution of the potential energy it is hidden in either the stress power or the external forcing terms. It is the stress power that is of interest because it is unclear at the moment what this is, and so it is assumed that there are no body forces.

8.13.1 Incompressible Viscous Fluid

Assuming that $\mathbf{T} = -p\mathbf{I} + 2\mu\mathbf{D}$, and $\nabla \cdot \mathbf{v} = 0$, then

$$\begin{aligned}\operatorname{tr}(\mathbf{TD}) &= \operatorname{tr}\left((-p\mathbf{I} + 2\mu\mathbf{D})\mathbf{D}\right)\\ &= -p\ \operatorname{tr}(\mathbf{D}) + 2\mu\ \operatorname{tr}(\mathbf{DD})\\ &= 2\mu\ \operatorname{tr}(\mathbf{D}^2).\end{aligned}$$

In fluid mechanics it is conventional to set

$$\Phi = 2\mu\ \operatorname{tr}(\mathbf{D}^2), \tag{8.110}$$

which is known as the viscous dissipation function. With this, (8.109) becomes

$$\frac{d}{dt}\iiint\limits_{R(t)} \frac{1}{2}\rho\mathbf{v}\cdot\mathbf{v}\,dV = \iint\limits_{\partial R(t)} (\mathbf{Tv})\cdot\mathbf{n}\,dS - \iiint\limits_{R(t)} \Phi\,dV. \tag{8.111}$$

Given that $\Phi \geq 0$, then the above expression shows that the kinetic energy decreases due to this term. Physically, this means that the stress power in this particular example accounts for the loss of energy due to viscosity. For this reason, a viscous fluid does not conserve energy. As a final comment, note that for an incompressible viscous fluid, with no body forces, there is no potential energy term in the energy equation.

8.13.2 Elasticity

To apply the above arguments to an elastic material we first need to derive the energy equation in material coordinates. Proceeding in a similar manner as before, we take the dot product of the momentum equation (8.85) with the velocity. Remembering our earlier assumption that there are no body forces, the result is

$$\frac{\partial}{\partial t}\left(\frac{1}{2}R_0\mathbf{V}\cdot\mathbf{V}\right) = \nabla_A\cdot(\mathbf{PV}) - \operatorname{tr}(\mathbf{P}\nabla_A\mathbf{V}). \tag{8.112}$$

The above equation states that the rate of change in the kinetic energy density is balanced by the energy flux and the negative of the stress power $\mathrm{tr}(\mathbf{P}\nabla_A\mathbf{V})$. It is the latter term that needs to be sorted out, and this requires the constitutive law for the stress.

To express the stress power in terms that make its contribution more evident we need to first derive a few identities involving the derivative of the trace. The first step is finding the derivative of a product of tensors

$$\begin{aligned}\frac{\partial}{\partial t}\left(\mathbf{P}^2\right) &= \frac{\partial}{\partial t}\left(\mathbf{P}\mathbf{P}\right)\\ &= \frac{\partial \mathbf{P}}{\partial t}\mathbf{P} + \mathbf{P}\frac{\partial \mathbf{P}}{\partial t}\,.\end{aligned}$$

This shows the usual power rule does not apply to tensors, but in taking the trace of the above expression we get

$$\begin{aligned}\frac{\partial}{\partial t}\,\mathrm{tr}\big(\mathbf{P}^2\big) &= \mathrm{tr}\left(\frac{\partial}{\partial t}\mathbf{P}^2\right)\\ &= \mathrm{tr}\left(\frac{\partial \mathbf{P}}{\partial t}\mathbf{P} + \mathbf{P}\frac{\partial \mathbf{P}}{\partial t}\right)\\ &= \mathrm{tr}\left(\frac{\partial \mathbf{P}}{\partial t}\mathbf{P}\right) + \mathrm{tr}\left(\mathbf{P}\frac{\partial \mathbf{P}}{\partial t}\right)\\ &= \mathrm{tr}\left(\mathbf{P}\frac{\partial \mathbf{P}}{\partial t}\right) + \mathrm{tr}\left(\mathbf{P}\frac{\partial \mathbf{P}}{\partial t}\right)\\ &= 2\,\mathrm{tr}\left(\mathbf{P}\frac{\partial \mathbf{P}}{\partial t}\right).\end{aligned} \tag{8.113}$$

Although the stress was used in the derivation of the above formula, the result holds for any differentiable tensor. In a similar manner one can show that

$$\frac{\partial}{\partial t}\,\mathrm{tr}\big(\mathbf{F}^T\mathbf{F}\big) = \frac{\partial}{\partial t}\,\mathrm{tr}\big(\mathbf{F}\mathbf{F}^T\big) = 2\,\mathrm{tr}\left(\mathbf{F}^T\frac{\partial \mathbf{F}}{\partial t}\right).$$

Again, the above formula holds for any tensor, as long as it is differentiable.

Using the constitutive law in (8.104), and the results from Exercise (8.9), we have that

$$\begin{aligned}\mathrm{tr}(\mathbf{P}\nabla_A\mathbf{V}) &= \mathrm{tr}\left((\lambda_s \mathrm{I}_E\mathbf{F}^T + 2\mu_s\mathbf{E}\mathbf{F}^T)\frac{\partial}{\partial t}\mathbf{F}\right)\\ &= \lambda_s\mathrm{I}_E\,\mathrm{tr}\left(\mathbf{F}^T\frac{\partial}{\partial t}\mathbf{F}\right) + \mu_s\,\mathrm{tr}\left((\mathbf{F}^T\mathbf{F}-\mathbf{I})\mathbf{F}^T\frac{\partial}{\partial t}\mathbf{F}\right)\\ &= \frac{1}{2}\left(\lambda_s\mathrm{I}_E - \mu_s\right)\frac{\partial}{\partial t}\,\mathrm{tr}\big(\mathbf{F}^T\mathbf{F}\big) + \mu_s\,\mathrm{tr}\left(\mathbf{F}^T\mathbf{F}\mathbf{F}^T\frac{\partial}{\partial t}\mathbf{F}\right).\end{aligned} \tag{8.114}$$

It is convenient at this point to introduce the left Cauchy-Green deformation tensor, defined as

$$\mathbf{B} = \mathbf{F}\mathbf{F}^T. \tag{8.115}$$

From (8.113),

$$\begin{aligned}\frac{\partial}{\partial t}\,\mathrm{tr}\big(\mathbf{B}^2\big) &= 2\,\mathrm{tr}\left(\mathbf{B}\frac{\partial \mathbf{B}}{\partial t}\right)\\ &= 2\,\mathrm{tr}\big(\mathbf{F}\mathbf{F}^T(\mathbf{F}\mathbf{F}_t^T + \mathbf{F}_t\mathbf{F}^T)\big)\\ &= 4\,\mathrm{tr}\left(\mathbf{F}^T\mathbf{F}\mathbf{F}^T\frac{\partial}{\partial t}\mathbf{F}\right).\end{aligned}$$

Also, from (8.98), $\mathrm{I}_E = \frac{1}{2}\big(\mathrm{tr}(\mathbf{F}^T\mathbf{F}) - 3\big)$. Substituting these into (8.114) we obtain

$$\begin{aligned}\mathrm{tr}(\mathbf{P}\nabla_A\mathbf{V}) &= \frac{1}{4}\left(\lambda_s\,\mathrm{tr}\left(\mathbf{F}^T\mathbf{F}\right) - 3\lambda_s - 2\mu_s\right)\frac{\partial}{\partial t}\,\mathrm{tr}\big(\mathbf{F}^T\mathbf{F}\big) + \frac{1}{4}\mu_s\frac{\partial}{\partial t}\,\mathrm{tr}\big(\mathbf{B}^2\big)\\ &= \frac{\partial}{\partial t}\left[\frac{1}{8}\lambda_s\left(\mathrm{tr}\big(\mathbf{F}^T\mathbf{F}\big)\right)^2 - \frac{1}{4}(3\lambda_s + 2\mu_s)\,\mathrm{tr}\big(\mathbf{F}^T\mathbf{F}\big) + \frac{1}{4}\mu_s\,\mathrm{tr}\big(\mathbf{B}^2\big)\right].\end{aligned} \tag{8.116}$$

Inserting (8.116) into (8.112), the conclusion is that

$$\frac{\partial}{\partial t}\,(K + U) = \nabla_A \cdot (\mathbf{P}\mathbf{V}), \tag{8.117}$$

where

$$K = \frac{1}{2}R_0\mathbf{V}\cdot\mathbf{V} \tag{8.118}$$

is the kinetic energy density, and

$$U = \frac{1}{8}\lambda_s\left(\mathrm{tr}\big(\mathbf{F}^T\mathbf{F}\big)\right)^2 - \frac{1}{4}(3\lambda_s + 2\mu_s)\,\mathrm{tr}\big(\mathbf{F}^T\mathbf{F}\big) + \frac{1}{4}\mu_s\,\mathrm{tr}\big(\mathbf{B}^2\big) \tag{8.119}$$

is the potential energy density. Using the Green strain tensor (8.98), and the properties of the trace, the above expression can be written as

$$U = \frac{1}{2}\lambda_s\left[\mathrm{tr}(\mathbf{E})\right]^2 + \mu_s\,\mathrm{tr}\big(\mathbf{E}^2\big) - \frac{3}{8}(3\lambda_s + 2\mu_s). \tag{8.120}$$

The function U is the energy stored in the elastic material due to the deformation. For this reason it is often called the stored energy function, or the strain energy function. One conclusion coming from (8.117) is that the total energy $K + U$ only changes due to the energy flux $-\mathbf{P}\mathbf{V}$. It does not change because of a loss due to dissipation, such as happened with a viscous fluid.

Exercises

8.1. This exercise is based on the definition of the material and spatial coordinate systems.

(a) Suppose a particle that started at location $(2, 0, -1)$ is, at $t = 4$, located at $(1, 0, 1)$. What is $\mathbf{A}$ for this particle? For this particle, what is $\mathbf{U}(\mathbf{A}, 4)$? What is the spatial coordinate $\mathbf{x}$ at $t = 4$, for this particle, and what is the corresponding value of $\mathbf{u}(\mathbf{x}, 4)$?

(b) For another particle, $\mathbf{u}(\mathbf{x}, 3) = (1, 0, 0)$ for $\mathbf{x} = (2, 1, 1)$. What is $\mathbf{A}$ for this particle?

8.2. A motion of the form $x = \alpha(t)A_1$, $y = A_2/\alpha(t)$, $z = A_3$, where $\alpha(0) = 1$ and $\alpha > 0$, is an example of pure shear.

(a) Give a geometric interpretation of this motion by describing what happens to the unit cube $0 \le A_1 \le 1$, $0 \le A_2 \le 1$, $0 \le A_3 \le 1$.

(b) Find $\mathbf{U}$, $\mathbf{u}$, $\mathbf{V}$, and $\mathbf{v}$.

(c) Verify that $\mathbf{v} = \frac{D\mathbf{u}}{Dt}$.

(d) Show that $\mathbf{v}$ satisfies the continuity equation for an incompressible material.

(e) Find $\mathbf{D}$ and then calculate the invariants I_D, II_D, III_D.

8.3. Consider the motion $x = A_1 + \alpha(t)A_2$, $y = A_2 + \alpha(t)A_3$, $z = A_3$, where $\alpha(0) = 0$.

(a) Give a geometric interpretation of this motion by describing what happens to the unit cube $0 \le A_1 \le 1$, $0 \le A_2 \le 1$, $0 \le A_3 \le 1$.

(b) Find $\mathbf{U}$, $\mathbf{u}$, $\mathbf{V}$, and $\mathbf{v}$.

(c) Verify that $\mathbf{v} = \frac{D\mathbf{u}}{Dt}$.

(d) Show that $\mathbf{v}$ satisfies the continuity equation for an incompressible material.

(e) Find $\mathbf{D}$ and then calculate the invariants I_D, II_D, III_D.

8.4. Linear flow occurs when $\mathbf{v} = \mathbf{H}\mathbf{x} + \mathbf{h}$, where the matrix $\mathbf{H}$ and vector $\mathbf{h}$ can depend on t.

(a) Find $\mathbf{H}$ and $\mathbf{h}$ for uniform dilatation, for simple shear, and for rigid body motion.

(b) Show that linear motion is possible for an incompressible material only if $\mathrm{tr}(\mathbf{H}) = 0$.

(c) Does simple shear satisfy the condition in part (b)?

(d) Show that rigid body motion satisfies the condition in part (b). The results from Exercise 8.22 might be helpful.

8.5. This problem continues the study of linear flow, introduced in the previous exercise.

(a) Assuming there are no body forces, show that linear flow is a solution of the incompressible Navier-Stokes equation if $\mathbf{H}'(t)$ is symmetric. This assumes that $\mathrm{tr}(\mathbf{H}) = 0$, which was established in Exercise 8.4(b)

(b) Under what conditions, if any, does simple shear satisfy the conditions in part (a)?
(c) Under what conditions, if any, does rigid body motion satisfy the conditions in part (a)?

8.6. Suppose the stress tensor is

$$\mathbf{T} = \begin{pmatrix} x & yz & 2 \\ yz & y & x \\ 2 & x & z \end{pmatrix}.$$

(a) Assuming there are no body forces, explain why it is not possible that the material is at rest.
(b) What would the body force need to be so the material is at rest?

8.7. Suppose the stress tensor is

$$\mathbf{T} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 1 \end{pmatrix}.$$

(a) Consider the unit cube $0 \le x \le 1$, $0 \le y \le 1$, $0 \le z \le 1$. Find the stress vector on each face of the cube.
(b) Suppose $\mathbf{T}$ is the stress tensor for a viscous incompressible fluid, with $p = 0$. Find the velocity $\mathbf{v}$.

8.8. Suppose it is known that the stress is identically zero, and there are no body forces.
(a) What is the resulting displacement in material and in spatial coordinates?
(b) Suppose the constitutive law for a Newtonian fluid is used. What is the pressure?
(c) Explain why your conclusion from part (a) holds in the case of when the stress tensor is assumed to be constant.

8.9. This problem develops some of the formulas for spatial and material coordinates.
(a) Show that $\mathbf{F} = \mathbf{I} + \nabla_A \mathbf{U}$.
(b) Show that $\dfrac{\partial}{\partial t}\mathbf{F} = \nabla_A \mathbf{V}$.
(c) Assuming (8.15) holds, show that $\nabla_A F = \mathbf{F}^T(\nabla f)$.
(d) Let $\mathbf{G}(\mathbf{A}, t)$ be a vector function in material coordinates and let its spatial version be $\mathbf{g}(\mathbf{x}, t)$. This means that $\mathbf{G}(\mathbf{A}, t) = \mathbf{g}(\mathbf{X}(\mathbf{A}, t), t)$, and from this equation show that $\nabla_A \mathbf{G} = (\nabla \mathbf{g})\mathbf{F}$. Explain why this shows that $\nabla_A \mathbf{V} = (\nabla \mathbf{v})\mathbf{F}$.

8.10. This problem develops some of the connections between the displacement and velocity in spatial coordinates.

(a) Show that to find $\mathbf{v}$ given $\mathbf{u}$ one must solve $(\mathbf{I} - \nabla\mathbf{u})\mathbf{v} = \partial_t\mathbf{u}$, where $\mathbf{u} = (u_1, u_2, u_3)$ and

$$\nabla\mathbf{u} = \begin{pmatrix} \dfrac{\partial u_1}{\partial x} & \dfrac{\partial u_1}{\partial y} & \dfrac{\partial u_1}{\partial z} \\ \dfrac{\partial u_2}{\partial x} & \dfrac{\partial u_2}{\partial y} & \dfrac{\partial u_2}{\partial z} \\ \dfrac{\partial u_3}{\partial x} & \dfrac{\partial u_3}{\partial y} & \dfrac{\partial u_3}{\partial z} \end{pmatrix}.$$

(b) In the case of when the motion is one-dimensional, show that the formula in part (a) reduces to (6.14).
(c) Verify the result from part (a) for uniform dilatation, as given in Section 8.2.
(d) Suppose $\mathbf{v}$ is known and one wants to determine $\mathbf{u}$. Explain why one way this can be done is by solving the partial differential equation $\partial_t\mathbf{u} + (\mathbf{v} \cdot \nabla)\mathbf{u} = \mathbf{v}$, with $\mathbf{u}(\mathbf{x}, 0) = \mathbf{0}$.
(e) Another method for finding $\mathbf{u}$ can be derived by reverting to material coordinates. Given a particle that starts out at $\mathbf{A}$, explain why the position function $\overline{\mathbf{X}}(t)$ of that particle satisfies the ordinary differential equation $\overline{\mathbf{X}}' = \mathbf{v}(\overline{\mathbf{X}}, t)$, where $\overline{\mathbf{X}}(0) = \mathbf{A}$. Given the solution of this problem, assume the equation $\mathbf{x} = \overline{\mathbf{X}}$ is solved for $\mathbf{A}$, to obtain $\mathbf{A} = \mathbf{a}(\mathbf{x}, t)$. With this, explain why the displacement is $\mathbf{u} = \overline{\mathbf{X}} - \mathbf{A}$, where $\mathbf{A} = \mathbf{a}(\mathbf{x}, t)$.
(f) Suppose that $\mathbf{v} = (\gamma y, 0, 0)$, where γ is a constant. Show that $\mathbf{u} = (\gamma yt, 0, 0)$. Also, explain why the position of the particle that starts out at (x_0, y_0, z_0) is $(x_0 + \gamma yt, y_0, z_0)$.

8.11. If the densities are per unit mass, then the general balance law (8.30) takes the form

$$\frac{d}{dt}\iiint\limits_{R(t)} \rho f(\mathbf{x}, t)\, dV = -\iint\limits_{\partial R(t)} \mathbf{J} \cdot \mathbf{n}\, dS + \iiint\limits_{R(t)} \rho Q(\mathbf{x}, t)\, dV.$$

Show that in this case (8.31) is replaced with

$$\rho\frac{Df}{Dt} = -\nabla \cdot \mathbf{J} + \rho Q.$$

8.12. Show that the following cannot be constitutive laws.
(a) $\mathbf{T} = \mathbf{G}(\mathbf{x})$.
(b) $\mathbf{T} = \mathbf{G}(\mathbf{u})$.
(c) $\mathbf{P} = \alpha(\mathbf{F} + \mathbf{F}^T)$.
(d) $\mathbf{P} = \mathbf{G}(\mathbf{V})$.

8.13. This problem considers possible motions of an incompressible fluid, with no external body forces.

(a) Consider the simple shear motion described in Section 8.2. What is the corresponding spatial velocity $\mathbf{v}$? When will this be a solution of the equations of motion?
(b) Consider the rigid body motion described in Section 8.2. What is the corresponding spatial velocity $\mathbf{v}$? When will this be a solution of the equations of motion?

8.14. Suppose that the incompressible fluid equations (8.77), (8.78) are to be solved in a bounded domain D, and the impermeability condition (8.79) is used on the boundary ∂D. Show that not just any boundary velocity $\mathbf{v}_s$ can be used. Namely, show that the given velocity must satisfy

$$\iint_{\partial D} \mathbf{v}_s \cdot \mathbf{n}\, dS = 0.$$

What is the physical meaning of the above equation?

8.15. This problem concerns some of the connections between spatial and material variables.
(a) Show that the spatial and material stresses are equal at $t = 0$. Namely, show that $\mathbf{T} = \mathbf{P} = \mathbf{S}$ at $t = 0$.
(b) Using (8.27), show that in material coordinates the assumption of incompressibility results in the equation $\det(\mathbf{F}) = 1$.

8.16. In formulating the constitutive law for elasticity we used the right Cauchy-Green deformation tensor $\mathbf{C}$ given in (8.97). The left Cauchy-Green deformation tensor is $\mathbf{B} = \mathbf{F}\mathbf{F}^T$. This problem develops some of the similarities, and differences, of these two tensors.
(a) Show that $\mathbf{B}$ and $\mathbf{C}$ are symmetric.
(b) Show that $\mathbf{C}$ is not objective but it is invariant in the sense that $\mathbf{C}^* = \mathbf{C}$.
(c) Show that $\mathbf{B}$ is objective but not materially objective.
(d) Explain why a constitutive law of the form $\mathbf{T} = \alpha \mathrm{I}_B\, \mathbf{I} + 2\beta\mathbf{B}$ is permitted, but one of the form $\mathbf{T} = \alpha \mathrm{I}_C\, \mathbf{I} + 2\beta\mathbf{C}$ is not. Similarly, explain why $\mathbf{S} = \alpha \mathrm{I}_C\, \mathbf{I} + 2\beta\mathbf{C}$ is permitted, but $\mathbf{S} = \alpha \mathrm{I}_B\, \mathbf{I} + 2\beta\mathbf{B}$ is not. In these expressions, α and β are constants.
(e) Prove that $\mathbf{C}$ and $\mathbf{B}$ have the same eigenvalues but do not necessarily have the same eigenvectors.
(f) Prove that the eigenvalues of $\mathbf{C}$ and $\mathbf{B}$ are nonnegative.
(g) If the material is incompressible, use Exercise 8.15(b) to show that $\mathrm{III}_B = \mathrm{III}_C = 1$.

8.17. A strain tensor is symmetric, and it is zero if there is no deformation or if the deformation corresponds to a rigid body motion. Consequently, if $\mathbf{Z}$ is a strain tensor then $\mathbf{Z}$ must be symmetric, and $\mathbf{Z} = \mathbf{0}$ if $\mathbf{U} = \mathbf{0}$ or if $\mathbf{X} = \mathbf{Q}(t)\mathbf{A} + \mathbf{b}(t)$.
(a) Show that the Green strain tensor $\mathbf{E}$ satisfies the stated conditions.

(b) Does the Lagrangian strain tensor, given in (8.105), satisfy the stated conditions?
(c) Show that $\mathbf{Z} = \frac{1}{2}(\mathbf{I} - \mathbf{B})$ satisfies the stated conditions. It is known as the Finger strain tensor.
(d) Show that $\mathbf{Z} = \frac{1}{2}(\mathbf{I} - \mathbf{B}^{-1})$ satisfies the stated conditions. It is known as the Almansi strain tensor.
(e) Show that $\mathbf{Z} = \mathbf{B}$ does not qualify as a strain tensor.

8.18. The kinetic energy for a regular region $R(t)$ is

$$K = \iiint\limits_{R(t)} \frac{1}{2}\rho \mathbf{v} \cdot \mathbf{v}\, dV.$$

Let K_0 be the value of K when the motion is irrotational, which means that the velocity can be written as $\mathbf{v} = \nabla\phi$. Let $\mathbf{v}$ be any other velocity, not necessarily irrotational, but which has the same normal velocity at the boundary as the irrotational motion. This means that $\mathbf{v} \cdot \mathbf{n} = (\nabla\phi) \cdot \mathbf{n}$ on ∂R. Assuming the density is constant, show that $K_0 \leq K$. This observation that irrotational flows minimize the kinetic energy is known as Kelvin's Minimum Energy Theorem.

8.19. This problem extends some of the ideas developed with the energy equation.
(a) Using the second law of thermodynamics it can be shown that the stress power is nonnegative. Use this to show that if $\mathbf{T} = -p\mathbf{I} + \alpha_0\mathbf{I} + \alpha_1\mathbf{D} + \alpha_2\mathbf{D}^2$, and the material is incompressible, then $0 \leq \alpha_1 \mathrm{tr}(\mathbf{D}^2) + 3\alpha_2 \det(\mathbf{D})$. Hint: The Cayley-Hamilton theorem will be useful here.
(b) Use the result from part (a) to prove that the dynamic viscosity of an incompressible Newtonian fluid is nonnegative.
(c) Show that the potential energy density (8.119) can be written as

$$U = \frac{1}{8}(\lambda_s + 2\mu_s)\mathrm{I}_B^2 - \frac{1}{4}(3\lambda_s + 2\mu_s)\mathrm{I}_B - \frac{1}{2}\mu_s \mathrm{II}_B.$$

8.20. This problem provides the details related to the derivation of the energy equation.
(a) Show that $\mathbf{v} \cdot (\nabla \cdot \mathbf{T}) = \nabla \cdot (\mathbf{T}\mathbf{v}) - \mathrm{tr}(\mathbf{T}\nabla\mathbf{v})$. Also, explain why this equation holds even if $\mathbf{T}$ is not symmetric.
(b) Using the fact that $\mathbf{T}$ is symmetric show that $\mathrm{tr}(\mathbf{T}\nabla\mathbf{v}) = \mathrm{tr}(\mathbf{T}(\nabla\mathbf{v})^T)$. With this show that $\mathbf{v} \cdot (\nabla \cdot \mathbf{T}) = \nabla \cdot (\mathbf{T}\mathbf{v}) - \mathrm{tr}(\mathbf{T}\mathbf{D})$.
(c) Explain why part (a) can be used to show that $\mathbf{V} \cdot (\nabla_A \cdot \mathbf{P}) = \nabla_A \cdot (\mathbf{P}\mathbf{V}) - \mathrm{tr}(\mathbf{P}\nabla_A\mathbf{V})$.

8.21. This problem concerns the inverse of the constitutive law for a linear material.
(a) For a viscous compressible fluid show that

$$\mathbf{D} = -\frac{\lambda}{2\mu(3\lambda + 2\mu)}(3p + \mathrm{tr}(\mathbf{T}))\mathbf{I} + \frac{1}{2\mu}(\mathbf{T} + p\mathbf{I}).$$

(b) For a linearly elastic material show that

$$\mathbf{E} = -\frac{\lambda_s}{2\mu_s(3\lambda_s + 2\mu_s)}\mathrm{tr}(\mathbf{P})\mathbf{I} + \frac{1}{2\mu_s}\mathbf{P}.$$

8.22. This problem develops some of the properties of the rotation matrix $\mathbf{Q}(t)$ used in the definition of a rigid body motion (8.13). Therefore, $\mathbf{Q}$ has the properties that $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ and $\det(\mathbf{Q}) = 1$.

(a) Show that $(\mathbf{Q}^T)' = (\mathbf{Q}')^T$ and from this show that $\mathbf{Q}'\mathbf{Q}^T = -\mathbf{Q}(\mathbf{Q}')^T$.

(b) Setting $\boldsymbol{\Omega} = \mathbf{Q}'\mathbf{Q}^T$, show that the spatial velocity for rigid body motion is $\mathbf{v} = \boldsymbol{\Omega}\mathbf{x} + \mathbf{b}' - \boldsymbol{\Omega}\mathbf{b}$.

(c) Show that $\boldsymbol{\Omega}$ is skew symmetric and $\mathrm{tr}(\boldsymbol{\Omega}) = 0$.

(d) Let $\mathbf{Q} = \exp(\mathbf{M}t)$, where $\mathbf{M}$ is a skew-symmetric matrix and the exponential is defined using the Taylor series

$$\begin{aligned}\exp(\mathbf{M}t) &= \mathbf{I} + t\mathbf{M} + \frac{1}{2}t^2\mathbf{M}^2 + \frac{1}{3!}t^3\mathbf{M}^3 + \cdots \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}t^n\mathbf{M}^n.\end{aligned}$$

Show that $\mathbf{Q}$ is a rotation, and $\mathbf{Q}' = \mathbf{M}\mathbf{Q}$. Therefore, $\mathbf{Q}$ is a rotation that satisfies $\mathbf{Q}(0) = \mathbf{I}$ and $\mathbf{Q}'(0) = \mathbf{M}$. You can assume the above series has the same convergence properties as the Taylor series for e^x.

8.23. This problem derives formulas for vector and tensor quantities when making the rigid body change of variables given in (8.49).

(a) Show that $\mathbf{F}^* = \mathbf{Q}\mathbf{F}$

(b) Show that $\mathbf{D}^* = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$ and $\mathbf{W}^* = \mathbf{Q}\mathbf{W}\mathbf{Q}^T + \boldsymbol{\Omega}$.

(c) Show that $(\nabla \cdot \mathbf{T})^* = \mathbf{Q}(\nabla \cdot \mathbf{T})$.

(d) Setting $\mathbf{c} = \mathbf{Q}^{\mathbf{T}}\mathbf{b}$, show that $\mathbf{V} = \mathbf{Q}^T\mathbf{V}^* + (\mathbf{Q}^T)'\mathbf{X}^* - \mathbf{c}'$. From this show that

$$\frac{D\mathbf{v}}{Dt} = \mathbf{Q}^T\left(\frac{D\mathbf{v}}{Dt}\right)^* + 2(\mathbf{Q}^T)'\mathbf{v}^* + (\mathbf{Q}^T)''\mathbf{x}^* - \mathbf{c}''.$$

(e) Assuming the body force transforms as $\mathbf{f}^* = \mathbf{Q}\mathbf{f}$, and the density as $\rho^* = \rho$, show that the momentum equation in the $\mathbf{x}^*$ coordinate system is

$$\rho^*\left(\frac{D\mathbf{v}}{Dt}\right)^* = (\nabla \cdot \mathbf{T})^* + \rho^*\mathbf{f}^* - \rho^*\mathbf{z}^*,$$

where $\mathbf{z}^* = 2\mathbf{Q}(\mathbf{Q}^T)'\mathbf{v}^* + \mathbf{Q}\mathbf{Q}''\mathbf{x}^* - \mathbf{Q}\mathbf{c}''$ is an acceleration term that comes from the change of variables.

(f) Explain why the momentum equation in not Euclidian invariant but is Galilean invariant.

8.24. This problem develops some of the properties of the principal invariants of a symmetric matrix $\mathbf{R}$.

(a) Derive the formulas in (8.62)-(8.64).

(b) Show that the characteristic equation for $\mathbf{R}$ can be written as $\lambda^3 - \mathrm{I}_R\lambda^2 + \mathrm{II}_R\lambda - \mathrm{III}_R = 0$.

(c) By definition, I_R is a function of the components of $\mathbf{R}$. Show that

$$\frac{\partial}{\partial R_{ij}}\mathrm{I}_R = \delta_{ij},$$

where δ_{ij} is the Kronecker delta function. Explain why the above equation can be written as

$$\frac{\partial}{\partial \mathbf{R}}\mathrm{I}_R = \mathbf{I}.$$

(d) Using the ideas developed in part (c) show that

$$\frac{\partial}{\partial \mathbf{R}}\mathrm{II}_R = -\mathbf{R} + \mathrm{I}_R\mathbf{I}.$$

8.25. This problem derives the Rivlin-Ericksen representation theorem for a special case. The assumption is that the constitutive law for the stress is $\mathbf{T} = \mathbf{G}(\mathbf{R})$, where $\mathbf{R}$ is a symmetric tensor. Also, assume that the function $\mathbf{G}$ can be expanded using a Taylor series to give

$$\mathbf{G} = \sum_{n=0}^{\infty} \kappa_n \mathbf{R}^n,$$

where the κ_n's are constants and $\mathbf{R}^n = \mathbf{I}$ for $n = 0$.

(a) Use the Cayley-Hamilton theorem to show that the constitutive law can be written as $\mathbf{T} = \alpha_0\mathbf{I} + \alpha_1\mathbf{R} + \alpha_2\mathbf{R}^2$, where the coefficients α_0, α_1, and α_2 are functions of the three invariants of $\mathbf{R}$.

(b) Suppose $\mathbf{R}$ is objective, so that $\mathbf{R}^* = \mathbf{Q}\mathbf{R}\mathbf{Q}^T$. Does the assumed constitutive law satisfy the Principle of Material Frame-Indifference?

8.26. This problem outlines the proof of the Rivlin-Ericksen representation theorem. The assumption is that the constitutive law for the stress is given in (8.52), where $\mathbf{R}$ is a symmetric, objective tensor. Because of symmetry, there is a basis so $\mathbf{R}$ is diagonal, where the diagonal entries are its eigenvalues λ_1, λ_2, and λ_3. In this problem, this basis is used for the $\mathbf{x}$-system. Also, because $\mathbf{R}$ is objective, it satisfies the transformation law $\mathbf{R}^* = \mathbf{Q}\mathbf{R}\mathbf{Q}^T$.

(a) Show that $\mathbf{Q}_1$, given below, is a rotation. With this rotation, show that (8.53) implies that $\mathbf{T}^* = \mathbf{T}$, while (8.51) implies something else. Use this to prove that $T_{12} = T_{13} = 0$.

(b) Find a rotation $\mathbf{Q}_2$ that shows that $T_{23} = 0$.

(c) The conclusion, so far, is that if $\mathbf{R}$ is diagonal then so is $\mathbf{T}$. With this, we write $T_{11} = f_1(\lambda_1, \lambda_2, \lambda_3)$, $T_{22} = f_2(\lambda_1, \lambda_2, \lambda_3)$, and $T_{33} = f_3(\lambda_1, \lambda_2, \lambda_3)$.

Show that $\mathbf{Q}_3$, given below, is a rotation. With this rotation, show that $f_2(\lambda_1, \lambda_2, \lambda_3) = f_1(\lambda_2, \lambda_3, \lambda_1)$ and $f_3(\lambda_1, \lambda_2, \lambda_3) = f_1(\lambda_3, \lambda_1, \lambda_2)$.

(d) Show that $\mathbf{Q}_4$, given below, is a rotation. With this rotation, show that $f_1(\lambda_1, \lambda_2, \lambda_3) = f_1(\lambda_1, \lambda_3, \lambda_2)$.

(e) Assume α, β, and γ are solutions of the following systems of equations:

$$\begin{aligned}
\alpha + \beta\lambda_1 + \gamma\lambda_1^2 &= f_1(\lambda_1, \lambda_2, \lambda_3),\\
\alpha + \beta\lambda_2 + \gamma\lambda_2^2 &= f_1(\lambda_2, \lambda_3, \lambda_1),\\
\alpha + \beta\lambda_3 + \gamma\lambda_3^2 &= f_1(\lambda_3, \lambda_1, \lambda_2).
\end{aligned}$$

In this case, α, β, and γ are functions of λ_1, λ_2, and λ_3. Use parts (c) and (d) to show that they are symmetric functions, that is, their values do not change if any pair of λ_1, λ_2, and λ_3 are interchanged. From the theory of symmetric functions, it can be shown that α, β, and γ can be written as functions of $\mathrm{tr}(\mathbf{R})$, $\mathrm{tr}(\mathbf{R}^2)$, and $\mathrm{tr}(\mathbf{R}^3)$. Use this to obtain the Rivlin-Ericksen representation theorem.

(f) Show that α, β, and γ are uniquely determined if λ_1, λ_2, and λ_3 are distinct.

As a comment, the solution of the system in part (e) is not necessarily unique, and this has repercussions for how smooth α, β, and γ are as functions of the invariants. A discussion of this can be found in Truesdell and Noll [2004].

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \mathbf{Q}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{Q}_4 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Chapter 9
Fluids

9.1 Newtonian Fluids

The equations of motion for an incompressible Newtonian fluid were derived in the previous chapter. The conclusion was that the constitutive law for the stress is

$$\mathbf{T} = -p\mathbf{I} + 2\mu\mathbf{D}, \tag{9.1}$$

where p is the pressure, $\mathbf{D}$ is the rate of deformation tensor given in (8.67), and μ is the dynamic viscosity. The SI unit for μ is the Pascal-second (Pa s), and to help provide some perspective on this, the viscosities of some well-known fluids are given in Table 9.1. Not unexpectedly, the viscosity of air is significantly less than the viscosity of water, which in turn is less viscous than olive oil. What might seem odd is that there is an entry for peanut butter, which was determined experimentally in Baker et al. [2004]. You might think that a substance like peanut butter behaves more as a solid than a fluid. This is partly due to the length of time it takes peanut butter to flow. It is so slow that it seems to have more of the characteristics of a solid. As it turns out, peanut butter is not a Newtonian fluid, but this is not due to its slow flow characteristic. It has properties similar to toothpaste and ketchup, two materials that are discussed in more depth in the next section.

Fluid	Viscosity (Pa s)	Density (kg/m^3)
Air	1.8×10^{-5}	1.18
Water	0.89×10^{-3}	0.997×10^3
Mercury	1.5×10^{-3}	1.3×10^4
Olive Oil	0.8×10^{-1}	0.92×10^3
Peanut Butter	1.2×10^5	1.02×10^3

Table 9.1 Viscosity and density of various substances at 25° C.

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_9,
© Springer Science+Business Media, LLC 2009

The question that arose about peanut butter is one of the objectives of this chapter, namely how can you determine if a substance can be modeled as a Newtonian fluid? This same question came up in Chapters 6 and 7 when studying elasticity and viscoelasticity, and the answer is the same as before. Namely, we will derive solutions to the equations of motion and then compare them with what is found experimentally. Assuming they agree then we should be able to use the experimental data to determine the viscosity. We will also use this approach to investigate various simplifications that can be made in the Newtonian model. For example, the viscosity of air is so small, it would seem that it might be possible to simply assume it is zero. This assumption produces what is known as an inviscid fluid, and the resulting mathematical problem gives the appearance of being simpler than what is obtained for a viscous fluid. In this chapter a progression of such simplifying assumptions is examined, with the goal of better understanding fluid motion.

9.2 Steady Flow

One of the more studied problems in fluids involves steady flow. This means that the fluid velocity and pressure are independent of time. Assuming there are no body forces then the equations of motion for a steady incompressible fluid, coming from (8.77) and (8.78), are

$$\rho(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \mu\nabla^2\mathbf{v}, \tag{9.2}$$

$$\nabla\cdot\mathbf{v} = 0. \tag{9.3}$$

As always, with incompressible motion, it is assumed that ρ is constant.

We will solve several fluid problems, and it is always of interest on such occasions to be able to visualize the flow. One method is to find the paths of individual fluid particles as the fluid moves, what are known as pathlines. Once the velocity is known, then the pathline $\mathbf{x} = \mathbf{X}(t)$ that starts out at $\mathbf{x} = \mathbf{A}$ is found by solving

$$\frac{d\mathbf{X}}{dt} = \mathbf{v}(\mathbf{X}, t), \tag{9.4}$$

where

$$\mathbf{X}(0) = \mathbf{A}.$$

As is probably evident, a pathline is just the position function used to define material coordinates introduced in Sections 6.2 and 8.2. As demonstrated in Exercise 6.3, even for one dimensional motion it is not particularly easy to find an analytical solution of (9.4). For steady motion, which is what we are currently investigating, the problem is a bit easier as the velocity does not depend explicitly on time. However, for most problems numerical methods are usually needed to find the solution.

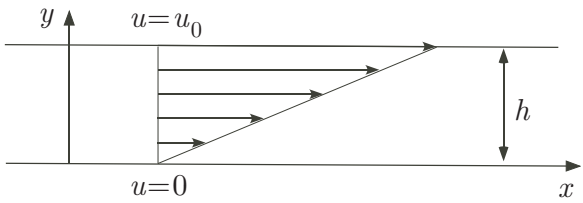


Figure 9.1 In plane Couette flow the lower plate is stationary, while the top plate moves in the x-direction. Solving this problem shows that the velocity of the fluid varies linearly between the two plates.

9.2.1 Plane Couette Flow

One of the more basic flows arises when studying the motion of a fluid between two parallel plates. A cross-section of this configuration is shown in Figure 9.1. The lower plate, located at $y = 0$, is fixed, while the upper plate, at $y = h$, moves with a constant velocity u_0 in the x-direction. The associated boundary conditions are

$$\mathbf{v} = (u_0, 0, 0) \quad \text{on } y = h, \tag{9.5}$$

$$\mathbf{v} = \mathbf{0} \quad \text{on } y = 0. \tag{9.6}$$

It is assumed that the upper plate has been moving with this constant velocity for a long time, so the flow is steady. It is also assumed that the fluid is incompressible, so (9.2), (9.3) apply.

At first glance, given that (9.2) is a nonlinear partial differential equation, finding the velocity and pressure would seem to be an almost impossible task. However, some useful insights on the properties of the solution can be derived from the boundary conditions and the geometry. In particular, given that the upper and lower boundaries are flat plates, and the upper one moves with a constant velocity in the x-direction, it is not unreasonable to guess that there is no dependence on, or motion in, the z-direction. In other words $\mathbf{v} = (u, v, 0)$, where u, v, and p are independent of z. In this case, (9.2), (9.3) reduce to

$$\begin{aligned} \rho(u\partial_x + v\partial_y)u &= -\partial_x p + \mu\nabla^2 u, \\ \rho(u\partial_x + v\partial_y)v &= -\partial_y p + \mu\nabla^2 v, \\ \partial_x u + \partial_y v &= 0. \end{aligned}$$

This is still a formidable problem, so we need another insight into the form of the solution. Given that the upper plate is sliding in the x-direction, it is not unreasonable to expect that there is no flow in the y-direction. In this case, $v = 0$ and the above system reduces to

$$\rho u \partial_x u = -\partial_x p + \mu \partial_y^2 u,$$
$$0 = -\partial_y p,$$
$$\partial_x u = 0.$$

From the last two equations we have that $p = p(x)$ and $u = u(y)$. In this case, the first equation reduces to $p'(x) = \mu u''(y)$. The only way for a function of x to equal a function of y is that both functions are constants. Consequently, $p'(x)$ constant means $p(x) = p_0 + xp_1$, where p_0 and p_1 are constants. It is assumed that the pressure remains bounded, and so $p_1 = 0$. With this, the solution of $\mu u''(y) = p'(x)$ is $u = ay + b$. Imposing the boundary conditions $u(0) = 0$ and $u(h) = u_0$, it follows that $u = u_0 y/h$.

The solution of the plane Couette flow problem is, therefore,

$$\mathbf{v} = (\gamma y, 0, 0), \tag{9.7}$$

where $p = p_0$ is constant, and

$$\gamma = \frac{u_0}{h} \tag{9.8}$$

is known as the *shear rate*. This shows that the fluid velocity in the x-direction increases linearly between the two plates, from zero to u_0. This dependence is illustrated in Figure 9.2. Also, the resulting fluid stress tensor (9.9) is

$$\mathbf{T} = -p_0 \mathbf{I} + \mu \begin{pmatrix} 0 & \frac{u_0}{h} & 0 \\ \frac{u_0}{h} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{9.9}$$

The above solution gives us something we sorely need, and that is a method for checking on the assumption that a fluid is Newtonian. The solution shows that for a Newtonian fluid the shear stress is $T_{12} = \mu u_0/h$. Therefore, the shear stress depends linearly on the shear rate $\gamma = u_0/h$, with the slope of the curve equal to the viscosity. This is the basis for one of the more important experiments in fluid dynamics, where the shear stress is measured as a function of the shear rate. Results from such tests are shown in Figures 9.2 and 9.3, for fluids most people have experience with. Based on the linearity of the data in Figure 9.2, the assumption that water and oil are Newtonian is reasonable. For the same reason, from Figure 9.3, ketchup and toothpaste are not Newtonian, or only Newtonian for very small shear rates. They are examples of what are called nonlinear power-law fluids, where $T_{12} = \alpha\gamma^\beta$. Based on the data in Figure 9.3, for ketchup, $\beta = 0.55$, where as for toothpaste, $\beta = 0.44$. Some of the implications of such a constitutive law are investigated in Exercise 9.7.

Before moving on to another topic, a comment needs to be made about our solution of the plane Couette flow problem. The assumptions we made in

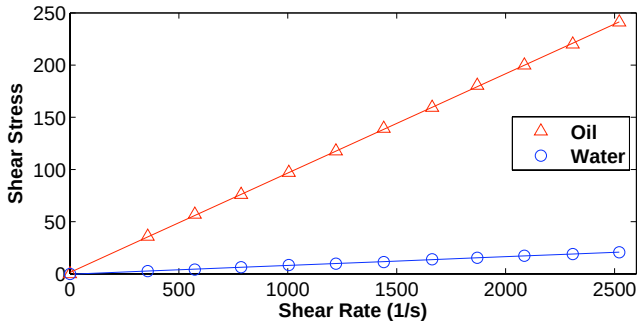


Figure 9.2 Shear stress, as a function of shear rate γ, for water and oil at 25° C (Ellenberger et al. [1976]).

deriving the solutions worked in the sense that we found pressure and velocity functions that satisfy the original steady flow problem given in (9.2), (9.3), along with the stated boundary conditions (9.5), (9.6). So, if the solution to this problem is unique then we have found it. The question is uniqueness. We saw in Chapter 3 that nonlinear problems often have multiple solutions. In such cases the question that arose was whether the solution was asymptotically stable, because even if there are multiple solutions, those that are unstable are effectively unachievable. This question arises in all but the simplest fluid problems because of the inherent nonlinear nature of fluid flow. It has been shown that the solution we have derived for plane Couette flow is linearly stable (Drazin and Reid [2004]). However, it has been found experimentally that as the shear rate increases there is a value where the flow changes dramatically, from the unidirectional flow we found to one that is three-dimensional and turbulent. The appearance of a solution different than the one we derived is due to the experimental setup, where the flow is perturbed sufficiently to cause the turbulence to appear. This means that for higher shear rates the solution we derived is not globally asymptotically sta-

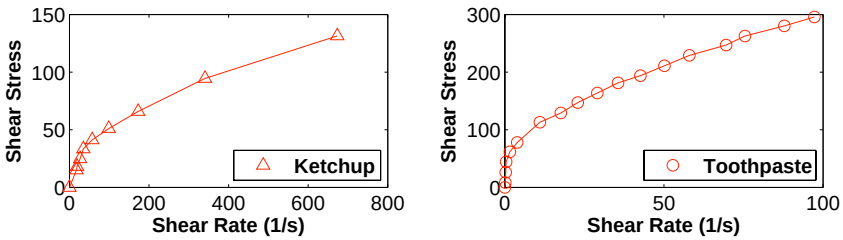


Figure 9.3 Shear stress, as a function of shear rate γ, for ketchup and toothpaste at room temperature (Leong and Yeow [2003]).

ble, and an article discussing the ramifications of this issue can be found in Tillmark and Alfredsson [1992].

9.2.2 Poiseuille Flow

A second steady flow that is often studied involves the fluid motion through a long pipe due to a pressure difference between the two ends of the pipe. The pipe is assumed to have length L, and radius R. Given the geometry, it is easier to use cylindrical coordinates, with the pipe oriented as shown in Figure 9.4. In this case the spatial coordinate system is (r, θ, z), and the associated velocity vector is $\mathbf{v} = (v_r, v_\theta, v_z)$. It should be noted that the subscripts on the components of this vector do not indicate differentiation, but identify the coordinate of the particular velocity component. So, for example, v_r is the velocity in the r-direction.

The boundary conditions for Poiseuille flow are

$$\mathbf{v} = \mathbf{0} \quad \text{on } r = R, \tag{9.10}$$

$$p = p_0, v_r = v_\theta = 0 \quad \text{on } z = 0, \tag{9.11}$$

$$p = p_1, v_r = v_\theta = 0 \quad \text{on } z = L. \tag{9.12}$$

To explain these, (9.10) is the no-slip condition and it applies because the pipe does not move. The conditions at $z = 0, L$ account for the prescribed pressures at these ends, and the assumption that the fluid velocity is only in the axial direction as it enters and leaves the pipe.

To find the solution, we will first consider some of the basic properties of the flow. Given the boundary conditions (9.10) and (9.11), both v_r and v_θ are zero on the pipe and at both ends. Based on this, it is expected that $v_r = v_\theta = 0$ everywhere. Also, there is no θ dependence in the boundary, or the boundary conditions. Because of this it is expected that the axial velocity v_z and pressure p do not depend on the angular coordinate θ. In other words, it is expected that $v_z = v_z(r, z)$ and $p = p(r, z)$. The equations of motion in

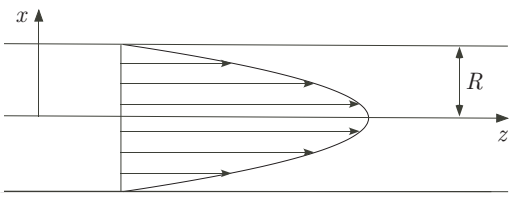


Figure 9.4 In Poiseuille flow, fluid moves through a pipe due to a pressure difference across the ends. Solving this problem shows that the axial velocity of the fluid has a parabolic distribution, as given in (9.14).

cylindrical coordinates, which are given in Appendix E, in this case reduce to

$$\begin{aligned}
\frac{\partial p}{\partial r} &= 0, \\
\rho v_z \frac{\partial v_z}{\partial z} &= -\frac{\partial p}{\partial z} + \mu \left(\frac{\partial^2 v_z}{\partial r^2} + \frac{1}{r}\frac{\partial v_z}{\partial r} + \frac{\partial^2 v_z}{\partial z^2} \right), \\
\frac{\partial v_z}{\partial z} &= 0.
\end{aligned} \tag{9.13}$$

From the first and third equation we conclude that $p = p(z)$ and $v_z = v_z(r)$. In this case (9.13) reduces to

$$\frac{dp}{dz} = \mu \left(\frac{d^2 v_z}{dr^2} + \frac{1}{r}\frac{dv_z}{dr} \right).$$

The left hand side is only a function of z, while the right-hand side is only a function of r. The only way that this can happen is that $p'(z)$ is constant. Given the boundary conditions on the pressure we conclude that $p = p_0 + z(p_1 - p_0)/L$. The remaining equation (9.13) reduces to

$$\mu \left(\frac{d^2 v_z}{dr^2} + \frac{1}{r}\frac{dv_z}{dr} \right) = p_1/L.$$

This is a first order equation for $\frac{d}{dr}v_z$. Using this observation to solve the equation, one finds that the general solution is

$$v_z = \frac{p_1 - p_0}{4\mu L} r^2 + a \ln(r) + b.$$

The solution must be bounded, so $a = 0$, and it must also satisfy the no-slip boundary condition $v_z = 0$ at $r = R$. The resulting axial velocity is therefore

$$v_z = \frac{p_0 - p_1}{4\mu L}(R^2 - r^2). \tag{9.14}$$

This shows that the velocity has a parabolic distribution in the pipe, and this is illustrated in Figure 9.4. The fact that pipe flow has this parabolic shape is demonstrated in Figure 9.5.

It is important to make a point that was also made after solving the plane Couette flow problem. Several simplifying assumptions were made about the velocity and pressure functions, based on the given boundary conditions and geometry of the pipe, to reduce the momentum equations down to (9.13). These assumptions might be better described as educated guesses on the form of the solution. They worked in the sense that we found pressure and velocity functions that satisfy the original steady flow problem given in (9.2), (9.3), along with the stated boundary conditions (9.10)-(9.12). So, if the solution

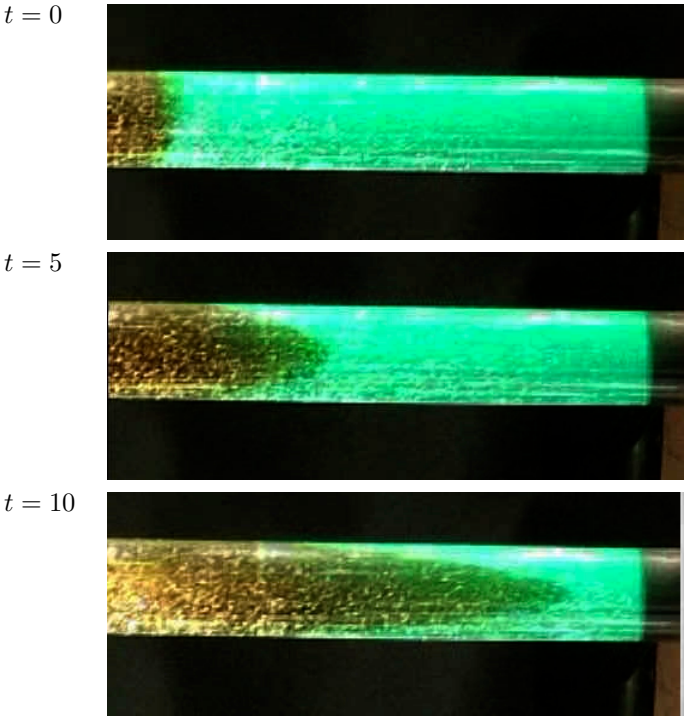


Figure 9.5 Two fluids flowing, from left to right, in a clear pipe (Kunkle [2008]). At $t = 0$ the darker fluid is located at the left end. At $t = 10$ sec the darker fluid shows the parabolic shape predicted by the solution given in (9.14).

to this problem is unique then we have found it. Moreover, an experimental demonstration that the solution has the predicted parabolic profile is shown in Figure 9.5.

As it turns out, experiments show that non-parabolic flow can be obtained in pipe flow. As with the plane Couette problem, for large enough perturbations in the flow, it is found that at high velocities the flow in the pipe can be three-dimensional and turbulent. This does not mean our solution is in question, it just means that it is not globally asymptotically stable at high flow rates. A great deal of effort has been invested into understanding the properties of flow in a pipe, and a recent review of this work can be found in Eckhardt et al. [2007].

9.3 Vorticity

If you float on an inner tube on a river you notice that not only do you move downstream, the moving water also causes you to spin. It is the rotational component of the motion that we are now interested in exploring. The first step is to derive a variable that can be used to measure the rotation, at least locally.

To explain how this is done, consider three fluid particles located on the coordinate axis, at $t = 0$, as shown in Figure 9.6. For simplicity the flow is assumed to be two-dimensional, and the positions of the three particles at $t = \Delta t$ are also shown in the figure. The velocity, at $t = 0$, of the particle located at the origin is $\mathbf{v}_0 = (u_0, v_0)$, where $u_0 = u(0,0)$ and $v_0 = v(0,0)$. The initial velocity of the particle located at $x = \Delta x$ is $\mathbf{v}_1 = (u_1, v_1)$, where $u_1 = u(\Delta x, 0)$ and $v_1 = v(\Delta x, 0)$. We are interested in the case of when Δx and Δt are small. In this case, Taylor's theorem gives us

$$\begin{aligned} u_1 &= u(\Delta x, 0) \\ &= u(0,0) + \Delta x\, u_x(0,0) + \cdots \\ &= u_0 + \Delta x\, u_x + \cdots . \end{aligned}$$

Similarly, $v_1 = v_0 + \Delta x\, v_x + \cdots$. With this, $\mathbf{v}_1 \approx \mathbf{v}_0 + \Delta x(u_x, v_x)$. Now, at $t = \Delta t$ the particle that started at the origin is located at approximately $\Delta t \mathbf{v}_0$, and the one that started at $x = \Delta x$ is at approximately $\Delta t \mathbf{v}_1 \approx \Delta t(\mathbf{v}_0 + \Delta x(u_x, v_x))$. With this

$$\begin{aligned} \tan(\theta_1) &\approx \frac{\Delta t(v_0 + \Delta x v_x) - \Delta t v_0}{\Delta x + \Delta t(u_0 + \Delta x u_x) - \Delta t u_0} \\ &= \frac{\Delta t v_x}{1 + \Delta t u_x} \\ &\approx \Delta t v_x . \end{aligned}$$

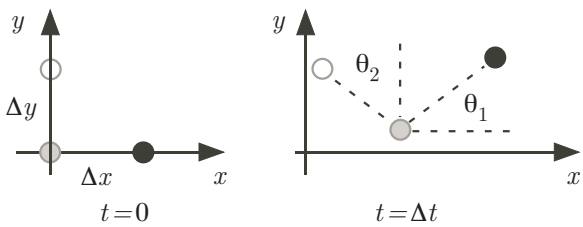


Figure 9.6 Three nearby fluid particles used to introduce the concept of vorticity. Their motion from $t = 0$ to $t = \Delta t$, causes both a translation and relative rotation in their configuration.

Given that the angle is small, so $\tan(\theta_1) \approx \theta_1$, we have that $\theta_1 \approx \Delta t\, v_x$. Carrying out a similar analysis using the particle that started at $y = \Delta y$ one finds that $\theta_2 \approx -\Delta t\, u_y$. The average angular velocity around the z-axis is therefore $(\theta_1 + \theta_2)/(2\Delta t) \approx \frac{1}{2}(v_x - u_y)$. Similar expressions can be derived for the rotation around the other two axes. This is the motivation for introducing the *vorticity* $\boldsymbol{\omega}$, which is defined as

$$\begin{aligned}\boldsymbol{\omega} &= \nabla \times \mathbf{v} \\ &= \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right)\mathbf{i} + \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right)\mathbf{j} + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)\mathbf{k},\end{aligned} \tag{9.15}$$

where $\mathbf{v} = (u, v, w)$. Consequently, $\boldsymbol{\omega}$ is twice the average angular velocity in the three coordinate planes. This also helps explain why $\mathbf{W}$ in (8.68) is known as the vorticity tensor.

Example: Plane Couette Flow

As shown in Figure 9.1, the fluid particles in Couette flow move in straight lines, and, consequently, appear to have no rotational component. However, using the solution (9.7) in (9.15), one obtains $\boldsymbol{\omega} = (0, 0, -\gamma)$. In other words, the vorticity is nonzero. To explain this, plane Couette flow can be thought of using traffic flow on a multilane road, where the fluid particles are the cars. This is shown in Figure 9.7. The slowest lane is at $y = 0$, and the fastest lane is at $y = h$. Given a line of cars that start out at $x = 0$, after a short time they will have a linear distribution as shown in Figure 9.7. A driver in one of middle lanes will see the car on the left a bit farther ahead, and the one on the right a bit farther behind. Hence, from the driver's perspective there has been a rotation in the orientation, the rotation being in the clockwise direction. This gives rise to a negative angular velocity, and this is why the z-component of the vorticity is negative for this flow. This example also shows that nonzero vorticity does not necessarily mean that the fluid particles themselves are rotating. The definition of vorticity assumes nothing about how the fluid particles interact, it only measures their respective orientations as they flow past each other. ■

9.3.1 Vortex Motion

A vortex is a circular flow around a center, and is similar to what is seen in a tornado, hurricane, and in the swirling flow through a drain. To study such motions, it is often convenient to use cylindrical coordinates, and the equations of motion in this coordinate system are given in Appendix E. The coordinates in this system are (r, θ, z), with corresponding velocity $\mathbf{v} = (v_r, v_\theta, v_z)$. Assuming the center of the vortex is the z-axis then to have

circular motion around the z-axis we assume that $v_r = v_z = 0$. This means there is no motion in either the z- or r-direction, and so the fluid particles move on circles centered on the z-axis. Making the additional assumption that $v_\theta = v_\theta(r,t)$ then the equations of motion reduce to

$$\frac{\partial v_\theta}{\partial t} = \nu \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial (r v_\theta)}{\partial r}\right), \tag{9.16}$$

$$\frac{\partial p}{\partial r} = \frac{\rho}{r} v_\theta^2. \tag{9.17}$$

This assumes the fluid is incompressible, and that there are no body forces. The vorticity for this flow is

$$\boldsymbol{\omega} = \left(0, 0, \frac{1}{r}\frac{\partial (r v_\theta)}{\partial r}\right). \tag{9.18}$$

The momentum equation (9.16) is an old friend because it is the radially symmetric diffusion equation given in (4.78). In this case, the kinematic viscosity is the diffusion coefficient. The point source solution given in (4.81) gives rise to what is known as the Taylor vortex. The analysis of this vortex is carried out in Exercise 9.10, while we will investigate a related vortex in the following example.

Example: Oseen-Lamb Vortex

In this flow $v_r = v_z = 0$, and

$$v_\theta = \frac{\alpha}{r}\left(1 - \exp\left(\frac{-r^2}{\beta^2 + 4\nu t}\right)\right). \tag{9.19}$$

It is not hard to show that this function satisfies (9.16), and is therefore an exact solution of the incompressible fluid equations. The pressure is found by integrating (9.17), and the vorticity is calculated using (9.18). One finds that

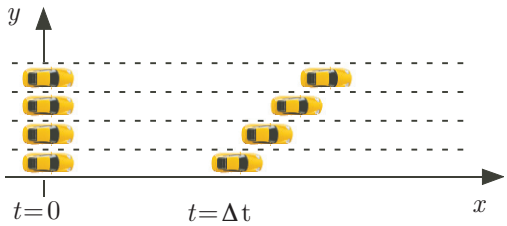


Figure 9.7 Multilane traffic flow analogy used to explain vorticity in plane Couette flow.

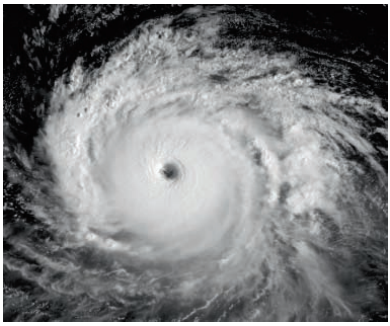


Figure 9.8 The rotational motion of a hurricane is an example of vortex type motion, with the eye containing the central axis.

$$\boldsymbol{\omega} = \left(0, 0, \frac{2\alpha}{\beta^2 + 4\nu t} \exp\left(\frac{-r^2}{\beta^2 + 4\nu t}\right)\right). \tag{9.20}$$

The velocity (9.19) is shown in Figure 9.9 at different time points, for $\alpha = \beta = 4\nu = 1$. This shows that when the vortex starts out, it is confined to the region near $r = 0$. As time passes the vortex slows down, with the maximum velocity moving outward from the center and decreasing in the process. This is due entirely to the viscosity of the fluid, and the result is that in the limit of $t \to \infty$, the vortex disappears. Also note that when this vortex starts out, there is a region near $r = 0$ where there is little motion, which is reminiscent of the eye of the hurricane shown in Figure 9.8. ■

9.4 Irrotational Flow

One of the ideas underlying the introduction of vorticity is that the motion of a fluid can be split into two components, a rotational part and a non-rotational part. How this can be done is obtained from the Helmholtz Representation Theorem, and this will be presented shortly. In preparation for this we introduce the concept of an irrotational flow.

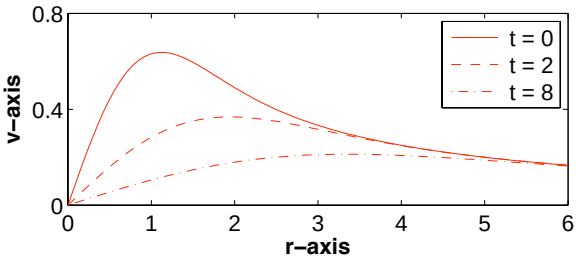


Figure 9.9 Circumferential velocity (9.19) for a Oseen-Lamb vortex.

Definition 9.1. A fluid for which the vorticity is identically zero is said to be *irrotational*. The flow is rotational if the vorticity is nonzero anywhere in the flow.

One might guess that a flow which moves in a straight line is irrotational, but the plane Couette flow example shows that statement is incorrect. It is also incorrect to assume that if the flow is a vortex then it must be rotational. The next example explains why.

Example: Line Vortex

In the special case of when $v_r = v_z = 0$, and $v_\theta = v_\theta(r,t)$, then the vorticity is given in (9.18). This will be zero if rv_θ is constant. Consequently, an irrotational flow is achieved by taking

$$v_\theta = \frac{\alpha}{r}\,, \tag{9.21}$$

where α is a constant. The flow in this case is circular motion around the z-axis, just as it is for the Oseen-Lamb vortex shown in Figure 9.9. This is called a line vortex, and it produces irrotational flow. ■

As the above example clearly demonstrates, rotational motion around the origin does not necessarily mean that the vorticity is nonzero. The reason this is confusing is that vorticity is a local property of the flow, and it is determined by the relative movement of nearby fluid particles. This is not necessarily the same as what is happening to the flow on the macroscopic level. This is why the conclusion coming from the line vortex example, that this particular rotational flow around the origin is irrotational, is not self-contradictory.

One of the difficulties with assuming a flow is irrotational is that it is a statement about the absence of a property, namely no vorticity. The question arises whether it might be possible to characterize the solutions of the Navier-Stokes equation that are irrotational. To answer this, we will make use of the following result, which is known as the Helmholtz Representation Theorem.

Theorem 9.1. *Assume* $\mathbf{q}(\mathbf{x})$ *is a smooth function of* $\mathbf{x}$ *in a domain* D*. In this case, there exists a scalar function* $\phi(\mathbf{x})$ *and a vector function* $\mathbf{g}(\mathbf{x})$ *so that for* $\mathbf{x} \in D$*,*

$$\mathbf{q}(\mathbf{x}) = \nabla\phi + \nabla \times \mathbf{g}, \tag{9.22}$$

where $\nabla \cdot \mathbf{g} = 0$*. The function* ϕ *is called the scalar potential, and* $\mathbf{g}$ *is the vector potential, for* $\mathbf{q}$*.*

The proof of this theorem involves two vector identities and a result from partial differential equations. The first identity is that, given any smooth vector function $\mathbf{h}(\mathbf{x})$,

$$\nabla^2\mathbf{h} = \nabla(\nabla\cdot\mathbf{h}) - \nabla\times(\nabla\times\mathbf{h}).$$

The right hand side of this equation resembles the result in (9.22), where $\phi = \nabla\cdot\mathbf{h}$ and $\mathbf{g} = -\nabla\times\mathbf{h}$. What is needed is to find $\mathbf{h}$ so that $\mathbf{q} = \nabla^2\mathbf{h}$. This is where the result from partial differential equations comes in. Solving $\nabla^2\mathbf{h} = \mathbf{q}$ for $\mathbf{h}$ is known as Poisson's equation, and a particular solution is (Weinberger [1995])

$$\mathbf{h}(\mathbf{x}) = -\frac{1}{4\pi}\iiint\limits_D \frac{\mathbf{q}(\mathbf{s})}{||\mathbf{x}-\mathbf{s}||}\,dV_s, \tag{9.23}$$

where the subscript s indicates integration with respect to $\mathbf{s}$. With this choice for $\mathbf{h}$ we have derived an expression of the form given in (9.22). The only thing left to show is that $\nabla\cdot\mathbf{g} = 0$. This follows because $\mathbf{g} = -\nabla\times\mathbf{h}$ and the vector identity that states, given any smooth vector function $\mathbf{h}(\mathbf{x})$, $\nabla\cdot(\nabla\times\mathbf{h}) = 0$.

The above proof relied on the solution of Poisson's equation, and this requires certain conditions to be satisfied. If the closure of D is a bounded regular region, then the stated assumption that $\mathbf{q}$ is smooth is sufficient. Specifically, what this means is that $\nabla\times\mathbf{q}$ and $\nabla\cdot\mathbf{q}$ are continuous functions. If D is not bounded then the integral requires an additional condition, which is that $\mathbf{q}$ goes to zero faster than $||\mathbf{x}||^{-2}$ as $||\mathbf{x}|| \to \infty$. It is, however, possible to modify the proof so this latter condition is not needed. The details concerning the extension to unbounded domains can be found in Gregory [1996].

Another comment concerning the proof is that it is constructive in the sense that it provides formulas for the potential functions. To explore some of the consequences of this, suppose for the moment that $D = \mathbb{R}^3$. Given that $\phi = \nabla\cdot\mathbf{h}$, then from (9.23) the scalar potential function can be written as

$$\begin{aligned}
\phi &= -\frac{1}{4\pi}\nabla\cdot\iiint \frac{\mathbf{q}(\mathbf{s})}{||\mathbf{x}-\mathbf{s}||}\,dV_s \\
&= -\frac{1}{4\pi}\iiint \mathbf{q}\cdot\nabla_x\frac{1}{||\mathbf{x}-\mathbf{s}||}\,dV_s \\
&= -\frac{1}{4\pi}\iiint \mathbf{q}\cdot\nabla_s\frac{1}{||\mathbf{x}-\mathbf{s}||}\,dV_s \\
&= -\frac{1}{4\pi}\iiint \nabla_s\cdot\frac{\mathbf{q}}{||\mathbf{x}-\mathbf{s}||} - \frac{\nabla_s\cdot\mathbf{q}}{||\mathbf{x}-\mathbf{s}||}\,dV_s \\
&= \frac{1}{4\pi}\iiint \frac{\nabla_s\cdot\mathbf{q}}{||\mathbf{x}-\mathbf{s}||}\,dV_s.
\end{aligned}$$

Carrying out a similar calculation for $\mathbf{g}$ one finds that the vector potential can be written as

$$\mathbf{g} = -\frac{1}{4\pi}\iiint \frac{\nabla_s\times\mathbf{q}}{||\mathbf{x}-\mathbf{s}||}\,dV_s. \tag{9.24}$$

It should be remembered that this is for $D = \mathbb{R}^3$, so any contribution from the boundary of the domain is not accounted for in this formula.

9.4.1 Potential Flow

We are interested in irrotational fluid motion, and what can be learned using the Helmholtz Representation Theorem. Taking $\mathbf{q}$ to be the fluid velocity then $\nabla \times \mathbf{q} = \boldsymbol{\omega}$. For an irrotational flow, so $\boldsymbol{\omega} = \mathbf{0}$, the conclusion we derive from (9.22) and (9.24) is that the velocity has the form

$$\mathbf{v} = \nabla \phi. \tag{9.25}$$

Any flow in which the velocity can be written in this way is called potential flow. To investigate the consequences of this, we will assume that the fluid is incompressible and there are no body forces. The continuity equation $\nabla \cdot \mathbf{v} = 0$ in this case reduces to

$$\nabla^2 \phi = 0. \tag{9.26}$$

This means that the velocity can be found by simply solving Laplace's equation, and this is one of the reasons why potential flow is a centerpiece in most fluid dynamics textbooks. It is important to point out here that nothing has been said about the boundary conditions. These have major repercussions for potential flow, and this will be discussed in more detail shortly.

The pressure p for potential flow is determined by solving the momentum equations. In the x-direction, as given in (8.77), we have that

$$\rho \left(\frac{\partial^2 \phi}{\partial x \partial t} + \frac{\partial \phi}{\partial x} \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial \phi}{\partial y} \frac{\partial^2 \phi}{\partial x \partial y} + \frac{\partial \phi}{\partial z} \frac{\partial^2 \phi}{\partial x \partial z} \right) = -\frac{\partial p}{\partial x} + \mu \nabla^2 \frac{\partial \phi}{\partial x}. \tag{9.27}$$

Given (9.26), then the viscous stress term $\mu \nabla^2 \partial_x \phi$ in the above equation is zero. In other words, for irrotational flow the viscosity does not contribute to the momentum equation. With this, it is possible to rewrite (9.27) in the form

$$\frac{\partial}{\partial x} \left[\frac{\partial \phi}{\partial t} + \frac{1}{2} \left(\frac{\partial \phi}{\partial x} \right)^2 + \frac{1}{2} \left(\frac{\partial \phi}{\partial y} \right)^2 + \frac{1}{2} \left(\frac{\partial \phi}{\partial z} \right)^2 + \frac{1}{\rho} p \right] = 0.$$

Not too surprisingly, the y and z momentum equations show that the y and z derivatives of the above quantity in the square brackets are zero. The conclusion is that

$$\frac{\partial \phi}{\partial t} + \frac{1}{2} \left(\frac{\partial \phi}{\partial x} \right)^2 + \frac{1}{2} \left(\frac{\partial \phi}{\partial y} \right)^2 + \frac{1}{2} \left(\frac{\partial \phi}{\partial z} \right)^2 + \frac{1}{\rho} p$$

is only a function of time. In other words,

$$p = p_0(t) - \rho \left[\frac{\partial \phi}{\partial t} + \frac{1}{2} \left(\frac{\partial \phi}{\partial x} \right)^2 + \frac{1}{2} \left(\frac{\partial \phi}{\partial y} \right)^2 + \frac{1}{2} \left(\frac{\partial \phi}{\partial z} \right)^2 \right], \tag{9.28}$$

or equivalently

$$p = p_0(t) - \rho \left(\frac{\partial \phi}{\partial t} + \frac{1}{2} \nabla \phi \cdot \nabla \phi \right). \tag{9.29}$$

This is known as Bernoulli's theorem for irrotational flow. Once Laplace's equation is solved to find the potential function, (9.25) is used to find the velocity and (9.29) is used to find the pressure.

Example: Line Vortex (cont'd)

Using cylindrical coordinates, then $\mathbf{v} = (v_r, v_\theta, v_z)$, and

$$\nabla \phi = \left(\frac{\partial \phi}{\partial r}, \frac{1}{r} \frac{\partial \phi}{\partial \theta}, \frac{\partial \phi}{\partial z} \right). \tag{9.30}$$

To obtain $v_r = v_z = 0$ it is required that $\phi = \phi(\theta, t)$. To have (9.21) hold it is required that $\frac{\partial \phi}{\partial \theta} = \alpha$. Therefore, the scalar potential function for the line vortex is $\phi = \alpha\theta$. The pressure, obtained from (9.29), is $p = p_0 - \frac{1}{2}\rho\alpha^2/r^2$. ■

One question that has not been addressed is, how realistic is it to assume a flow is irrotational? In applications, in addition to the equations of motion, there are boundary and initial conditions, and these were conveniently ignored when deriving (9.24). The fact is that they can easily ruin the assumption of irrotationality. To explain why, consider the no-slip condition (8.7). This prescribes all three components of the velocity vector on the boundary. The equation to solve for an irrotational flow is Laplace's equation (9.26), from which the velocities are determined using (9.25). Mathematically, for Laplace's equation, one can only impose one condition on the boundary, and

Figure 9.10 The motion of the airflow around a plane generates vorticity into the flow (Morris [2006]). This is evident in the motion of the clouds behind the plane in the photograph.

not three as required from the no-slip condition. The usual choice is to have the solution satisfy the impermeability condition (8.79). Therefore, if the flow is to be irrotational, the other two boundary conditions making up the no-slip condition would have to be selected to be consistent with the resulting solution of Laplace's equation. What this means is that irrotational flow in a viscous fluid is possible, but the boundary conditions have to be just right. An example is the line vortex above, where there are no boundaries, and hence no difficulties trying to satisfy the no-slip condition. Physically, what happens in most fluid problems is that the boundaries generate vorticity, which then spreads into the flow and causes it to be rotational. An example of this is shown in Figure 9.10. One way to avoid this from happening, in addition to adjusting the boundary conditions, is to assume the fluid viscosity is zero. This produces what is known as an inviscid fluid, and this is the subject of the next section.

Because of the complications of the no-slip condition, most textbooks associate potential flow with an inviscid fluid. In fact, to overcome this association, in the research literature the above discussion would be referred to as viscous potential flow, just to make sure to point out that the viscosity has not been assumed to be zero. Those interested in learning about some of the consequences of keeping the viscosity in a potential flow should consult Joseph [2006].

9.5 Ideal Fluid

As seen in Table 9.1, the viscosity of air is much less than it is, for example, for water. It is for this reason that when studying air flow that it is often assumed to be inviscid, which means the viscosity is zero. If, in addition, the fluid is assumed to be incompressible then one has what is called an ideal fluid. The equations of motion in this case are

$$\rho(\partial_t + \mathbf{v} \cdot \nabla)\mathbf{v} = -\nabla p, \tag{9.31}$$

$$\nabla \cdot \mathbf{v} = 0. \tag{9.32}$$

The above system is known as the Euler equations. The absence of viscosity means that the no-slip condition is inappropriate, but the impermeability boundary condition still applies.

Two assumptions are made in this section. One is relatively minor, and it is that there are no body forces. The formulas derived below can be extended in a straightforward manner to include body forces, and the results are given in Exercise 9.15. The second assumption is not minor, and it is that there is a unique solution of the ideal fluid problem and it is smooth. This issue will be discussed again at the end of this section.

Example: Plane Couette Flow Revisited

The lack of viscosity has some interesting consequences. As an example, suppose in the plane Couette flow problem the fluid starts from rest. With a viscous fluid, because of the no-slip condition, when the upper surface starts to move it pulls the nearby fluid with it. After a short amount of time the fluid between the two plates approaches a steady flow, and the solution given in (9.7) applies. This does not happen if the fluid is inviscid. The only boundary condition at $y = h$ is the impermeability condition, which is that the velocity in the vertical direction is zero. The motion of the plate has no effect on the fluid, and so the fluid remains at rest. Therefore, the solution of the plane Couette flow problem for an ideal fluid is simply $\mathbf{v} = \mathbf{0}$ and $p = p_0$. ■

9.5.1 Circulation and Vorticity

An important property of an ideal fluid is that if it starts out irrotational, it is irrotational for all time. To explain why, we start with the surface integral

$$\iint_S \boldsymbol{\omega} \cdot \mathbf{n}\, dA, \tag{9.33}$$

where S is an oriented smooth surface that is bounded by a simple, closed, smooth boundary curve C with positive orientation. This is the vorticity flux across the surface S, and it can be used to measure the vorticity. The first step is to recall Stokes' theorem, which states that

$$\iint_S \nabla \times \mathbf{v} \cdot \mathbf{n}\, dA = \int_C \mathbf{v} \cdot \mathbf{dx}.$$

Using this in (9.33) we have

$$\iint_S \boldsymbol{\omega} \cdot \mathbf{n}\, dA = \int_C \mathbf{v} \cdot \mathbf{dx}. \tag{9.34}$$

It is the last integral that we will work with, and so let

$$\Gamma(t) = \int_C \mathbf{v} \cdot \mathbf{dx}. \tag{9.35}$$

The function $\Gamma(t)$ is called the *circulation*.

Example: Oseen-Lamb Vortex Revisited

The vorticity for the Oseen-Lamb vortex is given in (9.20). Suppose we want to calculate the circulation when the curve C is the circle in the x,y-plane, with radius R and centered at the origin (see Figure 9.11). It is easier, in this case, to use (9.34) and write

$$\Gamma(t) = \iint_S \boldsymbol{\omega} \cdot \mathbf{n}\, dA.$$

Using cylindrical coordinates, the surface S is the disk $r \leq R$ in the plane $z = 0$, and $\mathbf{n} = (0, 0, 1)$. From (9.20), $\boldsymbol{\omega} \cdot \mathbf{n} = 2\alpha q(t) \exp(-q(t)r^2)$, where $q(t) = 1/(\beta^2 + 4\nu t)$, and so

$$\begin{aligned}\Gamma &= \int_0^{2\pi} \int_0^R 2\alpha q(t) \exp(-q(t)r^2) r dr d\theta \\ &= 2\pi\alpha \left[1 - \exp(-q(t)R^2)\right].\end{aligned}$$

Consequently, for a viscous fluid, the circulation starts out with the value $\Gamma_0 = 2\pi\alpha \left[1 - \exp(-R^2/\beta^2)\right]$, and decays to zero as $t \to \infty$. In contrast, for an ideal fluid, so $\nu = 0$, the circulation has the constant value Γ_0. It is this property, that the circulation is constant for an ideal fluid, that is the central idea of the next theorem. ■

Before stating the theorem, the concept of a material curve needs to be explained. Suppose one starts out, at $t = 0$, with a simple closed curve. As time progresses, the material points making up this initial curve move with the fluid, deforming the original shape. Due to the impenetrability of matter assumption, the points never intersect, so the shape remains a simple closed curve. This is what is known as a material curve. With this, we can now state what is known as Kelvin's Circulation Theorem.

Theorem 9.2. *For an ideal fluid, if C is a material curve, then $\frac{d\Gamma}{dt} = 0$.*

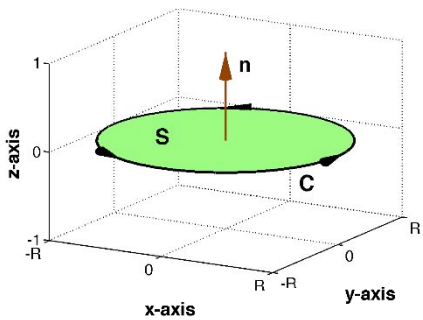


Figure 9.11 Surface, and contour, used to calculate the circulation for an Oseen-Lamb vortex.

To prove this, we need what effectively is a Reynolds transport theorem for line integrals. The first step is to use material coordinates to get the time dependence out of the limits of integration, and so, at $t = 0$ assume the curve C is given as $\mathbf{A} = \mathbf{G}(s)$, for $a \leq s \leq b$. At later times the curve is described as $\mathbf{x} = \mathbf{X}(\mathbf{G}(s), t)$. With this, $\mathbf{dx} = \mathbf{F}\mathbf{G}'(s)\mathbf{ds}$, and so (9.35) becomes

$$\Gamma = \int_a^b \mathbf{V} \cdot \mathbf{F}\mathbf{G}'(s) ds,$$

where $\mathbf{V}$ and $\mathbf{F}$ are evaluated at $\mathbf{A} = \mathbf{G}$ in the above integral. Taking the time derivative yields

$$\frac{d\Gamma}{dt} = \int_a^b \left(\frac{\partial \mathbf{V}}{\partial t} \cdot \mathbf{F}\mathbf{G}'(s) + \mathbf{V} \cdot \frac{\partial \mathbf{F}}{\partial t} \mathbf{G}'(s) \right) ds. \tag{9.36}$$

Using the results from Exercise 8.9, and remembering that $\mathbf{V}$ is evaluated at $\mathbf{A} = \mathbf{G}$, it follows that

$$\begin{aligned} \mathbf{V} \cdot \frac{\partial \mathbf{F}}{\partial t} \mathbf{G}'(s) &= \mathbf{V} \cdot \nabla_A \mathbf{V} \mathbf{G}'(s) \\ &= \frac{d}{ds} \frac{1}{2} (\mathbf{V} \cdot \mathbf{V}). \end{aligned}$$

Given that the curve is closed then (9.36) reduces to

$$\begin{aligned} \frac{d\Gamma}{dt} &= \int_a^b \frac{\partial \mathbf{V}}{\partial t} \cdot \mathbf{F}\mathbf{G}'(s) ds \\ &= \int_C \frac{D\mathbf{v}}{Dt} \cdot \mathbf{dx}. \end{aligned} \tag{9.37}$$

What remains is to recall a property of line integrals. Specifically, given any smooth function ϕ, and a closed curve C, the following holds $\int_C \nabla \phi \cdot \mathbf{dx} = 0$. From (9.32) we have that $\frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho} \nabla p$, where ρ is constant. Therefore, from (9.37), we have that $\frac{d\Gamma}{dt} = 0$.

The above result will enable us to make the stated conclusion about the irrotationality of an ideal fluid. The following result is known as Helmholtz's Third Vorticity Theorem.

Theorem 9.3. *If an ideal fluid is irrotational at $t = 0$, then it is irrotational for all time.*

The proof of this starts with using Stokes' theorem to write the circulation as

$$\Gamma = \iint_S \boldsymbol{\omega} \cdot \mathbf{n} \, dA.$$

This shows that because $\boldsymbol{\omega} = \mathbf{0}$ at $t = 0$, then $\Gamma = 0$ at $t = 0$. Given that Γ is constant, it follows that $\Gamma = 0$ for all time. To use this observation to prove the vorticity is always zero, suppose $\boldsymbol{\omega}$ is nonzero at some point in the flow. In this case, given that $\boldsymbol{\omega}$ is continuous, it is possible to find a small surface containing this point for which the above integral is nonzero. This is a contradiction, and therefore $\boldsymbol{\omega}$ must be zero everywhere.

As an example, the above theorem shows that an ideal fluid which starts at rest is irrotational for all time. The reason is that because $\mathbf{v} = \mathbf{0}$ at $t = 0$ then $\boldsymbol{\omega} = \mathbf{0}$ at $t = 0$.

9.5.2 Potential Flow

What we have been able to show is that if the fluid is irrotational at $t = 0$, then it is possible to introduce a potential function ϕ so that

$$\mathbf{v} = \nabla\phi, \tag{9.38}$$

and

$$p = p_0(t) - \rho\left(\frac{\partial\phi}{\partial t} + \frac{1}{2}\nabla\phi\cdot\nabla\phi\right). \tag{9.39}$$

To find ϕ one solves

$$\nabla^2\phi = 0, \tag{9.40}$$

along with the appropriate boundary conditions. For example, at a solid boundary surface the impermeability boundary condition (8.79) is imposed. If the boundary S is not moving then, given (9.38), the resulting boundary condition is

$$\nabla\phi\cdot\mathbf{n} = 0 \quad \text{on } S, \tag{9.41}$$

or equivalently

$$\frac{\partial\phi}{\partial n} = 0 \quad \text{on } S. \tag{9.42}$$

If the problem involves a pressure boundary condition, then the corresponding boundary condition for ϕ is obtained using (9.39). However, this can make the problem much harder to solve because the $\nabla\phi\cdot\nabla\phi$ term causes the problem to be nonlinear.

Any fluid flow in which the velocity satisfies (9.38) is known as potential flow. Although this might seem obvious, it differs from the definition used in some textbooks on fluid dynamics, where a potential flow is defined as "an irrotational flow in an inviscid and incompressible fluid." The reason for including these additional qualifications, as explained at the end of Section 9.4.1, is the difficulty of obtaining a potential flow when the fluid is viscous. However, it is inappropriate to include them. The reason is that potential

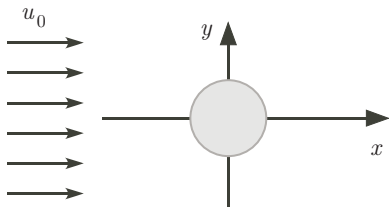


Figure 9.12 Cross-section for uniform flow past a cylinder.

flow is a statement about a fluid's motion, while the statement that it is inviscid is an assumption about its material properties.

We have been making a series of simplifying assumptions in this chapter, attempting to obtain a more tractable mathematical problem. By this measure, we have been extraordinarily successful because we have reduced a coupled system of nonlinear equations down to the single linear equation in (9.40). This has been done by excluding the effects of viscosity, and assuming the flow is irrotational. This degree of simplification helps explain the interest in potential flow. It is also why textbooks on the applications of complex variables inevitably have a chapter on fluid flow, although they must limit their analysis to flow in two dimensions. The question is, however, just how realistic is it to assume a potential flow? The next example will shed some light on this topic.

Example: Potential Flow Past a Cylinder

Consider air flow over a solid cylinder of radius R centered on the z-axis, as shown in Figure 9.12. It is assumed that the flow is from left to right, and the specific condition is that $\mathbf{v} = (u_0, 0, 0)$ as $x \to -\infty$. The flow must also satisfy the impermeability condition on the surface of the cylinder, and this means that $\mathbf{v} \cdot \mathbf{n} = 0$ on $||\mathbf{x}|| = R$, where $\mathbf{n}$ is the unit normal to the boundary. Given the geometry and flow at infinity it is reasonable to expect there is no flow in the z-direction, and the potential function is independent of z. With this, Laplace's equation in cylindrical coordinates becomes

$$\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r}\frac{\partial \phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \phi}{\partial \theta^2} = 0, \text{ for } r > R. \tag{9.43}$$

The impermeability condition (9.42) takes the form

$$\frac{\partial \phi}{\partial r} = 0, \text{ for } r = R, \tag{9.44}$$

and the flow at infinity requires that

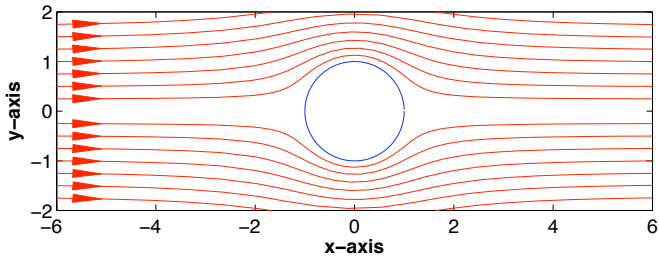


Figure 9.13 Flow lines around a cylinder in uniform flow, calculated using the potential function given in (9.46). In this calculation, $u_0 = R = 1$.

$$\left.\begin{aligned} \cos\theta\frac{\partial\phi}{\partial r} + \sin\theta\frac{1}{r}\frac{\partial\phi}{\partial\theta} &= u_0, \\ \sin\theta\frac{\partial\phi}{\partial r} + \cos\theta\frac{1}{r}\frac{\partial\phi}{\partial\theta} &= 0. \end{aligned}\right\} \quad \text{for } r \to \infty \tag{9.45}$$

This is one of the few unbounded domain problems for which the method of separation of variables can be used. So, assuming that $\phi(r,\theta) = F(r)G(\theta)$ one finds from (9.43) that $F = \alpha r^n + \beta r^{-n}$ and $G = A\cos(n\theta) + B\sin(n\theta)$. Because the solution must be 2π periodic in θ, it is required that n be a positive integer. From (9.44) it follows that $\beta = \alpha R^{2n}$. Imposing (9.45) yields $n = 1$, $\alpha A = u_0$, and $B = 0$. The resulting potential is

$$\phi = u_0\cos\theta\left(r + \frac{R^2}{r}\right). \tag{9.46}$$

The velocity field is therefore

$$\begin{aligned} v_r &= \frac{\partial\phi}{\partial r} = u_0\cos\theta\left(1 - \frac{R^2}{r^2}\right), \\ v_\theta &= \frac{1}{r}\frac{\partial\phi}{\partial\theta} = -u_0\sin\theta\left(1 + \frac{R^2}{r^2}\right). \end{aligned}$$

It is possible to determine the paths of individual fluid particles by solving (9.4) using the above velocity functions. This is easily done numerically, and the results from this calculation are shown in Figure 9.13. ■

If air can be assumed to be an ideal fluid, then it would seem that potential flow could be used in aerodynamics to help understand flight. As an example, you could think of Figure 9.13 as the flow around an airplane wing that has a circular cross-section. You also might think that this is not particularly

realistic because cross-sections of airplane wings are relatively thin, to help reduce the drag and increase lift. Well, let's see about this. The pressure is determined by substituting (9.46) into (9.39), yielding

$$p = \frac{1}{2}\rho\left(\frac{u_0 R}{r}\right)^2\left(4\cos^2\theta - 2 - \frac{R^2}{r^2}\right). \tag{9.47}$$

The force on the circular-cross section is

$$\mathbf{F} = -\int_C p\mathbf{n}ds,$$

where C is the boundary circle $x^2 + y^2 = R^2$, and $\mathbf{n}$ is the unit outward normal to the circle. The x and y components of this force are

$$F_x = -R\int_0^{2\pi} p(R,\theta)\cos(\theta)d\theta,$$
$$F_y = -R\int_0^{2\pi} p(R,\theta)\sin(\theta)d\theta.$$

A straightforward calculation shows that both integrals are zero. In other words, the drag F_x, and the lift F_y, are both zero. As it turns out, this happens with any shape, as long as the fluid is ideal and the flow is steady and irrotational. This is clearly at odds with what is expected, and it is known as d'Alembert's paradox. It is possible to produce lift, an essential requirement to be able to fly, if the flow is rotational. This result is known as the Kutta-Joukowski theorem, but as we saw earlier, it is impossible to get an ideal fluid to be rotational if you start from rest. In other words, if you strap on a pair of wings and starting running in still air, there is no way you are taking off, no matter how fast you are able to run. What this means is that if you want your airplane to fly it is essential that the fluid is viscous. Or, more precisely, that the contribution of the viscosity in generating vorticity from the solid surface of the wing is accounted for in the model. One method how this can be done is explained in the next section.

9.5.3 End Notes

An important issue that arises when assuming the fluid is inviscid concerns the regularity of the solution. Viscosity acts to smooth out jumps and other irregular behavior. By not having viscosity, we have equations similar to those used to model traffic flow. This means that shock wave solutions are possible, and the uniqueness of the solution is an issue. In traffic flow we introduced the entropy condition to determine uniqueness, but for multidimensional fluid problems there are still questions related to the appropriate condition. In fact,

there are several open problems associated with the Euler equations. One that has generated considerable interest is the Euler blow-up problem. It is suspected that the solution of the three-dimensional Euler equations develops a singularity in finite time, but no one has been able to prove this assertion. This means that most of the evidence has come from numerical solutions, but even this has been contradictory. An interesting survey of the blow-up problem, as well as other aspects of the Euler equations, can be found in the proceedings of the conference, Euler Equations: 250 Years On (Eyink et al. [2008]).

9.6 Boundary Layers

The assumption that a fluid is inviscid corresponds to setting the viscosity equal to zero in the Navier-Stokes equation (8.77). What drops out of the equation in this case is the highest spatial derivative in the problem. As we found in Section 2.4, this is the type of limit that is associated with the appearance of a boundary layer. The fact that boundary layers might occur in the flow of a viscous fluid is not surprising given the rapid transitions shown in Figure 8.7. However, the situation is not as straightforward as what occurred in Chapter 2, because the viscous fluid problem is time dependent, and it is not clear what exactly the assumption "small viscosity" means. To get started, we will consider a example that illustrates what happens in a time-dependent flow.

9.6.1 Impulsive Plate

This example is known as Stokes' first problem, and it is one of the few time dependent solutions known for the Navier-Stokes equation. It is assumed that the fluid is incompressible, has no body forces, and it occupies the region $y > 0$. Also, it is at rest for $t < 0$, and at $t = 0$ the lower boundary, at $y = 0$, is given the constant velocity $\mathbf{v} = (u_0, 0, 0)$. This situation is similar to the plane Couette flow problem, in the sense that a planar boundary surface produces a flow in the x-direction. For this reason, the argument used to solve the Couette flow problem can be used here. Assuming that $\mathbf{v} = (u(y,t), 0, 0)$, then the problem reduces to solving

$$\begin{aligned} \rho(u_t + u\partial_x u) &= -\partial_x p + \mu \partial_y^2 u, \\ 0 &= -\partial_y p, \\ \partial_x u &= 0. \end{aligned}$$

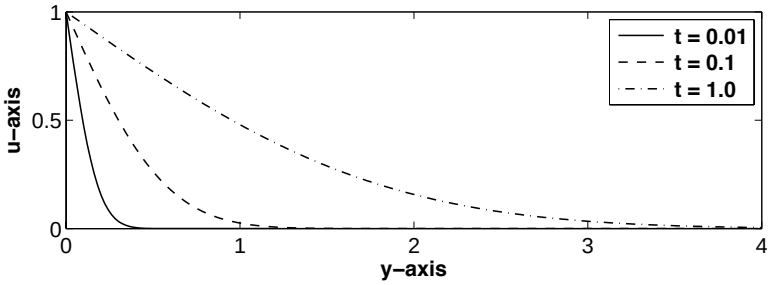


Figure 9.14 Solution (9.49) of the impulsive plate problem, at three time values.

As before, it follows that p is constant, and the entire problem reduces to solving

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial y^2}\,, \tag{9.48}$$

where $u(y,0) = 0$, $u(0,t) = u_0$, and $u(\infty,t) = 0$. Also, $\nu = \mu/\rho$ is the kinematic viscosity. This diffusion problem was solved in Section 1.4 using a similarity variable. The solution, given in (1.61), is

$$u(y,t) = u_0 \operatorname{erfc}\left(\frac{y}{2\sqrt{\nu t}}\right), \tag{9.49}$$

where

$$\operatorname{erfc}(\eta) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{\eta} e^{-s^2}\, ds. \tag{9.50}$$

This solution is shown in Figure 9.14 at three time values, in the case of when $u_0 = \nu = 1$. What is seen is that the effect of the moving plate is initially located near $y = 0$, which is expected of a boundary layer. However, as time passes the effects spread through the fluid domain, and this is due to the diffusive nature of the viscous stress. This gives rise to what is known as a diffusive boundary layer. To quantify what this means, in the engineering literature the boundary layer thickness is defined to be the distance between the boundary and the point where the velocity is 1% of the imposed value. Given that $\operatorname{erfc}(\eta) = 0.01$ for $\eta \approx 1.8$, then the boundary layer thickness in this problem is approximately $y = 3.6\sqrt{\nu t}$. Consequently, this layer grows and spreads through the fluid region.

The existence of a boundary layer separates the flow into an inner and outer region. In the outer region the fluid can be approximated to be inviscid, and the viscous effects are confined to the inner, or boundary layer, region. This observation is routinely used in the numerical solution of the Navier-Stokes equation, because the resolution needed in the inviscid region is usually much less than what is needed in the boundary layer. This is seen in Figure

??, where the grid structure near the surface of the plane is much finer than the one used in the outer, inviscid, flow region.

9.6.2 Blasius Boundary Layer

The next boundary layer example involves the steady flow over a stationary flat plate (see Figure 9.15). The plate occupies the plane $y = 0$, for $0 < x < L$, and the flow is coming in from the left. Assuming that the flow is steady, and that there is no flow in the z-direction, then the fluid equations are

$$\rho\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right) = -\frac{\partial p}{\partial x} + \mu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right),$$
$$\rho\left(u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right) = -\frac{\partial p}{\partial y} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right),$$
$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0.$$

The boundary conditions are

$$\mathbf{v} = (0, 0, 0) \quad \text{on } y = 0, 0 < x < L,$$
$$\mathbf{v} = (u_0, 0, 0) \quad \text{for } y \to -\infty.$$

To undertake a boundary layer analysis we must nondimensionalize the problem. This is done by letting $x = L\bar{x}$, $y = L\bar{y}$, $u = u_0\bar{u}$, $v = u_0\bar{v}$, and $p = p_c\bar{p}$, where $p_c = \rho u_0^2$. In this case the equations of motion become

$$\left(\bar{u}\frac{\partial \bar{u}}{\partial \bar{x}} + \bar{v}\frac{\partial \bar{u}}{\partial \bar{y}}\right) = -\frac{\partial \bar{p}}{\partial \bar{x}} + \epsilon^2\left(\frac{\partial^2 \bar{u}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{u}}{\partial \bar{y}^2}\right), \tag{9.51}$$
$$\left(\bar{u}\frac{\partial \bar{v}}{\partial \bar{x}} + \bar{v}\frac{\partial \bar{v}}{\partial \bar{y}}\right) = -\frac{\partial \bar{p}}{\partial \bar{y}} + \epsilon^2\left(\frac{\partial^2 \bar{v}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{v}}{\partial \bar{y}^2}\right), \tag{9.52}$$
$$\frac{\partial \bar{u}}{\partial \bar{x}} + \frac{\partial \bar{v}}{\partial \bar{y}} = 0, \tag{9.53}$$

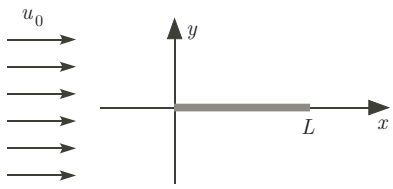


Figure 9.15 Flow over a flat plate used to study viscous boundary layers.

where

$$\epsilon^2 = \frac{\mu}{\rho L u_0}\,. \tag{9.54}$$

From (1.20), we have that $\epsilon^2 = 1/Re$. In other words, ϵ^2 is the inverse of the Reynolds number for the flow. Our assumption that the viscosity is small translates into the assumption that the Reynolds number is large. As an example, consider the flow over an airplane wing. The width of the wing on the Boeing 787 is 18 ft (5.5 m) and cruises at a speed of 561 mph (903 km/h). In this case, $Re = 4 \times 10^7$, which certainly qualifies as high Reynolds number flow.

The reduction of the above problem will closely follow the format used in Section 2.4, although the calculations are a bit more involved.

Outer Solution

The expansion in this region is assumed to have the form $\bar{\mathbf{v}} \sim \bar{\mathbf{v}}_0 + \epsilon\bar{\mathbf{v}}_1 + \cdots$ and $\bar{p} \sim \bar{p}_0 + \epsilon\bar{p}_1 + \cdots$. The problem for the first term, obtained by setting $\epsilon = 0$ in (9.51) - (9.53), is the problem for an inviscid flow. The solution is just $\mathbf{v}_0 = (u_0, 0, 0)$, and $\bar{p}_0$ is a constant. It is assumed, for simplicity, that $\bar{p}_0 = 0$.

Boundary Layer Solution

The boundary layer coordinate is

$$Y = \frac{\bar{y}}{\epsilon}\,.$$

As in Section (2.4), capitals will be used to designate the dependent variables in the boundary layer region. With this, (9.51) - (9.53) take the form

$$\left(\bar{U}\frac{\partial \bar{U}}{\partial \bar{x}} + \frac{1}{\epsilon}\bar{V}\frac{\partial \bar{U}}{\partial Y}\right) = -\frac{\partial \bar{P}}{\partial \bar{x}} + \epsilon^2\frac{\partial^2 \bar{U}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{U}}{\partial Y^2}\,, \tag{9.55}$$

$$\left(\bar{U}\frac{\partial \bar{V}}{\partial \bar{x}} + \frac{1}{\epsilon}\bar{V}\frac{\partial \bar{V}}{\partial Y}\right) = -\frac{1}{\epsilon}\frac{\partial \bar{P}}{\partial Y} + \epsilon^2\frac{\partial^2 \bar{V}}{\partial \bar{x}^2} + \frac{\partial^2 \bar{V}}{\partial Y^2}\,, \tag{9.56}$$

$$\frac{\partial \bar{U}}{\partial \bar{x}} + \frac{1}{\epsilon}\frac{\partial \bar{V}}{\partial Y} = 0, \tag{9.57}$$

The appropriate expansions in this case are $\bar{U} \sim \bar{U}_0 + \cdots$, $\bar{V} \sim \epsilon(\bar{V}_0 + \cdots)$, and $\bar{P} \sim \bar{P}_0 + \cdots$. Introducing these into (9.55) - (9.57), and letting $\epsilon \to 0$ we obtain

$$\left(\bar{U}_0\frac{\partial \bar{U}_0}{\partial \bar{x}} + \bar{V}_0\frac{\partial \bar{U}_0}{\partial Y}\right) = -\frac{\partial \bar{P}_0}{\partial \bar{x}} + \frac{\partial^2 \bar{U}_0}{\partial Y^2}\,, \tag{9.58}$$

$$\frac{\partial \bar{P}_0}{\partial Y} = 0\,, \tag{9.59}$$

$$\frac{\partial \bar{U}_0}{\partial \bar{x}} + \frac{\partial \bar{V}_0}{\partial Y} = 0. \tag{9.60}$$

From the no-slip condition on the plate, it is required that

$$(\bar{U}_0, \bar{V}_0) = (0, 0) \quad \text{on } Y = 0, 0 < \bar{x} < 1. \tag{9.61}$$

Moreover, the solution must match with the outer solution, and for this reason it is required that

$$\bar{U}_0 \to 1 \quad \text{and} \quad \bar{P}_0 \to 0 \quad \text{as } Y \to \infty, 0 < \bar{x} < 1. \tag{9.62}$$

There is a matching condition for $\bar{V}_0$, but it is not needed at the moment and this will be explained after the solution is derived.

From (9.59) and (9.62) it follows that $\bar{P}_0 = 0$. The usual method for finding the velocity functions is to introduce a stream function $\psi(\bar{x}, Y)$, which is defined so that

$$\bar{U}_0 = \frac{\partial \psi}{\partial Y}, \tag{9.63}$$

$$\bar{V}_0 = -\frac{\partial \psi}{\partial \bar{x}}. \tag{9.64}$$

By doing this, the continuity equation (9.57) is satisfied automatically. This leaves the momentum equation (9.55), which reduces to

$$\frac{\partial \psi}{\partial Y}\frac{\partial^2 \psi}{\partial Y \partial \bar{x}} - \frac{\partial \psi}{\partial \bar{x}}\frac{\partial^2 \psi}{\partial Y^2} = \frac{\partial^3 \psi}{\partial Y^3}. \tag{9.65}$$

The boundary (9.61) and matching (9.62) conditions transform into the following

$$\frac{\partial \psi}{\partial Y} = \frac{\partial \psi}{\partial \bar{x}} = 0, \quad \text{on } Y = 0, \tag{9.66}$$

and

$$\frac{\partial \psi}{\partial Y} \to 1, \quad \text{as } Y \to \infty. \tag{9.67}$$

Something that was not explained above is where the idea of using a stream function comes from. The answer is the Helmholtz Representation Theorem (9.22). When the flow is incompressible, and two-dimensional as in the present example, then the velocity vector can be written as $\mathbf{v} = \nabla \times \mathbf{g}$, where $\mathbf{g} = (0, 0, \psi)$. Expanding the curl, one obtains $\mathbf{v} = (\partial_y \psi, -\partial_x \psi, 0)$, and this gives rise to the stream function.

It is not possible to find an analytical solution of the above problem for the stream function. However, it is possible to come close if we make one more assumption. Instead of a plate of finite length, we assume that the plate is semi-infinite and occupies the interval $0 \le \bar{x} < \infty$. This gives rise to what is known as the Blasius boundary layer problem, and it can be reduced by introducing a similarity variable. Specifically, assuming that $\psi = \sqrt{\bar{x}} f(\eta)$, where $\eta = Y/\sqrt{\bar{x}}$, then (9.65) reduces to

$$f''' + \frac{1}{2} f f'' = 0, \text{ for } 0 < \eta < \infty, \tag{9.68}$$

where (9.66) and (9.67) become

$$f(0) = f'(0) = 0, \quad \text{and} \quad f'(\infty) = 1. \tag{9.69}$$

One might argue that we have not made much progress, because the solution of the above problem is not known. However, the ordinary differential equation (9.68) is certainly simpler than the partial differential equation (9.65), and this does provide some benefit. For example, it is much easier to solve (9.68) numerically than it is to solve (9.65) numerically. Just one last comment to make here, before working out an example, is that once the function f is determined then the velocity functions are calculated using the formulas

$$\bar{U}_0 = f'(\eta), \tag{9.70}$$

$$\bar{V}_0 = -\frac{1}{2\sqrt{\bar{x}}}(f - \eta f'). \tag{9.71}$$

These expressions are obtained by substituting the similarity solution into (9.63) and (9.64).

Example: Numerical Solution

To use a numerical method to solve (9.68) it is a bit easier to rewrite the equation as a system by letting $g = f'$. In this case the equation can be written as

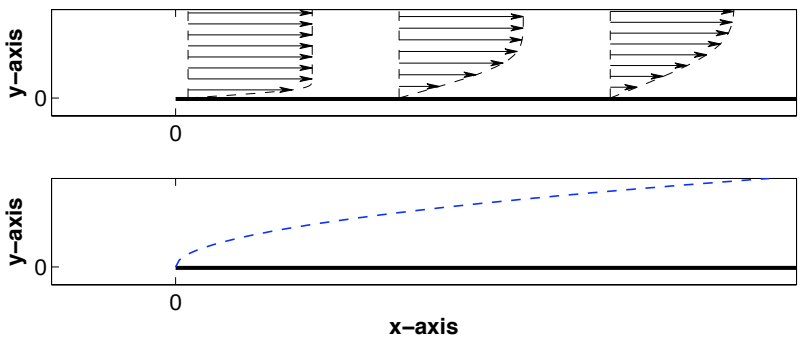


Figure 9.16 Flow over a flat plate, as determined from solving (9.68), (9.69). The upper graph shows $\bar{U}_0$, as a function of Y, at three points on the plate. The dashed curve in the lower graph is where $\bar{U}_0 = 0.99$.

$$f' = g,$$
$$g'' = -\frac{1}{2} f g'.$$

The boundary conditions (9.69) become $g(0) = 0$, $g(\infty) = 1$, and $f(0) = 0$. With this, it is relatively straightforward to use finite differences to solve the problem (Holmes [2005]). The result of such a calculation is given in Figure 9.16. The upper graph shows the horizontal velocity u at three locations along the plate. As required, the velocity is zero on the plate, and as the vertical distance from the plate increases it approaches the constant velocity of the outer region. It is also evident that the velocity reaches this constant value fairly quickly for a point on the plate that is near the leading edge, where $\bar{x} = 0$, and less so as the distance from the leading edge increases. The reason is that the boundary layer on the plate grows with distance from the leading edge. Using the engineering definition that the boundary layer thickness is where the flow reaches 99% of the outer flow value, the dashed curve shown in the lower graph is obtained. The shape of this curve can be explained using (9.71). By definition, the dashed curve is where $\bar{U}_0 = 0.99$, and this means that $f'(\eta) = 0.99$. Letting the solution of this equation be η_0 then, because $\eta = Y/\sqrt{\bar{x}}$, we have that the dashed curve is $Y = \eta_0\sqrt{\bar{x}}$. ■

The above example illustrates how a flow can be separated into an outer, inviscid, region, and a boundary layer where the viscous affects play an important role. This requires a large value for the Reynolds number, and does not hold for a low Reynolds number. It is also based on the solution for an infinitely long plate, something that is rather rare in the real world. When the plate has finite length, a wake is formed downstream from the plate. An example of this is shown in Figure 9.17. The pattern seen in the wake is known as a Karman vortex street. It is also possible to see the boundary layer on the plate in the upper figure. What is interesting is that the fluid used in this experiment is water, and not air. This is indicative of the fact that the separation of the flow into inviscid and boundary layer domains is a charac-

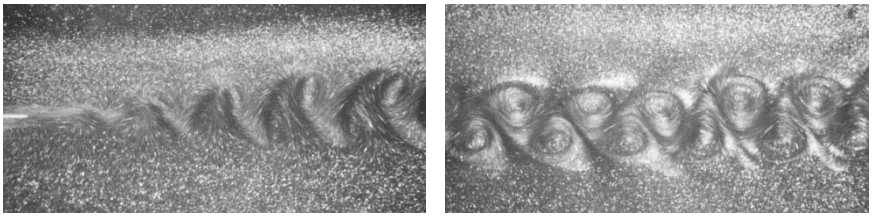


Figure 9.17 Wake behind a flat plate, showing the vortices generated in the flow (Tanaka [1986]). The photograph of the left is the flow immediately behind the plate, and the one on the right is further downstream. The vortices are evident because aluminum particles are suspended in the flow. In this experiment, $Re = 15800$.

teristic of any fluid governed by the Navier-Stokes equations, assuming the Reynolds number is sufficiently large. It is also evident, given the complexity of the flow, that finding the solution for the finite plate requires numerical methods. Some of the issues that arise with this are discussed in Cebeci and Cousteix [2005].

Before closing this section, a couple of comments are needed about the boundary layer reduction. First, the flow in the immediate vicinity of the leading edge requires a more refined boundary layer analysis that was used here. The same comment applies to the trailing edge for the finite length plate. Second, there are questions remaining about the matching requirement for the vertical velocity. In particular, there must be a matching condition, yet it is not included in (9.62). This is an issue, because according to (9.71), it appears that the vertical velocity is unbounded when one moves out of the boundary layer into the outer region. Namely, given that $f'(\infty) = 1$ then $\eta f'$ is unbounded as $\eta \to \infty$. In comparison, we know that the vertical velocity in the inviscid region is just zero. Therefore, to guarantee that the vertical velocity matches it must be that $f \sim \eta$ as $\eta \to \infty$. If the solution of (9.68) does not do this then the whole approximation fails. It is found, from the numerical solution, that f does indeed have the correct limiting behavior, and so the expansions match.

Exercises

9.1. Suppose an incompressible viscous fluid has velocity $\mathbf{v} = (u, v, 0)$, with $u = ax^2 + bxy + cy^2$, where a, b, and c are constant.
(a) Find v assuming that $v(x, 0, z) = 0$.
(b) Find $\mathbf{T}$.
(c) For what values of a, b, and c, if any, is the flow irrotational?

9.2. Suppose the velocity for an incompressible fluid is $\mathbf{v} = (-\alpha y, \alpha x, \beta)$, where α and β are constants.
(a) Show v satisfies the continuity equation.
(b) Assuming no external body forces, find the pressure.
(c) Is this flow rotational or irrotational?
(d) Find the pathlines.
(e) This is known as steady helical flow. Why?

9.3. Suppose the velocity for an incompressible fluid is $\mathbf{v} = (x + y, 3x - y, 0)$.
(a) Show v satisfies the continuity equation.
(b) Assuming no external body forces, find the pressure.
(c) Is this flow rotational or irrotational?
(d) Find the pathlines.
(e) Use the result from part (d) to find the material description of the flow.
(f) Find the invariants for $\mathbf{D}$.

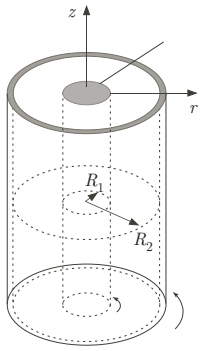


Figure 9.18 Concentric rotating cylinders used in the Taylor-Couette problem in Exercise 9.6.

9.4. This problem considers some of the limitations on the method used to solve the steady flow equations.

(a) Suppose in the Poiseuille flow problem in Section 9.2.2 that the pipe has an elliptical cross-section. What assumptions about the solution used to derive (9.13) no longer apply? What assumptions should still be valid?

(b) Suppose in the plane Couette flow problem in Section 9.2.1 that gravity is included. This means that a forcing function must be included in (9.2), as determined from (8.77), of the form $\mathbf{f} = (0, -g, 0)$. What assumptions about the solution used to derive (9.7) no longer apply? What assumptions should still be valid?

9.5. As a modification of the plane Couette flow problem, suppose there are two fluids between the plates. One fluid occupies the region $0 < y < h_0$, and has density ρ_1 and viscosity ν_1. The second fluid occupies the region $h_0 < y < h$ and has density ρ_2 and viscosity ν_2.

(a) In plane Couette flow the velocity has the form $\mathbf{v} = (u(y), 0, 0)$. Also, at the interface, where $y = h_0$, the velocity and stress are assumed to be continuous. Use this to show that p, u and $u'(y)$ are continuous at $y = h_0$.

(b) Using the results from part (a), solve this plane Couette problem.

9.6. An incompressible viscous fluid occupies the region between two concentric cylinders of radii R_1 and R_2, where $R_1 < R_2$. Assume the cylinders are infinitely long, and centered on the z-axis (see Figure 9.18). The inner cylinder is assumed rotating around the z-axis with angular velocity ω_1, while the outer cylinder rotates around the z-axis with angular velocity ω_2. The flow is assumed to be steady, and there are no body forces. This is known as the Taylor-Couette problem.

(a) Using cylindrical coordinates, explain why the boundary conditions on the cylinders are $(v_r, v_\theta, v_z) = (0, \omega_i R_i, 0)$ for $r = R_i$.

(b) Explain why it is reasonable to assume that the solution has $v_z = 0$ and $v_r = 0$.

(c) Find v_θ and p.
(d) What is the vorticity for this flow? With this show that the flow is irrotational if $R_1^2\omega_1 = R_2^2\omega_2$.

9.7. This problem examines a model for power-law fluids. It is based on the observation coming from Figure 9.3 that the shear stress for plane Couette flow has the form $T_{12} = \alpha(\frac{\partial u}{\partial y})^\beta$. It is assumed here that the fluid is incompressible.
(a) In plane Couette flow the velocity has the form $\mathbf{v} = (u(y), 0, 0)$. What are $\mathbf{D}$ and its three invariants in this case?
(b) As shown in Section 8.10.2.1, the general form of the constitutive law for a nonlinear viscous fluid is $\mathbf{T} = -p\mathbf{I} + \mathbf{G}$, where $\mathbf{G} = \alpha_0\mathbf{I} + \alpha_1\mathbf{D} + \alpha_2\mathbf{D}^2$. Explain how the power-law

$$T_{12} = \alpha \left|\frac{\partial u}{\partial y}\right|^m \frac{\partial u}{\partial y}$$

is obtained by assuming that $\alpha_0 = \alpha_2 = 0$ and α_1 depends on II_D in a particular way. It is assumed that $m > -1$, which guarantees that $\mathbf{G} = \mathbf{0}$ if $\mathbf{D} = \mathbf{0}$.
(c) Assuming that $\frac{\partial u}{\partial y} > 0$, and using the constitutive law from part (b), solve the resulting plane Couette flow problem. From this show that $T_{12} = \alpha\gamma^{m+1}$, where γ is given in (9.8).
(d) On the same axes, sketch T_{12} as function of γ when $-1 < m < 0$, when $m = 0$, and when $1 < m$. Use this to compare the differences in the behavior of the shear stress for large values of γ. Would the $-1 < m < 0$ case be called a shear-thickening or a shear-thinning situation?

9.8. This problem examines the vorticity for a linear flow, which means that $\mathbf{v} = \mathbf{H}\mathbf{x} + \mathbf{h}$, where the matrix $\mathbf{H}$ and vector $\mathbf{h}$ can dependent on t. Other properties of linear flows were developed in Exercises 8.4 and 8.5.
(a) Show that $\boldsymbol{\omega} = (H_{32} - H_{23}, H_{13} - H_{31}, H_{21} - H_{12})$. What is the vorticity when $\mathbf{H}$ is symmetric?
(b) Show that for rigid body motion, as given in (8.13), $\mathbf{H} = \mathbf{Q}'\mathbf{Q}^T$ and $\mathbf{h} = \mathbf{b}' - \mathbf{Q}'\mathbf{Q}^T\mathbf{b}$. Therefore, rigid body motion is a special case of a linear flow.
(c) What is $\mathbf{v}$ in the case when $\mathbf{Q}$ is given in (8.14) and $\mathbf{b} = \mathbf{0}$? For this flow show that $\boldsymbol{\omega} = (0, 0, 2\omega)$.
(d) The equations for vortex motion are given in Section 9.3.1. Show that the only vortex with a smooth velocity and constant vorticity has $v_\theta = \frac{1}{2}r\omega$. In this case, show that $\mathbf{h} = \mathbf{0}$ and

$$\mathbf{H} = \begin{pmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(e) For the $\mathbf{H}$ given in part (d) find a rotation $\mathbf{Q}$ so that $\mathbf{H} = \mathbf{Q}'\mathbf{Q}^T$. Do this by showing that this equation reduces to solving $\mathbf{Q}'' = \mathbf{H}^2\mathbf{Q}$, with $\mathbf{Q}(0) = \mathbf{I}$ and $\mathbf{Q}'(0) = \mathbf{H}$, and then solving for $\mathbf{Q}$. Make sure to verify that the solution is a rotation matrix.
(f) Explain why it is possible to conclude that a vortex motion with a smooth velocity and constant vorticity must be a rigid body motion.

9.9. Suppose that $\mathbf{v} = \alpha||\mathbf{x}||^k\mathbf{x}$, where k and α are real numbers.
(a) Show that the flow is irrotational.
(b) Find a potential function ϕ for this flow.
(c) Show that this velocity function does not correspond to incompressible fluid motion, unless $\alpha = 0$.

9.10. For a Taylor vortex, $v_r = v_z = 0$, and

$$v_\theta = \frac{\alpha r}{t^2}\exp\bigl(-r^2/(4\nu t)\bigr)\,.$$

Show this satisfies the equations of motion, assuming the fluid is incompressible and there are no body forces. In doing this also determine the pressure.

9.11. For Burger's vortex, $v_r = -\alpha r$, $v_z = 2\alpha z$, and

$$v_\theta = \frac{\beta}{r}\left(1 - e^{-\alpha r^2/(2\nu)}\right).$$

Show this satisfies the equations of motion, assuming the fluid is incompressible and there are no body forces. In doing this also determine the pressure.

9.12. This exercise explores the connections between vorticity and energy dissipation in a viscous fluid.
(a) The viscous dissipation function Φ is given in (8.110). Show that

$$\Phi = 2\mu(D_{xx}^2 + D_{yy}^2 + D_{zz}^2 + 2D_{xy}^2 + 2D_{xz}^2 + 2D_{yz}^2),$$

where the D_{ij}'s are the components of the rate of deformation tensor given in (8.67).
(b) Show that for an incompressible fluid,

$$\Phi = \mu\boldsymbol{\omega}\cdot\boldsymbol{\omega} + 2\mu\nabla\cdot\mathbf{q},$$

where $\mathbf{q} = (\nabla\mathbf{v})\mathbf{v}$.
(c) Let B is a bounded region in space. Use the result from part (b) to derive what is known as the Bobyleff-Forsyth formula, given as

$$\iiint\limits_B \Phi\,dV = \mu\iiint\limits_B \boldsymbol{\omega}\cdot\boldsymbol{\omega}\,dV + 2\mu\iint\limits_{\partial B}\mathbf{n}\cdot\mathbf{q}\,dS.$$

(d) If $\mathbf{v} = \mathbf{0}$ on ∂B show that

$$\iiint_B \Phi \, dV = \mu \iiint_B \boldsymbol{\omega} \cdot \boldsymbol{\omega} \, dV.$$

This shows that the total energy dissipation in the region is determined by the magnitude of the vorticity vector.

(e) Show that for an incompressible fluid, with no body force,

$$\frac{d}{dt} \iiint_{R(t)} \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v} \, dV = \iint_{\partial R(t)} \mathbf{g} \cdot \mathbf{n} \, dS - \mu \iiint_{R(t)} \boldsymbol{\omega} \cdot \boldsymbol{\omega} \, dV,$$

where $\mathbf{g} = -p\mathbf{v} - \mu \boldsymbol{\omega} \times \mathbf{v} + 2\mu \mathbf{q}$, and $\mathbf{q}$ is given in part (b).

(f) If the fluid is compressible, show that the generalization of the Bobyleff-Forsyth formula is

$$\iiint_B \Phi \, dV = \iiint_B \left[(\lambda + 2\mu)\Theta^2 + \mu \boldsymbol{\omega} \cdot \boldsymbol{\omega} \right] dV + 2\mu \iint_{\partial B} \mathbf{n} \cdot \mathbf{q} \, dS,$$

where $\Theta = \nabla \cdot \mathbf{v}$ and $\mathbf{q} = (\nabla \mathbf{v})\mathbf{v} - \Theta \mathbf{v}$.

9.13. There are three known principal invariants, or conserved quantities, for an ideal fluid. One is the circulation, which comes directly from Kelvin's Circulation Theorem. This problem derives the other two. Assume $R(t)$ is a material volume, as used for the Reynolds Transport Theorem, and there are no body forces.

(a) If $\mathbf{v} \cdot \mathbf{n} = 0$ on ∂R, show that

$$\frac{d}{dt} \iiint_R \mathbf{v} \cdot \mathbf{v} \, dV = 0.$$

This is the energy invariant, and states that the kinetic energy of the material volume is constant.

(b) If $\boldsymbol{\omega} \cdot \mathbf{n} = 0$ on ∂R, show that

$$\frac{d}{dt} \iiint_R \boldsymbol{\omega} \cdot \mathbf{v} \, dV = 0.$$

This is called the helicity invariant, and it measures the extent the pathlines coil around each other.

(c) Explain why the conclusions of parts (a) and (b) hold if the body force has the form $\mathbf{f} = \nabla \Psi$.

9.14. Suppose that the velocity of an incompressible fluid is $\mathbf{v} = \mathbf{v}_0 + \frac{1}{2} \boldsymbol{\Omega} \times \mathbf{x}$, where $\mathbf{v}_0$ and $\boldsymbol{\Omega}$ are constant vectors. Consequently, $\mathbf{v}$ consists of a constant

velocity $\mathbf{v}_0$ added to the velocity for circular motion in the plane perpendicular to $\boldsymbol{\Omega}$.

(a) Show that $\boldsymbol{\omega} = \boldsymbol{\Omega}$.

(b) The helicity density is defined as $h = \boldsymbol{\omega} \cdot \mathbf{v}$, and it gives rise to the invariant derived in Exercise 9.13(b). Using the result from part (a), show that $h = \boldsymbol{\Omega} \cdot \mathbf{v}_0$.

(c) Assuming $\boldsymbol{\Omega} = (0, 0, \Omega)$ and $\mathbf{v}_0 = (0, 0, w_0)$, find the pathlines and from this show that the flow is helictical.

(d) From the description of $\mathbf{v}$ given above, one might think that it corresponds to rigid body motion. Prove this using the results from Exercises 8.22(d) and 9.8(b).

9.15. In this problem assume the body force in the Navier-Stokes equations can be written as $\mathbf{f} = \nabla \Psi$.

(a) Assuming that the fluid is ideal show that

$$\frac{\partial \mathbf{v}}{\partial t} + \nabla \left(\frac{1}{2} \mathbf{v} \cdot \mathbf{v} + \frac{1}{\rho} p - \Psi \right) + \boldsymbol{\omega} \times \mathbf{v} = \mathbf{0}.$$

In the case where the fluid is also irrotational show that

$$p = p_0(t) - \rho \left(\frac{\partial \phi}{\partial t} + \frac{1}{2} \nabla \phi \cdot \nabla \phi \right) + \rho \Psi.$$

This is a generalization of Bernoulli's theorem given in (9.29).

(b) Suppose the fluid is inviscid and irrotational. Also, assume it satisfies the equation of state for a polytropic fluid, which is $p = k\rho^\gamma$, where $\gamma > 1$. Adapt the argument of part (a) to show that

$$\frac{\partial \phi}{\partial t} + \frac{\gamma}{\gamma - 1} \frac{p}{\rho} + \frac{1}{2} \nabla \phi \cdot \nabla \phi - \Psi = c(t).$$

9.16. For the impulsive plate problem in Section 9.6.1, suppose the lower plate moves with velocity $\mathbf{v} = (u_0 f(t), 0, 0)$. Assuming that $f(t)$ is a smooth function of t, use the Laplace transform to show that

$$u(y, t) = u_0 f(0) \operatorname{erfc}\left(\frac{y}{2\sqrt{\nu t}} \right) + u_0 \int_0^t f'(t - r) \operatorname{erfc}\left(\frac{y}{2\sqrt{\nu r}} \right) dr.$$

9.17. For the impulsive plate problem in Section 9.6.1, suppose the lower plate moves with velocity $\mathbf{v} = (u_0 \cos(\omega t), 0, 0)$. This is known as Stokes' second problem. The exact solution can be found using the formula from the previous problem, but a different approach is taken here.

(a) After a sufficiently long period of time the solution should be approximately periodic. Assume that $u = e^{i\omega t} q(y)$, where it is understood that the real part of this expression is used. This expression should satisfy the momentum equation, and the two boundary conditions. Show that this results in the solution of the form $u = u_0 e^{-\sigma y} \cos(\sigma y - \omega t)$.

(b) Sketch the solution as a function of y, and describe the basic characteristics of the solution.

(c) Show that the boundary layer thickness is approximately $5\sqrt{2\nu/\omega}$. What happens to the thickness as the frequency increases?

Appendix A
Taylor's Theorem

A.1 Single Variable

The single most important result needed to develop an asymptotic approximation is Taylor's theorem. The single variable version of the theorem is below.

Theorem A.1. *Given a function $f(x)$ assume that its $(n+1)st$ derivative $f^{(n+1)}(x)$ is continuous for $x_L < x < x_R$. In this case, if a and x are points in the interval (x_L, x_R) then*

$$f(x) = f(a) + (x-a)f'(a) + \frac{1}{2}(x-a)^2 f''(a) + \cdots + \frac{1}{n!}(x-a)^n f^{(n)}(a) + R_{n+1}, \tag{A.1}$$

where the remainder is

$$R_{n+1} = \frac{1}{(n+1)!}(x-a)^{n+1} f^{(n+1)}(\eta), \tag{A.2}$$

and η is a point between a and x.

There are different, but equivalent, ways to write the above result. One is

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + \cdots + \frac{1}{n!}h^n f^{(n)}(x) + R_{n+1}, \tag{A.3}$$

The requirement here is that x and $x+h$ are points in the interval (x_L, x_R).

A.2 Two Variables

The two-variable version of the expansion in (A.3) is

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_BM1,
© Springer Science+Business Media, LLC 2009

$$f(x+h,t+k) = f(x,t) + Df(x,t) + \frac{1}{2}D^2 f(x,t) + \cdots + \frac{1}{n!}D^n f(x,t) + R_{n+1}. \tag{A.4}$$

where

$$D = h\frac{\partial}{\partial x} + k\frac{\partial}{\partial t}\,.$$

Writing this out, through quadratic terms, yields

$$\begin{aligned} f(x+h,t+k) = {} & f(x,t) + hf_x(x,t) + kf_t(x,t) \\ & + \frac{1}{2}h^2 f_{xx}(x,t) + hkf_{xt}(x,t) + \frac{1}{2}k^2 f_{tt}(x,t) + \cdots\,. \end{aligned}$$

The subscripts in the above expression denote partial differentiation. So, for example,

$$f_{xt} = \frac{\partial^2 f}{\partial x \partial t}\,.$$

It is assumed that the function f has continuous partial derivatives up through order $n+1$.

The above expansion can be expressed in a form similar to the one in (A.1), and the result is

$$\begin{aligned} f(x,t) = {} & f(a,b) + (x-a)f_x(a,b) + (t-b)f_t(a,b) \\ & + \frac{1}{2}(x-a)^2 f_{xx}(a,b) + (x-a)(t-b)f_{xt}(a,b) + \frac{1}{2}(t-b)^2 f_{tt}(a,b) \\ & + \cdots\,. \end{aligned}$$

A.3 Multivariable Versions

For more than two variables it is convenient to use vector notation. In this case (A.4) takes the form

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + Df(\mathbf{x}) + \frac{1}{2}D^2 f(\mathbf{x}) + \cdots + \frac{1}{n!}D^n f(\mathbf{x}) + R_{n+1},$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_k)$, $\mathbf{h} = (h_1, h_2, \cdots, h_k)$ and

$$\begin{aligned} D &= \mathbf{h}\cdot\nabla \\ &= h_1\frac{\partial}{\partial x_1} + h_2\frac{\partial}{\partial x_2} + \cdots + h_k\frac{\partial}{\partial x_k}\,. \end{aligned}$$

Writing this out, through quadratic terms, yields

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}\cdot\nabla f(\mathbf{x}) + \frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h} + \cdots\,,$$

where $\mathbf{H}$ is the Hessian and is given as

$$\mathbf{H} = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_k \partial x_1} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_k \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_k} & \dfrac{\partial^2 f}{\partial x_2 \partial x_k} & \cdots & \dfrac{\partial^2 f}{\partial x_k^2} \end{pmatrix}.$$

Taylor's theorem can also be extended to vector functions, although the formulas are more involved. To write down the expansion through the linear terms, assume that $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ and $\mathbf{x} = (x_1, x_2, \ldots, x_k)$. In this case,

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + (\nabla \mathbf{f})\mathbf{h} + \cdots ,$$

where

$$\nabla \mathbf{f} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_k} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_k} \end{pmatrix}.$$

Appendix B
Fourier Analysis

B.1 Fourier Series

It is assumed here that the function $f(x)$ is piecewise continuous for $0 \leq x \leq \ell$. Recall that this means $f(x)$ is continuous on the interval $0 \leq x \leq \ell$ except at a finite number of points within the interval at which the function has a jump discontinuity.

The Fourier sine series for $f(x)$ is defined as

$$S(x) = \sum_{n=1}^{\infty} \beta_n \sin(\lambda_n x), \tag{B.1}$$

where $\lambda_n = n\pi/\ell$ and

$$\beta_n = \frac{2}{\ell} \int_0^{\ell} f(x) \sin(\lambda_n x) dx. \tag{B.2}$$

The Fourier cosine series for $f(x)$ is defined as

$$C(x) = \frac{1}{2}\alpha_0 + \sum_{n=1}^{\infty} \alpha_n \cos(\lambda_n x), \tag{B.3}$$

where

$$\alpha_n = \frac{2}{\ell} \int_0^{\ell} f(x) \cos(\lambda_n x) dx. \tag{B.4}$$

A certain amount of smoothness is required of the function $f(x)$ so the above series are defined. For example, $f(x)$ must be smooth enough that the integrals in (B.2) and (B.4) exist. Certainly assuming $f(x)$ is continuous is enough for the integrals, but, unfortunately, this is not enough to guarantee that the series in (B.1) and (B.3) converge. They will converge, however, if

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_BM2,
© Springer Science+Business Media, LLC 2009

$f(x)$ and $f'(x)$ are piecewise continuous. The question naturally arises as to what they converge to, and for this we have the following result.

Theorem B.1. *Assume $f(x)$ and $f'(x)$ are piecewise continuous for $0 \leq x \leq \ell$. On the interval $0 < x < \ell$, the Fourier sine series, and the Fourier cosine series, converge to $f(x)$ at points where the function is continuous, and they converge to $\frac{1}{2}(f(x+) + f(x-))$ at points where the function has a jump discontinuity. At the endpoints, $S(0) = S(\ell) = 0$, while $C(0) = f(0)$ and $C(\ell) = f(\ell)$.*

When using a Fourier series to solve a differential equation one usually needs the expansion of the solution as well as its derivatives. The problem is that it is not always possible to obtain the series for $f'(x)$ by differentiating the series for $f(x)$. For example, given a sine series as in (B.1) one might be tempted to conclude that

$$S'(x) = \sum_{n=1}^{\infty} \beta_n \lambda_n \cos(\lambda_n x).$$

The issue is that the differentiation has resulted in λ_n appearing in the coefficient. As an example, for the function

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 2 & \text{if } 1 < x \leq 2, \end{cases}$$

one finds that

$$\beta_n \lambda_n = \frac{2}{\ell} \left[1 - 2(-1)^n + \cos(n\pi/2)\right].$$

The general term $\beta_n \lambda_n \cos(\lambda_n x)$ of the series does not converge to zero as $n \to \infty$, and this means that the series does not converge. Consequently, additional restrictions must be imposed on $f(x)$ to guarantee convergence. Basically what are needed are conditions that will give us $\beta_n = O(1/n^2)$, and this brings us to the next result.

Theorem B.2. *Assume $f(x)$ is continuous, with $f'(x)$ and $f''(x)$ piecewise continuous, for $0 \leq x \leq \ell$. If $f(x)$ is expanded in a cosine series then the series for $f'(x)$ can be found by differentiating the series for $f(x)$. If $f(x)$ is expanded in a sine series, and if $f(0) = f(\ell) = 0$, then the series for $f'(x)$ can be found by differentiating the series for $f(x)$.*

The question of convergence for integration is much easier to answer. As long as the Fourier series of $f(x)$ converges then the series for the integral of f can be found by simply integrating the series for f.

B.2 Fourier Transform

To derive the formula for the Fourier transform from the Fourier series, it is convenient to use the symmetric interval $-\ell < x < \ell$. Generalizing (B.2) and (B.3), the Fourier series of a continuous function $f(x)$ is

$$f(x) = \frac{1}{2}\alpha_0 + \sum_{n=1}^{\infty} \left[\alpha_n \cos(\lambda_n x) + \beta_n \sin(\lambda_n x)\right],$$

where $\lambda_n = n\pi/\ell$,

$$\alpha_n = \frac{1}{\ell}\int_{-\ell}^{\ell} f(x)\cos(\lambda_n x)dx,$$

and

$$\beta_n = \frac{1}{\ell}\int_{-\ell}^{\ell} f(x)\sin(\lambda_n x)dx.$$

By using the identities $\cos(\theta) = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$ and $\sin(\theta) = \frac{1}{2i}(e^{i\theta} - e^{-i\theta})$, the Fourier series can be written in exponential form as

$$f(x) = \sum_{n=-\infty}^{\infty} \gamma_n e^{i\lambda_n x},$$

where

$$\gamma_n = \frac{1}{2\ell}\int_{-\ell}^{\ell} f(\bar{x})e^{-i\lambda_n \bar{x}}d\bar{x}.$$

Combining these two expressions

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2\ell}\int_{-\ell}^{\ell} f(\bar{x})e^{i\lambda_n(x-\bar{x})}d\bar{x}.$$

The sum in the above equation is reminiscent of the Riemann sum used to define integration. To make this more evident, let $\Delta\lambda = \lambda_{n+1} - \lambda_n = \frac{\pi}{\ell}$. With this

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2\pi}\int_{-\ell}^{\ell} f(\bar{x})e^{i\lambda_n(x-\bar{x})}d\bar{x}\Delta\lambda.$$

The argument originally used by Fourier is that in the limit of $\ell \to \infty$, the above expression yields

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\bar{x})e^{i\lambda(x-\bar{x})}d\bar{x}d\lambda.$$

Fourier then made the observation that the above equation can be written as $f(x) = \mathcal{F}^{-1}(\mathcal{F}(f))$, where $\mathcal{F}$ is the Fourier transform defined in Section 7.2.5. With this, the Fourier transform was born.

To say that the above derivation is heuristic would be more than generous. However, it is historically correct, and it does show the origin of the Fourier transform and its inverse. The formal proof of the derivation can be found in Weinberger [1995].

Appendix C
Stochastic Differential Equations

The steps used to solve the Langevin equation look routine, and the solutions in (4.85) and (4.86) are not particularly remarkable. However, on closer inspection, the randomness of the forcing function raises some serious mathematical questions. An example of $\mathbf{R}$ is shown in Figure 4.27 using 400 points along the t-axis. As will be discussed in more detail in Section 4.7.1, the value of $\mathbf{R}(t_1)$ is independent of the value of $\mathbf{R}(t_2)$ if $t_1 \neq t_2$. This means that if more than 400 points are used, the graph would appear even more random than in Figure 4.27. The question that immediately arises is whether the non-differentiability of this function causes the differential equation (4.84), or its solution (4.84), to be meaningless. One approach for addressing this issue rests on denial, where the calculations are carried out as if everything is just fine. This is, in fact, what was done to derive (4.85), and this approach almost works. To have it succeed, all that is needed is to make sense of the solution, and then use this to justify the entire process.

The question is, therefore, how to define the integrals in (4.85) and (4.86). The exponentials are not an issue, and so to simplify the discussion we will concentrate on the expression

$$\mathbf{W}(t) = \int_0^t \mathbf{R}(\tau) d\tau. \tag{C.1}$$

The definition of this integral employs the same Riemann sum used in Calculus. With this in mind, we introduce a partition $0 < t_1 < t_2 < \cdots < t_m < t$, where $t_0 = 0$ and $t_{m+1} = t$. For simplicity, it is assumed the points are equally spaced, and so $t_{j+1} - t_j = \Delta t$. Letting s_j be a point from the interval $[t_j, t_{j+1}]$, then we introduce the partial sum

$$\mathbf{S}_m = \sum_{j=0}^{m-1} \mathbf{R}(s_j) \Delta t. \tag{C.2}$$

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_BM3,
© Springer Science+Business Media, LLC 2009

The question is, if $\Delta t \to 0$, does $\mathbf{S}_m$ converge? The answer is yes, although convergence is measured in the mean-square sense. Knowing that it converges then the limit of $\mathbf{S}_m$ serves as the definition of the integral in (C.1). This definition preserves most, but not all, of the properties associated with standard integration. In particular, $\mathbf{W}$ is a continuous function of t, and the integral is additive in the sense that if $t_1 < t_2$ then

$$\int_0^{t_2} \mathbf{R}(\tau)d\tau = \int_0^{t_1} \mathbf{R}(\tau)d\tau + \int_{t_1}^{t_2} \mathbf{R}(\tau)d\tau.$$

Moreover, the partial sums in (C.2) provide a method for numerically evaluating the stochastic integrals in (4.85) and (4.86).

Now that integration has been put onto a solid mathematical footing, we turn to the differential equation (4.84). In the case of when $\mathbf{R}$ is smooth, this equation can be integrated to yield

$$\mathbf{v}(t) = \mathbf{v}(0) - \lambda \int_0^t \mathbf{v}(\tau)d\tau + \frac{1}{m}\int_0^t \mathbf{R}(\tau)d\tau. \tag{C.3}$$

For smooth functions this integral equation is equivalent to the differential equation (4.84). This fact is used to explain what happens when a random forcing is used. Specifically, the interpretation of the differential equation (4.84) is that $\mathbf{v}$ satisfies (C.3). It is for this reason that in the subject of stochastic differential equations, (4.84) is conventionally written using differentials as

$$d\mathbf{v} = -\lambda \mathbf{v}dt + \frac{1}{m}\mathbf{R}dt.$$

The implication in using this notation is that the stochastic differential equation is being interpreted as the solution of the associated integral equation. With this viewpoint, (C.1) can be written as $d\mathbf{W} = \mathbf{R}dt$. Those interested in pursuing the theoretical foundation of the stochastic differential equations should consult Oksendal [2003].

Appendix D
Identities

D.1 Trace

In the following, $\mathbf{A}$ and $\mathbf{B}$ are 3×3 matrices, and α and β are scalars.

$$\begin{aligned}
\operatorname{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) &= \alpha \operatorname{tr}(\mathbf{A}) + \beta \operatorname{tr}(\mathbf{B}) \\
\operatorname{tr}(\mathbf{AB}) &= \operatorname{tr}(\mathbf{BA}) \\
\operatorname{tr}(\mathbf{A}^T) &= \operatorname{tr}(\mathbf{A})
\end{aligned}$$

If $\mathbf{A}$ is symmetric and $\mathbf{B}$ is skew-symmetric then $\operatorname{tr}(\mathbf{AB}) = 0$.

D.2 Determinant

In the following, $\mathbf{A}$ and $\mathbf{B}$ are 3×3 matrices, and α and β are scalars.

$$\begin{aligned}
\det(\mathbf{AB}) &= \det(\mathbf{BA}) = \det(\mathbf{A})\det(\mathbf{B}) \\
\det(\alpha\mathbf{A}) &= \alpha^3 \det(\mathbf{A}) \\
\det(\mathbf{A}^T) &= \det(\mathbf{A}) \\
\det(\mathbf{A}^{-1}) &= 1/\det(\mathbf{A}) \\
\det(\mathbf{I}) &= 1
\end{aligned}$$

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_BM4,
© Springer Science+Business Media, LLC 2009

D.3 Vector Calculus

In the following, ϕ is a scalar, $\mathbf{u} = (u, v, w)$ is a vector, and $\mathbf{A}(\mathbf{x})$ is a 3×3 matrix. They are all smooth functions of $\mathbf{x} = (x, y, z)$.

$$\begin{aligned}
\nabla \cdot \mathbf{u} &= \mathrm{tr}(\nabla \mathbf{u}) \\
\nabla \cdot (\phi \mathbf{u}) &= \mathbf{u} \cdot \nabla \phi + \phi(\nabla \cdot \mathbf{u}) \\
\nabla \cdot (\mathbf{A}\mathbf{u}) &= \mathbf{u} \cdot (\nabla \cdot \mathbf{A}) + \mathrm{tr}(\mathbf{A}^T \nabla \mathbf{u}) \\
\nabla \cdot (\phi \mathbf{A}) &= \mathbf{A}^T \nabla \phi + \phi(\nabla \cdot \mathbf{A}) \\
(\mathbf{v} \cdot \nabla)\mathbf{u} &= (\nabla \mathbf{u})\mathbf{v} \\
\nabla \times (\nabla \phi) &= \mathbf{0} \\
\nabla \cdot (\nabla \times \mathbf{u}) &= 0
\end{aligned}$$

In the above identities

$$\nabla \mathbf{u} = \begin{pmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} & \dfrac{\partial u}{\partial z} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} & \dfrac{\partial v}{\partial z} \\ \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} & \dfrac{\partial w}{\partial z} \end{pmatrix},$$

and

$$\nabla \cdot \mathbf{A} = \begin{pmatrix} \dfrac{\partial A_{11}}{\partial x} + \dfrac{\partial A_{21}}{\partial y} + \dfrac{\partial A_{31}}{\partial z} \\ \dfrac{\partial A_{12}}{\partial x} + \dfrac{\partial A_{22}}{\partial y} + \dfrac{\partial A_{32}}{\partial z} \\ \dfrac{\partial A_{13}}{\partial x} + \dfrac{\partial A_{23}}{\partial y} + \dfrac{\partial A_{33}}{\partial z} \end{pmatrix}.$$

Appendix E
Equations for a Newtonian Fluid

E.1 Cartesian Coordinates

Letting $\mathbf{v} = (u, v, w)$ and $\mathbf{f} = (f, g, h)$, then for an incompressible Newtonian fluid in Cartesian coordinates:

$$\rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z}\right) = -\frac{\partial p}{\partial x} + \mu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + \rho f$$

$$\rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z}\right) = -\frac{\partial p}{\partial y} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) + \rho g$$

$$\rho\left(\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z}\right) = -\frac{\partial p}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right) + \rho h$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0$$

E.2 Cylindrical Coordinates

Letting $\mathbf{v} = (v_r, v_\theta, v_z) = v_r\mathbf{e}_r + v_\theta\mathbf{e}_\theta + v_z\mathbf{e}_z$ and $\mathbf{f} = f_r\mathbf{e}_r + f_\theta\mathbf{e}_\theta + f_z\mathbf{e}_z$, then for an incompressible Newtonian fluid in cylindrical coordinates:

$$\rho\left(\frac{\partial v_r}{\partial t} + v_r\frac{\partial v_r}{\partial r} + \frac{v_\theta}{r}\frac{\partial v_r}{\partial \theta} - \frac{v_\theta^2}{r} + v_z\frac{\partial v_r}{\partial z}\right)$$
$$= -\frac{\partial p}{\partial r} + \mu\left[\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}(rv_r)\right) + \frac{1}{r^2}\frac{\partial^2 v_r}{\partial \theta^2} - \frac{2}{r^2}\frac{\partial v_\theta}{\partial \theta} + \frac{\partial^2 v_r}{\partial z^2}\right] + \rho f_r$$

$$\rho\left(\frac{\partial v_\theta}{\partial t} + v_r\frac{\partial v_\theta}{\partial r} + \frac{v_\theta}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{v_r v_\theta}{r} + v_z\frac{\partial v_\theta}{\partial z}\right)$$
$$= -\frac{1}{r}\frac{\partial p}{\partial \theta} + \mu\left[\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}(rv_\theta)\right) + \frac{1}{r^2}\frac{\partial^2 v_\theta}{\partial \theta^2} + \frac{2}{r^2}\frac{\partial v_r}{\partial \theta} + \frac{\partial^2 v_\theta}{\partial z^2}\right] + \rho f_\theta$$

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,*
Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_BM5,
© Springer Science+Business Media, LLC 2009

$$\rho\left(\frac{\partial v_z}{\partial t}+v_r\frac{\partial v_z}{\partial r}+\frac{v_\theta}{r}\frac{\partial v_z}{\partial \theta}+v_z\frac{\partial v_z}{\partial z}\right)$$
$$=-\frac{\partial p}{\partial z}+\mu\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial v_z}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 v_z}{\partial \theta^2}+\frac{\partial^2 v_z}{\partial z^2}\right]+\rho f_z$$
$$\frac{1}{r}\frac{\partial(rv_r)}{\partial r}+\frac{1}{r}\frac{\partial v_\theta}{\partial \theta}+\frac{\partial v_z}{\partial z}=0$$

Transformation laws for velocities:

$$u=v_r\cos\theta-v_\theta\sin\theta \qquad v_r=\frac{1}{\sqrt{x^2+y^2}}(xu+yv)$$
$$v=v_r\sin\theta+v_\theta\cos\theta \qquad v_\theta=\frac{1}{\sqrt{x^2+y^2}}(-yu+xv)$$
$$w=v_z \qquad v_z=w$$

Transformation laws for derivatives:

$$\frac{\partial}{\partial x}=\cos\theta\frac{\partial}{\partial r}-\frac{\sin\theta}{r}\frac{\partial}{\partial \theta} \qquad \frac{\partial}{\partial r}=\frac{x}{\sqrt{x^2+y^2}}\frac{\partial}{\partial x}+\frac{y}{\sqrt{x^2+y^2}}\frac{\partial}{\partial y}$$
$$\frac{\partial}{\partial y}=\sin\theta\frac{\partial}{\partial r}+\frac{\cos\theta}{r}\frac{\partial}{\partial \theta} \qquad \frac{\partial}{\partial \theta}=-y\frac{\partial}{\partial x}+x\frac{\partial}{\partial y}$$
$$\frac{\partial}{\partial z}=\frac{\partial}{\partial z} \qquad \frac{\partial}{\partial z}=\frac{\partial}{\partial z}$$

$$\frac{D}{Dt}=\frac{\partial}{\partial t}+v_r\frac{\partial}{\partial r}+\frac{v_\theta}{r}\frac{\partial}{\partial \theta}+v_z\frac{\partial}{\partial z}$$

Formulas from vector analysis:

$$\nabla\times\mathbf{v}=\left(\frac{1}{r}\frac{\partial v_z}{\partial \theta}-\frac{\partial v_\theta}{\partial z}\right)\mathbf{e}_r+\left(\frac{\partial v_r}{\partial z}-\frac{\partial v_z}{\partial r}\right)\mathbf{e}_\theta+\frac{1}{r}\left(\frac{\partial(rv_\theta)}{\partial r}-\frac{\partial v_r}{\partial \theta}\right)\mathbf{e}_z$$
$$\nabla^2\phi=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \phi}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2\phi}{\partial \theta^2}+\frac{\partial^2\phi}{\partial z^2}$$
$$\nabla\phi=\frac{\partial \phi}{\partial r}\mathbf{e}_r+\frac{1}{r}\frac{\partial \phi}{\partial \theta}\mathbf{e}_\theta+\frac{\partial \phi}{\partial z}\mathbf{e}_z$$
$$\nabla\cdot\mathbf{v}=\frac{1}{r}\frac{\partial(rv_r)}{\partial r}+\frac{1}{r}\frac{\partial v_\theta}{\partial \theta}+\frac{\partial v_z}{\partial z}$$

References

AAVSO. American association of variable star observers. Website, 2009. http://www.aavso.org.

M. Van Aerde and H. Rakha. Multivariate calibration of single-regime speed-flow-density relationships. In *Vehicle Navigation and Information Conference (VNIS)*, pages 334–341. IEEE, Piscataway NJ, 1995.

A. Ansorge. What does the entropy condition mean in traffic flow theory? *Trans Res Part B: Methodological*, 24:133–143, 1990.

G. L. Aranovich and M. D. Donohue. Eliminating the mean-free-path inconsistency in classical phenomenological model of diffusion for fluids. *Physica A*, 373:119–141, 2007.

R. Aris. Review of rational thermodynamics. *Am Math Monthly*, 94:562–564, 1987.

T. Asai, K. Seo, O. Kobayashi, and R. Sakashita. Fundamental aerodynamics of the soccer ball. *Sports Eng*, 10:101–110, 2007.

J. S. Bader, R. W. Hammond, S. A. Henck, M. W. Deem, G. A. McDermott, J. M. Bustillo, J. W. Simpson, G. T. Mulhern, and J. M. Rothberg. Dna transport by a micromachined brownian ratchet device. *Proc Natl Acad Sci*, 96:13165–13169, 1999.

D. R. Baker, C. Dalpe, and G. Poirier. The viscosities of foods as analogs for silicate melts. *J Geoscience Edu*, 52:363–367, 2004.

R. W. Balluffi, S. M. Allen, and W. C. Carter. *Kinetics of Materials*. Wiley-Interscience, New York, 2005.

J. R. Bamforth, S. Kalliadasis, J. H. Merkin, and S. K. Scott. Modelling flow-distributed oscillations in the cdima reaction. *Phys Chem Chem Phys*, 2: 4013–4021, 2000.

O. K. Baskurt. Red blood cells flowing in an arteriol. Website, 2009. http://www.rheology.org/sor/publications/rheology_b/Jul04/default.htm.

R. C. Batra. Universal relations for transversely isotropic elastic materials. *Math Mech Solids*, 7:421–437, 2002.

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,* Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_BM6,

© Springer Science+Business Media, LLC 2009

A. Bayon, F. Gascon, and A. Varade. Measurement of the longitudinal and transverse vibration frequencies of a rod by speckle interferometry. *IEEE Trans Ultrason Ferroelectr Freq Control*, 40:265–269, 1993.

A. Bayon, A. Varade, and F. Gascon. Determination of the elastic constants of isotropic solids by optical heterodyne interferometry. *J Acoust Soc Am*, 96:2589–2592, 1994.

T. A. Blackledge and C. Y. Hayashi. Silken toolkits: biomechanics of silk fibers spun by the orb web spider argiope argentata (fabricius 1775). *J Exp Biology*, 209:2452–2461, 2006.

J. Bluck. Nasa tunnels test tennis balls. Press Release, 00-58AR, 2000. `http://www.nasa.gov/centers/ames/news/releases/2000/00_58AR.html`.

J. Blum, S. Bruns, D. Rademacher, A. Voss, B. Willenberg, and M. Krause. Measurement of the translational and rotational brownian motion of individual particles in a rarefied gas. *Phys Rev Lett*, 97:230601, 2006.

G. Bluman and J. Cole. *Similarity Methods for Differential Equations.* Springer-Verlag, New York, 1974.

G. W. Bluman and S. Anco. *Symmetry and Integration Methods for Differential Equations.* Springer-Verlag, New York, 2002.

Bodner, Smith, Keys, and Greenbowe. The blue/amber/colorless oscillating reaction. Website, 2009. `http://chemed.chem.purdue.edu/demos/main_pages/22.8.html`.

C. Booth, T. Beer, and J. D. Penrose. Diffusion of salt in tap water. *Am J Physics*, 46:525–527, 1978.

N. Bourbaki. *Functions of a Real Variable.* Springer, New York, 2004.

W. E. Boyce and R. C. DiPrima. *Elementary Differential Equations and Boundary Value Problems.* Wiley, New York, 2004.

M. Braun. *Differential Equations and Their Applications: An Introduction to Applied Mathematics.* Springer, New York, 4th edition, 1993.

G. E. Briggs and J. B. S. Haldane. A note on the kinetics of enzyme action. *Biochem J*, 19:338–339, 1928.

B. Brixner. Trinity: 16 july 1945. Website, 2009. `http://www.radiochemistry.org/history/nuke_tests/trinity/index.html`.

A. J. Brown. Enzyme action. *J Chem Soc*, 81:373–386, 1902.

F. N. M. Brown. See the wind blow. *Dept. Aerosp. Mech. Eng. Rep., Univ. of Notre Dame*, 1971.

T. Cebeci and J. Cousteix. *Modeling and Computation of Boundary-layer Flows.* Springer, New York, 2nd edition, 2005.

C. Van den Broeck, R. Kawai, and P. Meurs. Microscopic analysis of a thermal brownian motor. *Phy Rev Lett*, 93:0906011–0906014, 2004.

K. J. Devlin. *The Millennium Problems: The Seven Greatest Unsolved Mathematical Puzzles of Our Time.* Basic, New York, 2002.

P. G. Drazin and W. H. Reid. *Hydrodynamic Stability.* Cambridge University Press, Cambridge, 2nd edition, 2004.

D. Drew. *Traffic flow theory and control.* McGraw Hill, New York, 1968.

B. Eckhardt, T. M. Schneider, B. Hof, and J. Westerweel. Turbulence transition in pipe flow. *Annu Rev Fluid Mech*, 39:447–468, 2007.

C. J. Efthimiou and M. D. Johnson. Domino waves. *SIAM Rev*, 49:111–120, 2007.

J. Ellenberger, P. J. Klijn, M. Tels, and J. Vleggaar. Construction and performance of a cone-and-plate rheogoniometer with air bearings. *J Phys E: Sci Instrum*, 9:763–765, 1976.

R. Engbert and F. Drepper. Chance and chaos in population biology, models of recurrent epidemics and food chain dynamics. *Chaos, Solutions Fractals*, 4:1147–1169, 1994.

A. C. Eringen. *Microcontinuum Field Theories II. Fluent Media.* Springer, New York, 2001.

L. C. Evans. *Partial Differential Equations.* American Mathematical Society, New York, 1998.

G. Eyink, U. Frisch, R. Moreau, and A. Sobolevski. Euler equations: 250 years on, proceedings of an international conference. *Physica D*, 237:xi–2250, 2008.

R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics, Vol. 1.* Addison Wesley, New York, 2nd edition, 2005.

A. Fick. On liquid diffusion. *Philos Mag*, 10:31–39, 1885.

R. J. Field and R. M. Noyes. Oscillations in chemical systems iv. limit cycle behavior in a model of a real chemical reaction. *J Amer Chem Soc*, 60: 1877–1884, 1974.

R. J. Field, E. Koros, and R. M. Noyes. Oscillations in chemical systems. ii. thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *J Amer Chem Soc*, 94:8649–8664, 1972.

M. Finnis. *Interatomic Forces in Condensed Matter.* Oxford University Press, Oxford, 2004.

M. Frewer. More clarity on the concept of material frame-indifference in classical continuum mechanics. *Acta Mech*, 202:213–246, 2009.

A. Friedman. *Generalized Functions and Partial Differential Equations.* Dover, New York, 2005.

Y. C. Fung. *Biomechanics: Mechanical Properties of Living Tissues.* Springer, New York, 2nd edition, 1993.

I. Gasser. On non-entropy solutions of scalar conservation laws for traffic flow. *ZAMM*, 83:137–143, 2003.

G. M. L. Gladwell. *Inverse Problems in Vibration.* Springer, New York, 2nd edition, 2004.

A. Gomez-Marin and J. M. Sancho. Ratchet, pawl and spring brownian motor. *Physica D*, 216:214–219, 2006.

S. R. Goodwill, S. B. Chin, and S. J. Haake. Aerodynamics of spinning and non-spinning tennis balls. *J Wind Eng Indust Aerodyn*, 92:935–958, 2004.

Inc Google Maps. Map of arlington memorial bridge. Website, 2007. `http://maps.google.com/`.

P. Gray and S. K. Scott. *Chemical Oscillations and Instabilities: Non-linear Chemical Kinetics*. Oxford University Press, Oxford, 1994.

H. Greenberg. An analysis of traffic flow. *Operations Res*, 7:79–85, 1959.

R. D. Gregory. Helmholtz's theorem when the domain is infinite and when the field has singular points. *Quart J Mech Appl Math*, 49:439–450, 1996.

R. Haberman. *Applied Partial Differential Equations*. Prentice Hall, New York, 2003.

J. K. Hale and H. Kocak. *Dynamics and Bifurcations*. Springer, New York, 1996.

P. Hanggi, F. Marchesoni, and F. Nori. Brownian motors. *Annalen der Physik*, 14:51–70, 2005.

N. E. Henriksen and F. Y. Hansen. *Theories of Molecular Reaction Dynamics: The Microscopic Foundation of Chemical Kinetics*. Oxford University Press, Oxford, 2008.

P. V. Hobbs and A. J. Kezweent. Splashing of a water drop. *Exper Fluids*, 155:1112–1114, 1967.

M. H. Holmes. *Introduction to Perturbation Methods*. Springer-Verlag, New York, 1995.

M. H. Holmes. *Introduction to Numerical Methods in Differential Equations*. Springer, New York, 2005.

M. H. Holmes, V. C. Mow, and W. M. Lai. The nonlinear interaction of solid and fluid in the creep response of articular cartilage. *Biorheology*, 20:422, 1983.

P. L. Houston. *Chemical Kinetics and Reaction Dynamics*. Dover, New York, 2006.

R. Hsu. Wind tunnel tests verify our design. Website, 2009. `http://www.pbworld.com/news_events/publications/network/issue_28/28_15_hsur_windtunnel.asp`.

K. Hutter and K. Johnk. *Continuum Methods of Physical Modeling*. Springer, New York, 2004.

D. D. Joseph. Potential flow of viscous fluids : Historical notes. *Inter J Multiphase Flow*, 32:285–310, 2006.

M. Kac. Can one hear the shape of a drum? *Am Math Month*, 73:1–23, 1966.

N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 3nd edition, 2007.

J. B. Keller. Diffusion at finite speed and random walks. *Proc Nat Acad Sci*, 101:1120–1122, 2004.

Y. Kimura, Y. Qi, T. Cagan, and W. A. Goddard. The quantum sutton-chen many-body potential for properties of fcc metals. *Caltech ASCI Reports*, 2000.003, 2000.

J. K. Knowles. On entropy conditions and traffic flow models. *ZAMM*, 88: 64–73, 2008.

S. V. Kryatov, E. V. Rybak-Akimova, A. Y. Nazarenko, and P. D. Robinson. A dinuclear iron(iii) complex with a bridging urea anion: implications for the urease mechanism. *Chem Commun*, 11:921–922, 2000.

C. Kunkle. Velocity field in a pipe. *Millersville University Physics: Experiment of the Month*, 2008.

M. Kwan. A finite deformation theory for nonlinearly permeable cartilage and other soft hydrated connective tissues and rheological study of cartilage proteoglycans. PhD Thesis, RPI, 1985.

R. S. Lakes. Viscoelastic measurement techniques. *Rev Sci Instru*, 75:797–810, 2004.

E. Lauga, M. P. Brenner, and H. A. Stone. Microfluidics: The no-slip boundary condition. In C. Tropea, A. L. Yarin, and J. F. Foss, editors, *Handbook of Experimental Fluid Dynamics*, New York, 2007. Springer.

NR-06-05-06 Lawrence Livermore National Laboratory. Nanotube membranes offer possibility of cheaper desalination. Website, 2009. `http://www.llnl.gov/pao/news/news_releases/2006/NR-06-05-06.html`.

P. D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*. SIAM, Philadelphia, 1973.

D. S. Lemons and A. Gythiel. Paul langevins 1908 paper, on the theory of brownian motion. *Am J Phys*, 65:1079–1081, 1997.

Y. K. Leong and Y. L. Yeow. Obtaining the shear stress shear rate relationship and yield stress of liquid foods from couette viscometry data. *Rheol Acta*, 42:365–371, 2003.

K. M. Liew, B. J. Chen, and Z. M. Xiao. Analysis of fracture nucleation in carbon nanotubes through atomistic-based continuum theory. *Phys Rev B: Condens Matter*, 71:235424:1–7, 2005.

M. J. Lighthill and G. B. Whitham. On kinematic waves; ii. a theory of traffic flow on long crowded roads. *Proc R Soc London, Ser A*, 229A: 317–345, 1955.

K. Lucas. *Molecular Models for Fluids*. Cambridge University Press, Cambridge, 2007.

J. E. Mark and B. Erman. *Rubberlike elasticity : a molecular primer*. Cambridge University Press, Cambridge, 2d edition, 2007.

J. A. Maroto, J. Duenas-Molina, and J. de Dios. Experimental evaluation of the drag coefficient for smooth spheres by free fall experiments in old mines. *Euro J Physics*, 26:323–330, 2005.

J. E. Marsden and T. J. R. Hughes. *Mathematical Foundations of Elasticity*. Dover, New York, 1994.

R. M. Mazo. *Brownian Motion: Fluctuations, Dynamics, and Applications*. Oxford University Press, Oxford, 2002.

R. D. Mehta. Aerodynamics of sports balls. *Ann Rev Fluid Mech*, 17:151–189, 1985.

H. Meinhardt. *Models of Biological Pattern Formation*. Academic Press, London, 1982.

L. Michaelis and M. Menten. Die kinetik der invertinwirkung. *Biochem Z*, 49:333–369, 1913.

S. G Mikhlin. *Mathematical physics, An advanced course*. North-Holland, New York, 1970.

Y. Morita, N. Tomita, H. Aoki, S. Wakitani, Y. Tamada, T. Suguro, and K. Ikeuchi. Visco-elastic properties of cartilage tissue regenerated with fibroin sponge. *Bio-Med Mater Eng*, 12:291298, 2002.

Steve Morris. Boeing 777-236/er aircraft. *AirTeamImages*, 2006.

A. Murdoch. Some primitive concepts in continuum mechanics regarded in terms of objective space-time molecular averaging: The key role played by inertial observers. *J Elasticity*, 84:69–97, 2006.

Ames Research Center NASA. Wind tunnel images. Website, 2008. `http://aeronautics.arc.nasa.gov/images.html`.

G. F. Newell. Nonlinear effects in the dynamics of car following. *Operations Res*, 9:209–229, 1961.

B. K. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, New York, 6th edition, 2003.

R. Penrose. *The Road to Reality : A Complete Guide to the Laws of the Universe*. Vintage, New York, 2007.

P. Raos. Modelling of elastic behaviour of rubber and its application in fea. *Plast Rubber Compos Process Appl*, 19:293–303, 1993.

P. I. Richards. Shock waves on the freeway. *Oper Res*, 4:42–51, 1956.

R. Rioboo, C. Bauthier, J. Conti, M. Voue, and J. De Coninck. Experimental investigation of splash and crown formation during single drop impact on wetted surfaces. *Exper Fluids*, 35:648–652, 2003.

H. Sawada and Tetsuya Kunimasu. Sphere drag measurements with the nal 60cm msbs. *J Wind Eng*, 98:129–136, 2004.

D. Schomburg and D. Stephan. *Enzyme Handbook*. Springer-Verlag, New York, 1997.

L. A. Segel and M. Slemrod. The quasi-steady-state assumption: A case study in perturbation. *SIAM Rev*, 31:446–477, 1989.

A. J. Smits and S. Ogg. *Aerodynamics of the golf ball*, chapter 1. Biomedical Engineering Principles in Sports. Kluwer Academic, Boston, 2004.

K. P. Soldatos. On universal relations in orthotropic material subclasses. *Inter J Eng Sci*, 46:306–324, 2008.

C. G. Speziale. Comments on the material frame-indifference controversy. *Phys Rev A: At Mol Opt Phys*, 36:4522–4525, 1987.

C. G. Speziale. A review of material frame-indifference in mechanics. *Appl Mech Rev*, 51:489504, 1998.

J. P. Steinbrenner and J.P. Abelanet. Anistropic tetrahedral meshing based on surface deformation techniques. In *Proceedings of the AIAA 45th Aerospace Sciences Meeting*, pages AIAA–2007–0554, Reno, NV, 2007.

W. J. Stronge and D. Shu. The domino effect: Successive destabilization by cooperative neighbours. *Proc Roy Soc A*, 418:155–163, 1988.

B. Svendsen and A. Bertram. On frame-indifference and form-invariance in constitutive theory. *Acta Mech*, 132:195–207, 1999.

S. Tanaka. Irregular flows. In T. Matsui, editor, *Proceedings of the Third Asian Congress of Fluid Mechanics*, pages 3–14, Toyko, 1986.

R. Temam. *Navier-Stokes Equations: Theory and Numerical Analysis*. American Mathematical Society, New York, 2001.

N. Tillmark and P. H. Alfredsson. Experiments on transition in plane couette flow. *J Fluid Mech*, 235:89–102, 1992.

T. T. Tran, A. Mittal, T. Aldinger, J. W. Polli, A. Ayrton, H. Ellens, and J. Bentz. The elementary mass action rate constants of p-gp transport for a confluent monolayer of mdckii-hmdr1 cells. *Biophys J*, 88:715–738, 2005.

C. Truesdell. *Rational Thermodynamics*. Springer, New York, 2d edition, 1984.

C. Truesdell and W. Noll. *The Non-Linear Field Theories of Mechanics*. Springer, New York, 3d edition, 2004.

A. Turing. The chemical basis of morphogenesis. *Phil Trans Roy Soc B*, 237: 37–72, 1952.

G. Turk. Generating textures on arbitrary surfaces using reaction-diffusion. *Comput Graphics*, 25(4):289–298, 1991.

T. Vanderbilt. *Traffic: Why We Drive the Way We Do (and What It Says About Us)*. Knopf, New York, 2008.

S. Velan and M. Florian. A note on the entropy solutions of the hydrodynamic model of traffic flow. *Trans Sci*, 36:435–446, 2002.

M. Verhaest, W. Lammens, K. Le Roy, B. De Coninck, C. De Ranter, A. Van Laere, W. Van den Ende, and A. Rabijns. X-ray diffraction structure of a cell-wall invertase from arabidopsis thaliana. *Acta Crystallogr, Sect D: Biol Crystallogr*, 62:1555–1563, 2006.

H. F. Weinberger. *A First Course in Partial Differential Equations: with Complex Variables and Transform Methods*. Dover, New York, 1995.

J. Zhou and H. Peng. Range policy of adaptive cruise control vehicles for improved flow stability and string stability. *IEEE Trans Intell Transport Syst*, 6:229–237, 2005.

Index

admissibility condition, 239
advection equation, 218
Alfven speed, 40
Almansi strain, 285, 308
Arrhenius equation, 97
articular cartilage, 172, 283
asymptotically stable, 115, 119, 121, 132
autocatalytic reaction, 95, 127
Avogadro's number, 152

balance law, 170, 210, 361
balancing, 59, 63, 69, 70
bell curve, 145
Belousov-Zhabotinskii reaction, 126
Bernoulli's theorem, 418, 439
Bessel function, 318, 341
binding energy, 310
Blasius boundary layer, 431
Bobyleff-Forsyth formula, 437
Bohr radius, 40
Boltzmann constant, 152, 191
Boltzmann distribution, 177
boundary layer coordinate, 63, 69, 430
boundary layer solution, 63, 69
boundary layer thickness, 428, 433
Bratu's equation, 84
brittle material, 284
Brownian motion, 141
Brownian ratchet, 155
Buckingham Pi Theorem, 16
bungie cord, 266, 288, 302, 306
Burgers' equation, 41

capture silk, 283
carbon nanotube, 287, 310
carburization, 153
Cauchy stress tensor, 366
Cauchy-Green deformation tensor, 306
cellular automata modeling, 248
characteristics, 221, 229
Clausius-Duhem inequality, 299
complementary error function, 25, 166, 318
composite expansion, 66, 71, 108
compressive strain, 292, 313
conservation law, 93, 99, 211
constitutive law, 172, 283, 294
  diffusion, 171
  elastic, 296
  Greenshields law, 213, 231
  linear elastic, 286, 311, 388
  viscoelastic, 331
  viscous fluid, 378
contact discontinuity, 234
continuity equation, 211, 276, 280
control volume, 209, 308
convolution theorem, 161, 321
cooperativity, 138
Couette flow, 405
creep, 282

d'Alembert's paradox, 426
deformation gradient, 305, 354, 387
density, 207, 275, 362
diffusion coefficient, 22, 151
diffusion equation, 42, 151, 182
  point source solution, 42, 181, 184, 413
  radially symmetric, 184, 413
diffusive boundary layer, 428
dimension matrix, 17
dimensionally complete, 17, 19
dimensionally homogeneous, 5, 19
dimensionless product, 8, 18

M.H. Holmes, *Introduction to the Foundations of Applied Mathematics,* Texts in Applied Mathematics 56, DOI 10.1007/978-0-387-87765-5_BM7,
© Springer Science+Business Media, LLC 2009

independent, 18
displacement gradient, 387
distinguished limit, 151
Divergence Theorem, 360
drag coefficient, 9
drag on sphere, 6
drift coefficient, 192
drift diffusion, 176
drift velocity, 196
drift-diffusion equation, 196
driver's ride impulse, 239
du Bois-Reymond lemma, 274
ductile material, 284
Duffing equation, 84

Einstein-Smoluchowski equation, 151, 197
elastic beam, 38
elastic limit, 293
elastic modulus, 4, 286
elastic string, 37
elastomer, 293
elementary reaction, 97, 112
Eley-Rideal mechanism, 134
entropy, 4, 239, 298
entropy condition, 239
epidemic equilibrium, 134, 136
error function, 318
Euclidean transformation, 296, 369
Euler equations, 419
Eulerian coordinates, 352
Eulerian strain, 285
expansion fan, 41, 238, 239
exponential horn, 308
exponential order, 320
extension ratio, 283, 307

Fick's law of diffusion, 171
first Piola-Kirchhoff stress tensor, 383
Fisher's equation, 30
fixed junction model, 307
FKN mechanism, 127
flux, 38, 171, 208, 361
form invariance, 370
Fourier law of heat conduction, 171
Fourier series, 445
Fourier transform, 158
fracture, 293
frame-indifference, 296, 369, 385
fundamental diagram, 216
fundamental dimension, 3, 16

Galilean transformation, 296, 369
gap, 249
geometric analysis, 115
geometric Brownian motion, 193
geometric linearity, 311, 328
globally asymptotically stable, 119
Goldilocks, 150
Green strain, 285, 388
Greenshields constitutive law, 213, 228

half-plane of convergence, 320
Hanes-Woolf plot, 138
Heaviside step function, 319
helical flow, 434
helicity, 438
Helmholtz free energy, 299
Helmholtz Representation Theorem, 415, 431
Helmholtz's Third Vorticity Theorem, 422
Hencky strain, 285
Hill's equation, 138
homogeneous material, 376
Hopf bifurcation, 124
hurricane, 414
hydrogen-bromine reaction, 139
hyperelasticity, 300, 302

ideal fluid, 419
ideal gas, 301, 378
impenetrability of matter, 273, 274, 354
impermeability boundary condition, 380, 423
impulsive plate, 427, 439
incompressibility
  material coordinates, 397
  spatial coordinates, 362
indicator function, 162
infinitesimal deformation, 328
initial layer, 103
inner solution, 63, 107
instantaneous elastic modulus, 344
integro-differential equation, 336, 340
internal energy, 298
interstitial diffusion, 153
Inverse Function Theorem, 274
inverse problems, 327
invertase, 102
inviscid fluid, 419
irrotational flow, 398, 415
isotropic material, 373

Jacobian matrix, 120, 354, 384
jam density, 256

Karman vortex street, 433

Kelvin's Circulation Theorem, 421, 438
Kelvin's Minimum Energy Theorem, 398
Kelvin-Voigt model, 330
Kermack-McKendrick model, 88
ketchup, 407
kinematic viscosity, 428
kinetic energy, 298, 390, 393, 398
Kutta-Joukowski theorem, 426

Lagrangian coordinates, 266, 352
Lagrangian strain, 285, 389
Lamè constants, 389
Langevin equation, 186
Laplace transform, 316
Law of Mass Action, 91
left Cauchy-Green deformation tensor, 393, 397
Leibniz's rule, 274
Lengyel-Epstein model, 20
Lennard-Jones potential, 293
limit cycle, 125
Lincoln Tunnel, 212
linear flow, 394, 436
linear stability analysis, 119

magnetonsonic waves, 40
Markov property, 143
Markovian forcing, 190
mass density, 4
mass, spring, dashpot, 37, 329
master equation, 150, 182, 196
matching condition, 65, 70, 107
material coordinate system, 266, 352
material derivative, 270, 357
material linearity, 311, 328, 343
material velocity gradient tensor, 360
Maxwell model, 330
mean free path, 147, 151
mean-square displacement, 189, 195
measles, 135
mechanical energy equation, 309, 390
merge density, 208, 256
Merritt Parkway, 212
metallic bonding, 290
method of characteristics, 313
  linear wave equation, 221
  nonlinear wave equation, 229
method of multiple scales, 76
Michaelis-Menten reaction, 101, 121
midpoint strain, 285
mobility, 176
momentum equation, 280
  angular, 367, 384
  material coordinates, 279, 308, 384
  spatial coordinates, 279, 367
Mooney-Rivlin model, 307
Morse potential function, 310

N-wave, 247
Nanson's formula, 385
Navier equations, 389
Navier-Stokes equation, 378
Nernst-Planck law, 177
Neubert-Fung relaxation function, 338
Newtonian fluid, 378, 403
no-slip condition, 381, 420
nominal stress tensor, 385
non-isotropic material, 373
non-Newtonian fluid, 406, 436
nondimensionalization, 26, 105
normal stress, 366
nuclear explosion, 37
nullcline, 116, 129

objective tensor, 369, 386
one-way wave equation, 220
Oregonator, 127
outer solution, 63, 68, 106, 128, 430
overlap domain, 64

P-glycoprotein, 102, 108
partial derivative notation, 270, 442
Pascal, 287
pathline, 404, 425
Pauli exclusion principle, 290
peanut butter, 403
pendulum, 33, 72
phantom traffic jam, 228, 245
piecewise continuous, 445
pipe flow, 33, 381, 408
Planck's constant, 40
plasticity, 293
plug-flow reactor, 19
point source solution of diffusion equation, 155, 181, 184, 413
Poiseuille flow, 382, 408
polyconvexity, 304
polytropic fluid, 439
potential energy, 298, 393
potential flow, 417, 423
power-law fluid, 406, 436
predator-prey model, 88, 113
pressure, 301
principal invariants, 373, 400
Principle of Dissipation, 299, 379
Principle of Material Frame-Indifference, 295, 369, 385
projectile problem, 1, 26, 53

pure shear, 394

quantum chromodynamics, 40
quasi-steady-state assumption, 104, 109

radioactive decay, 87
random walk, 142, 179
  biased, 195
  lazy, 197
  non-rectangular lattice, 199
  persistent, 198
  with loss, 199
  with memory, 198
Rankine-Hugoniot condition, 234, 344
rarefaction wave, 238
rate of deformation tensor, 375, 437
reaction analysis, 115
reaction-diffusion equations, 178
red blood cells, 206
red light - green light problem, 221, 240, 251
  modified, 230
reduced entropy inequality, 299
reduced problem, 28, 32, 43
reference configuration, 267, 353
regular perturbation problem, 43
Reiner-Rivlin fluid, 378
resonance, 326
Reynolds number, 9, 430
Reynolds Transport Theorem, 274, 358, 361
Riemann problem, 41, 236
right Cauchy-Green deformation tensor, 388
rigid body motion, 399
Rivlin-Ericksen representation theorem, 372, 377, 400
rotation matrix, 356, 369, 399
Rozenzweig-MacArthur model, 135
rubber, 283, 307

scale model testing, 12
Schnakenberg chemical oscillator, 139
SCTA model, 249
second law of thermodynamics, 239, 298
second Piola-Kirchhoff stress tensor, 387
secular term, 75
shear stress, 366, 407, 436
shock wave, 235, 242
similarity variable, 23, 174, 184, 261
simple shear, 355
singular perturbation problem, 58, 106
SIR model, 89
  SIER, 136
  with vaccination, 134
  with vital dynamics, 135
slinky, 312, 322
slip plane, 293
small disturbance approximation, 226
spatial coordinate system, 267, 352
spatial velocity gradient tensor, 360
spin tensor, 375
standard linear model, 330
steady flow, 404
steady-state, 94, 114, 280
Stirling's approximation, 148, 196
stochastic differential equation, 186
stoichiometric coefficients, 91, 98
stoichiometric matrix, 98
Stokes drag formula, 11, 177, 191
Stokes flow, 11
Stokes' first problem, 427
Stokes' Law, 36
Stokes-Einstein equation, 152, 191
stored energy function, 393
strain
  Almansi, 285, 398
  engineering, 284
  Eulerian, 285
  Finger, 398
  Green, 285, 388, 397
  Hencky, 285
  Lagrangian, 285, 286, 389, 398
  midpoint, 285
  nominal, 284
  true, 284
strain energy function, 393
strain tensor, 397
stream function, 431
stress, 4, 277, 286, 363, 383
stress power, 390
stress relaxation, 281
surface tension, 4
Sutton-Chen potential, 293

Tacoma Narrows Bridge, 327
tautochrone problem, 349
Taylor's theorem, 44, 441
Taylor-Couette problem, 435
Taylor-Sedov formula, 37
telegraph equation, 198
temperature, 299, 378
tensile strain, 292, 313
toothpaste, 407
traffic flow equation
  linear, 212, 218
  nonlinear, 214, 225, 247
  small disturbance approximation, 226

wave velocity, 214, 225
transcendentally small, 50
trimerization, 138
two-timing, 76

uniform approximation, 66, 71
uniform dilatation, 354
universal gas constant, 152

van der Pol equation, 123
van der Waals bonding, 293
velocity gradient tensor, 360, 375
viscoelasticity
  Burger model, 346
  creep function, 347
  Kelvin-Voigt model, 330
  Maxwell model, 330
  relaxation function, 336
  standard linear model, 330
viscosity, 4, 301, 378, 403
viscous dissipation function, 391, 437
viscous fluid, 301
volatility, 192
volume fraction, 261
vortex
  line, 415
  Oseen-Lamb, 413
  Taylor, 437
vorticity, 412, 420, 426, 437
vorticity tensor, 375, 412

wave velocity, 214, 225
weak nonlinearity, 30
Weber number, 34
Webster's equation, 343
well-ordering assumption, 45

Young's modulus, 286, 311

zebra stripes, 179

Texts in Applied Mathematics

(continued after page ii)

31. *Brémaud*: Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues.
32. *Durran*: Numerical Methods for Wave Equations in Geophysical Fluids Dynamics.
33. *Thomas*: Numerical Partial Differential Equations: Conservation Laws and Elliptic Equations.
34. *Chicone*: Ordinary Differential Equations with Applications.
35. *Kevorkian*: Partial Differential Equations: Analytical Solution Techniques, 2nd ed.
36. *Dullerud/Paganini*: A Course in Robust Control Theory: A Convex Approach.
37. *Quarteroni/Sacco/Saleri*: Numerical Mathematics.
38. *Gallier*: Geometric Methods and Applications: For Computer Science and Engineering.
39. *Atkinson/Han*: Theoretical Numerical Analysis: A Functional Analysis Framework, 2nd ed.
40. *Brauer/Castillo-Chávez*: Mathematical Models in Population Biology and Epidemiology.
41. *Davies*: Integral Transforms and Their Applications, 3rd ed.
42. *Deuflhard/Bornemann*: Scientific Computing with Ordinary Differential Equations.
43. *Deuflhard/Hohmann*: Numerical Analysis in Modern Scientific Computing: An Introduction, 2nd ed.
44. *Knabner/Angermann*: Numerical Methods for Elliptic and Parabolic Partial Differential Equations.
45. *Larsson/Thomée*: Partial Differential Equations with Numerical Methods.
46. *Pedregal*: Introduction to Optimization.
47. *Ockendon/Ockendon*: Waves and Compressible Flow.
48. *Hinrichsen*: Mathematical Systems Theory I.
49. *Bullo/Lewis*: Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Simple Mechanical Control Systems.
50. *Verhulst*: Methods and Applications of Singular Perturbations: Boundary Layers and Multiple Timescale Dynamics.
51. *Bondeson/Rylander/Ingelström*: Computational Electromagnetics.
52. *Holmes*: Introduction to Numerical Methods in Differential Equations.
53. *Pavliotis/Stuart*: Multiscale Methods: Averaging and Homogenization.
54. *Hesthaven/Warburton*: Nodal Discontinuous Galerkin Methods.
55. *Allaire/Kaber*: Numerical Linear Algebra.
56. *Mark H. Holmes*: Introduction to the Foundations of Applied Mathematics.