Alain Goriely

APPLIED MATHEMATICS

A Very Short Introduction

OXFORD

Applied Mathematics: A Very Short Introduction

VERY SHORT INTRODUCTIONS are for anyone wanting a stimulating and accessible way into a new subject. They are written by experts, and have been translated into more than 45 different languages.

The series began in 1995, and now covers a wide variety of topics in every discipline. The VSI library currently contains over 550 volumes—a Very Short Introduction to everything from Psychology and Philosophy of Science to American History and Relativity—and continues to grow in every subject area.

Very Short Introductions available now:

ACCOUNTING Christopher Nobes
ADOLESCENCE Peter K. Smith
ADVERTISING Winston Fletcher
AFRICAN AMERICAN RELIGION Eddie S. Glaude Jr
AFRICAN HISTORY John Parker and Richard Rathbone
AFRICAN RELIGIONS Jacob K. Olupona
AGEING Nancy A. Pachana
AGNOSTICISM Robin Le Poidevin
AGRICULTURE Paul Brassley and Richard Soffe
ALEXANDER THE GREAT Hugh Bowden
ALGEBRA Peter M. Higgins
AMERICAN HISTORY Paul S. Boyer
AMERICAN IMMIGRATION David A. Gerber
AMERICAN LEGAL HISTORY G. Edward White
AMERICAN POLITICAL HISTORY Donald Critchlow
AMERICAN POLITICAL PARTIES AND ELECTIONS L. Sandy Maisel
AMERICAN POLITICS Richard M. Valelly
THE AMERICAN PRESIDENCY Charles O. Jones
THE AMERICAN REVOLUTION Robert J. Allison
AMERICAN SLAVERY Heather Andrea Williams
THE AMERICAN WEST Stephen Aron
AMERICAN WOMEN'S HISTORY Susan Ware
ANAESTHESIA Aidan O'Donnell
ANALYTIC PHILOSOPHY Michael Beaney
ANARCHISM Colin Ward
ANCIENT ASSYRIA Karen Radner
ANCIENT EGYPT Ian Shaw
ANCIENT EGYPTIAN ART AND ARCHITECTURE Christina Riggs

ANCIENT GREECE Paul Cartledge
THE ANCIENT NEAR EAST Amanda H. Podany
ANCIENT PHILOSOPHY Julia Annas
ANCIENT WARFARE Harry Sidebottom
ANGELS David Albert Jones
ANGLICANISM Mark Chapman
THE ANGLO-SAXON AGE John Blair
ANIMAL BEHAVIOUR Tristram D. Wyatt
THE ANIMAL KINGDOM Peter Holland
ANIMAL RIGHTS David DeGrazia
THE ANTARCTIC Klaus Dodds
ANTISEMITISM Steven Beller
ANXIETY Daniel Freeman and Jason Freeman
THE APOCRYPHAL GOSPELS Paul Foster
APPLIED MATHEMATICS Alain Goriely
ARCHAEOLOGY Paul Bahn
ARCHITECTURE Andrew Ballantyne
ARISTOCRACY William Doyle
ARISTOTLE Jonathan Barnes
ART HISTORY Dana Arnold
ART THEORY Cynthia Freeland
ASIAN AMERICAN HISTORY Madeline Y. Hsu
ASTROBIOLOGY David C. Catling
ASTROPHYSICS James Binney
ATHEISM Julian Baggini
THE ATMOSPHERE Paul I. Palmer
AUGUSTINE Henry Chadwick
AUSTRALIA Kenneth Morgan
AUTISM Uta Frith
THE AVANT GARDE David Cottington
THE AZTECS Davíd Carrasco
BABYLONIA Trevor Bryce
BACTERIA Sebastian G. B. Amyes
BANKING John Goddard and John O. S. Wilson
BARTHES Jonathan Culler
THE BEATS David Sterritt
BEAUTY Roger Scruton
BEHAVIOURAL ECONOMICS Michelle Baddeley
BESTSELLERS John Sutherland
THE BIBLE John Riches
BIBLICAL ARCHAEOLOGY Eric H. Cline
BIG DATA Dawn E. Holmes
BIOGRAPHY Hermione Lee

BLACK HOLES Katherine Blundell
BLOOD Chris Cooper
THE BLUES Elijah Wald
THE BODY Chris Shilling
THE BOOK OF MORMON Terryl Givens
BORDERS Alexander C. Diener and Joshua Hagen
THE BRAIN Michael O'Shea
BRANDING Robert Jones
THE BRICS Andrew F. Cooper
THE BRITISH CONSTITUTION Martin Loughlin
THE BRITISH EMPIRE Ashley Jackson
BRITISH POLITICS Anthony Wright
BUDDHA Michael Carrithers
BUDDHISM Damien Keown
BUDDHIST ETHICS Damien Keown
BYZANTIUM Peter Sarris
CALVINISM Jon Balserak
CANCER Nicholas James
CAPITALISM James Fulcher
CATHOLICISM Gerald O'Collins
CAUSATION Stephen Mumford and Rani Lill Anjum
THE CELL Terence Allen and Graham Cowling
THE CELTS Barry Cunliffe
CHAOS Leonard Smith
CHEMISTRY Peter Atkins
CHILD PSYCHOLOGY Usha Goswami
CHILDREN'S LITERATURE Kimberley Reynolds
CHINESE LITERATURE Sabina Knight
CHOICE THEORY Michael Allingham
CHRISTIAN ART Beth Williamson
CHRISTIAN ETHICS D. Stephen Long
CHRISTIANITY Linda Woodhead
CIRCADIAN RHYTHMS Russell Foster and Leon Kreitzman
CITIZENSHIP Richard Bellamy
CIVIL ENGINEERING David Muir Wood
CLASSICAL LITERATURE William Allan
CLASSICAL MYTHOLOGY Helen Morales
CLASSICS Mary Beard and John Henderson
CLAUSEWITZ Michael Howard
CLIMATE Mark Maslin
CLIMATE CHANGE Mark Maslin
CLINICAL PSYCHOLOGY Susan Llewelyn and Katie Aafjes-van Doorn
COGNITIVE NEUROSCIENCE Richard Passingham

THE COLD WAR Robert McMahon
COLONIAL AMERICA Alan Taylor
COLONIAL LATIN AMERICAN LITERATURE Rolena Adorno
COMBINATORICS Robin Wilson
COMEDY Matthew Bevis
COMMUNISM Leslie Holmes
COMPLEXITY John H. Holland
THE COMPUTER Darrel Ince
COMPUTER SCIENCE Subrata Dasgupta
CONFUCIANISM Daniel K. Gardner
THE CONQUISTADORS Matthew Restall and Felipe Fernández-Armesto
CONSCIENCE Paul Strohm
CONSCIOUSNESS Susan Blackmore
CONTEMPORARY ART Julian Stallabrass
CONTEMPORARY FICTION Robert Eaglestone
CONTINENTAL PHILOSOPHY Simon Critchley
COPERNICUS Owen Gingerich
CORAL REEFS Charles Sheppard
CORPORATE SOCIAL RESPONSIBILITY Jeremy Moon
CORRUPTION Leslie Holmes
COSMOLOGY Peter Coles
CRIME FICTION Richard Bradford
CRIMINAL JUSTICE Julian V. Roberts
CRITICAL THEORY Stephen Eric Bronner
THE CRUSADES Christopher Tyerman
CRYPTOGRAPHY Fred Piper and Sean Murphy
CRYSTALLOGRAPHY A. M. Glazer
THE CULTURAL REVOLUTION Richard Curt Kraus
DADA AND SURREALISM David Hopkins
DANTE Peter Hainsworth and David Robey
DARWIN Jonathan Howard
THE DEAD SEA SCROLLS Timothy H. Lim
DECOLONIZATION Dane Kennedy
DEMOCRACY Bernard Crick
DEPRESSION Jan Scott and Mary Jane Tacchi
DERRIDA Simon Glendinning
DESCARTES Tom Sorell
DESERTS Nick Middleton
DESIGN John Heskett
DEVELOPMENTAL BIOLOGY Lewis Wolpert
THE DEVIL Darren Oldridge
DIASPORA Kevin Kenny
DICTIONARIES Lynda Mugglestone

DINOSAURS David Norman
DIPLOMACY Joseph M. Siracusa
DOCUMENTARY FILM Patricia Aufderheide
DREAMING J. Allan Hobson
DRUGS Les Iversen
DRUIDS Barry Cunliffe
EARLY MUSIC Thomas Forrest Kelly
THE EARTH Martin Redfern
EARTH SYSTEM SCIENCE Tim Lenton
ECONOMICS Partha Dasgupta
EDUCATION Gary Thomas
EGYPTIAN MYTH Geraldine Pinch
EIGHTEENTHCENTURY BRITAIN Paul Langford
THE ELEMENTS Philip Ball
EMOTION Dylan Evans
EMPIRE Stephen Howe
ENGELS Terrell Carver
ENGINEERING David Blockley
THE ENGLISH LANGUAGE Simon Horobin
ENGLISH LITERATURE Jonathan Bate
THE ENLIGHTENMENT John Robertson
ENTREPRENEURSHIP Paul Westhead and Mike Wright
ENVIRONMENTAL ECONOMICS Stephen Smith
ENVIRONMENTAL LAW Elizabeth Fisher
ENVIRONMENTAL POLITICS Andrew Dobson
EPICUREANISM Catherine Wilson
EPIDEMIOLOGY Rodolfo Saracci
ETHICS Simon Blackburn
ETHNOMUSICOLOGY Timothy Rice
THE ETRUSCANS Christopher Smith
EUGENICS Philippa Levine
THE EUROPEAN UNION John Pinder and Simon Usherwood
EUROPEAN UNION LAW Anthony Arnull
EVOLUTION Brian and Deborah Charlesworth
EXISTENTIALISM Thomas Flynn
EXPLORATION Stewart A. Weaver
THE EYE Michael Land
FAIRY TALE Marina Warner
FAMILY LAW Jonathan Herring
FASCISM Kevin Passmore
FASHION Rebecca Arnold
FEMINISM Margaret Walters
FILM Michael Wood

FILM MUSIC Kathryn Kalinak
THE FIRST WORLD WAR Michael Howard
FOLK MUSIC Mark Slobin
FOOD John Krebs
FORENSIC PSYCHOLOGY David Canter
FORENSIC SCIENCE Jim Fraser
FORESTS Jaboury Ghazoul
FOSSILS Keith Thomson
FOUCAULT Gary Gutting
THE FOUNDING FATHERS R. B. Bernstein
FRACTALS Kenneth Falconer
FREE SPEECH Nigel Warburton
FREE WILL Thomas Pink
FREEMASONRY Andreas Önnerfors
FRENCH LITERATURE John D. Lyons
THE FRENCH REVOLUTION William Doyle
FREUD Anthony Storr
FUNDAMENTALISM Malise Ruthven
FUNGI Nicholas P. Money
THE FUTURE Jennifer M. Gidley
GALAXIES John Gribbin
GALILEO Stillman Drake
GAME THEORY Ken Binmore
GANDHI Bhikhu Parekh
GENES Jonathan Slack
GENIUS Andrew Robinson
GEOGRAPHY John Matthews and David Herbert
GEOPOLITICS Klaus Dodds
GERMAN LITERATURE Nicholas Boyle
GERMAN PHILOSOPHY Andrew Bowie
GLOBAL CATASTROPHES Bill McGuire
GLOBAL ECONOMIC HISTORY Robert C. Allen
GLOBALIZATION Manfred Steger
GOD John Bowker
GOETHE Ritchie Robertson
THE GOTHIC Nick Groom
GOVERNANCE Mark Bevir
GRAVITY Timothy Clifton
THE GREAT DEPRESSION AND THE NEW DEAL Eric Rauchway
HABERMAS James Gordon Finlayson
THE HABSBURG EMPIRE Martyn Rady
HAPPINESS Daniel M. Haybron
THE HARLEM RENAISSANCE Cheryl A. Wall

THE HEBREW BIBLE AS LITERATURE Tod Linafelt
HEGEL Peter Singer
HEIDEGGER Michael Inwood
THE HELLENISTIC AGE Peter Thonemann
HEREDITY John Waller
HERMENEUTICS Jens Zimmermann
HERODOTUS Jennifer T. Roberts
HIEROGLYPHS Penelope Wilson
HINDUISM Kim Knott
HISTORY John H. Arnold
THE HISTORY OF ASTRONOMY Michael Hoskin
THE HISTORY OF CHEMISTRY William H. Brock
THE HISTORY OF CINEMA Geoffrey Nowell-Smith
THE HISTORY OF LIFE Michael Benton
THE HISTORY OF MATHEMATICS Jacqueline Stedall
THE HISTORY OF MEDICINE William Bynum
THE HISTORY OF PHYSICS J. L. Heilbron
THE HISTORY OF TIME Leofranc Holford-Strevens
HIV AND AIDS Alan Whiteside
HOBBES Richard Tuck
HOLLYWOOD Peter Decherney
HOME Michael Allen Fox
HORMONES Martin Luck
HUMAN ANATOMY Leslie Klenerman
HUMAN EVOLUTION Bernard Wood
HUMAN RIGHTS Andrew Clapham
HUMANISM Stephen Law
HUME A. J. Ayer
HUMOUR Noël Carroll
THE ICE AGE Jamie Woodward
IDEOLOGY Michael Freeden
THE IMMUNE SYSTEM Paul Klenerman
INDIAN CINEMA Ashish Rajadhyaksha
INDIAN PHILOSOPHY Sue Hamilton
THE INDUSTRIAL REVOLUTION Robert C. Allen
INFECTIOUS DISEASE Marta L. Wayne and Benjamin M. Bolker
INFINITY Ian Stewart
INFORMATION Luciano Floridi
INNOVATION Mark Dodgson and David Gann
INTELLIGENCE Ian J. Deary
INTELLECTUAL PROPERTY Siva Vaidhyanathan
INTERNATIONAL LAW Vaughan Lowe
INTERNATIONAL MIGRATION Khalid Koser

INTERNATIONAL RELATIONS Paul Wilkinson
INTERNATIONAL SECURITY Christopher S. Browning
IRAN Ali M. Ansari
ISLAM Malise Ruthven
ISLAMIC HISTORY Adam Silverstein
ISOTOPES Rob Ellam
ITALIAN LITERATURE Peter Hainsworth and David Robey
JESUS Richard Bauckham
JEWISH HISTORY David N. Myers
JOURNALISM Ian Hargreaves
JUDAISM Norman Solomon
JUNG Anthony Stevens
KABBALAH Joseph Dan
KAFKA Ritchie Robertson
KANT Roger Scruton
KEYNES Robert Skidelsky
KIERKEGAARD Patrick Gardiner
KNOWLEDGE Jennifer Nagel
THE KORAN Michael Cook
LAKES Warwick F. Vincent
LANDSCAPE ARCHITECTURE Ian H. Thompson
LANDSCAPES and GEOMORPHOLOGY Andrew Goudie and Heather Viles
LANGUAGES Stephen R. Anderson
LATE ANTIQUITY Gillian Clark
LAW Raymond Wacks
THE LAWS OF THERMODYNAMICS Peter Atkins
LEADERSHIP Keith Grint
LEARNING Mark Haselgrove
LEIBNIZ Maria Rosa Antognazza
LIBERALISM Michael Freeden
LIGHT Ian Walmsley
LINCOLN Allen C. Guelzo
LINGUISTICS Peter Matthews
LITERARY THEORY Jonathan Culler
LOCKE John Dunn
LOGIC Graham Priest
LOVE Ronald de Sousa
MACHIAVELLI Quentin Skinner
MADNESS Andrew Scull
MAGIC Owen Davies
MAGNA CARTA Nicholas Vincent
MAGNETISM Stephen Blundell
MALTHUS Donald Winch

MAMMALS T. S. Kemp
MANAGEMENT John Hendry
MAO Delia Davin
MARINE BIOLOGY Philip V. Mladenov
THE MARQUIS DE SADE John Phillips
MARTIN LUTHER Scott H. Hendrix
MARTYRDOM Jolyon Mitchell
MARX Peter Singer
MATERIALS Christopher Hall
MATHEMATICS Timothy Gowers
THE MEANING OF LIFE Terry Eagleton
MEASUREMENT David Hand
MEDICAL ETHICS Tony Hope
MEDICAL LAW Charles Foster
MEDIEVAL BRITAIN John Gillingham and Ralph A. Griffiths
MEDIEVAL LITERATURE Elaine Treharne
MEDIEVAL PHILOSOPHY John Marenbon
MEMORY Jonathan K. Foster
METAPHYSICS Stephen Mumford
THE MEXICAN REVOLUTION Alan Knight
MICHAEL FARADAY Frank A. J. L. James
MICROBIOLOGY Nicholas P. Money
MICROECONOMICS Avinash Dixit
MICROSCOPY Terence Allen
THE MIDDLE AGES Miri Rubin
MILITARY JUSTICE Eugene R. Fidell
MILITARY STRATEGY Antulio J. Echevarria
II MINERALS David Vaughan
MIRACLES Yujin Nagasawa
MODERN ART David Cottington
MODERN CHINA Rana Mitter
MODERN DRAMA Kirsten E. Shepherd-Barr
MODERN FRANCE Vanessa R. Schwartz
MODERN INDIA Craig Jeffrey
MODERN IRELAND Senia Pašeta
MODERN ITALY Anna Cento Bull
MODERN JAPAN Christopher Goto-Jones
MODERN LATIN AMERICAN LITERATURE Roberto González Echevarría
MODERN WAR Richard English
MODERNISM Christopher Butler
MOLECULAR BIOLOGY Aysha Divan and Janice A. Royds
MOLECULES Philip Ball
MONASTICISM Stephen J. Davis

THE MONGOLS Morris Rossabi
MOONS David A. Rothery
MORMONISM Richard Lyman Bushman
MOUNTAINS Martin F. Price
MUHAMMAD Jonathan A. C. Brown
MULTICULTURALISM Ali Rattansi
MULTILINGUALISM John C. Maher
MUSIC Nicholas Cook
MYTH Robert A. Segal
THE NAPOLEONIC WARS Mike Rapport
NATIONALISM Steven Grosby
NATIVE AMERICAN LITERATURE Sean Teuton
NAVIGATION Jim Bennett
NELSON MANDELA Elleke Boehmer
NEOLIBERALISM Manfred Steger and Ravi Roy
NETWORKS Guido Caldarelli and Michele Catanzaro
THE NEW TESTAMENT Luke Timothy Johnson
THE NEW TESTAMENT AS LITERATURE Kyle Keefer
NEWTON Robert Iliffe
NIETZSCHE Michael Tanner
NINETEENTHCENTURY BRITAIN Christopher Harvie and H. C. G. Matthew
THE NORMAN CONQUEST George Garnett
NORTH AMERICAN INDIANS Theda Perdue and Michael D. Green
NORTHERN IRELAND Marc Mulholland
NOTHING Frank Close
NUCLEAR PHYSICS Frank Close
NUCLEAR POWER Maxwell Irvine
NUCLEAR WEAPONS Joseph M. Siracusa
NUMBERS Peter M. Higgins
NUTRITION David A. Bender
OBJECTIVITY Stephen Gaukroger
OCEANS Dorrik Stow
THE OLD TESTAMENT Michael D. Coogan
THE ORCHESTRA D. Kern Holoman
ORGANIC CHEMISTRY Graham Patrick
ORGANIZATIONS Mary Jo Hatch
ORGANIZED CRIME Georgios A. Antonopoulos and Georgios Papanicolaou
PAGANISM Owen Davies
PAIN Rob Boddice
THE PALESTINIAN-ISRAELI CONFLICT Martin Bunton
PANDEMICS Christian W. McMillen
PARTICLE PHYSICS Frank Close
PAUL E. P. Sanders

PEACE Oliver P. Richmond
PENTECOSTALISM William K. Kay
PERCEPTION Brian Rogers
THE PERIODIC TABLE Eric R. Scerri
PHILOSOPHY Edward Craig
PHILOSOPHY IN THE ISLAMIC WORLD Peter Adamson
PHILOSOPHY OF LAW Raymond Wacks
PHILOSOPHY OF SCIENCE Samir Okasha
PHILOSOPHY OF RELIGION Tim Bayne
PHOTOGRAPHY Steve Edwards
PHYSICAL CHEMISTRY Peter Atkins
PILGRIMAGE Ian Reader
PLAGUE Paul Slack
PLANETS David A. Rothery
PLANTS Timothy Walker
PLATE TECTONICS Peter Molnar
PLATO Julia Annas
POLITICAL PHILOSOPHY David Miller
POLITICS Kenneth Minogue
POPULISM Cas Mudde and Cristóbal Rovira Kaltwasser
POSTCOLONIALISM Robert Young
POSTMODERNISM Christopher Butler
POSTSTRUCTURALISM Catherine Belsey
PREHISTORY Chris Gosden
PRESOCRATIC PHILOSOPHY Catherine Osborne
PRIVACY Raymond Wacks
PROBABILITY John Haigh
PROGRESSIVISM Walter Nugent
PROJECTS Andrew Davies
PROTESTANTISM Mark A. Noll
PSYCHIATRY Tom Burns
PSYCHOANALYSIS Daniel Pick
PSYCHOLOGY Gillian Butler and Freda McManus
PSYCHOTHERAPY Tom Burns and Eva Burns-Lundgren
PUBLIC ADMINISTRATION Stella Z. Theodoulou and Ravi K. Roy
PUBLIC HEALTH Virginia Berridge
PURITANISM Francis J. Bremer
THE QUAKERS Pink Dandelion
QUANTUM THEORY John Polkinghorne
RACISM Ali Rattansi
RADIOACTIVITY Claudio Tuniz
RASTAFARI Ennis B. Edmonds
THE REAGAN REVOLUTION Gil Troy

REALITY Jan Westerhoff
THE REFORMATION Peter Marshall
RELATIVITY Russell Stannard
RELIGION IN AMERICA Timothy Beal
THE RENAISSANCE Jerry Brotton
RENAISSANCE ART Geraldine A. Johnson
REVOLUTIONS Jack A. Goldstone
RHETORIC Richard Toye
RISK Baruch Fischhoff and John Kadvany
RITUAL Barry Stephenson
RIVERS Nick Middleton
ROBOTICS Alan Winfield
ROCKS Jan Zalasiewicz
ROMAN BRITAIN Peter Salway
THE ROMAN EMPIRE Christopher Kelly
THE ROMAN REPUBLIC David M. Gwynn
ROMANTICISM Michael Ferber
ROUSSEAU Robert Wokler
RUSSELL A. C. Grayling
RUSSIAN HISTORY Geoffrey Hosking
RUSSIAN LITERATURE Catriona Kelly
THE RUSSIAN REVOLUTION S. A. Smith
SAVANNAS Peter A. Furley
SCHIZOPHRENIA Chris Frith and Eve Johnstone
SCHOPENHAUER Christopher Janaway
SCIENCE AND RELIGION Thomas Dixon
SCIENCE FICTION David Seed
THE SCIENTIFIC REVOLUTION Lawrence M. Principe
SCOTLAND Rab Houston
SEXUALITY Véronique Mottier
SHAKESPEARE'S COMEDIES Bart van Es
SHAKESPEARE'S SONNETS and POEMS Jonathan F. S. Post
SHAKESPEARE'S TRAGEDIES Stanley Wells
SIKHISM Eleanor Nesbitt
THE SILK ROAD James A. Millward
SLANG Jonathon Green
SLEEP Steven W. Lockley and Russell G. Foster
SOCIAL AND CULTURAL ANTHROPOLOGY John Monaghan and Peter Just
SOCIAL PSYCHOLOGY Richard J. Crisp
SOCIAL WORK Sally Holland and Jonathan Scourfield
SOCIALISM Michael Newman
SOCIOLINGUISTICS John Edwards
SOCIOLOGY Steve Bruce

SOCRATES C. C. W. Taylor
SOUND Mike Goldsmith
THE SOVIET UNION Stephen Lovell
THE SPANISH CIVIL WAR Helen Graham
SPANISH LITERATURE Jo Labanyi
SPINOZA Roger Scruton
SPIRITUALITY Philip Sheldrake
SPORT Mike Cronin
STARS Andrew King
STATISTICS David J. Hand
STEM CELLS Jonathan Slack
STRUCTURAL ENGINEERING David Blockley
STUART BRITAIN John Morrill
SUPERCONDUCTIVITY Stephen Blundell
SYMMETRY Ian Stewart
TAXATION Stephen Smith
TEETH Peter S. Ungar
TELESCOPES Geoff Cottrell
TERRORISM Charles Townshend
THEATRE Marvin Carlson
THEOLOGY David F. Ford
THINKING AND REASONING Jonathan St B. T. Evans
THOMAS AQUINAS Fergus Kerr
THOUGHT Tim Bayne
TIBETAN BUDDHISM Matthew T. Kapstein
TOCQUEVILLE Harvey C. Mansfield
TRAGEDY Adrian Poole
TRANSLATION Matthew Reynolds
THE TROJAN WAR Eric H. Cline
TRUST Katherine Hawley
THE TUDORS John Guy
TWENTIETHCENTURY BRITAIN Kenneth O. Morgan
THE UNITED NATIONS Jussi M. Hanhimäki
THE U.S. CONGRESS Donald A. Ritchie
THE U.S. SUPREME COURT Linda Greenhouse
UTILITARIANISM Katarzyna de Lazari-Radek and Peter Singer
UNIVERSITIES AND COLLEGES David Palfreyman and Paul Temple
UTOPIANISM Lyman Tower Sargent
VETERINARY SCIENCE James Yeates
THE VIKINGS Julian Richards
VIRUSES Dorothy H. Crawford
VOLTAIRE Nicholas Cronk
WAR AND TECHNOLOGY Alex Roland

WATER John Finney
WEATHER Storm Dunlop
THE WELFARE STATE David Garland
WILLIAM SHAKESPEARE Stanley Wells
WITCHCRAFT Malcolm Gaskill
WITTGENSTEIN A. C. Grayling
WORK Stephen Fineman
WORLD MUSIC Philip Bohlman
THE WORLD TRADE ORGANIZATION Amrita Narlikar
WORLD WAR II Gerhard L. Weinberg
WRITING AND SCRIPT Andrew Robinson
ZIONISM Michael Stanislawski

Available soon:

DEVELOPMENT Ian Goldin
COMPARATIVE LITERATURE Ben Hutchinson
GENOMICS John Archibald
SEXUAL SELECTION Marlene Zuk and Leigh W. Simmons
CRIMINOLOGY Tim Newburn

For more information visit our website

www.oup.com/vsi/

Alain Goriely

APPLIED MATHEMATICS

A Very Short Introduction

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP, United Kingdom

Oxford University Press is a department of the University of Oxford. If furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

First edition published 2018
Impression: 1

Published in the United States of America by Oxford University Press 198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2017954542

ISBN 978–0–19–875404–6
Ebook ISBN 978–0–19–106888–1

Printed in Great Britain by Ashford Colour Press Ltd, Gosport, Hampshire

To Nita

Contents

Preface

Before I speak, I have something important to say.

–Groucho (attributed)

When meeting new people at a party, it is customary, out of politeness, to ask what one does. When it happens, I usually mumble that I am a doctor, but not the real kind, just the academic kind. Then, if further questioned, I admit that I am mathematician, but not the real kind, just the applied kind. Then, the dreadful question comes: what is applied mathematics? In desperation, I use one of the usual witty platitudes ('It is like mathematics but useful and more fun', 'We are the social kind, we look at other people's shoes', 'Applied mathematics is to pure mathematics, what pop music is to classical music'). After the awkward pause that sums up most of my human interactions, I look for the closest exit, convinced that further contact would inevitably deepen my existentialist crisis. As I walk out I ask myself if I could really make an honest statement about my own field of study that has also become my way of life. Why is applied mathematics so different from scientific disciplines and so clearly unlike pure mathematics? How could I possibly explain the constant excitement and fun that it brings to my intellectual life?

The decision to write a *Very Short Introduction* for applied mathematics is an attempt to answer this single question: what is applied mathematics? Rather than giving an encyclopaedic description, my aim here is to give a feeling for the problems that applied mathematicians face every day and how they shape their views of the world. In most cases, I use historical perspectives to tell the

story of how certain scientific or mathematical problems develop into modern mathematical theories and how these theories are still active fields of research with outstanding challenges.

Unavoidably, I introduce a few equations. It is always a danger but I do not apologize for it. You can hardly expect to open a book about French Literature and not expect to find a few French words. Equations are the language of mathematics. In the proper context, equations sum up concisely simple self-evident truths. The reader not familiar with such expressions should not unduly worry and can safely skip the technical parts. Equations are included as their absence would make them mythical quantities and their invocation without explicit mention would border on esotericism.

When I started writing this book, I still hoped that my sons could be convinced to study applied mathematics. How could they possibly resist the pressure of a perfect argument so nicely illustrated by so many examples? I certainly managed to convince myself that applied mathematics was indeed the queen of all sciences. However, it became clear that they are not likely to follow in my footsteps and that forcing them to read my writings isa form of cruel and unusual punishment. Yet, I have not given up hope that other mathematically inclined readers may learn about the topic and be charmed by its endless possibilities. There is great beauty in mathematics and there is great beauty in the world around us. Applied mathematics brings the two together in a way that it is not always beautiful but that is always interesting and exciting.

Reading playlist

Pure mathematics is often associated with classical music for its beauty and construction. When I work, I usually listen to classical music, but when I write I enjoy something with a little more energy and fun. Here are a few suggestions that are naturally associated with this book:

‘Should I stay or should I go?’ (The Clash)

‘What’s so funny ‘bout peace, love & understanding?’ (Elvis Costello)

‘Do you want to know a secret?’ (The Beatles)

‘Do you believe in magic?’ (The lovin’ spoonful)

‘Do you know the way to San Jose?’ (Dionne Warwick)

‘What’s the frequency, Kenneth?’ (R.E.M.)

‘Can you picture that?’ (Dr Teeth and the Electric Mayhem)

‘War, (what is it good for?)’ (Edwin Starr)

‘Where are we going?’ (Marvin Gaye)

‘Can you please crawl out your window?’ (Bob Dylan)

Oxford, UK
May 2017

Acknowledgements

I mostly blame Dyrol Lumbard for the existence of this book. It never crossed my mind to write it and I had no particular desire to do it until he twisted my arm and told me that I might actually enjoy doing it. I wish I could say he was wrong. Looking back, I am glad he pushed me out of my comfort zone.

I am most grateful to a number of colleagues and friends for reading and commenting on earlier drafts of this book: Michel Destrade, Nita Goriely, Thomas Lessinnes, Angela Mihai, Derek Moulton, Nick Trefethen, Dominic Vella, and John Wettlaufer. I also take this opportunity to thank the good folks at Oxford University Press (Joy Mellor, Latha Menon, and Jenny Nugee) for their work in seeing this *Very Short Introduction* through publication.

List of illustrations

15 A harmonic of the fundamental mode of a vibrating string

16 An example of D'Alembert's solution

17 As seismic P-waves travel through the Earth's mantle

18 Typically waves feel the effect of dispersion

19 Two solitons interacting

20 A box classification of applied mathematics

21 Traditional X-ray radiography

22 Computed tomography (CT)
Du Cane Medical Imaging Ltd / Science Source.

23 X-ray crystallography

24 X-ray crystallography of helices and DNA
Provided by The Archives at King's College, London.

25 Digital images are arrays of numbers

26 JPEG compression
Time Life Pictures / Mansell / The LIFE Picture Collection / Getty Images.

27 Wavelets

28 Wavelet compression
Time Life Pictures / Mansell / The LIFE Picture Collection / Getty Images.

29 Compressed sensing is used in MRI scanners

30 Complex numbers

31 A gimbal

32 A table of prime knots

33 The square knot

34 Surgery on knots

35 The unknotting problem

36 Topoisomerase 1 action on DNA

37 Action of the resolvase Tn3
Figure adapted from figure 14 in N.R. Cozzarelli, M.A. Krasnow, S.P. Gerrard, and J.H. White (1984), 'A topological treatment of recombination and topoisomerases', *Cold Spring Harbor Symposia on Quantitative Biology* 49: 383–400. Copyright holder: Cold Spring Harbor Laboratory Press.

38 A network of family relationships

39 Matrices and networks

40 The small-world effect

41 The airline network
Jpatokal, CC BY-SA 3.0.

42 The brain network
K.E. Stephan, C.C. Hilgetag, M.A. O'Neill, M.P. Young, and R. Kotter (2000), 'Computational Analysis of Functional Connectivity between Areas of Primate Cortex', *Philosophical Transaction of the Royal Society of London B Biological Sciences* 355(1393), 111–26. By permission of the Royal Society.

43 Brain folding
Figure courtesy of Ellen Kuhl, Stanford University.

44 Mammalian brains
Reprinted from A. Goriely, S. Budday, and E. Kuhl, 'Advances in Applied Mechanics', *Neuromechanics: From Neurons to Brain*, 48: 79–139. Copyright (2015) with permission from Elsevier.

Chapter 1

What's so funny 'bout applied mathematics? Modelling, theory, and methods

Please accept my resignation. I don't want to belong to any club that will accept me as a member.

–Groucho

The modern world of mathematics is divided into different categories and if you are so lucky as to meet real-life mathematicians and engage them in a conversation, they will typically tell you that they are either mathematicians or applied mathematicians. You have probably heard of mathematics, but what is applied mathematics? A quick look on the Internet will give you conflicting definitions. It will also reveal that applied mathematics has found its place in modern academia. As such it is recognized by international scientific societies, journals, and the usual conferences. What is so special about applied mathematics? How is it different from mathematics, or any other scientific discipline?

Mathematics

Let us start with mathematics itself. Whereas philosophers still ponder the best definition, most scientists and mathematicians agree that modern mathematics is an intellectual discipline whose aim is to study idealized objects and their relationships, based on formal logic. Mathematics stands apart from scientific disciplines because it is not restricted by reality. It proceeds solely through logic and is only restricted by our imagination. Indeed, once structures and operations have been defined in a formal setting, the possibilities are endless. You can think of it as a game with very precise rules. Once the rules are laid out, the game of proving or disproving a statement proceeds.

For example, mathematicians have enjoyed numbers for millennia. Take, for instance, the natural numbers (0,1,2, …) and the familiar multiplication operation (×). If we take two numbers p and q together, we obtain a third one as $n = p \times q$. A simple question is then to do the reverse operation: given a number n can we find two numbers p and q such that $n = p \times q$? The simple answer is: of course! Take $p = 1$ and $q = n$. If this is the only possible way that a natural number n larger than 1 can be written as a product of two numbers, then n is called *a prime number*. Mathematicians love prime numbers and their wonderful, and oftentimes, surprising properties. We can now try to prove or disprove statements about these numbers. Let us start with simple ones. We can prove that *there exist prime numbers* by showing that the natural numbers 2, 3, and 5 have all the required properties to be prime numbers. We can disprove the naive statement that *all odd numbers are prime* by showing that $9 = 3 \times 3$. A more interesting statement is that *there are infinitely many prime numbers*. This was first investigated *c*.300 BC by Euclid who showed that new larger prime numbers can always be constructed from the list of all known prime numbers up to a certain value. As we construct new prime numbers the list of prime numbers increases indefinitely. Prime numbers have beautiful properties and play a central role in number theory and pure mathematics. Mathematicians are still trying to establish simple relationships between them. For instance, most mathematicians believe *there are infinitely many pairs of prime numbers that differ by 2*, the so-called *twin-prime conjecture* (a *conjecture* is a statement believed to be true but still unconfirmed). For example, (5,7), (11,13), and (18369287,18369289) are all pairs of primes separated by 2, and many more such pairs are known. The burning question is: are there infinitely many such

pairs? Mathematicians do believe that it is the case but demonstrating this seemingly simple property is so difficult that it has not yet been proved or disproved. However, at the time of writing, a recent breakthrough has taken place. It was established that there exist infinitely many pairs of prime numbers that differ by 246. This result shook the mathematical community and the subject is now a hot topic of modern mathematics.

Through centuries of formalization and generalization, mathematics has evolved into a unified field with clear rules. The rules have been systematically codified by the introduction of formal ideas such as the notions of definition, axiom, proposition, lemma, theorem, and conjecture. These, in turn, are the guidelines through which a stated truth may be systematically verified, step by step, given enough time and patience. Nowadays, mathematics presents itself as a well-organized discipline with well-defined sub-disciplines, some of which may even have familiar names. For instance, number theory, algebra, and geometry are all accepted branches of mathematics. Mathematics is sometimes referred to as pure mathematics to further reinforce its ethereal quality. Pure mathematics has reigned supreme in the pantheon of human intellectual constructs for more than 2,000 years. As the Greek philosopher Aristotle said: ‘The mathematical sciences particularly exhibit order, symmetry, and limitation; and these are the greatest forms of the beautiful.’

A Study in Contrast

I wish I could stand on the shoulders of 2,000 years of history and intellectual achievement, and affirm with pride that I am a mathematician in the tradition of the ancient Greeks. Unfortunately, the situation is not so pristine when it comes to mathematics’ oxymoronic sibling: applied mathematics. Applied mathematics is an intellectual bastard, born from an illicit relationship between the grand noble discipline of mathematics and its plebeian neighbour, the natural sciences; it is the illegitimate offspring of pure rationality and empirical wisdom. This misbegotten hybrid belongs in neither world. It lives in limbo, between the two, in a mathematical and scientific purgatory. Due to its unfortunate heritage, it is difficult to find a universal definition of applied mathematics, even among its practitioners and forebears.

Richard Courant, who founded one of the first centres of applied mathematics, the Courant Institute at New York University, once said:

> Applied mathematics is not a definable scientific field but a human attitude. The attitude of the applied scientist is directed towards finding clear cut answers which can stand the test of empirical observations.

The distinction between pure and applied mathematics is a relatively new concept. It will come as no surprise to the reader that mathematics' origins, not unlike society's nobility, can be traced to modest origins. Indeed, most branches of modern mathematics arose from earthly and human concerns. Arithmetic and numbers were developed for the purpose of trade and tax collection. Many ideas of geometry originated in problems related to measuring distances and making maps. The field of analysis is connected to physical or engineering problems related to mundane concerns such as the design of pendulum clocks, pipe flows, steam engines, or the construction of bridges. The great mathematicians of the past like Descartes, Newton, Euler, Gauss, Kovalevskaya, Maxwell, and Kelvin are all pure and applied mathematicians. Indeed, their work has been of equal importance to mathematics, physics, and engineering. The distinction between sciences and mathematics is a modern construct born from our eagerness to draw borders around disciplines and place ourselves into well-defined boxes. Before the 20th century, scientists or mathematicians were known simply as *natural philosophers*.

Natural philosophy is the general study of mathematics, nature, and the physical universe. It promotes a unified view of the sciences through mathematics, and the development of new mathematics through science that is still dear to many modern applied mathematicians.

Rather than engage in futile epistemological discussions, I will take a pragmatic approach. I will argue that applied mathematics includes the modelling of natural phenomena and human endeavours, the study of mathematical ideas originating from these models, and the systematic development of theoretical and computational tools to probe models, handle data, and gain insight into any problem that has been properly quantified (with special emphasis on *properly* and *quantified*, as there are many abuses

of mathematics in areas where problems are neither quantified nor properly defined).

Applied mathematics is best characterized by three intertwined areas: modelling, theory, and methods.

Modelling refers to the intellectual steps that start with a question, phenomenon, set of data, or any process that attracts our curiosity; then, through suitable assumptions, identifies the key elements (the *variables*) that can be described mathematically, and derives the relations that these variables satisfy (the *equations*). In the best situations, modelling provides us with a well-defined mathematical problem from which one can hope to make predictions, extract information, or simply gain insight.

Theory is the conceptual framework that provides a systematic way to endow data and equations with a clear meaning based on fundamental or phenomenological principles. It includes mathematics itself as its central organizing language and logical framework as well as the theory of statistics that provides precise meaning to data. The formulation of mathematical models also relies on scientific and engineering theories that have been developed over past centuries.

Methods are the tools necessary to extract useful information from equations or data. It includes both the theoretical tools that help us solve equations, as well as the algorithms and computational techniques that are used to solve equations and manipulate data.

Any work in applied mathematics fits in one of these three categories or combines them judiciously. To illustrate these different aspects, let us briefly consider an example.

Burning Candles

In the 19th century the great English scientist Michael Faraday made fundamental contributions to electromagnetism and electrochemistry. He was

also known as a gifted lecturer. In 1825, he initiated the Royal Institution Christmas Lectures, a series of public lectures that are still held every year (see Figure 1). In ‘The Chemical History of a Candle’, he begins:

1. Faraday and his lectures at the Royal Institution are celebrated on the Series E of the £20 British note, in circulation between 1991 and 2001.

> There is no better, there is no more open door by which you can enter into the study of natural philosophy than by considering the physical phenomena of a candle. There is not a law under which any part of this universe is governed which does not come into play, and is not touched upon, in these phenomena. I trust, therefore, I shall not disappoint you in choosing this for my subject rather than any newer topic, which could not be better, were it even so good.

Following in Faraday’s footsteps and hoping that I shall not disappoint the reader, I shall enter the study of applied mathematics with the humble candle rather than any newer topic and ask a simple question: how quickly does a candle burn?

I know from experience that the candles we use on birthday cakes burn so quickly that it is hard to light eighty of them before the first one is completely consumed. I have also seen large pillar candles in churches burning for days. We could perform a number of experiments with candles of different sizes

and obtain a fairly good understanding of the relationship between a candle's size (estimated from its diameter) and flame velocity (defined as the height of candle consumed per unit time). This empirical estimate will be useful but it would not give me insight into the underlying mechanism and I would remain forever curious about the beautiful phenomenon of flames.

Modelling the burning of a candle is surprisingly difficult. It requires a knowledge of fluid dynamics, combustion, and chemistry. However, simple mathematical reasoning can be applied to partially answer the question. To model a burning candle, we need to identify the key variables. But what are they? Many effects could play a role: the diameter of the candle, its composition, the width of the wick, the atmospheric pressure, the size of the room, its temperature, and so on. The crucial step in modelling is to retain important effects, and, as a first step, ignore minor contributions. We can therefore make the following assumptions:

- The candles are all placed in the same large chamber at the same atmospheric pressure and temperature. A candle is cylindrical with radius R and height H, measured in centimetres. We also assume that they all have the same wick's type and size.
- For comparison, candles are made of the same wax. So, the chemical energy density, E, stored in the wax is independent of the candle size and type. The variable E is defined as the energy per unit volume, let's say in joules per cubic centimetre. This means that a candle has volume $\pi R^2 H$ cubic centimetres and contains $E\pi R^2 H$ joules that can be eaten by the flame and transformed into heat and light.
- The rate of energy dissipation, that is how fast energy is lost, is independent of the candle height and radius. We call this quantity P and it denotes the rate at which energy is consumed through light and heat emission, say in joules per minute. We assume that all candles, big or small, have the same wick and release the same amount of energy per unit time. A lit candle dissipates PT joules during T minutes.

Calling on basic physics, we can model the flame dynamics as a process between the two flows of energy. Let T be the time taken for a candle of size H to be consumed. In that time, the candle consumes $EH\pi R^2$ joules and releases PT joules: the energy goes from wax to flame. The balance of these two processes leads to an *equation*:

$$PT = EH\pi R^2.$$

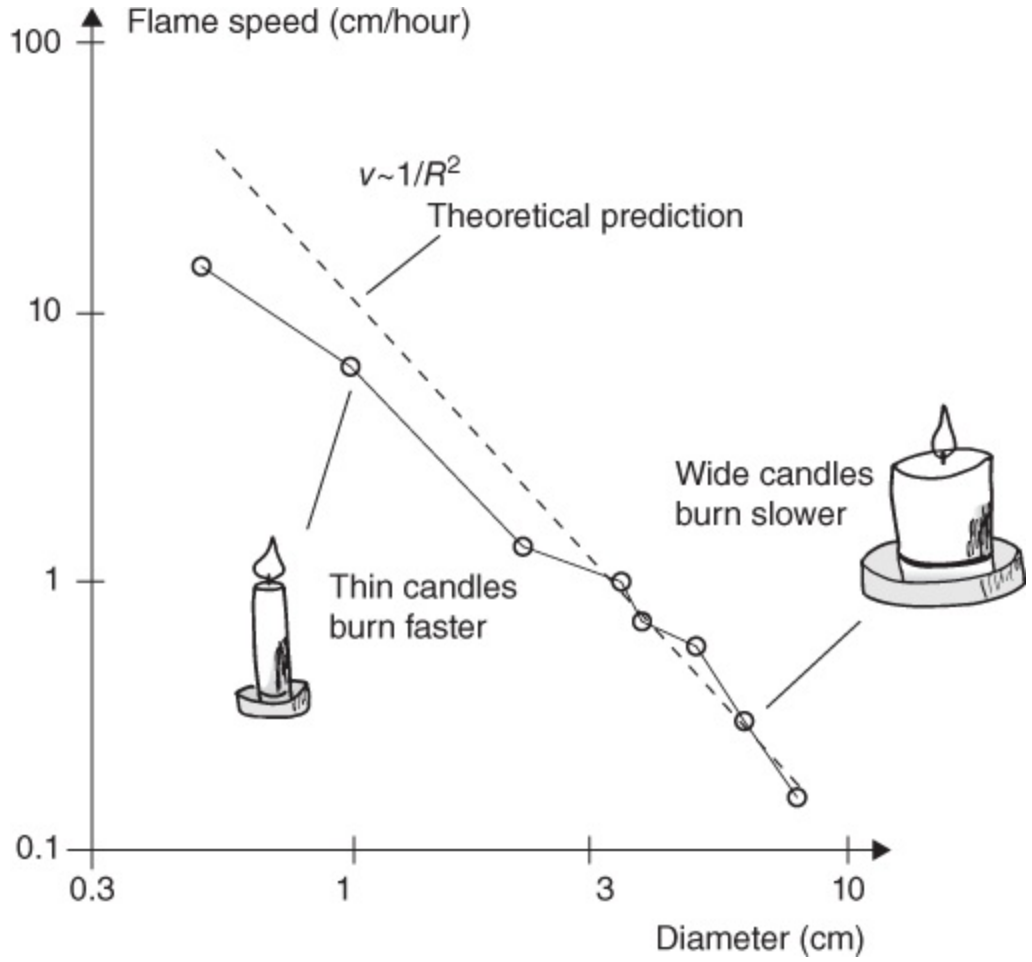
Since a candle of height H burns in a time T, the flame velocity is $\upsilon = H/T$

and using this model, we can express this velocity as

$$v = \frac{H}{T} = \frac{P}{\pi E} \times \frac{1}{R^2}.$$

This simple model does not tell us much about the velocity of the flame since we do not know P and E and, at first sight, it looks like we have replaced the unknown velocity by two unknown and mysterious quantities. Nevertheless, it provides interesting information when we realize that P and E are independent of the size of the candle. Therefore, we can extract information about the *scaling* of the process. We know that the velocity depends on the inverse of the radius squared, which implies that a candle of twice the diameter would burn four times slower; a candle ten times the diameter would burn a hundred times slower, and so on.

How good is this prediction? It is fairly easy to run a series of experiments with candles of different sizes and measure both the velocity and radius for each candle as shown in Figure 2. This figure is a log-log plot, a particular way to plot data that transforms any power relationship into a straight line as detailed in Chapter 2. The only important thing to know at this point is that if the data is close to a line of gradient −2, the candle follows our model. A look at the figure tells us that our prediction is quite good for large candles but not so good for smaller ones where other effects may be playing a role (the wick may be smaller, the wax may be melting away, …). Nevertheless, this model helps us understand candle flame propagation as a basic energy conversion problem. It also provides a simple law that we can test. This approach is often referred to as a *back-of-the-envelope computation*: a simple computation that aims to understand the basic mechanisms involved in a physical process. Other examples will be studied in Chapter 2.

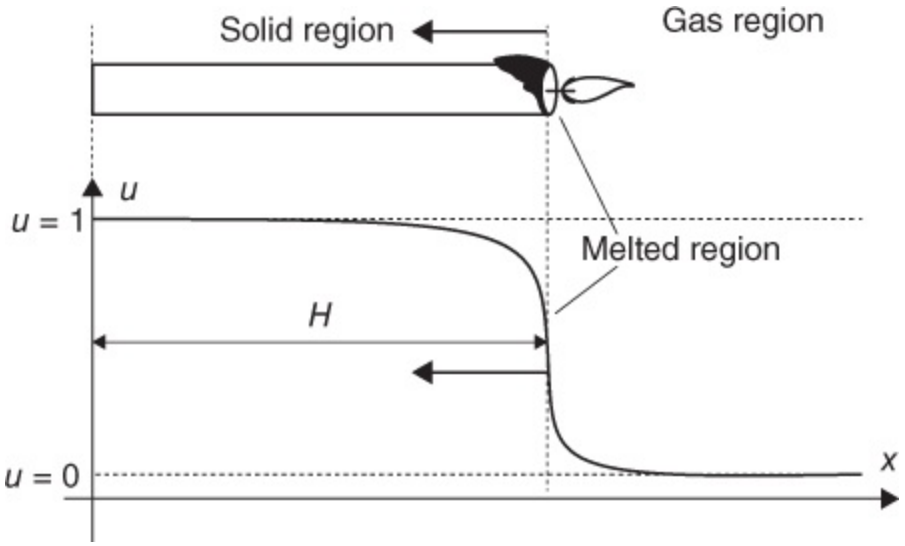


2. Experiments on candles: if we take the logarithm of the velocity and plot it against the logarithm of the radius, a relationship of the form $v \sim R^{-2}$ is, in a log-log plot, the dashed line of gradient −2. Therefore, if our prediction is correct, the data should sit on a line of gradient −2. Comparing data with prediction, we see that, indeed, the burning velocity of large candles is proportional to $1/R^2$ as predicted from the model.

The candle flame is a beautiful example of an autocatalytic process, that is, a process that literally fuels itself once started. Further insight into the evolution of a candle flame can be obtained by considering a more sophisticated model to describe the evolution of the candle in both space and time. For example, we could consider the composition u of the candle as going from a solid to a gas. In this description, the solid is given the value $u =$

1 and the gas a value $u = 0$. In the flame region, the wax is partially melted and partially consumed so u varies between the two values. Since the variable u changes both in location and time, it is a function of the position x (distance from one side of the candle) and time t.

Based on the fundamental physical principles of combustion and energy release, an equation for u can be obtained. This equation is a *differential equation*. We will encounter in later chapters many such equations. Differential equations are the workhorses of most of applied mathematics, physics, and engineering. Many applied mathematicians dedicate their lives towards understanding the properties of these equations and finding clever ways to solve them analytically or numerically. An example of a solution for the problem of the candles is shown in Figure 3.



3. A second model for flame propagation. The propagation of the flame is described by a variable *u* as a function of both position and time. In the extreme case, we have *u* = 1 for the solid part, and *u* = 0 in the gas part (where there is no candle left), and *u* takes values between 0 and 1 in the melted region. So *u* measures how melted the candle is. The height *H* of the candle at any given time is determined by the place where that transition occurs.

Our first example of applied mathematics follows the general description of a

problem motivated by a natural phenomenon that can be *modelled* mathematically based on some *theoretical* principles and analysed by the appropriate *methods*.

Reflection

Our definition of applied mathematics as a combination of modelling, theory, and methods, is very broad. We have casta very wide net by including in our informal description any work from any discipline that has ever used any bit of mathematics, however simple. This net is so wide that we may declare any intellectual discipline using numbers, mathematics, or logic to be part of applied mathematics, and by the same token that ‘pure mathematics is a subfield of applied mathematics’, a mischievous dictum attributed to the remarkably creative applied mathematician, Joseph Keller. Quip aside, we are back to square one and the question remains: what is applied mathematics? What sets it apart as a field of study from say, economics, engineering, physics, or pure mathematics?

The fundamental difference between applied mathematics and other scientific disciplines is that its principal object of study is not a tangible object. Particle physicists study elementary particles, nuclear physicists study nuclei, atomic physicists study atoms, chemists study molecules, cell biologists study the cell, and so on up to larger scales including geologists, climate scientists, planetary scientists, and astrophysicists. All of these disciplines have particular systems that they study through a combination of observations, experiments, theory, and mathematical methods. The problems they study are generated by these systems and scientists devote their lives to answering them.

To return to the burning candle, we considered this as a problem of combustion theory, a field of enquiry studied by both physical chemists and fluid dynamicists. Admittedly, if applied mathematicians were to spend their lives looking only at burning candles, applied mathematics would be a rather dull discipline. It is not the candle we care about! Flame dynamics is but one simple example of the general phenomenon of autocatalytic activation that can be modelled by reaction-diffusion equations. The same basic processes

and equations appear in many other physical and chemical systems such as in genetics (describing the spread of a genotype in a population), biology (species invasion in a habitat or pattern formation on seashells), or physiology (propagation of nerve impulse).

The goal of applied mathematics is to establish the connections between separate academic fields, to develop mathematical tools to analyse general classes of equations independently of their origin, and to answer scientific questions for particular classes of systems. For the applied mathematician, the candle is not the object of study, but a paradigm for a large class of autocatalytic phenomena described by various physical principles leading to the same mathematical description. Once we understand the candle, we understand a multitude of other systems. On the one hand, understanding these equations from a mathematical perspective provides insights on the phenomena. On the other hand, studying the details of these different systems motivates us to better understand the underlying mathematical structure that they share.

As Einstein famously said: ‘God does not play dice’. Natural phenomena are not arbitrary, they all come equipped with a mathematical structure and each instance and insight from one provides a better understanding of another.

Maybe there is no outstanding need to define formally applied mathematics. A way forward is to follow the timeless wisdom of the United State Supreme Court. In 1964, the justices were asked whether a French film, *Les Amants* from Louis Malle, was hard-core pornography and, as a result, should be banned from cinemas in Ohio. Justice Potter Stewart opposed the ban in the following words:

> I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description, and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and the motion picture involved in this case is not that.

I will follow Justice Stewart’s lead here. I shall not attempt further to define applied mathematics but view it as a discipline that is easier to recognize than to define. The vagueness that comes from the absence of a rigorous definition leaves it unshackled and allows its disciples to explore new areas of

knowledge. But, to learn how to recognize good applied mathematics, we must first see it in action. And to see it, it must be shown.

Chapter 2

Do you want to know a secret? Turkeys, giants, and atomic bombs

Those are my principles,
and if you don't like them … well, I have others.

–Groucho (attributed)

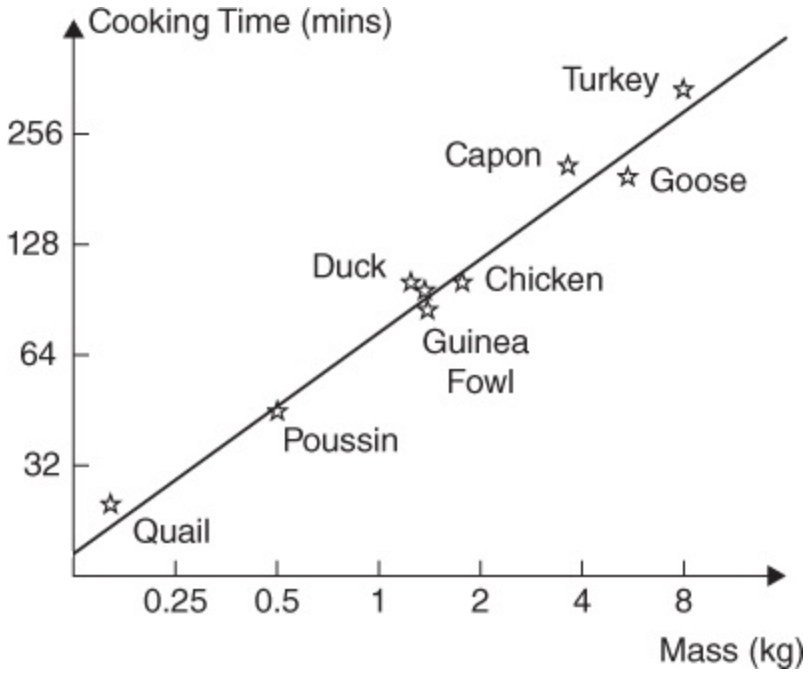
Science is based on observation and quantification. When observing and trying to quantify the world around us, two simple facts can be appreciated and readily agreed upon. First, different physical quantities are characterized by different quantifiers: time is not a length, a length is not a weight, and so on. Different physical quantities such as time intervals, lengths, areas, volumes, weights, and forces cannot be compared or added together. If you add an apple to an orange, you have two fruits; but adding a metre to a second is meaningless. Length and time are examples of *dimensions* that are used to characterize different physical quantities. Dimension matters!

Second, objects come in different *sizes*; some are minuscule, some are small, some are large, and some are gigantic. From mice to men and men to whales, mammals have different lengths. From the blink of an eye to a lifetime, and from the age of the oldest tree to the age of the universe, time spans have different extent. Size matters!

These two apparently innocuous facts have far-reaching consequences that can be appreciated when we understand their mathematical implications and study the constraints that dimensions and size impose on physical processes. Obtaining information by considering dimensions or scales is known as *dimensional* or *scaling analysis*. In Chapter 1, we encountered our first scaling analysis when we computed how the life span of a candle depends on its radius. Here are some other examples from different disciplines.

Cooking a turkey

When using a kitchen oven, I am always perplexed by the time it takes to cook food, whether a chicken, a goose, or even a turkey. Despite their all being poultry, these birds come in different sizes and have different tissue compositions. The packages provided by large food suppliers typically come with helpful instructions that typically read: cook the bird at a given temperature (say around 180°C) for a time that depends on the weight. For instance, we are advised to cook a 5 kg turkey for $2\frac{1}{2}$ hours and a 10 kg turkey for 4 hours (see Figure 4). Where do these numbers come from? How does the cooking time depend on the bird weight? Maybe there are secret warehouses with rows of well-calibrated ovens and an army of testers with chef's hats, large aprons, and industrial size thermometers trying various combinations of weight and poultry for the good of public safety. What if I want to cook an ostrich, with a typical weight varying between 60 kg and 150 kg? How about finding a cooking time estimate for the Bedouin delicacy consisting of a whole camel stuffed with a few sheep that are in turn stuffed with about twenty chickens themselves stuffed with fish? Clearly the problem of cooking poultry could be solved by empirical testing, but this practical solution will never satisfy my curiosity.



4. Cooking time versus mass of various poultry in a log-log plot (the star between the Duck and the Guinea Fowl is a Free-range Chicken). Data obtained from commercially available sources. The line indicates the theoretical prediction of a 2/3 scaling law (that is, the cooking time increases as $t_M = CM^{2/3}$ and appears on a log-log plot as a line of slope 2/3).

The modelling process often starts with simple questions. In this case: how does the cooking time, at a given fixed temperature, depend on the weight? Or consider an even simpler question. If I know the cooking time for a 5 kg turkey, what would be the cooking time for a 10 kg or 100 kg bird? The next step is to simplify the problem by stating a number of assumptions about the system. This is where we first encounter the art and craft of modelling. We have to decide what are the relevant quantities affecting our problem. We could consider among others: the age of the bird, its shape and size, whether it was corn-fed or free range, the size and type of oven, the cooking cycle, the basting method and type of stuffing, and so on. While all these effects may be important, if there is any hope of progress, we will need to make stringent simplifications. The key is to make these simplifications completely explicit so that they can be challenged and modified when considering more

realistic situations. Therefore, as a starting point for our first model, we will make the following assumptions:

- The bird is a sphere of radius R with mass M and uniform constant density ρ (its weight per unit volume).
- The bird has an initial temperature T_i, and is cooked at a constant oven temperature T_O. That is, I assume that the temperature at the surface of the sphere is maintained at a constant T_O.
- The transfer of heat in the bird is described by a unique constant coefficient of thermal diffusivity k. This coefficient describes the ability of a material to conduct heat. A larger value of k means that the bird cooks faster compared to a bird with a smaller value. Importantly, the thermal diffusivity k has dimensions of length squared divided by time.

Since heating is a physical process, these assumptions are based on the physical theory of heat transfer. In the process of writing our assumptions we have identified the key variables, namely the radius R, the mass M, the temperatures T_i and T_O, and the thermal diffusivity k. Essentially, we have replaced our bird by a sphere of homogeneous material subject to a constant temperature at the surface. Mathematicians jokingly refer to this type of assumption as the *spherical-cow assumption*,a drastic assumption that transforms a problem into a much simpler one for the convenience of mathematical analysis. Remarkably, in many contexts, these simplifications provide excellent results, even for the unfortunate cows that are not perfectly spherical.

Next, we want to know the mathematical relationship that relates the cooking time t_M, defined as the earliest time at which a safe temperature is reached at all points of the mass M, as a function of all the variables:

$$t_M = f(R, T_\mathrm{i}, T_0, k).$$

Note that we do not list the density ρ or the mass M as variables. We assume that the density is the same for all birds. Therefore, the mass is $M = \rho \times 4\pi R^3/3$ (mass is density times volume).

Quite remarkably, within this set of assumptions and without writing any further equations for the physical process taking place, we can already extract some useful information and answer some of our questions. To do so, we will

use the method of *dimensional analysis*, which consists in finding relationships between different quantities with the same physical dimensions.

A digression about scales, dimensions, and units

In our problem, each variable has a dimension: the radius is a length, the initial bird and oven temperatures are temperatures (obviously), the mass is a mass, and the thermal diffusivity has the scale of a length squared divided by a time. There are two closely associated concepts when looking at scales. *Dimensions* are the physical quantities of an object. For instance, length, time, and mass are all dimensions. *Units* depend on the particular system that we use to express these dimensions. They have a history that depends on technology and politics. For instance, we can express lengths as metres, mass as grams, and time in seconds.

When doing science, it is important to choose a system of units and consistently stick to it. On 23 September 1999, Mars Climate Orbiter was lost in the Mars atmosphere due to unintended mixture of imperial and metric units. The most convenient system at our disposal is the so-called international system of units or SI (for *Système international d'unités* as it finds its root in the French Revolution). In this system, we can express each dimension in terms of seven basic units. Here we will use four units, namely the metre for length, the second for time, the kilogram for mass, and the kelvin for temperature (for completeness, the three other SI units are the ampère for electric current, the mole for the amount of a chemical substance, and the candela for light intensity). Any physical quantity can be expressed in terms of these units. For instance, a velocity is a length per unit of time with units in metres per second and an acceleration is a velocity per unit of time (meter per second squared in SI units). A force is a mass times an acceleration (think $F = ma$) and is measured by the introduction of a derived unit, the newton, defined as the force required to accelerate a mass of 1 kg by one metre per second squared.

Lastly, the thermal diffusivity has dimensions of a length squared by a time, and is therefore expressed in SI units as metre squared divided by second. If I use another system of units (say, fathom for length, Réaumur for temperature,

lustrum for time, and slug for mass), the thermal diffusivity would have the same dimension but a different numerical value that would be expressed in fathom squared divided by lustrum (for the inquisitive mind: a lustrum is five years and a fathom is two yards). This choice is clearly preposterous but, from a scientific point of view, no more preposterous than the antiquated use of pints, pounds, perches, and pennyweights that were in use in the British Empire.

Back to the kitchen

The specific choice of units is not crucial for our problem. What matters is to compare the different quantities that share common dimensions. Since I want to obtain an estimate, some computations will be shown but the reader should not worry unduly about these details and could safely skip to the conclusion.

For our problem, the only variable that contains the dimension of time is the thermal diffusivity, dependent on time through its inverse (explicitly, the inverse of the diffusivity is a time divided per length squared). Since we are interested in the time needed to cook a bird, any time in the problem must be related to the thermal diffusivity through its inverse. That is, the cooking time must be of the form:

$$t_M = \frac{1}{k} \times g(R, T_\mathrm{i}, T_0),$$

where g is a function that depends only on the arguments listed. But, since the quantity t_M on the left side of the equation has dimension of time, so must the quantity on the right side. Therefore, the function g must have dimensions of length squared to cancel the length dependence of the thermal diffusivity. But again, since the only length available in the variables of g is R, we must have

$$t_M = \frac{R^2}{k} \times h(T_\mathrm{i} / T_0).$$

The remaining unknown function *h* only depends on the temperature but is itself dimensionless and therefore it can only depend on the ratio of the temperatures, expressed in kelvins (a *dimensionless number* is a number without unit). In particular, this dimensionless number must be independent of the mass of the bird. What does this relation tell us? If we fix the thermal diffusivity, oven, and initial temperatures, the cooking time increases as the square of the radius. But, at fixed density, the radius varies as the cube root of the mass,

$$R = \left(\frac{3}{4\pi} \times \frac{M}{\rho} \right)^{1/3}.$$

So we can write

$$t_{\mathrm{M}} = C \times M^{2/3},$$

where

$$C = \frac{1}{k} \left(\frac{3}{4\pi} \times \frac{M}{\rho} \right)^{2/3} \times h(T_{\mathrm{i}} / T_{\mathrm{o}}),$$

does not depend on the mass. According to our assumptions, the density ρ is constant and so are the temperatures T_{i} and T_{o}. Therefore, C is constant in our experiment. Wisely choosing SI units, I have for a 5 kg and a 10 kg turkey

$$t_5 = C \times 5^{2/3},$$
$$t_{10} = C \times 10^{2/3}.$$

Without knowing C, I cannot find these times. However, the ratio of the two times is very simple and allows me to find one from the other:

$$\frac{t_{10}}{t_5} = 2^{2/3} \quad \Rightarrow \quad t_{10} = 2^{2/3} t_5.$$

I conclude that if I know that the cooking time of a 5 kg turkey is about $2\frac{1}{2}$ hours, the cooking time t_{10} of a 10 kg turkey is $t_{10} = 2^{2/3}t_5 \approx 4$ hours. Following the same reasoning, I learn with great relief that I will only need about $t_{100} = t_{10}10^{2/3} \approx 4 \times 4.6 \approx 18\frac{1}{2}$ hours before my plump and juicy 100 kg ostrich is nicely cooked and safe to eat. Now, how long would it take to cook my 500 kg stuffed camel in the same oven?

Using very simple dimensional arguments, I have found a partial answer to my question. And although it is a good start, I am still far from understanding the entire picture. If I want to know the actual cooking time as a function of the parameters, I need a more sophisticated model that describes the evolution of the temperature at all points in the sphere and for all time. The basic differential equation that describes this process is known as the *heat equation* as it expresses mathematically the transport of heat in a material. Finding its solution requires some methods of applied mathematics that will be discussed in Chapter 4.

The giants who did not walk among us

In his 18th-century political satire *Gulliver's Travels*, Jonathan Swift describes Gulliver's visit to Lilliput, where the inhabitants are twelve times smaller than normal humans. Apart from their size, these Lilliputians are in all aspects identical to humans, even in their predictable political habits. Is the existence of such small humans reasonable?

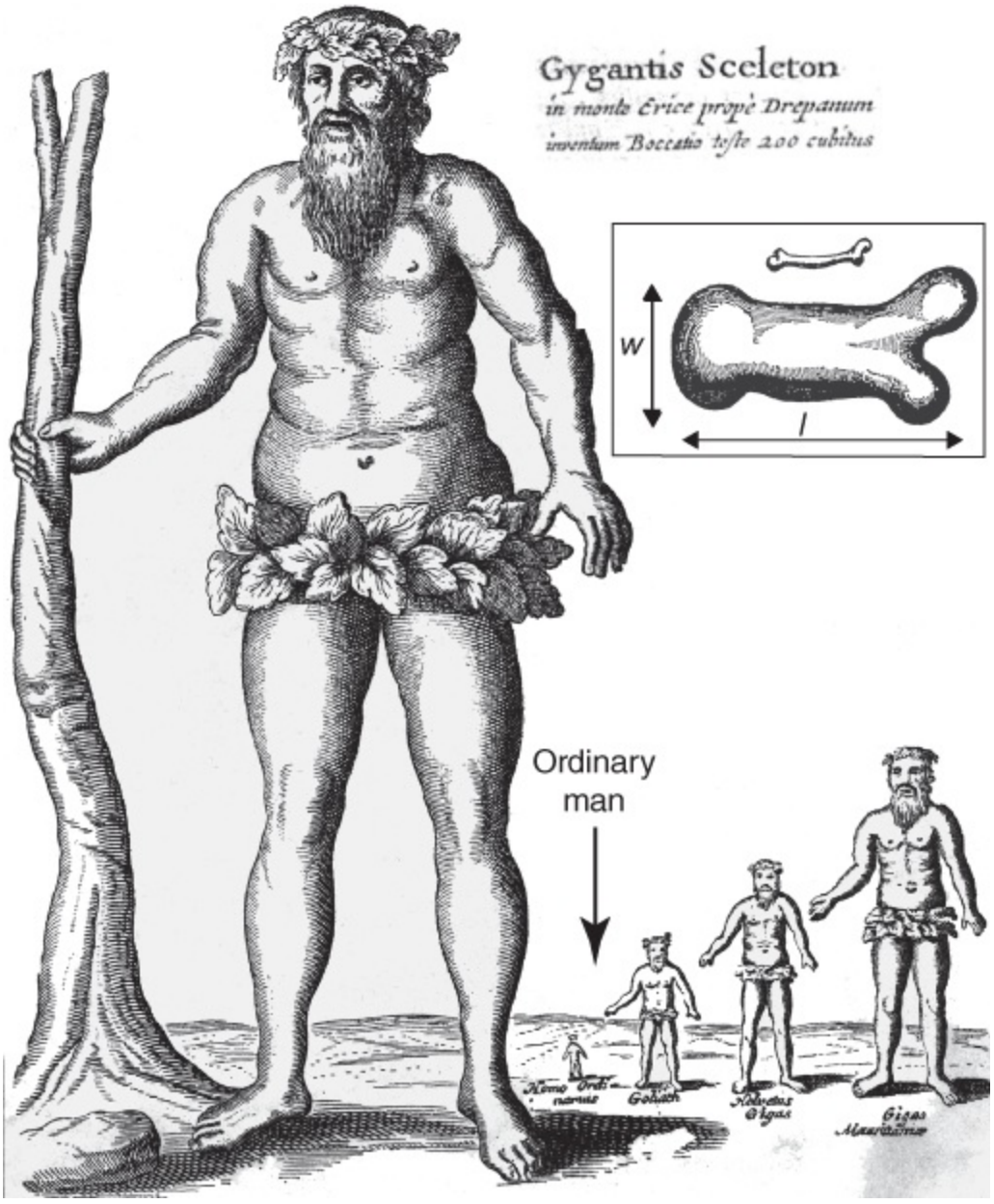
Mammals are warm-blooded animals that use energy to keep a constant temperature in a cooler environment. From a thermal point of view, we can think of a mammal as a turkey in an inverted oven, where the heat source would be placed inside the turkey. We lose heat through our body surface in the same way that turkeys acquire heat through their exposed surface. We learned from our turkey experiments that heat gain is proportional to their surface area. Therefore, smaller mammals lose more heat than larger mammals per unit mass because their area per unit mass is larger. In the same way as it takes less time to cook a kilogram of a small turkey than a kilogram

of a large turkey, a smaller human would lose heat more quickly than a larger one. Indeed, in the absence of an internal heat source, a Lilliputian would cool down to the external temperature $12^2 = 144$ times faster than an average-sized human (where the square expresses the fact that heat is exchanged through the surface of an object). Hence, an object twelve times smaller loses heat 12^2 times faster. This simple scaling argument explains why the smallest mammals, such as shrews and mice, must have an extremely fast metabolism (compared to larger mammals) in order to keep their body temperature constant. This also explains why the unfortunate Lilliputians with a human physiology would quickly die from hypothermia in a cold climate and from over-heating in a warm climate.

Having escaped the thermally unrealistic Lilliput, Gulliver finds himself in Brobdingnag, where the inhabitants are about twelve times larger than humans, but in all other aspects identical. Is the existence of such giants physically reasonable? The good news for the Brobdingnagians is that, proportionally to their size, they do not have to eat as much as we do. However, they will be subject to another dimensional frustration: their bones will not be able to support them. Indeed, if we consider a bone of length l and width w, a classic computation reveals that the force needed to buckle or bend the bone scales as w^4/l^2. This means that if we consider two bones of same width w but where one of them is twice as long as the other, the long bone requires a force four times smaller for buckling than the force required to buckle the shorter one (expressing the intuitive notion that long and thin objects are easier to bend than short and stubby ones). However, the force acting on long bones such as the femur depends on the total weight of the individual; it increases as lw^2. To understand this scaling, consider a bone of length l and width w that resists buckling. If we double its length $l \rightarrow 2l$ we must change $w \rightarrow 2^{3/2}w$ in order to offer the same resistance (this relation is obtained by realizing that the buckling force is proportional to both lw^2 and w^4/l^2 and therefore to offer the same resistance to buckling the width must increase as $w \sim l^{3/2}$). In other words, longer bones must be proportionally wider to provide the same protection against buckling and breaking. This analysis is well appreciated in civil engineering when it comes to building bridges and towers. It is also nicely verified when looking at bones in similar species: for instance, the ratio w/l of a 750 kg antelope femur bone is about

$(750/3)^{1/2}$ times larger than the same bone of a 3 kg antelope. It also explains why the aspect ratio of elephant bones is significantly larger than that of smaller quadrupeds. For the Brobdingnagians, since the ratio *w/l* should have increased by a factor $12^{3/2} \approx 42$ to ensure safety, their long bones would offer no more resistance than straws in the hands of a toddler, and these giants would sadly perish due to their inability to support their own weight.

The influence of size on the resistance of bones had already been discussed by Galileo in 1638, almost a century before Gulliver, in one of the first known instances of dimensional analysis. Galileo's study was inspired by the many discoveries of large fossil bones of dinosaurs or mammals that were wrongly attributed to giants of the past as illustrated in Figure 5. In his *Discoursi*, Galileo argued:



5. Improbable giants from *Mundus Subterraneus* by Athanasius Kircher (1678). The biggest giant in the figure, next to an 'ordinary man', is based on bones discovered in Sicily. Through simple scaling arguments, Galileo correctly concluded that if giants did exist they would not be a simple dilation of an ordinary man. Insert: in his *Discoursi*, published in 1638 Galileo correctly noted that the aspect ratio, *w/l*, of bones changes with size.

Nature cannot produce a horse as large as twenty ordinary horses or a giant ten times taller than

> an ordinary man unless by miracle or by greatly altering the proportions of his limbs and especially of his bones.

From this profoundly insightful remark, it is clear that Galileo fully understood that size matters and similar bones in larger animals must be comparatively thicker than those in smaller animals in order to offer the same resistance to loads. The same arguments can be applied to the relative proportions of many animals, plants, and engineering structures. Once aware of the role of sizes and scales, we can only marvel at the universality of dimensional constraints in shaping our world.

Dr Taylor or: How I learned to love the bomb

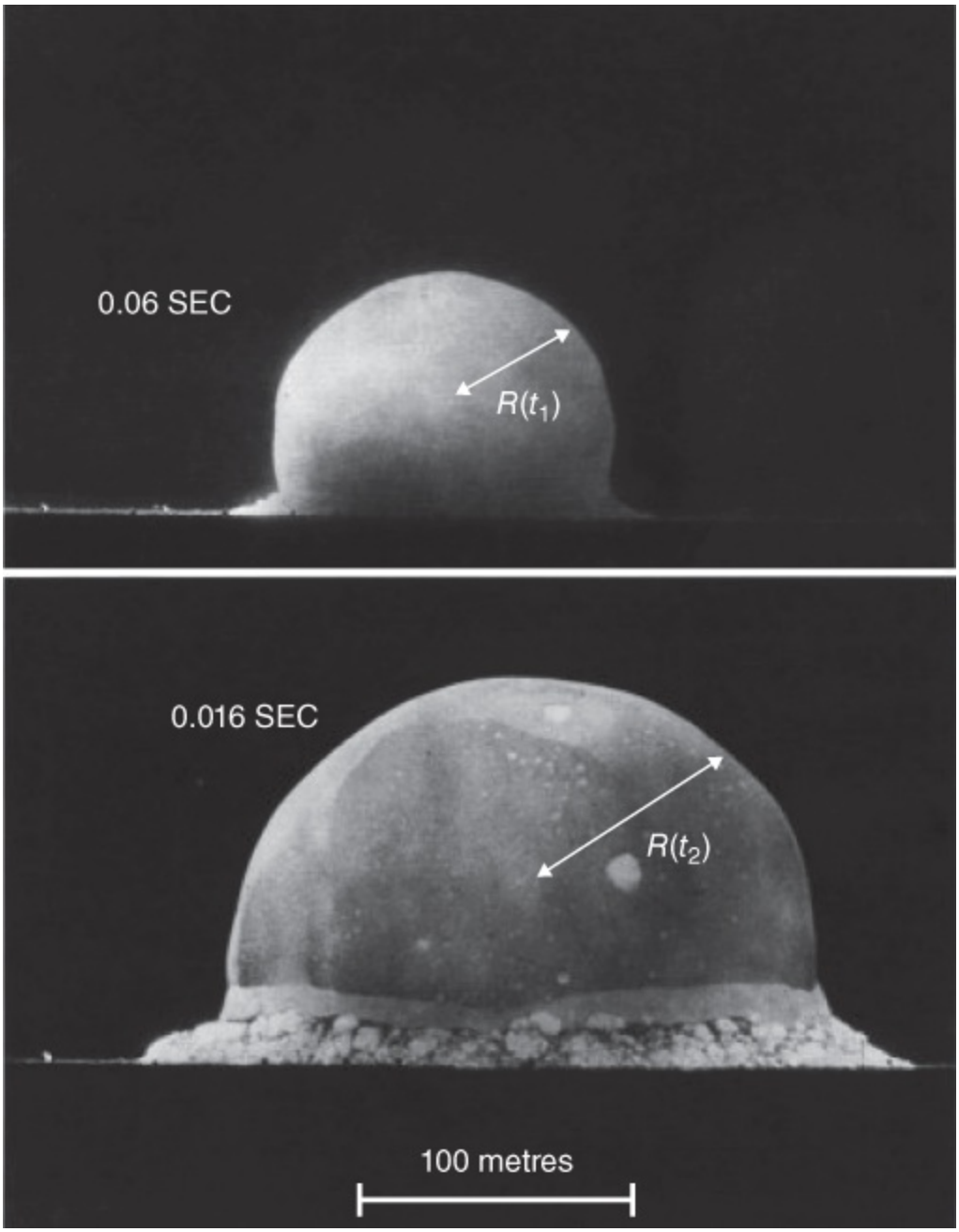
The most dramatic application of the method of dimensional analysis is found in the work of Geoffrey Ingram Taylor, a towering figure of 20th-century British applied mathematics. Atomic bombs not only played a central and tragic role at the end of World War II with the bombings of Hiroshima and Nagasaki, they were also a focal point during the Cold War. At that time, the arms race between the Soviets and the Americans centred around the secretive creation of increasingly powerful bombs.

The key descriptor in quantifying the destructive power of a bomb is how much energy is released when it explodes. This energy is generally expressed in the non-SI unit of tons of TNT (trinitrotoluene—a conventional chemical explosive).

By convention, a ton of TNT is the energy released by its explosion and is equal to 4.184 gigajoules (that is 4.184×10^9 joules, joules being the SI unit for energy). To have an idea of the tremendous power of nuclear weapons, the explosion at Hiroshima released the equivalent of about 15 kilotons of TNT (63×10^{12} joules), an event that killed more than 50,000 people within a day. A single hydrogen bomb developed during the Cold War has a yield of about 25 megatons of TNT (100×10^{15} joules, or about 1,600 Hiroshima bombs). The total global nuclear arsenal is estimated at a sobering 30,000 nuclear warheads with a destructive capacity of 7 billion tons of TNT (more than 450,000 Hiroshima disasters). The world is sitting on a mountain of

weapons of mass destruction that keeps us safe, according to the doctrine of *mutually assured destruction* (aptly called MAD for short), all of it controlled by the impeccable wisdom of a few heads of states.

During the Cold war, the entire atomic bomb programme was shrouded in mystery and subject to intense security measures. But, in an effort to inform the public, pictures of atomic explosions were released to the media without specific details. For instance, Figure 6 shows a sequence of high speed photography stills from the Trinity test of 1945, the first ever atomic explosion on Earth. These pictures were published in the 20 August 1950 edition of *Life* magazine with no other detail than the timestamps and physical scale of the blast radius. A seemingly impossible problem is to extract from these pictures the energy released by the blast.



6. High-speed photography of the Trinity test explosion in 1945 at two different times after the explosion.

Enter GI Taylor. During World War II Taylor was part of the famous Manhattan Project that developed the first atomic bombs at Los Alamos; he personally witnessed the Trinity test. As an applied mathematician, Taylor was particularly good at identifying both the key elements of a physical

phenomenon and at developing simple but extremely elegant models. His knowledge of fluid mechanics and atomic explosions was particularly useful in this instance. In 1950, he published two papers in which he explained that the energy, *E*, released in an atomic explosion can be estimated by using the method of dimensional analysis. Indeed, he argued, if the blast radius, $R(t)$ is known at a time t after the explosion, then the energy released is

$$E = C\frac{\rho R^5}{t^2},$$

where ρ is the air density before the explosion and C is a dimensionless constant. Most of Taylor's work consisted in estimating the constant C based on the equations describing shock waves in a gas. From this formula, GI Taylor estimated that the explosion pictured in *Life* magazine released between 16,800 and 23,700 thousands tons of TNT, a figure that was remarkably close to the official classified estimate of 18,000 to 20,000 thousands tons. The story goes that Taylor was properly admonished by the US Army for publishing his deductions from their (unclassified) photographs.

A word of caution from Lord Rayleigh

We have seen four examples of scaling laws. The first one related the burning velocity of a candle to its radius:

$$\upsilon = C_1 R^{-2}.$$

The second one gives the cooking time of turkey as a function of its mass:

$$t_M = C_2 M^{2/3}.$$

The third law explains how the width of a bone must increase with respect to its length:

$$\omega = C_3 l^{3/2},$$

and the fourth law gives the energy of an atomic explosion as a function of both the radius and the time.

$$E = C_4 R^5 t^{-2}.$$

In general, a scaling law is of the form

$$X = CY^{\alpha},$$

relating quantity X to Y through the scaling exponent α with a constant C that does not change when Y is varied. Many researchers try to uncover scaling laws directly from data. Indeed, a typical approach in dimensional analysis is to gather data at different scales. For instance, we could try to find a scaling law that relates the radius (X) of trees at their base to their height (Y). Once the data is collated, we can display it in log-log plot. In this plot, a relationship of the form $X = CY^{\alpha}$ becomes a line with slope α. Indeed, recalling that

$$\log(a^b) = b\log(a) and\, log(ab) = log(a) + log(b),$$

we have

$$log(X) = log(CY^{\alpha}) = \alpha \log Y + \log C.$$

By introducing, $x = \log X$, $y = \log Y$, and $c = \log C$, this last relationship reads

$$x = \alpha y + c,$$

which is a line of slope α in the $y - x$ plane. Therefore, if the data follows a law of the form $X = CY^{\alpha}$, we expect data points to cluster along a line in a log-log plot. By finding a best linear fit to the data points in this plot, an estimate for the value of α is obtained. If the fit is good, it suggests that there may be a physical reason for the existence of a scaling law. Data sets on trees often give an exponent close to $\alpha = 3/2$. The astute researcher would then try to find a mechanism or a physical constraint that would predict this exponent α. In the case of trees, it has been argued that they follow an elastic law and,

as they become taller, their diameters must increase proportionally to avoid buckling under their own weight, just like Galileo's giants.

This type of analysis has become a basic tool in understanding physical effects related to size and identifying key combinations of parameters describing a phenomenon. It is widely used in biology, ecology, physics, and engineering, with sometimes surprising results. Such analysis is particularly appealing because it does not rely on specific physical equations but only on the dimensions of physical quantities. However, it is limited in many ways. The benefits and pitfalls of this approach were described long ago by Lord Rayleigh, an intellectual giant of natural philosophy and the 1905 Nobel Physics Laureate for his discovery of Argon. In a 1915 article, Rayleigh gives no less than fifteen different examples where dimensional analysis can be used, such as in the sound of an Aeolian harp and in the vibration of a tuning fork. He also states that in the use of such methods, 'two cautions should be borne in mind'.

First, dimensional analysis only compares phenomena at different scales but says very little about a single scale. It is fascinating to compare the metabolism of a mouse to a human, but it will not shed much light on how the human metabolism works other than that it should be different than the mouse's. The main problem for dimensional laws of the form $X = CY^{\alpha}$ is that the power α can be easily identified but the constant C is hard to obtain. This constant is particularly important when trying to make predictions or using the law as part of a model. We assume that it does not vary between scales, but obtaining its value from basic principles is usually complicated and requires a deep understanding of the problem. Indeed, the real work of the researcher is in identifying this constant through detailed experiments or theory. For instance, GI Taylor obtained the correct scaling exponents for his problem in less than a page, but it took him another fourteen pages of detailed calculations to estimate the value of C appearing in his law. It is only through extensive mathematical modelling that the nature of these laws and their reason for existence can be appreciated.

Second, dimensional analysis requires either a supreme confidence in one's intuition or an almost supra-natural knowledge of a given phenomenon. How

do we know which quantities should be considered and which should be neglected? How do we know that we are not missing a key piece of information about the process since the analysis is not based on a particular mechanism? This point is nicely summed up by Rayleigh: ‘It is necessary as a preliminary step to specify clearly *all* the quantities on which the desired result may reasonably be supposed to depend.’

Dimensional analysis is only the first step in a journey of discovery and one should not shy away from elucidating the mechanisms responsible for these laws. Unfortunately, these words of advice from Rayleigh seem to have been forgotten by many modern researchers at the cost of falsidical results.

Reflections

Dimensional analysis relies on a simple universal principle, the *principle of dimensional homogeneity*, which states that every mathematical expression which expresses a physical law or describes a physical process must remain unchanged in relation to any change in the fundamental units of measure. From this principle, relationships between quantities can be obtained as Einstein explained in 1911:

> It is well known that from dimensional considerations it is possible to closely find general relationships between physical quantities when one knows all the physical quantities that participate in the referred relationship.

Mathematically, the principle of dimensional homogeneity can be turned into a theorem, the so-called *Buckingham Π-theorem*, where the Greek letter Π (capital ‘pi’) is used to represent *dimensionless* quantities. They are numbers that are independent of the units in which they are measured. This theorem automatically guarantees the existence of a certain number of scaling laws given a certain number of physical quantities.

Dimensional analysis is a simple example of the type of constraint imposed by the physical world on a mathematical theory. When developing models of the physical world, there are basic principles that one must abide by. Unlike pure mathematicians, the applied mathematician is handed strict constraints

dictated by the underlying structure of our world. When poets decide on a particular form, be it a sonnet or a haiku, they constrain the form of their poems at the gain of creativity. Similarly, the development of mathematical models is an exercise with rules of style dictated by mathematics but with internal constraints given by science.

Chapter 3

Do you believe in models? Simplicity and complexity

I'm not crazy about reality,
but it's still the only place to get a decent meal.

–Groucho (attributed)

Central to the world of applied mathematics is the notion of a model. In its simplest sense, a model, from the latin word *modus* (a measure), is an abstract representation of a system developed in order to answer specific questions or gain insight into a phenomenon. Models come in multiple flavours and sizes depending on the question and the discipline. For instance, for a biologist, a model might consist of the description of multiple biochemical processes summarized by a text and a picture establishing the influence of different players on the system. When graphically displayed, such models are organized into different blocks representing the various processes and quantities of the system and are connected by arrows indicating the mutual positive or negative influences of these quantities (e.g. the presence of a protein or a gene can up-regulate or down-regulate a process such as cell division or growth). A biological model is a summary of the current understanding of a process based on various experiments or theoretical relationships.

Models in applied mathematics have a completely different flavour. They are precise mathematical constructs developed to answer questions, to explore ideas, and to help us gain insights about a system. Much as microscopes help us make observations at different scales that are too small for our own eyes, mathematical models allow us to have access to quantities that may not be directly observable but are nonetheless important for the understanding of a problem. Setting them in mathematical terms forces the mind to stipulate every assumption explicitly and ensures that every consequence of the model is supported by a combination of mathematical rigour and numerical computations. Lord Kelvin, another great natural philosopher from Scotland whose name is attached to the SI temperature scale, summarized the need to quantify observations in 1883:

> I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.

Models are the ultimate form of quantification since all variables and parameters that appear must be properly defined and quantified for the equations to make sense.

In general, we expect a model to be based on sound principles, to be mathematically consistent, and to have some predictive or insight value. Let us take a look at these basic concepts.

Becoming rich

Typically, mathematical models take the form of a *set of equations* based on a combination of observations, experiments, and theory whose solutions inform us about the behaviour of the original system. These equations contain variables and parameters. The difference between the two is not always clear-cut but, essentially, *parameters* are inputs in the problem; they are given quantities. *Variables* are the quantities that we are trying to determine.

Consider the following elementary problem for illustration. Assume that I

have earned a certain amount of money from the sale of this book, say £10, to be on the generous side. I have been advised by serious people that I should place my earnings in the stock market rather than spend it on ice-cream. I would like to know how long it will take before I double my initial investment. In order to answer this question, I will have to make some basic assumptions about the nature of the stock market based on past data. Assuming that the stock market has an r% return on a yearly investment, I will have £10 × (1 + r/100) after a year, £10 × (1 + r/100)^2 after two years, and so on. It is easy to see the structure of this problem and generalize it to obtain a model for the increase of my wealth $w(x, r, N)$ in £ after N years starting with an initial investment x with an r% return:

$$\omega(x,r,N) = x \times (1 + r/100)^N.$$

In this problem, r is a parameter and x and N are the variables. This model is simple enough and is in a form that gives me w given x, r, and N. But I can also use this equation to predict the time it will take to double an initial investment (that is when $w = 2x$) by solving the equation

$$2x = x \times (1 + r/100)^N.$$

This equation can be solved by taking the logarithm of both sides and recalling that the logarithm of a product is the sum of the logarithms. I find

$$N = \log 2 / \log(1 + r/100).$$

For instance, with an optimistic 10% return, it would take log 2/ log(1.1) ≈ 7.27 years. I can also investigate the type of return I would need to double my investment after five years by solving for r in the previous equation with N = 5.I find that I need a return of about 15%. Now that I am fully educated about the problem, I can make an educated choice and I decide that I am better off buying ice-cream.

This model is based on a number of *assumptions*. Clearly, I do not know what the stock market will do in the future and I would not advise the reader to use this equation blindly to make any financial decisions. However, within

the framework of these assumptions, my model links the key variables of the problem. Importantly, its mathematical analysis leads me to understand the nature of compound interest and what I can expect under some simplistic scenarios.

Any prediction from small or large mathematical models, whether related to global climate change or the evolution of the economy, is based on a number of assumptions. A mathematical model does not exist independently of its assumptions. Once established, the mathematical equations that define the model represent a certain truth for the system under study that is only valid within the confines of these assumptions and great caution must be exercised when trying to apply the model to a wider context.

Throwing a ball

An example that may be familiar from school is the problem of determining the motion of a ball thrown by hand. We can use this simple problem to illustrate the increasing complexity of models based on the assumptions and the questions asked.

Our **first model** relies on Newton's equations of motion that states: the mass m times the acceleration a_{vert} of a point-mass in the vertical direction is equal to the vertical force f_{vert} applied to the object (often summarized as $f_{\text{vert}} = ma_{\text{vert}}$). Similarly, we know that the acceleration in the horizontal direction is related to the force applied to the mass in the same direction by $f_{\text{hor}} = ma_{\text{hor}}$. By assuming that the ball is a point-mass subject to a constant force of gravity, and neglecting friction, the horizontal force vanishes and the vertical force is constant. Then recalling that the acceleration is the second derivative of the position with respect to time, the model takes the mathematical form of two second-order linear differential equations for the vertical and horizontal position of the ball as a function of time. The *variables* in our problem are the positions and the *parameters* of our problem are the mass, the initial position and velocity, as well as the acceleration of gravity and time. Then, a simple computation will determine the time taken before hitting the ground as well as the distance travelled. The solution of the model would also describe the

velocity and the kinetic and potential energy of the ball at all times. In other words, this model can be solved exactly and completely.

The astute reader would note that this ideal description is not very realistic, as the air offers a drag resistance to the motion of the ball that we have neglected.A **second model** would then take into account the effect of air friction through a simple phenomenological law characterizing the force slowing down the ball. Here, the force acting on the ball is the sum of the force of gravity and a drag force proportional to the velocity. This equation is still simple enough so that it can be solved by traditional methods. This new model must contain an extra parameter, the drag coefficient, and its solution provides better estimates for all kinematic and mechanical quantities.

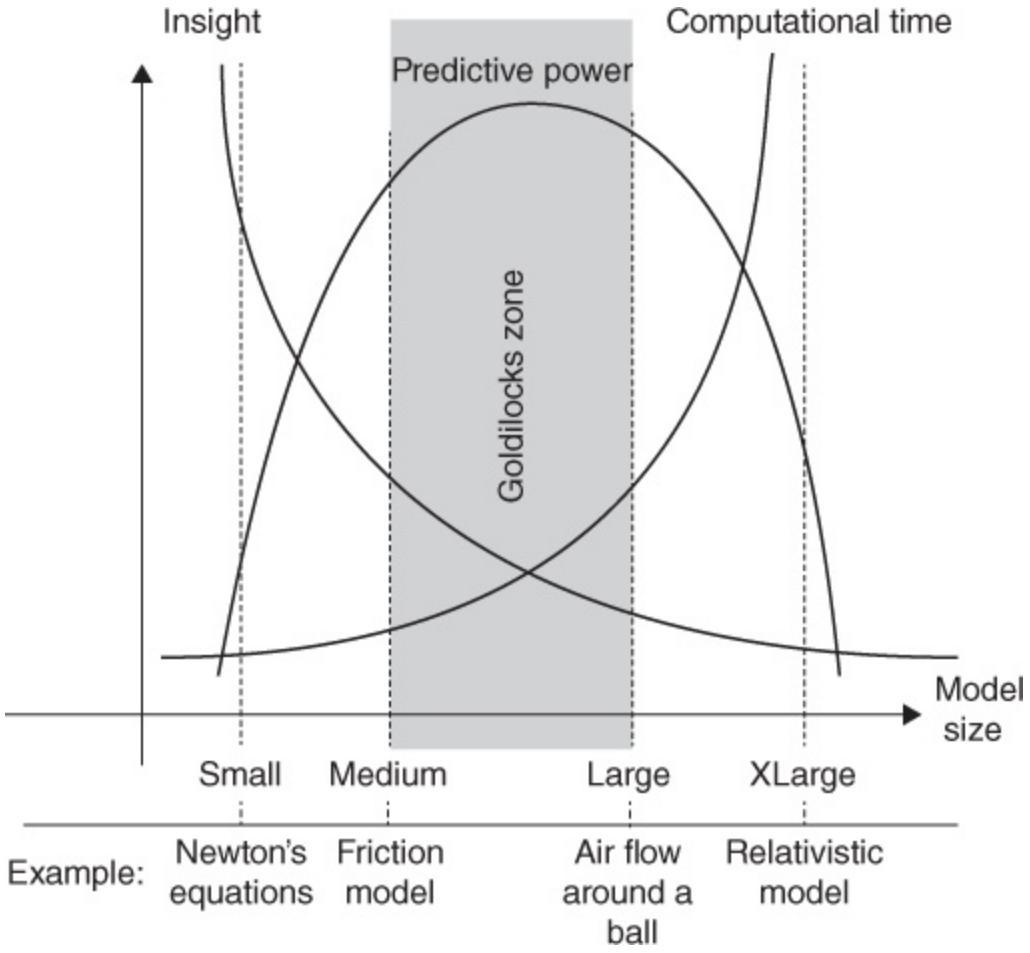
A perceptive reader might then object and correctly question the assumption that the force acting on the ball remains constant. Indeed, as the ball's vertical position increases, the force acting on it decreases according to Isaac Newton's law of universal gravity, which states that the force acting between two masses varies as the inverse of the square of the distance between their centre of gravity. While more complicated, a **third model** taking into account such a minute effect can easily be written down and solved. The analysis of this model will reveal that this effect is indeed negligible for the motion of a ball but that it may be relevant if one tries to understand the motion of satellites or inter-continental ballistic missiles.

Yet, the sporty reader is aware of another effect. When a cricket or baseball ball is spun during the throw, its position can wobble in the air, an effect that none of the previous models describe. In this case, experience (rather than experiments) guides us in considering a **fourth model** that will take into account the motion of an object of finite size that includes its spin and the interaction of the air with the stitches of the ball. This problem can be formulated but its solution is extremely complicated. A combination of advanced mathematical methods and extensive computational studies of such problems, coupling the air flow with the motion of the ball, has revealed how such balls can change direction multiple times in the air when properly spun. These methods, combined with experiments in wind tunnels, are used to design new sports balls.

Finally, a refined reader could further object that as the relative position of two masses change, they affect the underlying structure of space-time in which they sit by changing the metric used to measure both the distance between objects and the time interval between events. The problem of throwing a ball should then be solved within the theory of general relativity. A **fifth model** taking these effects into account could, in principle, be derived. But before doing so, it is important to question the relevance of such an approach. Modelling is not an exercise in abstract physics. It must rely on sound judgement and on the understanding of the underlying problem. In our case, unless the Earth and the ball have masses comparable to that of the Sun, we know from simple estimates that these effects do not play a role, and, indeed, could never be measured.

Model complexity

Just like the different lenses of a microscope, each of the models for the motion of a ball studies the same problem but at a different level of resolution or complexity in order to answer different questions. As the model complexity increases, the number of equations, variables, and parameters entering their mathematical formulation also increases. The equations generated by such a process increase in complexity and soon enough require a thorough numerical approach. As sketched in Figure 7, depending on the question, models can have different sizes that are roughly measured by the number of parameters and equations appearing in their formulation (a proper mathematical definition of complexity is elusive but most researchers would heartily agree that some models are more complex than others).



7. Model complexity showing the trade-off between model sizes, insight, and predictive power. Of course, for some problems the Goldilocks zone will lie further to the left or further to the right.

At one end of the spectrum, a simplistic approach to a problem may lead to a small model that can be easily solved but that still offers key insights. Our model for compound interest and our first model for throwing a ball would both fall into that category. They teach us about the fundamental nature of exponential growth and ballistic trajectories, respectively. However, these simple models typically have low predictive power as they ignore key properties of the system. Our financial model would be useless to predict the market and our Newtonian model would not help the design of new footballs or solve ballistic problems with any precision.

Larger models are more difficult to solve, and typically give less insight into a phenomenon, but when properly formulated and implemented they may have great predictive power. It is in that Goldilocks zone (*not too big*, *not too small!*), that modern mathematical modelling thrives as mathematics, computer science, sciences, and engineering come together to solve technological, societal, or scientific challenges.

At the other end of the spectrum, a blind adherence to all known physical processes or an excess in the number of parameters and variables leads to models that may be valid but are so large and poorly characterized that there is no hope of extracting any meaningful information from them. These models are both useless and impossible to solve. For instance, our preposterous fifth model for the motion of a ball may contain the correct physics, but it is completely disconnected from the original problem.

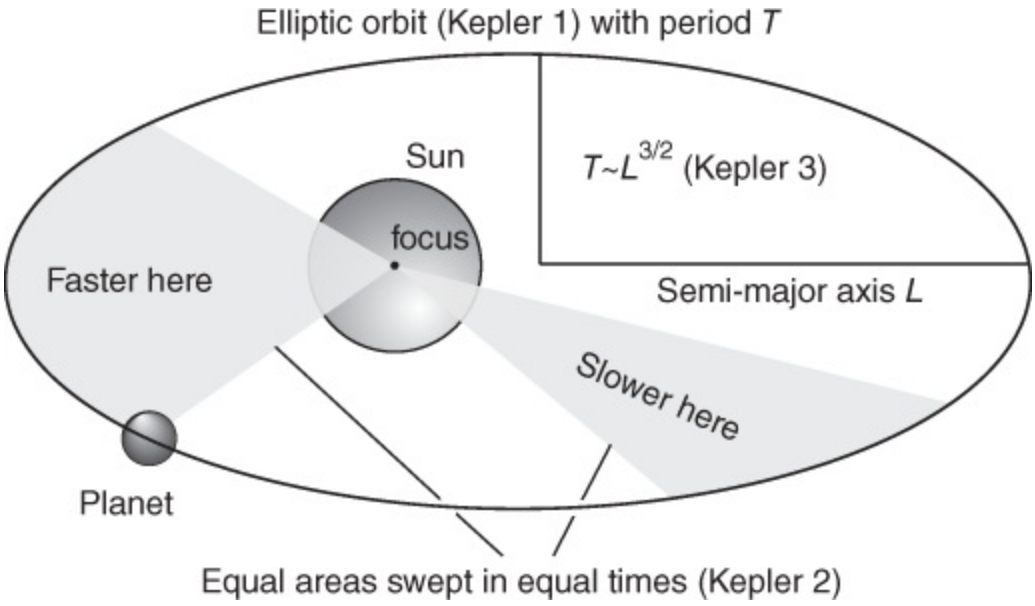
The tug-of-war between a model's complexity and how much information can be extracted from it is a delicate one and requires a good understanding of both the underlying scientific question and the methods that can be used to analyse a model. This is why modelling is often referred to as a craft and an art, as it must be based on knowledge, skills, intuition, and a certain sense of beauty and elegance.

The physics paradigm

What are the steps involved in developing a mathematical model for a given phenomenon? The textbook answer is provided by the so-called *physics paradigm*, a simplified apocryphal process through which the basic equations of physics were first established.

Consider, for instance, the motion of planets in our solar system. For millennia, astronomers have looked at the sky and dutifully tabulated the position of various objects. The outstanding question for natural philosophers was to find patterns in this data and formulate equations that capture key aspects of the orbits. The first major progress, marked by the publication of *de Revolutionibus* by Copernicus in 1543, was the realization that in order to

understand the motion of planets, it is easier to consider the Sun and not the Earth as the centre of the solar system. The second major area of progress was the extensive collection of data on orbits carefully recorded by Tycho Brahe in the second half of the 16th century. This data created a reference point that could be used by other researchers to test new hypotheses. The third step, thirty years after Brahe, was the discovery by Kepler of the laws of planetary motion that explains many astronomical data. These laws, sketched in Figure 8, are:



8. Kepler's laws of planetary motion. K1: closed orbits are ellipses with the Sun at one of its foci. K2: equal areas are swept in equal time. K3: the square of the orbital period of a planet is proportional to the cube of the semi-major axis.

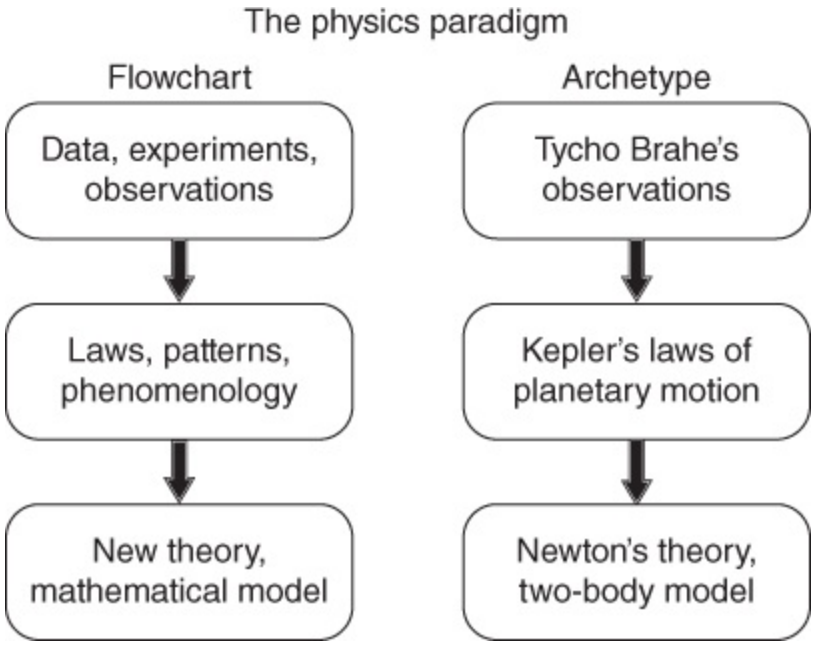
- K1. The orbit of a planet in the solar system is an ellipse with the Sun at one of the two foci.
- K2. Planets sweep out equal areas during equal intervals of time.
- K3. The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.

Kepler's laws are an example of *phenomenological laws* for planetary motion. By phenomenological, we mean an empirical model or a law that is not based on fundamental principles but explains the behaviour of a particular system and is corroborated by experimental data or observation.

A major revolution in the understanding of celestial mechanics took place with the publication of Isaac Newton's *Philosophiae Naturalis Principia Mathematica* (Latin for *Mathematical Principles of Natural Philosophy*). Published in 1687, this work is one of the most important treatises in the history of mankind. In the *Principia*, Newton postulated the fundamental laws for the motion of points under the action of forces, and proposed a universal law for gravity. Combining these two theories, Newton proposed a simple model to explain Kepler's laws. His main assumption was that the motion of a planet is only determined by the mutual gravity between the Sun and the planet. All other gravitational effects from other celestial bodies are ignored. Newton then showed that ellipses are possible solutions for the motion of two celestial bodies under mutual gravity and that Kepler's laws are a simple and direct consequence of his general theory.

The model for planetary orbits is just one of the many applications of Newton's theory and its importance is difficult to overstate. Apart from the extremely small (smaller than about a molecule) and the extremely large (say, larger than the solar system), Newton's laws of motion rule all physical objects and his equations are used routinely all over the world hundred of millions times every hour, in computer codes for the simulation or visualization of physical processes.

As illustrated by this example, the physics paradigm proceeds in three steps: observations or experiments that produce data, pattern identification that provides phenomenological laws, and theory building from first principles explaining within a few key assumptions all observations and laws observed, as depicted in Figure 9.



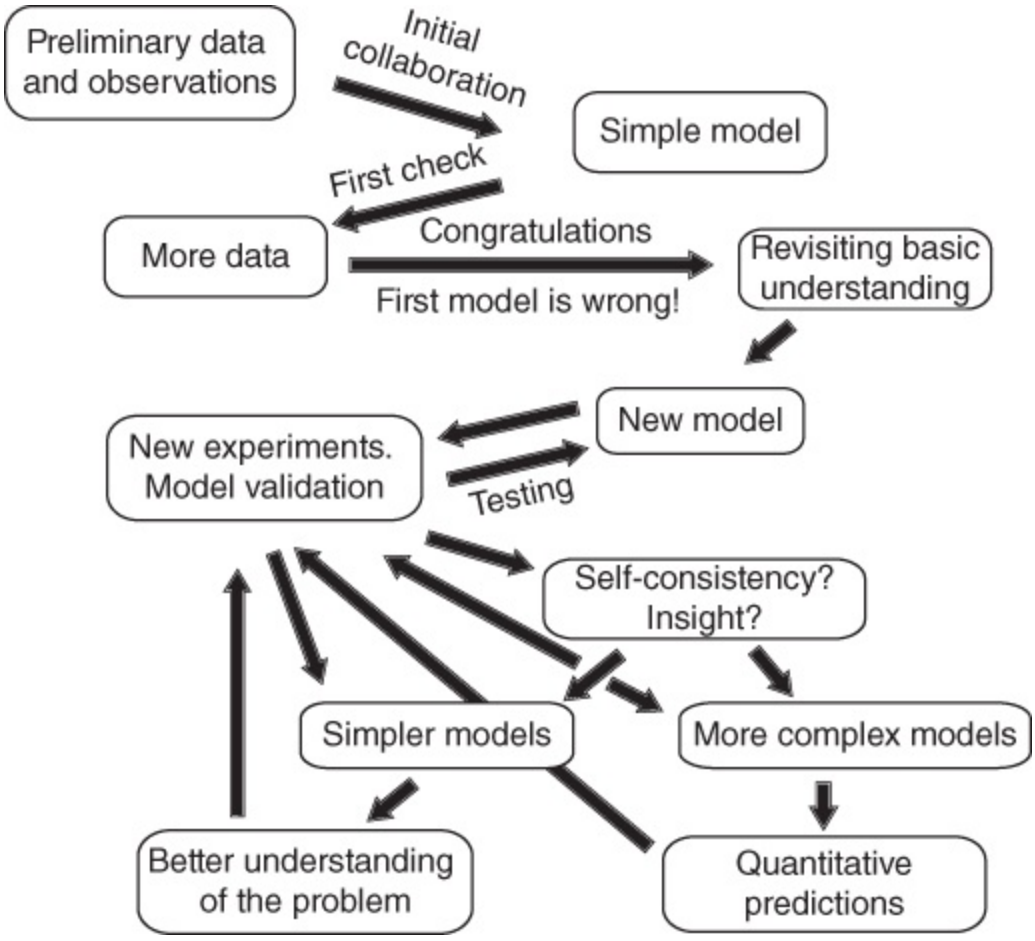
9. The physics paradigm. A possible way to build a model starts with data and observations that are summarized into phenomenological laws. In turn, both laws and data provide a basis for a general theory and more specific models applicable to other areas. The archetype of such a process is the formulation of Newton's theory of classical mechanics.

A modern view of mathematical modelling

The validity of the physics paradigm through history is open to debate. However, it is clear to practitioners that modern mathematical modelling is quite different from the classic view of physics and it will come as no surprise to the reader that applied mathematicians do not spend their lives idling under apple trees waiting for divine inspiration. There exists nowadays a large body of theory and methods that can be used to tackle problems. Most scientific experiments are not made in the ignorance of theory. They are designed hand-in-hand with theory to measure quantities that are precisely defined and to answer questions that are related to theoretical concepts. However, it would be wrong to assume that all scientific questions are amenable to mathematical modelling. In particular, research in biology, medicine, and chemistry typically relies on systematic experimentation and

statistical analyses to validate working hypotheses. In such studies, mathematical modelling may not be the best tool to approach the problem as the system may be too complex or not yet sufficiently understood to warrant a proper mathematical formulation. For instance, no mathematical modelling would ever replace a drug trial for cancer, no model would conclusively determine the role of a particular combination of genes on the development of an organism, and no chemist would believe a mathematical proof for the existence of new chemical compounds. Yet, even in these cases, mathematical modelling can provide insight, guidance, and inspiration.

In the areas where mathematical modelling is directly relevant, the process that leads to models, predictive or otherwise, is much more convoluted (see Figure 10). One such area is *collaborative modelling* where scientists or people from industry come together with applied mathematicians to tackle scientific questions. The scientist has a deep knowledge of the field and may come to a problem where they feel that modelling may be of some value. To the mathematician, the scientist says ‘I do not know if you can help me, but I believe that mathematics could be helpful for my problem.’ The mathematician has a poor knowledge of the problem but is easily impressed by nice colour pictures offering them a glimpse at the real world: ‘I have no idea if I can help you, but your problem seems fascinating. Please do tell me more.’ Out of this state of mutual respect and ignorance slowly emerges the fragment of a problem, a question, an idea, something worth trying either with a simple model or with an experiment. Typically, these first trials lead to simple but erroneous models and to irrelevant experiments that further reinforce the divide between the disciplines. Surprisingly, many mathematicians will freely admit that it is these early failures that are the most insightful as they force both parties to face and understand the basic concepts underlying the problem. If they manage to go beyond this state of maximum ignorance, many more iterations of these interactions may eventually lead to models that make sense and offer new perspectives on the problem. The seed is germinating and one can finally plan for the harvest. Mathematicians can write big models with many equations while scientists generate big data and new experiments to validate the model and extract useful parameters. It is in that state of mutual collaboration that collaborative mathematical modelling is at its best and that new properly quantified scientific disciplines emerge.



10. How modelling actually takes place. The development of a new mathematical model is a long and tortuous process that is continuously refined and checked for internal consistency and against data.

Chapter 4

Do you know the way to solve equations? Spinning tops and chaotic rabbits

A child of five could understand this. Send someone to fetch a child of five.

–Groucho (misattributed)

We learn at school that mathematics can be used to solve equations. We are typically presented with a relationship with numbers and variables and we are asked to isolate one of the variables, the mythical 'x'. For instance, we can ask what is the value of x that solves the equation $x + 3 = 7$. Isolating x on the left side of the equation by subtracting 3 from both sides leads to $x = 4$, the *solution* of the equation. We can generalize the problem by considering the equation $x + a = b$ where a and b are real numbers. Then, the solution is $x = (b - a)$. We have obtained a *formula* that gives us a solution to all equations of the type $x + a = b$.

As equations become more complicated, obtaining a formula for the unknown variable becomes increasingly hard. Soon, no solution can be found and one is forced to formulate more basic questions: does a solution even exist? If it does, is it unique? Is there a formula, even if that formula cannot

be found explicitly? Despite the popular image of the mathematician dealing with complicated expressions all day long, there is very little emphasis on solving equations in pure mathematics. The main emphasis is on the existence of solutions, their uniqueness, their properties, and how these properties relate to other equations or to other mathematical structures.

However, in applied mathematics it is of the greatest importance to solve equations. These solutions provide information on key quantities and allow us to give specific answers to scientific problems. If we want to predict tomorrow's cloud coverage and be able to write a series of equations to determine it, we may not be terribly satisfied to know that a solution for the problem exists in some properly defined space of functions if we cannot also provide an estimate for it. However, if we do obtain an estimate as a solution, we certainly would want to know if this solution is mathematically sound and whether it is unique. We would also like to be able to gauge the quality of our estimate. Developing methods to find exact, approximate, and numerical solutions for classes of equations sharing common features is at the heart of applied mathematics.

There are three ways to solve equations. The first consists in finding a general formula in terms of known functions. In this best-case scenario, the system is *explicitly solvable*. It means that the solution can be found explicitly in terms of known quantities. These exact solutions are rare but extremely useful.

When equations are not explicitly solvable, the second strategy is to find a reasonable approximation of the solution, valid within a certain range. These *asymptotic* or *perturbative methods* are central to applied mathematics and when properly used, allow the clever mathematician to extract information out of a seemingly unsolvable equations just like a magician would pull a rabbit out of a hat.

The third approach is to solve equations *numerically*. Since computers can deal with numbers, all parameters and variables must be specified and algorithms must be designed to manipulate numbers and extract numerical solutions out of equations. This approach has been extremely successful and

has revolutionized the way we handle models and solve equations. It also benefits from the rapid developments in computer power and algorithm design over the last sixty years. Computers can obtain solutions for systems of equations containing thousands to millions of variables. The development of modern applied mathematics has proceeded hand-in-hand with the development of computers and algorithms, each benefiting from this synergy.

Equations we cannot solve

The equation $x + 3 = 7$, which we started with, is an example of a linear equation. We call it *linear* because the unknown 'x' only appears by itself and not within a function or with a power (such as $1/x$, or x^2 or $\sin(x)$). The general form of these equations is $ax + b = 0$ where a and b are real numbers (and $a \neq 0$). Any equation of this form can be solved by using the formula $x = -b/a$. We are now ready to solve more equations. The next ones, in order of complexity, are *quadratic* equations of the form

$$ax^2 + bx + c = 0, \quad \text{with } a \neq 0,$$

which can be solved by the well-known formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \qquad (4.1)$$

Given the values of a, b, and c, and a calculator, we can obtain x. The similar problem for cubic and quartic equations

$$ax^3 + bx^2 + cx + d = 0,$$
$$ax^4 + bx^3 + cx^2 + dx + e = 0,$$

can also be solved by rather long but completely explicit expressions. These formulas contain many more terms but express all the solutions of x as explicit functions of the coefficients.

What about quintic or sextic equations? The groundbreaking work of the mathematicians Niels Henrik Abel and Evariste Galois in the 1820s tells us that there is no such general formula for higher order equations starting with the quintic equation

$$ax^5 + bx^4 + cx^3 + dx^2 + ex + f = 0.$$

It is not a case of being smart enough to find a formula. They proved rigorously that, in general, there is no formula like (4.1) that would give the solution of this equation in terms of the coefficients.

In the zoo of mathematical equations, quintic algebraic equations are rather simple. They involve only one variable, and no other complications such as differential quantities, integrals, or other functions. Yet, we are already faced with a simple truth: it is not possible, in general, to write formulas for the exact solutions of such equations.

Solving differential equations

The problem of finding formulas for equations has a long history in the theory of differential equations. These are equations that not only involve an unknown x but also its derivatives. Unlike the above polynomial equations where the answer is just a number, the solution of a differential equation is a function. For instance, consider the simple differential equation for the unknown $x = x(t)$:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x, \qquad (4.2)$$

where $\mathrm{d}x/\mathrm{d}t$ is the rate of change of $x(t)$ with respect to t and is called the *derivative* of $x(t)$. Unlike polynomial equations, the solution of this differential equation is not a number but the function

$$x(t) = C\exp(t),$$

where C is an arbitrary constant and $\exp(t) = e^t$ is the exponential function. The proof that a given function is the solution of a differential equation is performed by substituting it back into the differential equation. In our case, since the derivative of e^t is simply e^t, the left side of the equation is Ce^t and so is the right side, which concludes our proof (showing that this solution is unique requires a bit more work).

This differential equation was first proposed by Galileo as a model for uniform motion in the 16th century. The model turns out to be incorrect. Nevertheless, John Napier became interested in the mathematical properties of this equation's solution and its inverse. After twenty years of laborious work, his study eventually led him to the concept of exponential and logarithmic functions, summarized in his 1614 book *Mirifici Logarithmorum Canonis Descriptio (A Description of the Wonderful Law of Logarithms).* This book, the first table of logarithms, played a crucial role in the early development of science and engineering.

In previous sections, we have seen other differential equations cropping up in various places. In applied mathematics, these equations are found everywhere; front, centre, and middle, it is hard to find a model that does not involve them. Why are they so ubiquitous? What is so special about differential equations?

Isaac Newton was the first to recognize their importance. Remember that Newton's celebrated equation for the motion of a mass subjected to a force links force to acceleration. But, acceleration is also the rate of change of velocity with respect to time. It is the derivative of velocity with respect to time. Velocity itself is the rate of change of position with respect to time, which makes acceleration the second derivative of position with respect to time. If we know the force $f(x, t)$ applied to a mass, m (with $m \neq 0$), located at position $x(t)$, where t is time, Newton's equation is simply

$$ma = f \quad \Rightarrow \quad a = f / m \quad \underset{a=\frac{d^2x}{dt^2}}{\Rightarrow} \quad \frac{d^2x}{dt^2} = f(x(t),t) / m.$$

This last equation is a *second-order differential equation* for x (the term second-order simply refers to the fact that the highest derivative that appears in the equation is a second derivative). It expresses a very simple idea, that the change of velocity of an object in a small time interval change is due to the force that is applied to it during that interval.

Consider for instance the following scenario. If my car runs out of gas and I decide to push it along the road, the force that I exert on it will vary with time and position (since the road may not be flat). If I know this force and the force of friction exerted by the road on my car, I know $f(x(t), t)$ as a function of time. If I want to predict how far I can move the car in a few hours of fruitless effort, I will need to solve this differential equation to know what happens over a long period of time when the force is continuously applied. Solving Newton's equation bridges our knowledge from small-time intervals to long-time behaviour.

Newton opened an entire new direction of research. Based on his initial idea, it became conceivable to formulate equations of motion not only for apples falling off trees or planets moving around the Sun, but indeed for any object small or large that is subject to forces. And since, apart from light, the universe is made of matter, it applies to basically everything around us. This programme culminated in the 19th century with classical mechanics, a theoretical framework to describe the motion of fluids, solids, gases, and particles. There was a great sense of optimism by the middle of the 19th century through the work of the great mathematicians of the time such as Joseph-Louis Lagrange, Pierre-Simon Laplace, Siméon Denis Poisson, Augustin-Louis Cauchy, and William Rowan Hamilton. This is well summarized by a quote from Laplace's *Philosophical Essay on Probabilities*:

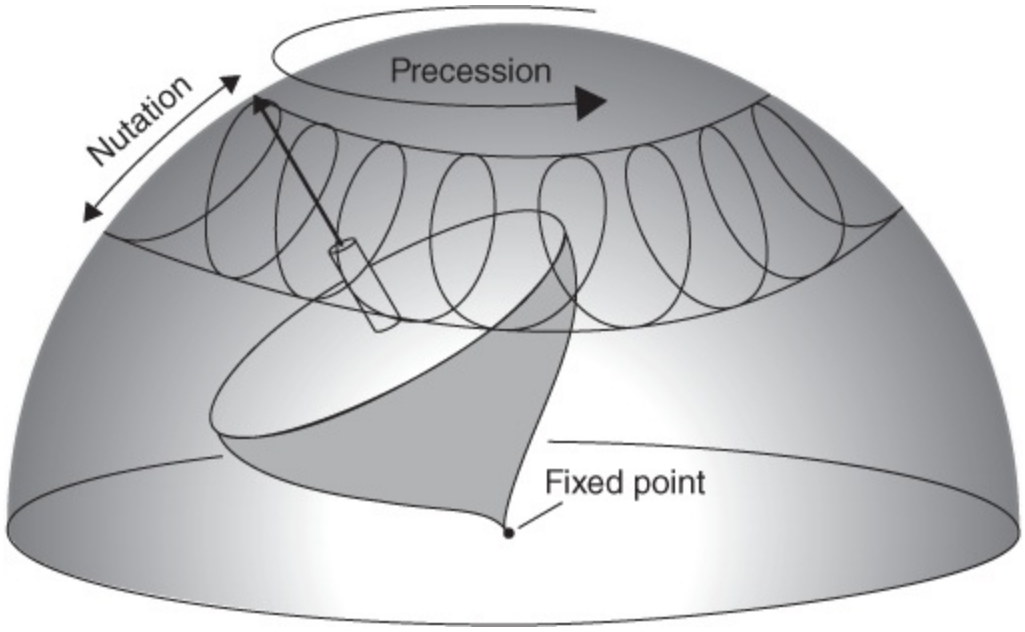
> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

This is called *determinism*; it is the idea that the trajectory of all objects can be determined as solutions to Newton's equations. Since the fundamental

equations of nature were known, the only remaining problem was to find their solutions.

We know now that with the advent of statistical and quantum physics in the early 20th century, Laplace's deterministic view is inconsistent with our understanding of microscopic physics. Even so, from a mathematical perspective, by the end of his century, Laplace's optimism was dampened by three important mathematical discoveries that changed the way we think about differential equations and their solutions.

The first milestone came with the remarkable work of Sofia Kovalevskaya, the first woman in Europe to earn a doctorate in mathematics (1874) and to be given the title of full professor (1889), at a time where most universities did not allow women to even attend lectures. The problem that Kovalevskaya studied consists in finding a solution for the motion of a rigid body around a fixed point, a central problem of classical mechanics. If you spin a top and fix its point of contact with the ground, the axis will rotate while moving up and down, creating intriguing motions (see Figure 11). The general equations describing this motion had already been obtained by Euler, who solved the equation for a top whose fixed point is the centre of gravity. Another solution had been found by Lagrange in the case of a heavy symmetric top, where the centre of gravity is on the symmetry axis (this is a long description for what is really the spinning top from our childhood). The natural question was then to find solutions of the Euler equations for other possible shapes and configurations of spinning tops. Kovalevskaya's contribution to the problem was truly spectacular. First, she found a new case for which there exists a complete solution of the problem. This is the so-called Kovalevskaya top, which can be realized artificially as a rather difficult arrangement of masses spinning around an axis. More importantly, she showed that no other mass distribution in a rigid body other than the ones known previously could be solved exactly. Again, we had a proof for the non-existence of solutions.



11. The Lagrange top is a spinning top held fixed at the base (fixed point). The solution found by Lagrange describes the curves that the axis draws on an imaginary sphere surrounding the top. Depending on the initial velocity and position, different types of motion are possible. For instance, the axis can draw interesting loops on the sphere during the motion. This particular motion is called *looping precession*.

Maybe Euler's equations were particularly difficult? What can be said about general equations? Soon after Kovalevskaya's solution, a group of French mathematicians headed by Paul Painlevé was trying to classify and solve all equations. Painlevé is yet another fascinating historical character. Aside from his mathematical achievements, he entered the political arena and was minister for Public Instruction and Inventions during World War I, before becoming prime minister. He was also an aviation enthusiast, promoted the development of civil aviation as a way to achieve global peace, and became, in 1908, the first airplane passenger by flying with the Wright brothers. In mathematics, by the end of the 19th century, Painlevé and his colleagues had started an ambitious programme of research to determine all those differential equations that can be solved exactly. The problem was not to try to solve a particular equation but to find all the equations that can be solved in terms of known functions. First-order equations had mostly been exhausted, so Painlevé focused on second-order equations. He managed to establish that

amongst all possible equations, there are only fifty that have the property of being simple enough to be called solvable. Out of these fifty, forty-four were known. The six new ones are called the Painlevé equations, the first one of which is

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = 6x^2 + t.$$

This equation cannot be solved in terms of simple functions, but its solution is regular enough that it can be used to define new functions, in the same way that Napier defined the logarithmic function through the solution of a differential equation. Aside from these equations, no other second-order differential equations are simple enough that they can be used to define new functions.

This abstract mathematical result has important consequences for mechanics. Recall that Newton's equations are second-order differential equations. Therefore, Painlevé's result implies that we cannot solve generic equations for the motion of particles under arbitrary forces.

Painlevé and Kovalevskaya conclusively showed that it is hopeless to try to find explicit solutions for generic equations. Therefore, new ideas were needed to be able to solve physical problems. Such ideas first began to appear at the turn of the 20th century with the work of Henri Poincaré, a giant amongst the giants in the pantheon of mathematics and physics (history textbooks tell us that the lives of the Poincaré and Painlevé were intertwined as Henri Poincaré's cousin, *Raymond* Poincaré was president of France when Painlevé was prime minister). One of the key challenges of mathematics at the time was to solve the three-body problem. The two-body problem is simply Kepler's problem encountered in Chapter 3, a problem that was fully solved by Newton. A natural question for mathematicians following in the footsteps of Newton was to solve the problem of three interacting bodies (the simplest example of which is the Sun–Earth–Moon system). While some progress was made by Euler and Lagrange, 200 years after Newton, a general solution of this problem still eluded mathematicians. In 1885, Oscar II, King of Sweden, raised the stakes by establishing an international prize for

anybody who could solve the problem. The winner would receive a gold medal an a hefty 2,500 crowns (equivalent at the time to £160).

Poincaré won the competition in 1889 despite the fact that he did not find a complete solution, but, in the words of the venerable German mathematician Karl Weierstrass: '[his work] is nevertheless of such importance that its publication will inaugurate a new era in the history of celestial mechanics'. There are three remarkable features to Poincaré's work. First, following Kovalevskaya and Painlevé, he proved that no exact solution could be found. Second, rather than throwing in the towel, he developed new methods to attack the problem. Essentially he changed the question. Instead of looking for formulas for the solution, Poincaré argues, we must look at the problem from a geometric

perspective. What are the different types of orbits that can exist? Are there closed orbits? If so, what do they look like? By doing so, Poincaré laid the foundation for a *qualitative* theory of differential equations (looking at global behaviour) rather than the *quantitative* approach (obtaining formulas) initiated by Newton and Leibniz.

The third remarkable and ironic feature of Poincaré's winning entry is that it contained a mistake, an outstanding, glorious, but serious, mistake. This mistake was spotted by a young Swedish mathematician, Edvard Phragmén, while editing Poincaré's memoir. Essentially, Poincaré had mistakenly assumed that certain exotic behaviours could not happen. When the hole in his proof was pointed out to him, Poincaré studied these solutions in more detail and realized that a completely new type of motion could exist. The complexity of these solutions was beyond anything that had ever been discovered in the history of mathematics. Of the shape of these solutions, Poincaré writes:

> The complexity of this figure is striking, and I shall not even try to draw it. Nothing is more suitable for providing us with an idea of the complex nature of the three-body problem, and of all the problems of dynamics in general.

What Poincaré discovered in the solutions of the three-body problem was the possibility of what modern scientists now understand as chaos. The solutions

can be so complicated that they would turn back onto themselves and form, in the words of Poincaré, ‘a type of trellis, tissue, or grid with infinitely fine mesh’. This initial discovery is connected to the birth of the theory of chaos and dynamical systems, as discussed later in this chapter.

Tautologies: the tale of rabbits and wolves

In physics, differential equations express the basic, local conservation of physical quantities such as energy and mass. When modelling a phenomenon, differential equations express simple fundamental relationships that are valid when small changes of a quantity are considered. Some of these relations are so trivial that they would seem useless at first sight. Yet, the cumulative effect of small trivial increments can lead to great complexity as was shown by Poincaré. Solving these equations allows us to understand how complex effects are born out of simple interactions.

Let us consider a situation that does not rely on any knowledge of physics: the highly simplified interaction of a prey with a predator in a homogeneous environment, say rabbits and wolves. In our ideal world, rabbits have as much food as they want and wolves can only rely on the rabbit population for nourishment. We assume that in the absence of wolves, the rabbit population would naturally grow without limitation. Since each pair of rabbits can create new rabbits, the rate of growth is proportional to the population itself. This is exactly what happened in 1859 in Australia when twenty-four rabbits were released. It was said at the time that ‘The introduction of a few rabbits could do little harm and might provide a touch of home, in addition to a spot of hunting.’

When the population is large enough, it is reasonable to assume that the number of rabbits can be modelled by a continuous variable $R(t)$ giving the rabbit population at time t. By themselves, the simplest model is to assume that *the rate of change of the population is proportional to the population itself*. The rate of change of the population is simply its derivative in time. Therefore, in terms of differential equations, this statement is written

$$\frac{dR}{dt} = \alpha R,$$

where α is a positive parameter. Since the rate of change of the population is positive, the population increases in time. Similarly, let $W(t)$ describe the population of wolves. Then, in the absence of any interaction between the two species, the population follows the law

$$\frac{dW}{dt} = -\beta W,$$

where β is positive. Now the rate of change of the population is negative and the population decreases in time.

Even though these two equations have nothing to do with Galileo's equation of motion (4.2), mathematically they are identical (with two extra parameters). Therefore, following the same steps, we obtain their solutions as

$$R = R_0 e^{\alpha t}, \quad \text{and} \quad W = W_0 e^{-\beta t},$$

where R_0 and W_0 are the levels of each population at a chosen time $t = 0$.

We conclude that in the absence of interactions, rabbits reproduce like rabbits, and the population happily grows exponentially as each new generation has more offspring, which in turn produces new offspring. This is just the compound law we encountered in Chapter 3 applied to population rather than savings. And, exponentially they grew! By 1869, the rabbit population of Australia without any predator had grown so fast that two million could be killed annually without making a dent on population growth. Despite all efforts, the population was estimated at 600 million by 1950. The wolf population, on the other hand, would disappear exponentially fast in a world where they cannot eat rabbits.

A nice interpretation of the parameter α is related to the doubling-time, r_2: the time it takes for the rabbit population to double. Setting $R(r_2) = 2R_0$ in the

previous equations leads to $r_2 = \log(2)/\alpha$. Similarly, the wolf population cannot thrive in a world without rabbits and the half-life for this hungry population, defined as the time it takes to decrease to half of its original size, is $w_2 = \log(2)/\beta$.

Sadly, the nature of life is such that things become more interesting when wolves and rabbits interact. Here, we assume that wolves eat rabbits at a rate proportional to their rate of meeting, RW. This product of R and W expresses the fact that it takes two to tango: both rabbits and wolves need to be present for predation to occur. This predation will decrease the population of rabbits and increase the population of wolves, leading to

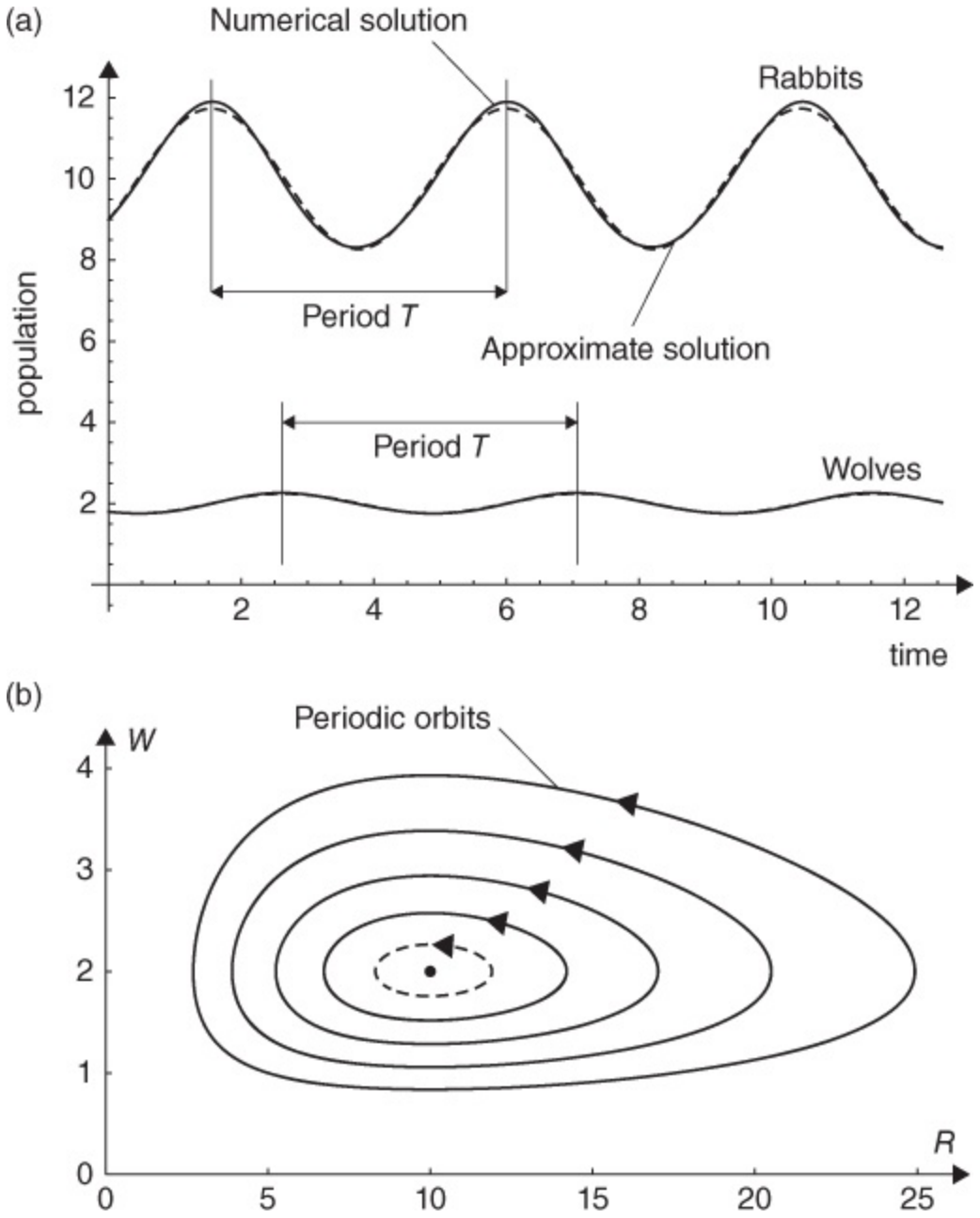
$$\frac{\mathrm{d}R}{\mathrm{d}t} = \alpha R - \gamma RW,$$
$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\beta W + \delta RW,$$

where $\gamma > 0$ and $\delta > 0$ quantify the net loss and gain for each population due to predation.

The reasoning that leads to this set of equations is almost trivial: rabbits reproduce, wolves eat rabbits, wolves die without rabbits to eat. We have mathematically expressed these assumptions in the simplest possible framework. Now, we can ask a non-trivial question: what happens to both populations? This question can only be answered by analysing this system of differential equations. In our case, the system is simple enough that it can be solved exactly in terms of complicated functions. However, this exact solution is not particularly illuminating. A more fruitful approach consists in looking at simple solutions where both populations are in equilibrium. This situation occurs when both R and W are constant, that is either $R = W = 0$ or $R_{\mathrm{eq}} = \beta/\delta$ and $W_{\mathrm{eq}} = \alpha/\gamma$. The first case is trivial and of no interest (no population). In the second case, both populations are exactly in agreement so that no change occurs in time. This situation is rather unlikely as the system must also start initially on this equilibrium to remain there. Therefore, a natural question is to ask what happens if the initial population is close to that special equilibrium solution? A mathematical analysis of the problem reveals

that in such a case, both population levels will oscillate periodically but out of phase: as the number of rabbits increases so does the number of wolves. But when too many wolves are present, the population of rabbits decreases and so does the wolf population that relies on them. A little bit more work would show that close to equilibrium, the period of oscillation is $T = 2\pi / \sqrt{\alpha\beta}$ (since the algebra requires more than two lines and a few more concepts, you will need to trust the author here).

The oscillatory behaviour appears in Figure 12 which shows both the approximate solution and the solution obtained numerically (see the next section about numerical solutions). To all practical purposes, the numerical is an exact solution and can be used to test the validity of the approximate solution. Our analysis, in the spirit of Poincaré, combines exact, approximate, and numerical results to give a full picture of the type of behaviours expected when two species interact. It also identifies a crucial combination of parameters $1 / \sqrt{\alpha\beta}$ as the key time scale for the problem.



12. Oscillations in the Lotka–Volterra model of prey-predator evolution. (a) the dashed line is the approximate solution and the solid curve is the numerical solution. Note that in the case of wolves, the two curves are indistinguishable. Both populations oscillate periodically but out of phase. The population of wolves is still increasing when the rabbit population reaches its maximum. The values chosen here are $\alpha = 2$, $\beta = \gamma = 1$, $\delta = 1/10$, corresponding to a period $T = \sqrt{2\pi} \approx 4.44$ and the initial conditions are $R0 = 9$, $W0 = 1.8$. (b) representation of the solutions in the

phase plane where at each time *t*, we plot the point *(R(t),W(t))* of a solution. A periodic solution forms a closed curve in this representation. The dashed curve corresponds to the same solution as the one shown in the upper graph.

This system of equations is known as the Lotka–Volterra model, after the two mathematicians who first studied it around 1910–20. The original motivation of Vito Volterra was to understand the oscillations observed in a fish population during World War I. The existence of oscillations in this model is an indication of the possible dynamics that can be obtained from nonlinear interactions in dynamical systems. The nonlinearity in our case is the product *RW* due to the predation term. It expresses the interaction between populations. Exquisite details about the solutions have been obtained through mathematical analysis, such as how the solution changes when the coefficients change due to seasonal forcing. This simple model is the starting point of a largebody of work that led to remarkable applications of mathematics in the modelling of ecosystems and the spread of diseases. Nowadays, the field of theoretical epidemiology and theoretical ecology are based on these ideas and the Lotka–Volterra model remains a paradigm in these studies.

Numerical solutions

What do we mean by a numerical solution? With the power of modern computers, scientists can obtain the numerical solution of very large models with millions of equations and make startling predictions. But how does it actually happen? And what is the difference between an exact and a numerical solution? To illustrate the basic idea, consider again our first differential equation for the exponential growth of the rabbit population,

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \alpha R.$$

We know that its *exact* solution is $R = R_0 \mathrm{e}^{\alpha t}$. Its *numerical* solution can be obtained by recalling that the rate of change of R is well approximated by

considering how the population changes over a small time increment Δt, where

$$\frac{dR}{dt} \approx \frac{R(t+\Delta t) - R(t)}{\Delta t}.$$

We can substitute this approximation in the differential equation to obtain

$$\frac{R(t+\Delta t) - R(t)}{\Delta t} = \alpha R(t) \quad \Rightarrow \quad R(t+\Delta t) = R(t) + \alpha R(t)\Delta t.$$

The key feature of this equation is that it can be used to find the value of the variable R at time $t + \Delta t$ by knowing R at time t. For instance, if I start with $R = 1$ at time $t = 0$ and use $\alpha = 1/2$, I can find the value of R at time $\Delta t = 0.1$ as $R(0+ 0.1) = 1\ (1+ 1/2 \times 0.1) = 1.05$. This process can now be iterated. I can find the value at $t = 0.2$ as $R(0.1 + 0.1) = R(0.1)(1 + 1/2 \times 0.1) = 1.05 \times 1.05$ and so on. From the definition of a derivative, I have derived a simple *numerical algorithm* to obtain a solution for a differential equation.

The same process can be used to compute the numerical solutions of the coupled system for R and W. Then, starting with an initial value for both populations, I can obtain, in principle, the solution at later times. But we have to be cautious here. In approximating the derivative dR/dt by a finite difference $(R(t + \Delta t) - R(t))/\Delta t$,I have made a small error. This error remains small after a few steps but if I keep computing the solution over larger times, these errors may accumulate. At some point the numerical solution may be completely different from the solution I was interested in.

We now have an interesting mathematical problem. How good is the numerical solution? Is it a decent approximation of the exact solution? Can I define properly what I mean by a ‘good’ solution or a ‘decent approximation’? How should I choose the small increment Δt to guarantee that I remain close to the actual solution after a long time? These questions are at the core of *numerical analysis*, the mathematical theory that underpins all numerical simulations. The main goals of numerical analysis are to find new algorithms, adapt existing ones to solve systems of equations and handle

data, and understand their mathematical properties. With good numerical methods and sufficient computer time, the errors developed in the computation can often be made as small as we want. Hence, numerical analysis provides an alternative way to generate reliable solutions whose deviations from the ideal exact solutions can be controlled so that, for all practical purposes, these solutions can also be considered as exact. Since many mathematical models involve differential equations and we know that most differential equations cannot be solved exactly through formulas, numerical analysis has become the principal tool for querying and solving scientific models.

Chaotic rabbits

A simple way to visualize the solution of the Lotka–Volterra equation is to draw them in the Rabbits–Wolves plane. Indeed at each time t, the solution is a pair of values $(R(t), W(t))$. This pair represents a single point in the *phase plane* $R - W$ as shown in Figure 12. As t increases, the point $(R(t), W(t))$ traces out a curve in that plane. Poincaré's revolutionary idea was to consider the shape of these curves. If a solution is constant, it will be a single *fixed point* in the phase plane. If it is periodic, it will create a *periodic orbit* in the same plane, that is a closed curve. A simple general question is then: what are the possible shapes in the phase-plane for solutions of differential equations? This problem is one of the central problems of the mathematical theory of dynamical systems developed in the 20th century as an alternative to finding exact solutions. A particularly important aspect of the theory is that for a large class of systems of two variables that do not depend explicitly on time (such as the Lotka–Volterra system), the answer is remarkably simple. The only possible solutions that remain finite are periodic orbits, fixed points, or curves that connect fixed points and periodic orbits (fixed point to fixed point, fixed point to periodic orbit, or periodic orbit to periodic orbit).

What about systems of three variables? One of the great surprises of applied mathematics was the discovery that for dynamical systems with three variables, the curves drawn in phase space can be extremely complex. Apart from the fixed points and periodic orbits, there are complex solutions known as *chaotic orbits*. To illustrate the possibility of such orbits, we return to our

Lotka–Volterra system and add another predator, say foxes. Our new system now reads

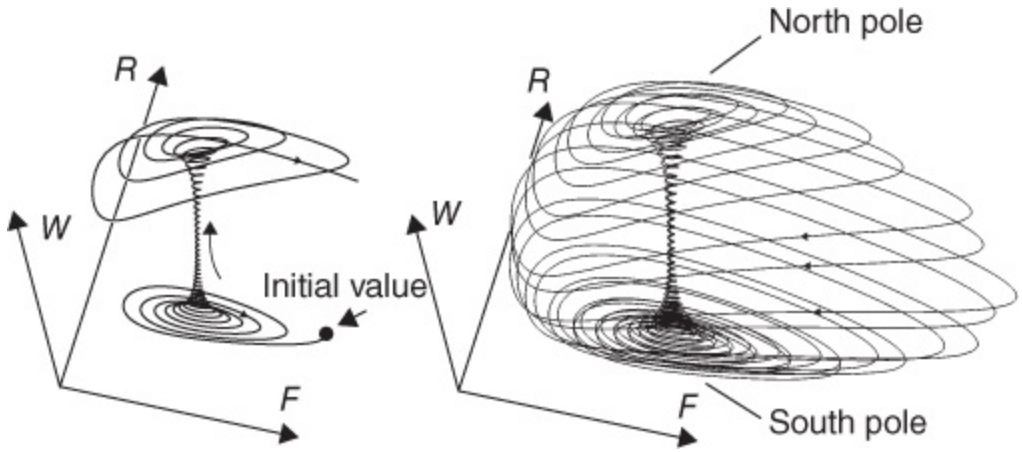
$$\frac{\mathrm{d}R}{\mathrm{d}t} = R - RW + \alpha R^2 - \beta FR^2,$$

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -W + RW,$$

$$\frac{\mathrm{d}F}{\mathrm{d}t} = -\gamma F + \beta FR^2,$$

where α, β, and γ are constant positive parameters and F represents the fox population. If these parameters vanish, we recover our initial Lotka–Volterra system. This system is not meant to be a model of a particular ecological system. Instead, it is used here to understand the nature of solutions to be found in dynamical systems.

Depending on the values of parameters and the initial population size, various behaviours, such as fixed points and periodic orbits, can be discovered. More interestingly, in some parameter regimes, the solutions become chaotic, as shown in Figure 13. For these values of the parameters, the solution does not settle on a periodic orbit, but instead is seen to wrap around a spheroidal structure until it reaches its south pole. Then the orbit is propelled directly from the south to the north pole before it starts winding around the ball again. This orbit is called chaotic because of its complexity. For instance, it becomes impossible to predict after a while whether the orbit will be close to the north pole or to the south pole, or how many times it is going to wind around the ball before reaching the north pole. Whereas the solution is fully determined by the initial values, many of its characteristics are similar to those found in random systems. This also implies that two trajectories that start arbitrarily close to each other will eventually look completely different.



13. A chaotic attractor for the three-dimensional Lotka–Volterra equations ($\alpha = 2$, $\beta = 2.9$, $\gamma = 3$). Left: for small times, an initial solution spirals around and downward until it reaches the south pole. The orbit follows the central core up to the north pole where it starts spiralling down again. Right: for longer times, the solution repeats this motion. However, the number of turns in the spiral changes in an unpredictable manner.

In the 1960s, the discovery that simple dynamical systems could have complex chaotic solutions marked the beginning of a new era in the study of differential equations. It was also a refutation of Laplace's deterministic view. Yes, the equations of classical mechanics are simple and deterministic, but their solutions can be as complex as random systems. The accumulation of small and simple interactions over long durations can lead to extraordinarily complex behaviours. Equations may be simple, but their solutions can be devilishly complex.

Reflections

When looking at the history of equations, and in particular, differential equations, we note an interesting interplay between mathematics, applied mathematics, and other sciences. Differential equations first emerged through the modelling of natural phenomena. They were then studied in detail by mathematicians who used them to develop new branches of mathematics.

These results and the understanding that mathematics brought were central to analysing many models.

In turn, new models raised new questions and were the source of new mathematical disciplines. For instance, it was the work of people such as Edward Lorenz in meteorology and Robert May in ecology that led to the discovery of chaos and fascinating new mathematical objects such as *strange attractors* and *wild hyperbolic sets*. As a result dynamical systems theory has emerged as a new branch of mathematics. These ideas come from the systematic analyses of very specific problems aided by a combination of rigorous tools, approximate methods, and, most importantly, numerical advances. Nowadays, applied mathematics remains a fertile ground and a source of inspiration for new mathematical ideas. And, mathematics brings not only new methods but also allows for a deeper understanding of natural phenomena. In the words of Maurice Biot, a particularly prolific applied mathematician of the 20th century:

> What should be referred to as applied mathematics does not exist on its own, but describes essentially a function and a craft by which the science of mathematics finds its nourishment.

Chapter 5

What's the frequency, Kenneth? Waves, quakes, and solitons

Time flies like an arrow. Fruit flies like a banana.

–Groucho (attributed)

When somebody talks to you, you hear the sound of their voice. Sound is a pressure wave that travels through the air. The information is transmitted by a change of air pressure. These variations of pressure are then decoded by your cochlea and processed by the brain. At any given time, the intensity of a sound wave, measured by the local pressure field, will vary from point to point and at any given position, this intensity will vary in time. The pressure $P = P(\mathbf{x}, t)$ is a function of both the position in space $\mathbf{x} = (x, y, z)$ and the time t. Similarly, most physical quantities of interest evolve both in space and in time. For instance, the temperature in a room varies not only throughout the day, but also at different places, for example next to a window or a radiator. Therefore, the temperature $T = T(\mathbf{x}, t)$ is a function of both position and time. Our main problem is to obtain mathematical models that describe quantities such as temperature or pressure that change both in time and position. By comparison, the models for planets and populations that we studied in the previous chapters were relatively simple. They could be described by functions that only varied in time (e.g. the position of the Earth or the number

of animals in interaction). At the mathematical level, these systems could be described by *ordinary differential equations* relating the rate of change of these functions to the functions themselves.

Since we live in a world where interesting things happen both in time and space, the description of many natural phenomena requires the study of functions that depend on multiple variables. These laws can also be expressed as differential equations, but since both space and time derivatives now appear in these equations, the corresponding equations are called *partial differential equations*. Partial differential equations are the generalization of ordinary differential equations for which more than one type of derivative appears in the equation. Just like ordinary differential equations, these equations are the natural mathematical tool for modelling because they express, locally, basic conservation laws such as the conservation of mass and energy. For instance, we can express the fundamental physical concept of mass conservation by stating that the change of mass in a small volume is equal to the mass going through the boundary of that volume (since mass cannot appear or disappear, any extra mass must come from outside that volume and any mass loss must also have gone through the boundary of the domain!). In the limit when the volume is taken to be an infinitely small quantity, this statement leads to an equation for the density that depends both on its time derivative (expressing the local change at one point) and its spatial derivative (expressing its variation in space).

We saw that ordinary differential equations can capture complex dynamics and their study has kept mathematicians busy for many centuries. By comparison, partial differential equations can be so complicated that next to them ordinary differential equations look like child's play. These equations occupy a particular place in applied and pure mathematics. They are so important that they have developed into a main topic of research. Indeed, any progress on extracting information out of these equations holds the promise of solving some of the fundamental problems of science, engineering, and finance. It is fair to say that a very large part of quantitative science rests on the foundation and analysis of partial differential equations. Here as a first introduction to the wonders of partial differential equations, we will discuss two generic spatiotemporal behaviours.

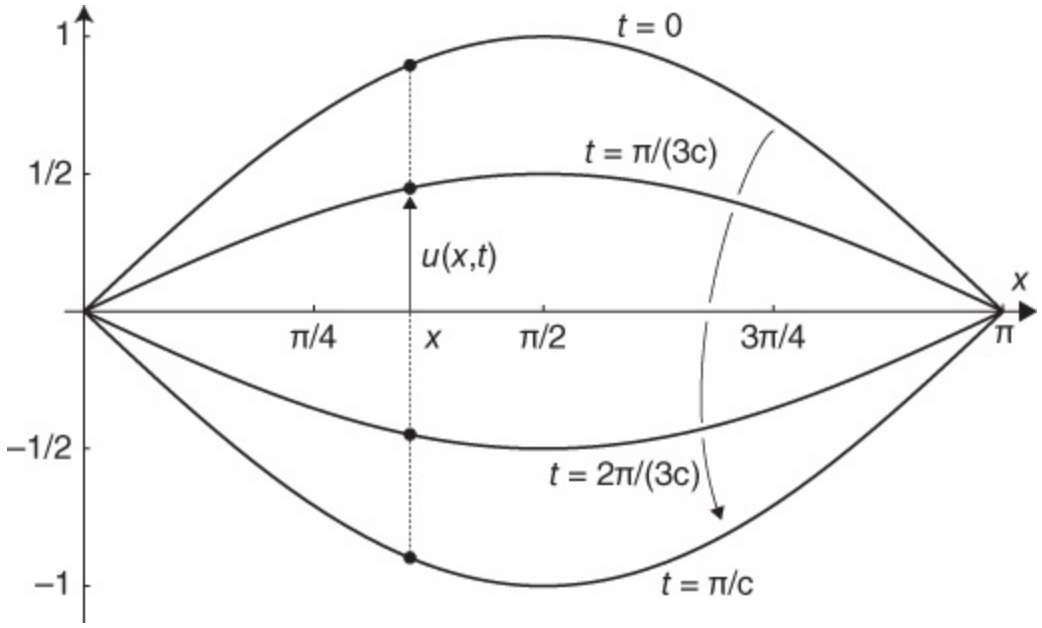
Linear waves on a string

Wave propagation is a generic phenomenon in physics. There are many types of waves: sound waves and seismic waves are examples of mechanical waves travelling through a medium via oscillations of nearby particles. Radio waves, mobile phone signals, X-rays, and visible light are electromagnetic waves that can travel through a vacuum. Large and rapid oscillations of masses in the universe lead to gravitational waves, which travel through space at the speed of light by changing the curvature of spacetime. In effect, wave motion can be understood as the transfer of a perturbation through space and is typically associated with oscillation of particles or a field with little mass transport. For instance, the ripples created by throwing a rock into a pond form a wave that propagates the initial disturbance. In this motion, most of the fluid oscillates up and down rather than moving with the wave.

A simple example of a wave is generated by plucking a taut guitar string. The string oscillates up and down, forming a standing wave since the string is attached on both ends. A problem for the mathematician is to describe the possible motion of the vibrating strings. The simplest motions, first discovered by the mathematician Joseph Sauveur at the end of the 17th century, are vertical oscillations known as the *fundamental mode* and its *harmonics*. In modern form, these oscillations are easily described by the vertical displacement $u(x, t)$ at time t of a point a distance x from one end on a string of length π. The formula for the fundamental mode in terms of u is

$$u(x,t) = \sin x \cos ct,$$

where c is the speed at which waves can travel along the string. Explicitly, c is equal to the square root of the tension in the string divided by the mass per unit length. As shown in Figure 14, the string oscillates in time and returns to its original position after a period $T = 2\pi/c$. Therefore, the frequency of the sound produced by the string is $\nu = 1/T = c/(2\pi)$. As guitar players know well, a larger tension leads to a larger velocity c, hence a higher frequency ν.



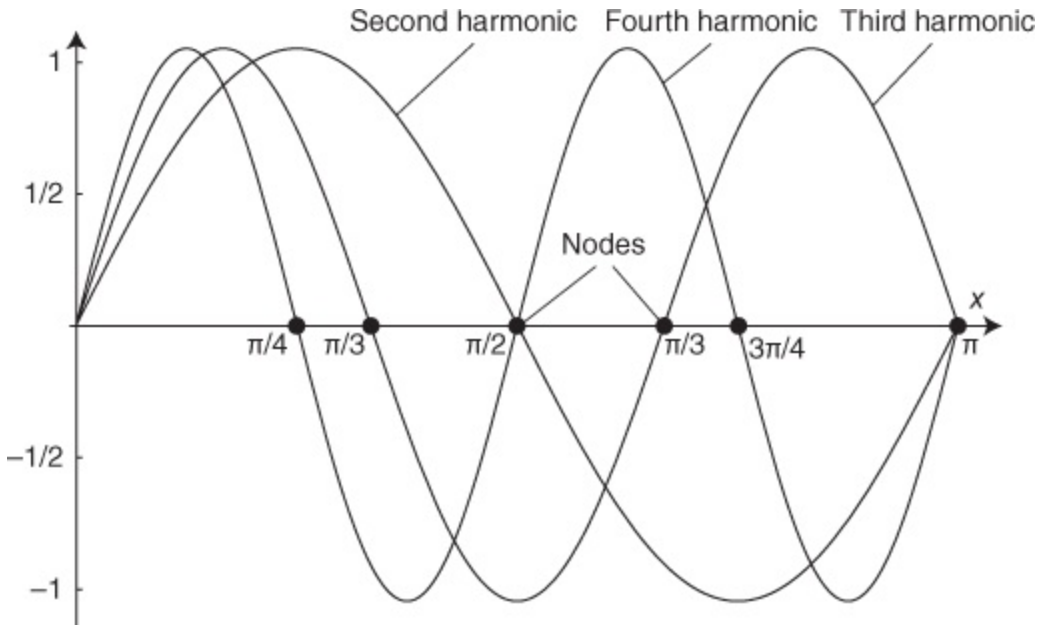
14. The fundamental mode of a vibrating string is a simple oscillation at a single frequency where the only points that remain fixed in time are the end points. A half period from $t = 0$ to $t = \pi/c$ is shown. In this motion a material point only moves vertically. The motion is then fully determined by the function $u(x, t)$ describing the height at time t of a material point found initially at position x.

The fundamental mode is not the only possible motion of the string. The harmonics of the strings are oscillations whose frequencies are multiple of the fundamental frequency v. They are:

$$
\begin{aligned}
u(x,t) &= \sin 2x \cos 2ct, && \text{the second } \textit{harmonic},\\
u(x,t) &= \sin 3x \cos 3ct, && \text{the third } \textit{harmonic},\\
&\vdots\\
u(x,t) &= \sin kx \cos kct, && \text{the } k\text{-th } \textit{harmonic}.\\
&\vdots
\end{aligned}
$$

We observe that the time structure of the motion is closely related to the spatial structure. Indeed, we see in Figure 15 that the second harmonic has one node between the end points (a *node* is a point that remains fixed during

the motion). The third harmonic has two nodes and the k-th harmonic has (k − 1) nodes.



15. A harmonic of the fundamental mode of a vibrating string is a solution with a single frequency that is an integer multiple of the fundamental frequency. The three first harmonics are shown (corresponding to k = 2, 3, 4). The k-th mode oscillates at a frequency kv and has (k − 1) nodes.

This initial mathematical insight took place at a time of tremendous musical development. While Bernoulli and Euler were busy understanding the combination of harmonics, Johann Sebastian Bach was also experimenting with harmonic modulation. In the *Well-Tempered Clavier*, Bach tuned the octave in twelve equal semitones so that they can be played in all keys. The problem of adding and subtracting harmonics was a central concern for mathematicians and musicians alike.

The motion of the fundamental mode and its harmonics requires that the initial shape has a very particular form. What happens if the initial shape of the string is not one of these modes? For instance, the string of a harpsichord, the main musical instrument of Bach's time, is plucked by a quill and the

initial shape of the string is triangular. How will the string move when released? What is the general formula for the motion? It took more than a hundred years as well as the greatest minds of the 18th century to fully answer these questions. As Bach was setting the basic rules for modern Western music, Euler, Bernoulli, and D'Alembert built the foundation of a theory of partial differential equations that revolutionized modern sciences.

In 1747, Jean Le Rond D'Alembert showed that the vertical displacement of the string, $u(x, t)$ is the solution of a partial differential equation, the *wave* equation:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad (5.1)$$

where we have introduced a special symbol for the partial derivative, the beloved curly 'd'. The partial derivative of a function is the derivative of a function with respect to an argument assuming all the other ones are constant. That is,

$\frac{\partial u}{\partial x}$ is the variation of u with respect to x at a fixed time,

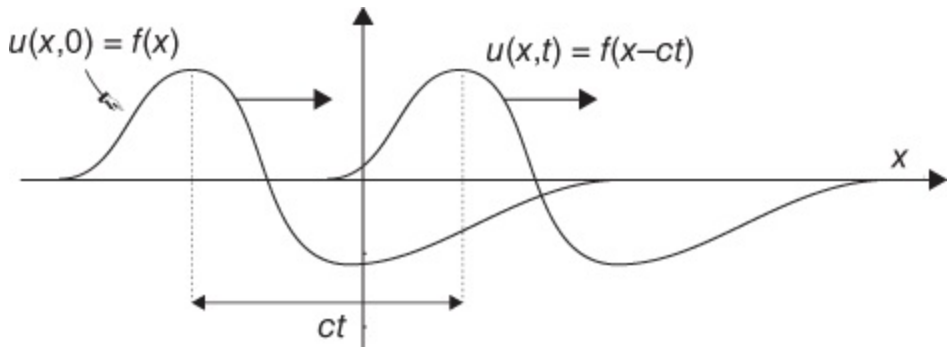
$\frac{\partial u}{\partial t}$ is the variation of u with respect to t at a fixed position.

The wave equation is actually a slightly disguised version of Newton's equation for a continuum. The left side of the equation represents the vertical acceleration of a material point on the string whereas the right side is the vertical force acting on that point by the stretched string. It is a linear equation in the sense that the variable u appears linearly in the equation (it appears by itself and not as a power or in a function). This property is fundamental and makes it possible to give a solution to the equation. Indeed D'Alembert showed that the general solution of the wave equation is

$$u(x,t) = f(x - ct) + g(x + ct),$$

where f and g are arbitrary functions. Physically, this solution can be

interpreted as follows: any solution of the wave equation can be understood as the superposition of two waves, one with a shape *f* going right and one with a shape *g* going left. An example with $g = 0$ is shown in Figure 16.



16. An example of D'Alembert's solution. A solution of the wave equation can be obtained by starting with any function $u(x, 0) = f(x)$ and translating it in time to obtain the solution $u(x, t) = f(x - ct)$.

This solution does not look at all like the fundamental and harmonic modes that we discussed previously. However, if we recall that $\sin(a + b) = \sin a \cos b + \sin b \cos a$, we can rewrite the harmonic mode as

$$\sin kx \times \cos ckt = \frac{1}{2}\left(\sin k(x - ct) + \sin k(x + ct)\right),$$

and we notice that the standing wave on the vibrating string is the sum of two travelling waves by choosing $f(x - ct) = (1/2) \sin k(x - ct)$ and $g(x + ct) = (1/2) \sin k(x + ct)$. D'Alembert's idea was then used by Euler to create solutions of the vibrating string problem when the initial shape is a triangle.

The second crucial result is that the general solution of the vibrating string can also be written as an infinite sum of harmonic modes with increasing mode number *k*. In 1753, Daniel Bernoulli proposed that the general solution be written as an infinite series

$$u(x,t) = a_1 \sin x \cos ct + a_2 \sin 2x \cos 2ct + a_3 \sin 3x \cos 3ct + \ldots,$$
$$= \sum_{k=1}^{\infty} a_k \sin kx \cos kct.$$

Each mode is a building block and by adding them judiciously, the general solutions can be constructed. This infinite sum of functions raised many deep mathematical questions at the time. When infinitely many numbers are added together, the result is either a number itself, meaning that the series is *convergent* (e.g $\sum_{i=1}^{\infty} n^{-2} = \pi^2 / 6$); or the series is *divergent* (e.g. $\sum_{i=0}^{N} 1 / n$ becomes larger and larger as *N* increases). Functions are much more complicated objects than numbers and it was not clear what happens when infinitely many of them are added together. The problem of evaluating the coefficients *ak* was eventually solved by Joseph Fourier in 1807 and a precise mathematical proof that these infinite series are indeed convergent was given by Peter Gustav Lejeune Dirichlet in 1829, more than 130 years after the initial work of Sauveur.

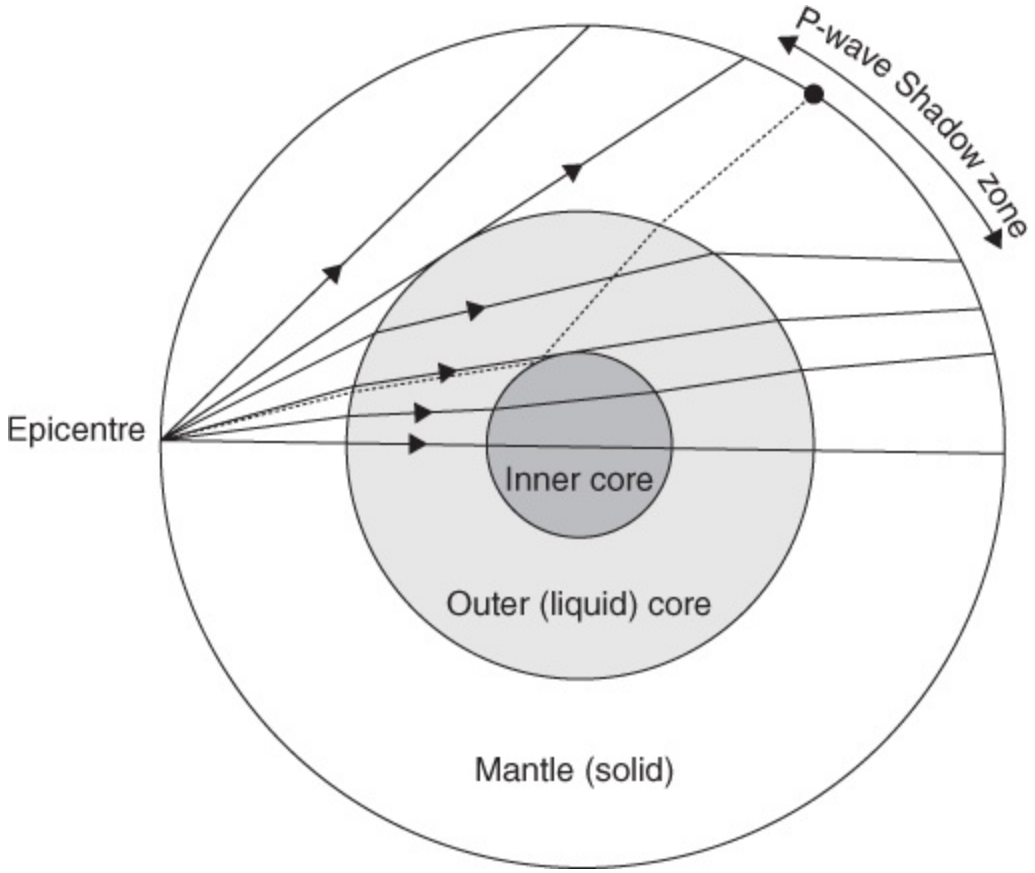
Unlike D'Alembert's solution, the method proposed by Fourier to solve a partial differential equation as an infinite sum of functions (known respectively as *Fourier series* and *Fourier modes*) is not restricted to the wave equation. It can also be used to solve many other linear problems arising in physics and is at the heart of many algorithms to solve partial differential equations numerically.

The wave equation for the vibrating string is the simplest model for wave propagation. Yet, most wave phenomena encountered in nature share common features with the vibrating string and the mathematical tools developed for the string can be readily exploited to understand waves in other contexts.

Earthquakes

A spectacular example of wave propagation occurs during an earthquake. The initial formidable release of energy at the epicentre propagates through the

Earth’s layers as seismic waves. In a three-dimensional body, vibrations can travel in different ways, each with a different velocity. The so-called *P-waves* or *primary waves* are pressure waves. Like sound waves, the displacement of the medium is in the same direction as the wave itself. These are the fastest seismic waves, travelling through water and rocks at velocities of up to several kilometres per second. The *S-waves* or *secondary waves* are transverse. The displacement is in a direction perpendicular to the motion. For instance, a wave on a string moves along the string in the horizontal direction but the displacement takes place in the vertical direction. Typically, these waves are about 60 per cent slower than the P-waves but they are very destructive because they induce shear (*shear* arises when forces acting on a body are unaligned, pushing one part of a body in one direction, and another part of the body in the opposite direction). These waves can be understood through the proper analysis of differential equations forming the basis of modern seismology. Their study is particularly useful for early warning systems, identifying the epicentre, and studying the composition of the Earth’s mantle (see Figure 17).



17. As seismic P-waves travel through the Earth's mantle, they are refracted at the interface between different materials. In the absence of an inner core, a simple tracing of rays predicts the existence of a shadow region where no waves are observed. The observation of P-waves by Inge Lehmann in 1936 in that shadow region (dashed line) indicated the existence of a solid inner core with different material property (her predictions were eventually confirmed in the 1970s). Note that actual P-waves are not straight, as drawn here, but bent due to the inhomogeneity of material properties within the Earth.

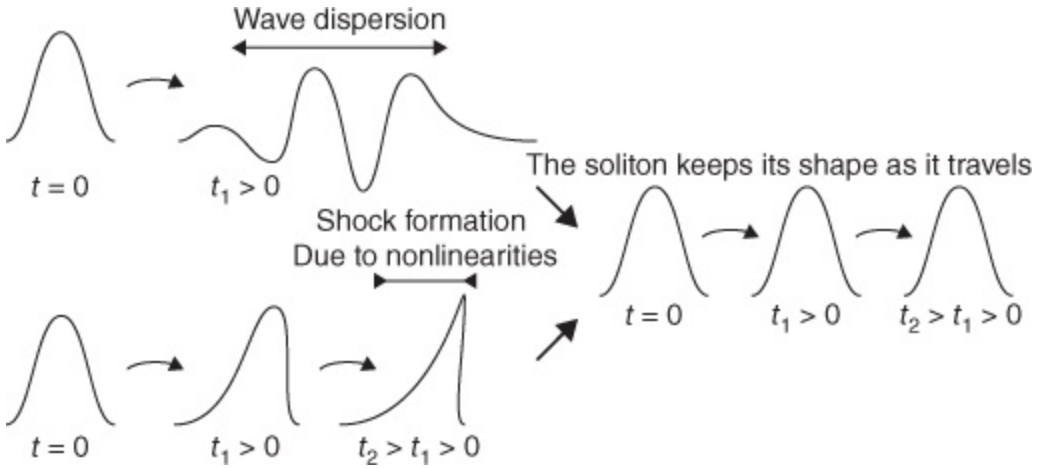
Nonlinear waves: the remarkable soliton

In the previous section, we described the behaviour of linear waves, that is, waves that are the solution of a linear wave equation such as Equation (5.1). Typically, the velocity of a single-frequency wave in a medium depends on

its frequency. In general, a wave is the superposition of multiple waves with different frequencies. For instance, sunlight is composed of many waves of different colours (i.e. different frequencies). When such a wave packet travels through a medium, each single-frequency wave travels at a different velocity and the packet changes its shape, the fastest wave pulling ahead of the other ones like the fastest horse in a race. This *dispersion phenomenon* is generic and, together with dissipation, would quickly alter the shape of any initial wave packet. It was therefore a great surprise to the Scottish engineer John Scott Russell when, in 1834, he observed a wave in a narrow boat channel

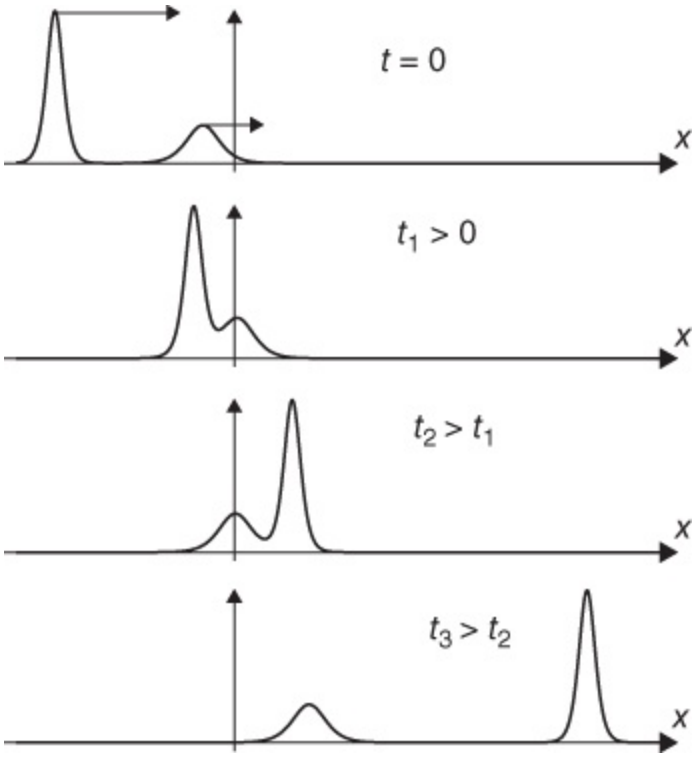
> apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some thirty feet long and a foot to a foot and a half in height. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phenomenon which I have called the Wave of Translation.

This unexpected behaviour does not fit with our understanding of linear waves, dispersion, and dissipation. Despite some work by Korteweg and de Vries at the end of the 19th century, this mysterious ‘Wave of Translation’ remained a bit of an oddity. The situation changed dramatically in 1965 when Norman Zabusky of Bell Labs and Martin Kruskal at Princeton University showed, through numerical analyses with the first digital computers, that these waves of translations occur in different systems. With a flair for good marketing, they rebranded the waves as ‘solitons’ to emphasize the particle-like nature of the phenomenon. They demonstrated that these solitons are produced through a combination of dispersion, tending to spread the wave-packet, and nonlinearity, tending to sharpen the wave-packet (see Figure 18). Under particular circumstances, these two effects balance exactly and a new type of wave is created: the soliton.



18. Typically waves feel the effect of dispersion (top) spreading a wave packet in time and/or wave breaking due to nonlinear effects creating a shock (bottom). The balance of these two effects leads to a soliton keeping its shape as it travels.

What is so special about solitons? The main spectacular property of solitons is that they behave very much like discrete particles: they are localized in space, they retain their shape as they move at constant velocity (in a homogeneous medium), and they interact with each other without destroying their shape. For instance, in Figure 19, we see the evolution of two solitons. After collision, the two solitons have exactly the same shape as they do before collision. This property remains true for N solitons (where N is any positive integer). Another remarkable property is that many of the equations that support soliton solutions can be solved exactly. We saw in Chapter 4 that algebraic equations of degree larger than four cannot be solved exactly, that ordinary differential equations can have chaotic solutions, and that partial differential equations are much more complicated than ordinary ones. The most complicated of all are nonlinear partial differential equations. Soliton equations are nonlinear partial differential equations, yet the great and beautiful surprise was that their solutions can be found. These equations have an extremely rich mathematical structure that has been explored for the last fifty years.



19. Two solitons interacting. The tall soliton is faster and catches up with the small one. After interaction, the two solitons recover their shape.

Aside from their mathematical beauty, it was soon realized that the particular behaviour of solitons could be harnessed. In particular, optical solitons can be understood as ‘light bullets’ that can propagate through optical fibres. Regular wave packets suffer from dispersion and when a signal is transmitted through a fibre, it has to be repeated periodically. Since solitons do not disperse and they can sustain their shapes over long distances, they can be used to send information (a single soliton being used in a digital communication as a single digit). In 1998, a team at France Télécom demonstrated that solitons could be used to create data transmission of 1 terabit per second (1012 bits per second). The full power of optical solitons for data transmission and optical switching holds great promise for a new generation of optical devices under investigation in the field of *nonlinear optics*, created in the 1960s.

Solitons are found in liquids, plasmas, condensed matter, and DNA. They have also been used to model *rogue waves* on the ocean. These freak or killer

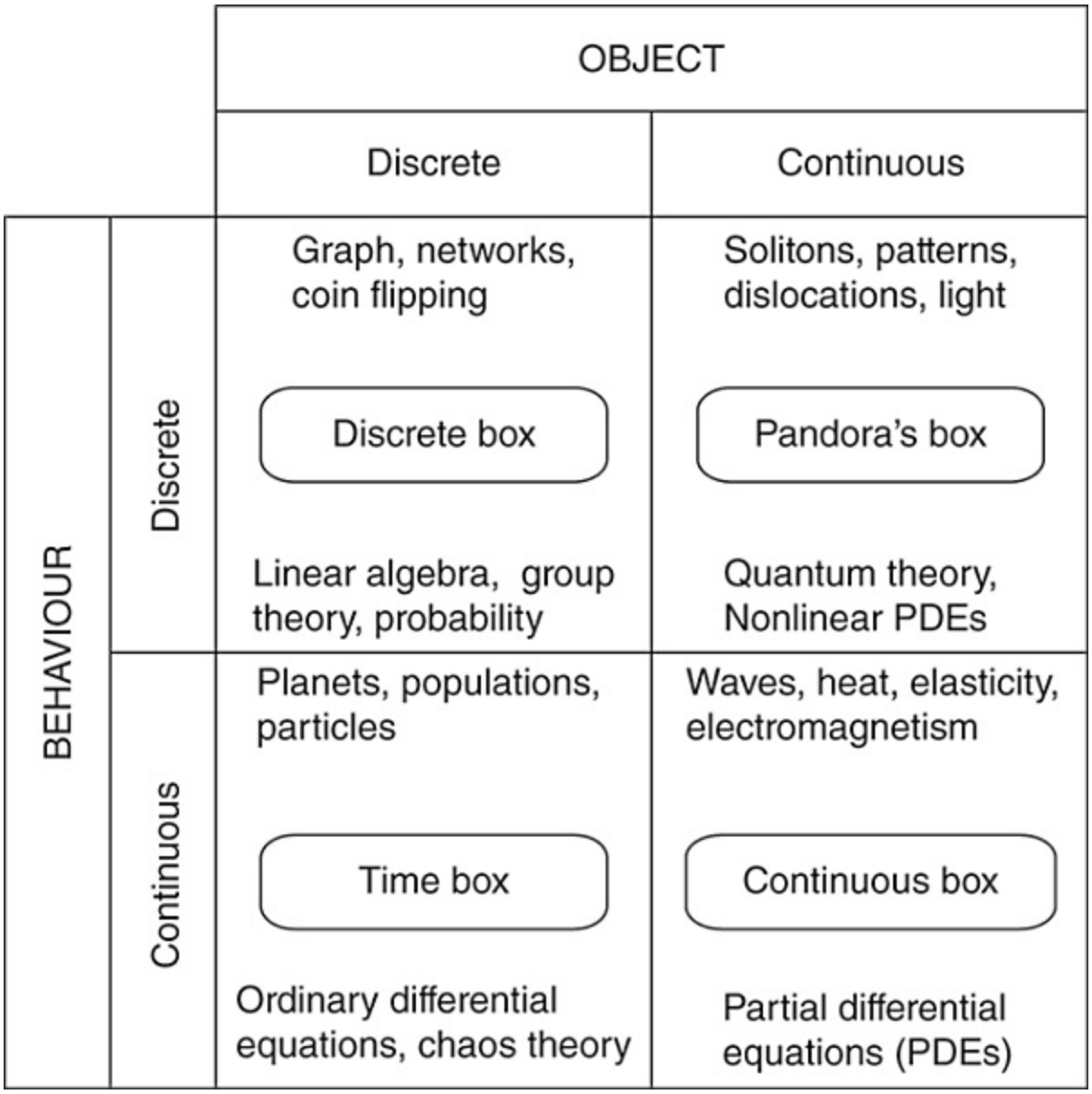
waves are large surface waves that can be extremely dangerous to ships. They can be greater than twice the size of surrounding waves and often come from directions other than that of the prevailing wind and other regular waves.

Reflection: Pandora's box

The discovery of solitons, and more generally nonlinear waves, created several new fields of study in mathematics, fluid dynamics, optics, and engineering. It is, undoubtedly, a fine example of applied mathematics. The discovery of nonlinear waves came through a combination of physical experiments, numerical analysis, and mathematical modelling. These results created a paradigm shift: once the initial structure was understood, the field rapidly expanded through mathematical and numerical analyses. It also influenced the design of new experiments and the modelling of other physical and biological systems.

A possible way to think about various theories of applied mathematics is to look at both the objects under study and their behaviour. Objects can be discrete (such as particles or planets) or continuous (such as pressure or temperature). Similarly, their evolution of behaviour can be discrete (as the sequences of heads or tails when flipping a coin) or continuous (as in the position of planets, or the motion of a piano string). In 1993, Christopher Zeeman, an influential and at times controversial British applied mathematician, proposed that mathematical theories could be classified by placing them in different boxes according to both their behaviour and the thing that they model, as shown in Figure 20. We already recognize a few boxes. The time box contains discrete objects that behave continuously in time and was investigated in Chapter 4. The continuous box is the classical field theories where continuous objects (fields) behave continuously in time. The interesting box for our discussion is Pandora's box. Here, we discover new objects whose behaviour is unexpected and often sparks controversy. The prime example in this box is quantum theory with its duality between (continuous) waves and (discrete) particles. But the box also contains more delightful surprises: everyday objects that should behave continuously according to the 19th-century theory, but actually behave like discrete particles. Solitons are one example of the treasures to be found here. They are

all beautiful and mathematically exciting and invariably useful.



20. A box classification of applied mathematics in terms of objects and their behaviour. Examples for each box are given (top) as well as the theories they generate (bottom).

Chapter 6

Can you picture that? X-rays, DNA, and photos

Well, who you gonna believe, me or your own eyes?

–Chico

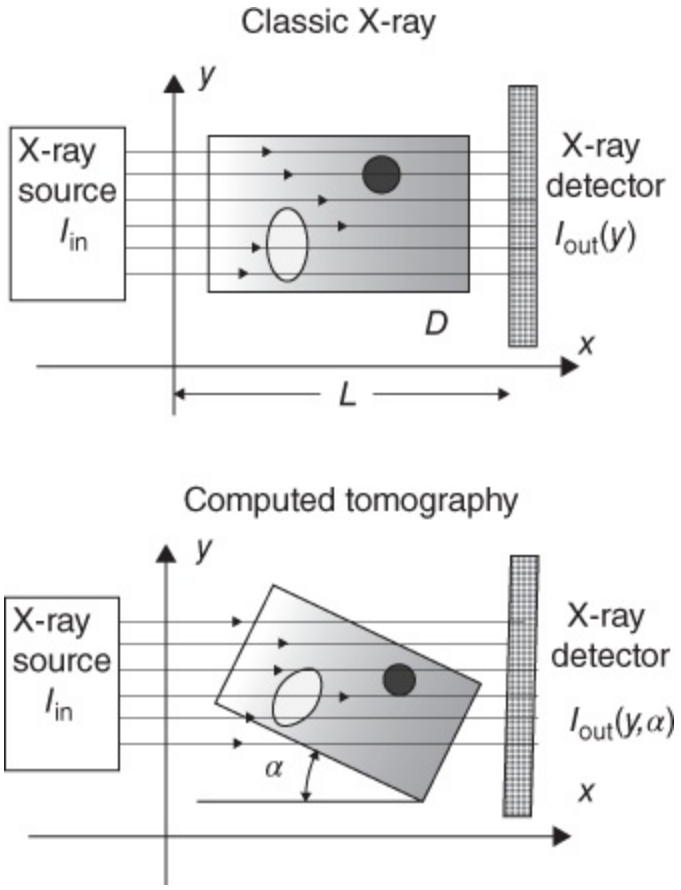
Applied mathematics is not just about models and equations. It is also concerned with the manipulation and analysis of signals and data. We live in a digital world where huge amounts of data are being routinely collected, transmitted, processed, analysed, compressed, encrypted, decrypted, and stored. For better or for worse, many aspects of modern life involve the handling of signals or data. None of it would happen without efficient mathematical methods and numerical algorithms.

Computer tomography and Radon transform

An important part of medicine is the ability to diagnose patients non-invasively. Since the invention of X-ray radiography in 1895 by the German physicist Wilhelm Röntgen, many different techniques have been developed in medical imaging to obtain images from organs and tissues. Most of them rely on a simple idea: a wave is sent through the body and captured by a

detector as it leaves. The interaction of the wave with the matter inside the body carries crucial diagnostic information about the way the matter is organized. In particular it gives the location and composition of the tissues and organs. The key mathematical question is how to extract this information from the data. In the simplest case high-energy waves are sent through the body and their intensity is measured as they exit the body. This is the familiar X-ray technique. The intensity of an exiting wave that penetrates through matter depends on the matter's atomic number, its density, and the distance travelled. In particular, in bones, it is the high amount of calcium that makes them appear clearly on X-rays.

In Figure 21, we see a highly idealized version of the problem. A monochromatic, planar wave I_{in} is sent through a sample occupying a domain D in the x–y plane. At each point (x,y) of the sample, we define an attenuation coefficient $\mu = \mu(x, y)$ that describes how much energy is lost by the wave as it moves through that point. The attenuation coefficient depends on both the material density at that point and its composition (e.g. a big block of lead would attenuate considerably an incoming wave). The intensity I_{out} of the outgoing wave is measured by a detector on the right. Theoretically, we know that a wave decays exponentially through a material. Explicitly, it is given by the Beer-Lambert law:



21. Traditional X-ray radiography. An X-ray wave is sent through a domain *D*. As it travels through the body, its intensity decays depending on the objects it encounters. The X-ray detector records this shadow. In computed tomography, the sample is rotated and X-rays at various angles *α* are recorded.

$$I_{\text{out}}(y) = I_{\text{in}} \exp\left(-\int_0^L \mu(x,y)\mathrm{d}x \right),$$

where L is the distance of the detector from the *y-axis*, and we assume that $\mu(x, y) = 0$ everywhere outside the domain D. Recalling that integral sign ∫ appearing in this expression is simply a sum on small elements, this expression reflects the accumulated effect of the decay as the wave goes through the material. If the material is homogeneous, μ is constant in the domain and we have $I_{\text{out}} = I_{\text{in}} \exp(-\mu l)$. In other words, the waves decay

exponentially with the length l of the domain. In general, the material is not homogeneous and the intensity I_{out} depends on the position y. As in Plato's cave allegory, the detector only records a shadow of the object which limits our ability to grasp the reality of the internal structure. However, if we rotate the sample and take a series of X-rays at different angles α, we obtain more information about the sample:

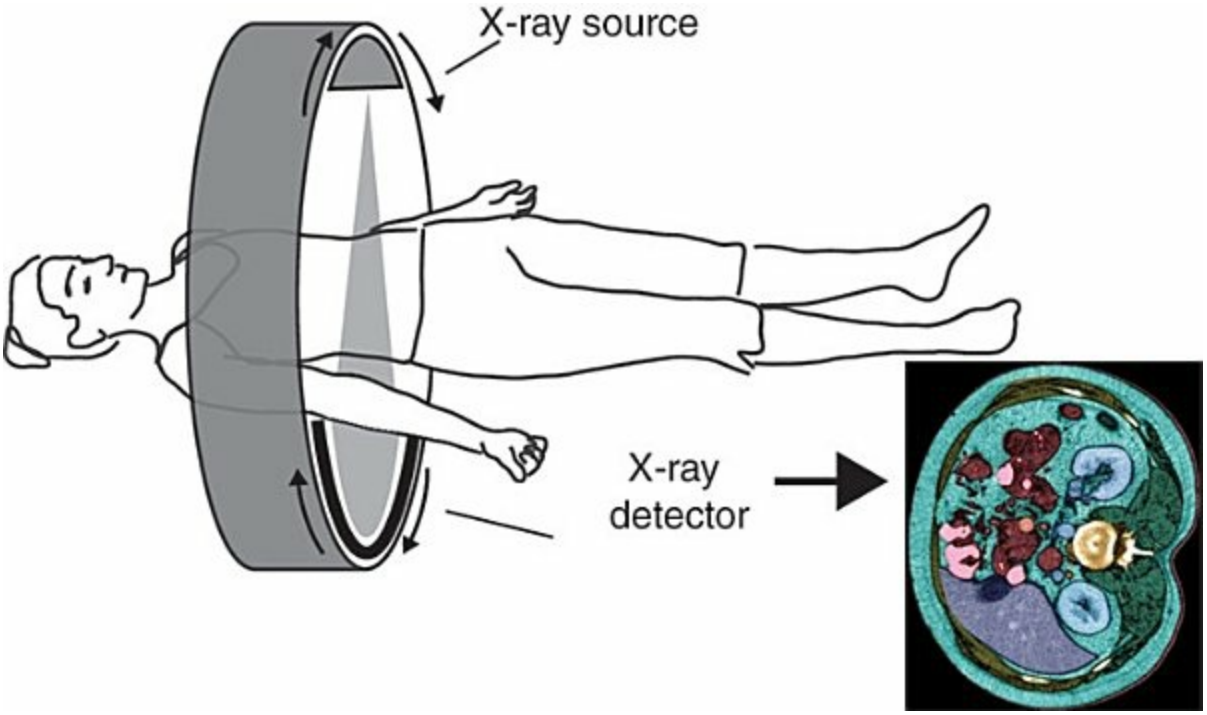
$$I_{out}(y,\alpha) = I_{in} \exp\left(-\int_0^l \mu(x,y,\alpha)\mathrm{d}x \right),$$

where $\mu(x, y, \alpha)$ is the attenuation coefficient of the same object but rotated by an angle α. The details of these mathematical expressions may appear difficult but the structure of the problem is actually quite simple. To see it appear clearly, we write this transformation in a compact form by calling $\boldsymbol{T}$ the mathematical operation that produces the function I_{out} from μ:

$$I_{out} = \mathcal{T}(\mu).$$

In principle, if I know the attenuation coefficient μ for a material, then I can plug the formula in a computer and I know how a wave intensity I_{out} would be affected by going through the sample. However, in an experiment, I measure I_{out} and I want to know μ. Can I reconstruct an object if I have many shadows of it? At first sight, it appears that I have not made much progress in solving the problem. But at least, the problem can now be formulated mathematically: I want to solve this equation for $\mu(x,y)$ as a function of the data $I_{out}(y,\alpha)$, that is to invert the relation $I_{out} = T(\mu)$. This is exactly the problem that the physicist Allan Cormack and the engineer Godfrey Hounsfield solved in the 1960s when they developed computed tomography for physiological use (see Figure 22 where the X-ray source and detector rotate around a patient). The mathematical method that Cormack uses in his 1963 paper relies on Fourier series, the same method that was developed to solve partial differential equations as discussed in Chapter 5. Fourteen years after he proposed this method, Cormack realized that the problem had been solved completely by the Austrian mathematician Johann Radon in 1917. The Radon transform R is a mathematical operation that takes $I_{out}(y, \alpha)$ as an input

and produces $\mu(x, y)$:



22. Computed tomography (CT). A CT scan is obtained by rotating both an X-ray source and an X-ray detector to obtain X-rays at multiple angles.

$$\mu = \mathcal{R}(I_{\text{out}}).$$

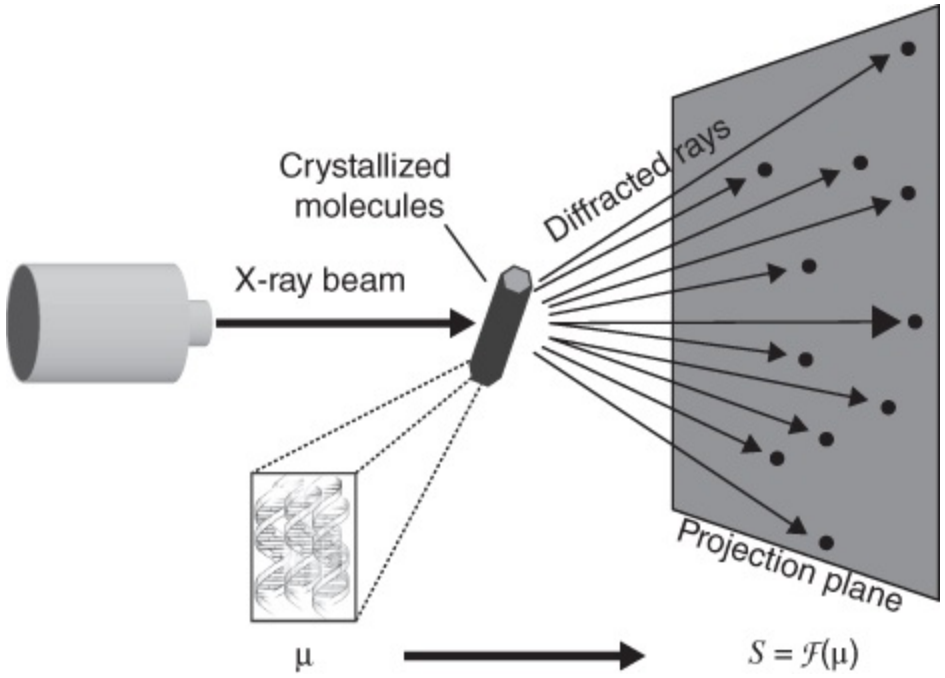
Mathematically, the transformation $\mathcal{R}$ is in the *inverse* of the transformation T in the sense that $\mathcal{R}(\mathcal{T}(\mu)) = \mu$. The implication for Plato's grotto is that a single X-ray shadow only gives us partial information, but multiple X-rays can be used to provide a perfect reconstruction of reality.

Most medical imaging is based on similar ideas: a number of data sets are recorded by sending waves through the body and the problem is to *invert* this data to obtain a three-dimensional image of the internal structure. In practice, the data is not a function but a series of discrete values given by each detector at a number of angles. This data is typically noisy and to date there is a large collaborative community of mathematicians, computer scientists, physicists, and medical doctors addressing these issues. At the mathematical level, the

problem is to develop clever methods and numerical algorithms to extract the information needed from partial, noisy data.

The discovery of the structure of DNA

X-rays can also be used to probe the atomic structure of macromolecules. Again, the basic idea is to send a beam of X-rays into a crystal made of repeated units of the same molecules. At the atomic scale, diffraction occurs between the incoming wave and the atoms, changing the wave direction. The multiple diffracted waves in the crystal interact with each other (see Figure 23). These interactions can be constructive (reinforcing the signal) or destructive (attenuating the signal) resulting in an image with a series of spots. These diffraction patterns give an indication of the general structure of the object. The mathematical challenge is then to decode the structure from these spots.



23. X-ray crystallography. The diffraction pattern is obtained by shining a beam of X-rays into a crystal made of identical molecules. The interaction between the outgoing waves depends on the molecular structure of the crystal and creates the diffraction pattern with

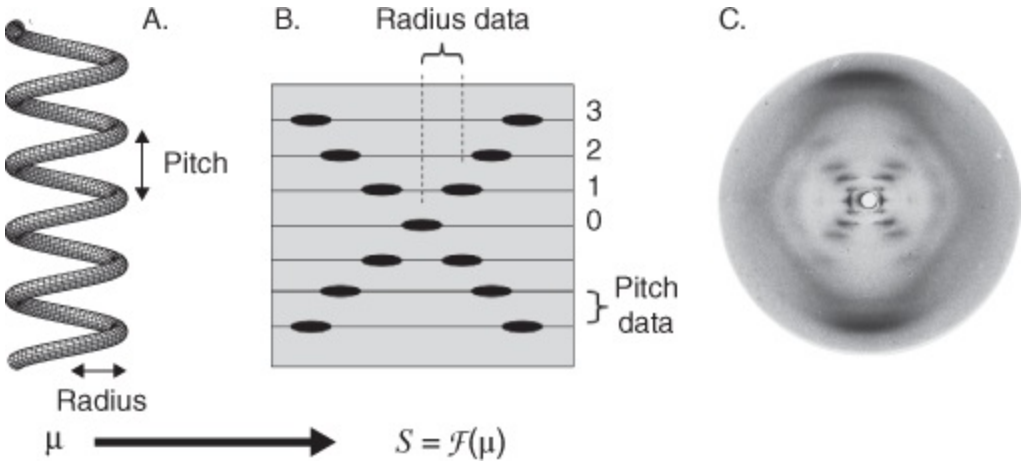
characteristic spots appearing on the projection screen.

The structure of a molecule can be described mathematically as a spatial density $\mu(x, y, z)$ representing the location of matter at a point (x, y, z) in space; in the simplest case, $\mu(x, y, z) = 1$ at points where there is an atom and 0 everywhere else. Then, the diffraction pattern can be expressed as a transformation of this distribution. Using a compact notation, we can write

$$S = \mathcal{F}(\mu),$$

where S is the intensity of the signal on the projection plane. The transformation $\mathcal{F}$ is different than the one obtained for the problem of computed tomography but, again, is based on Fourier's theory. The X-ray diffraction pattern is a two-dimensional projection of a three-dimensional object. Therefore, it cannot be inverted (i.e. μ cannot be obtained in terms of S). This is where the art and craft of biochemistry comes into play.

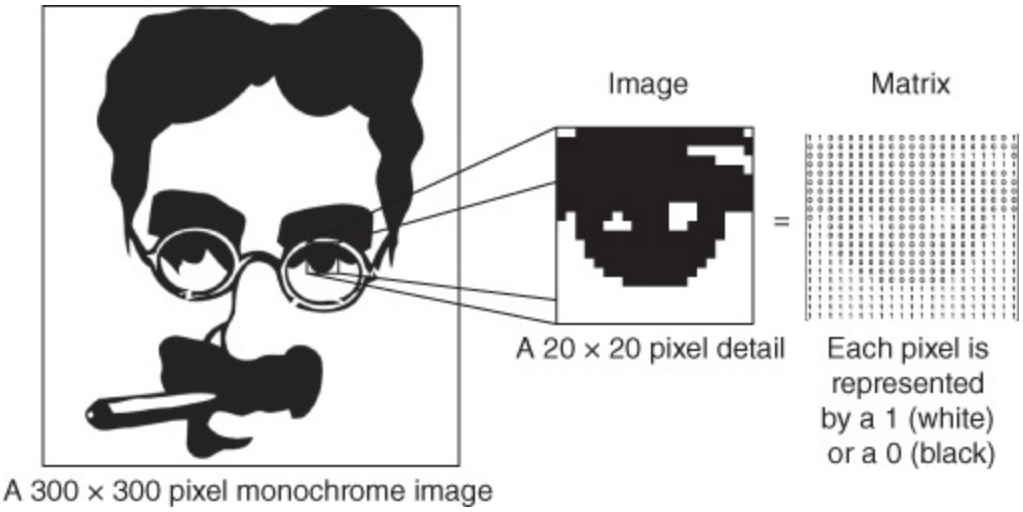
In 1952, Rosalind Franklin and her student Raymond Gosling produced a series of diffraction patterns for DNA (Deoxyribonucleic acid) shown in Figure 24. These photographs showed a characteristic X-shape. By then, several scientists had studied the structure of proteins such as keratin (the main fibre found in hair and nails). In particular, Francis Crick and his colleagues had computed theoretically the transform $\mathcal{F}(\mu)$ when μ represents a helix or a double helix. It is the mathematical understanding of Fourier transforms of helical structures that was crucial in the discovery of DNA. The rest of the story has already entered the realm of scientific legend: when Francis Crick and James Watson were shown the DNA diffraction patterns by Maurice Wilkins, they instantly recognized the structure as a double helix. Based on that information, a great deal of ingenuity, and further help from Franklin, they built the first model of DNA, one of the greatest discoveries of the 20th century. As a result, Watson, Crick, and Wilkins were awarded the 1962 Nobel Prize in Physiology or Medicine 'for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material'. Sadly, Franklin had died five years earlier.



24. X-ray crystallography of helices and DNA. The theoretical diffraction pattern of a helical structure (A) has the characteristic X-shape with a number of spots of decaying strengths (B). The radius and pitch of the helix can be obtained by measuring the vertical and horizontal distances between these spots. (C): the famous photograph 51 showing the X-ray diffraction pattern of B-DNA obtained by Franklin and Gosling in May 1952.

Digital images and compression

Another important use of mathematical methods is in the handling of digital data. One can think of data as strings of numbers. For instance, the digital black-and-white picture shown in Figure 25 is an array of 300 × 300 pixels. Each pixel is represented in binary by a 1 or a 0. In mathematics, these arrays are called *matrices* and any digital operation on an image is done by manipulating the numbers in these matrices. For instance, inverting from black and white to white and black can be done simply by creating a new matrix where each original 1 is replaced by a 0 and each original 0 is replaced by a 1.



25. Digital images are arrays of numbers. In the simplest case, each pixel is represented by one bit (0 or 1).

Similarly, when a digital camera takes a picture, its sensor captures levels of light intensity at different frequencies (for instance, red, green, and blue), and produces three numbers at each pixel (RGB). Typically, the intensity of each pixel is represented by a 12 or 14 bit number (that is a number between 0 and 2^{12} or 2^{14}). For instance, a 12 megapixel image in the format 4:3 is a 4,000 (wide) by 3,000 (high) array of pixels. A single digital image of this type is therefore made of three matrices (one for each colour), each containing twelve million entries. For most applications, such as viewing or posting on social media, these files are too large and contain too much information. The problem is then to reduce the size of the image while keeping most of the relevant information. While it is possible to compress a file without losing any data, most methods remove data. One of the most common compression formats is the so-called JPEG format (developed by the Joint Photographic Experts Group in 1992), with the typical file extension '.jpg'. During compression, some of the information is thrown away and the challenge is to find a balance between storage size and image quality. A typical compression factor without noticeable loss of image quality is 10:1, meaning that the JPEG files are ten times smaller than the original ones. How is this achieved?

The key mathematical idea behind this type of compression is that a signal can be decomposed into different modes at different frequencies as proposed

by Fourier. By analogy, consider the sound created by a vibrating string. We know from our analysis in Chapter 5 that a sound can be decomposed into different frequencies, with higher frequencies giving higher pitches. As discussed, the sound in the air is directly related to the motion of the string. If u is the displacement of the string, its amplitude u can be written in terms of the Fourier series

$$u(x,t) = \sum_{k=1}^{\infty} a_k \sin(kx)\cos(kct).$$

A digital recorder captures a signal by taking its value at different times, creating a large sequence of numbers $u_1 = u(x, t_1)$, $u_2 = u(x, t_2)$, …. Fourier analysis tells us that this audio signal can also be represented equivalently by the amplitude of infinitely many modes a_k. However, the human ear cannot perceive very low or very high frequencies (typically, the ear's dynamic range is between 20Hz and 20kHz). Therefore, the high frequency modes can be thrown away without any perceptible loss of information (essentially getting rid of the part of the signal that cannot be heard). To achieve data compression, long and complex signals can be chopped into small time intervals. At each time interval, the amplitudes of a few important modes are computed and recorded. Essentially, a sheet of music for a piano piece is a form of data compression. Rather than giving a recording (in time) of the sound, it prescribes on small time intervals (given by the tempo) which frequencies (that is, which combination of notes) should be played. The same idea is at work in the compression of digital images. Each line of an image can be seen as a signal, the intensity of the signal as a function of the position. Just like an audio signal can be decomposed into time frequencies, this signal can be decomposed into spatial frequencies. High spatial frequencies correspond to fast variations of black and white and small spatial frequencies are associated with slow changes along the line. The human eye cannot appreciate very high spatial frequencies in small sections of an image. The JPEG algorithm decomposes the big matrix that makes up the image into small eight by eight pixel blocks. After some initial manipulations, each block is decomposed into its spatial (horizontal and vertical) frequencies, creating a new frequency matrix. Then, by dividing the amplitude of each entry in that new frequency matrix by a constant Q and rounding it off to the

nearest integer, amplitudes that are small are replaced by the number zero. Keeping only the non-zero components reduces greatly the amount of information that characterizes the image. This information is then further encoded for easy transmission and processing. In practice, psychophysical studies have shown that little perceptual distortion occurs when removing high frequencies. Therefore, a different number Q is chosen for each frequency. Increasing the constant Q for high frequencies sends more entries to 0, further improving compression as shown in Figure 26. In broad terms, this procedure removes high spatial frequencies.

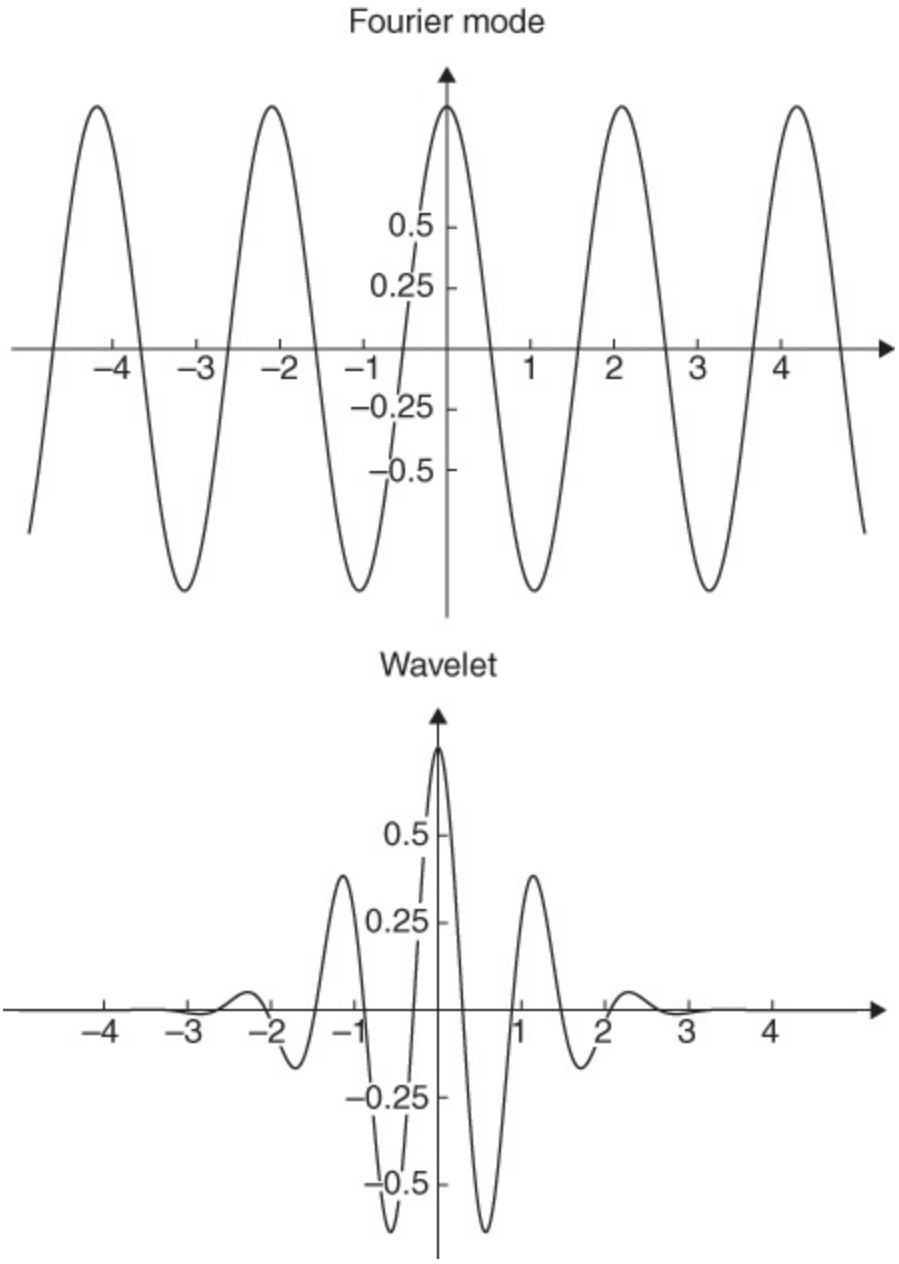
26. JPEG compression. The original image is 1,200 × 1,200 pixels. The

grey amplitude in each pixel is stored by an 8 bits = 1 byte number. The total size of the image is 1.44 megabytes. For small values of *Q*, large compression is achieved without any noticeable loss in image quality. For larger values of *Q*, the gain in compression is smaller and the images show reduction in quality.

Modern developments

The key mathematical idea behind data treatment is that data has an internal structure. This structure depends on the origin of the signal or information and, once understood, it can be used to manipulate the data to one's advantage. In the past fifty years, our understanding of data through mathematics has greatly improved as applied mathematicians have discovered new methods to extract information and process it efficiently. I will briefly give two examples.

First, the theory of wavelets (Figure 27). The fundamental idea behind Fourier's theory is that a signal can be represented as a sum of periodic modes with increasing frequencies. This decomposition works particularly well for signals that vary smoothly (in time or space). These signals can be efficiently described by a few frequencies. However, many signals do not have this structure. Instead, the information can be sporadic and localized. For instance, a jump in a picture from a white region to a black region is highly localized in space and requires many Fourier modes to be accurately represented. Similarly, the signal produced by an earthquake comes in random bursts and is localized in both space and time. The idea behind wavelet theory is that this type of signal can be represented by a superposition of functions that are also localized, hence capturing the key properties of the signal. These functions, called *wavelets*, were introduced initially by geophysicists to analyse seismographs. But their full potential was only understood in the 1980s when a complete mathematical theory was developed. Wavelets are short wavelike functions that can be scaled and translated.



27. Wavelets. In classical Fourier theory, a signal is represented as the sum of periodic modes with different frequencies. In wavelet theory, a signal is represented as series of wavelets that are localized in time. The particular wavelet shown here is a Morlet wavelet but many other functions are used for the same purpose.

Wavelets are particularly good for extracting features at different scales. They are used to recover weak signals from noise and are especially useful in

the processing of X-ray images in medical applications. They are also used in image compression in the JPEG2000 format and typically improve compression over JPEG by a factor of at least four (see Figure 28). Compression methods based on wavelets are particularly good on noisy images, sharp edges, and fine-scale structures. For instance the FBI use both JPEG2000 and Wavelet Scalar Quantization Compression algorithms for their database of about thirty million fingerprints.

28. Wavelet compression. The format JPEG2000 incorporates compression algorithms based on wavelets. A comparison between the traditional JPEG and JPEG2000 with files of the same size clearly shows the superiority of this new method.

A second major breakthrough in data handling came in the early 2000s with the invention of *compressed sensing* or *compressive sampling* through the unlikely collaboration of the statisticians Emmanuel Candès and David Donoho with the pure mathematician (and Fields medal winner) Terence Tao, known for his work in number theory. We saw that one of the central problems in medical imaging, and more generally in all areas of signal processing, is how to reconstruct a signal from partial and noisy measurements. Better reconstruction techniques allow the reduction of the number of samples needed, which in medical imaging implies a shorter scanning time. The classical approach to data acquisition is based on the

Nyquist/Shannon theory. In broad terms, this theory states that to obtain perfect reconstruction of the signal, the sampling rate (how often you acquire a signal) must be at least twice the maximum frequency of the signal. Compressed sensing is a different approach to both data acquisition and data reconstruction. It relies on the property that most signals are sparse. *Sparsity* is a property of the information content in a signal. A signal is *sparse* if the actual information content in the signal is much lower than the full signal itself. If a signal, such as an image, can be represented mathematically by a matrix of 1s and 0s, sparsity means that a large proportion of the entries are 0. The JPEG2000 format uses this idea and a wavelet basis to represent a signal with few components; written in terms of wavelets, the signal is sparse (hence highly compressible). The way most digital cameras work is that they first acquire the complete data set (three colour channels on every pixel), then compress the signal and store a fraction of it. Why spend all that time and energy acquiring information to eventually throw 90 per cent of it away? The central idea of compressed sensing is that, assuming sparsity, it is possible to acquire the data in compressed forms so that one does not need to throw away anything.

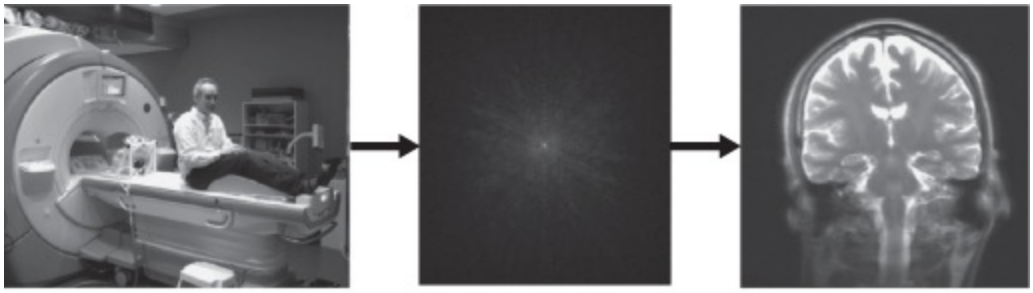
Mathematically, the problem of reconstructing a signal from sparse data can be understood as a problem of matrices and vectors. We assume that the full signal of interest (a sound or an image for instance) is written as a vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ where each component denotes the amplitude of the signal sampled at a given time. But if the full information is not known, we do not know all the components of **x**. Assuming a *linear measurement model*, the recorded information is not **x** but the vector **b** given by the application of a matrix *A*. That is,

$$\mathbf{b} = A\mathbf{x}.$$

Here, **b** is the partially recorded information. The length of this vector is much smaller than **x**. The central mathematical question is then to reconstruct the full information **x** from the partial information **b**. It is again a problem of *inversion*. We know from elementary mathematics that this problem cannot be solved exactly (in general). However, if the signal **x** is sparse (that is, if in a well-chosen representation, it mostly contains 0s), efficient numerical methods can be used to extract most of the information and an excellent

approximation can be found: a good guess that is very close to the exact solution. These ideas can be formalized mathematically and rigorous estimates can be obtained to describe how close the approximate solution is to the original one. In practice, this method can be used to reconstruct an image from noisy or partial data. More importantly, the understanding gained from this theory can be used to design a sampling protocol: one only needs a number of measurements proportional to the compressed size of the signal, rather than the uncompressed size. Therefore, hardware and software technologies can be adapted to create a suitable matrix *A*. Based on the analysis of this matrix, it can be decided what information should be recorded so as to increase the quality of the final reconstruction.

Within a few years after the initial theoretical work, these groundbreaking ideas have been quickly implemented in medical imaging. For instance, magnetic resonance imaging (MRI) scanners use this technology to shorten scanning time without losing image quality, as shown in Figure 29.



29. Compressed sensing is used in MRI scanners. For full three-dimensional reconstructions of the brain,multipleMRI scans are needed. To shorten the scanning time, compressed sensing technology is used to only acquire the data necessary to reconstruct individual brain slices (shown here: the author and a slice of his brain).

Reflection

At the mathematical level, the theories of compressed sensing, JPEG compression, and wavelets are based on functions and calculus. Yet,

computers only handle discrete information. Therefore, the algorithms built on these methods rely on a series of manipulations of matrices. The general mathematical theory that deals with the properties and manipulation of matrices is *linear algebra*. It has, over the years, become the central tool for the development of new methods and clever algorithms, and lies at the heart of all the digital devices around us.

The history of these groundbreaking technologies shows an interesting pattern. Initially, a new idea is proposed in an applied field: Fourier tried to solve heat problems, the wavelets were first introduced to understand seismographs, and Candès was trying to de-noise images using tools from statistics. Then, the mathematical properties of these methods are studied and, eventually, their general mathematical structure comes to light. Once their full potential is revealed through mathematics, these ideas then become routine and find their ways into a myriad of applications and technologies.

For their work on computed tomography, Cormack and Hounsfield received the 1979 Nobel Prize in Physiology or Medicine. In his address Cormack reflected on his work:

> What is the use of these results? The answer is that I don't know. They will almost certainly produce some theorems in the theory of partial differential equations, and some of them may find application in imaging with MRI or ultrasound, but that is by no means certain. It is also beside the point. Quinto and I are studying these topics because they are interesting in their own right as mathematical problems, and that is what science is all about.

Chapter 7

Mathematics, what is it good for? Quaternions, knots, and more DNA

Very interesting theory—it makes no sense at all.

–Groucho (attributed)

Modern mathematics is an ever-expanding field. Like the tower of Babel, it is continuously reaching for higher skies by pushing the boundary of human knowledge. Yet, it has reached such a point of extreme sophistication that one may question whether it has become irrelevant for the world in which we live and whether even its own language has become esoteric. It is important to emphasize that the goal of mathematics is not to be relevant or even accessible. We accept the basic principle that understanding the motion of galaxies or the constituents of matter is a worthwhile human activity that society should support despite the fact that it does little to increase directly the gross domestic product of any country. Similarly, mathematics is an important human enterprise that should be supported in itself for knowledge's sake. Yet, that argument by itself would be a poor justification for mathematics and might confine it to some obscure corner of academia, especially given that it has its own language which can exclude those who do not speak it. Mathematics has an important place in society, but one that is not always obvious. Paraphrasing William Cowper, mathematics moves in

mysterious ways, its wonders to perform. By its very nature, it is an abstract branch of knowledge and while many of its disciplines have earthly roots, its abstract form has always appeared disconnected from ordinary human pursuits and other sciences. One could argue that, as far as engineering and sciences are concerned, the mathematical knowledge from the end of the 19th century should be enough. Is there really a need for new mathematics? Yes, in many unpredictable ways, yesterday's mathematics is today's science, and today's mathematics is tomorrow's technology and science.

Quaternions

The history of arithmetic and numbers is a good place to start exploring the unreasonable applicability of mathematics. In *Arithmetica*, Diophantus of Alexandria tried to solve the equation (written in modern terminology), $4 = 4x + 20$. We know that the solution of this equation is $x = -4$, but Diophantus, realizing that this equation would give a negative number, calls this result absurd. Clearly, there was no reason to think at the time that negative numbers would be of any use in solving practical problems. Yet, mathematicians have systematically extended abstract notions guided by simple principles of self-consistency and intrinsic harmony.

Negative numbers are not just an abstract construct that helps us solve equations. The reason we teach them at school at an early age is that they are fundamental to our ability to quantify the world around us. Once negative numbers are accepted as a concept, it is easy to show that the equation $ax + b = 0$ has a solution for any a a non-vanishing integer and b any integer. These solutions are rational numbers (the ratio of two integers) of the form $x = -b/a$.

Following the same construction, one can look for multiple solutions of equations and realize that an equation such as $x^2 = 4$ has two solutions (2 and –2) despite the fact that the –2 solution does not have a direct geometric interpretation. It just makes good sense to enlarge the notion of admissible solutions rather than rejecting 'absurd', 'fictitious', or 'impossible' solutions.

What about $x^2 = 2$? This equation caused a lot of trouble for the Pythagoreans

and forced them to introduce the notion of irrational numbers (i.e. numbers that cannot be written as the ratio of two integers, such as $\sqrt{2}$). These numbers are now well understood and are an essential part of basic mathematics.

The next existential crisis in number construction came with the solution of the equation $x^2 = -1$. Again, we have a choice. Either, we declare that this equation has no solution because, obviously, there is no real number whose square is a negative number. Or, we generalize our definition of numbers to accommodate a new number. While they were first studied by Girolamo Cardano in 1545, it was René Descartes, a century later who called such solutions *imaginary* numbers. It took another few centuries and the genius of mathematical giants such as Euler, Gauss, and Cauchy to provide a clear mathematical basis and terminology for *complex* numbers.

With the help of the notation for the unit imaginary number $\mathrm{i} = \sqrt{-1}$, and the simple rule $\mathrm{i} \times \mathrm{i} = -1$, any complex number can be written $z = a + \mathrm{i}b$, where a and b are real numbers (the real and imaginary part of z, respectively). The introduction of complex numbers in mathematics had far-reaching consequences. Soon, it was realized that functions of complex numbers can be defined, such as the celebrated exponential related to the usual trigonometric functions through Euler's formula

$$e^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta.$$

Hidden in this one formula are most of the trigonometric identities that we learn at school. Also, when evaluated at $\theta = \pi$, it reads simply

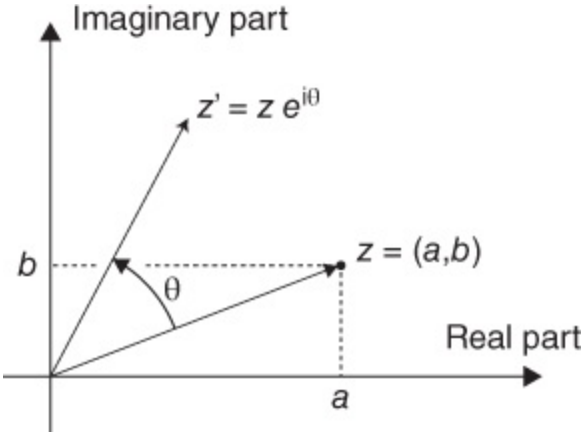
$$e^{\mathrm{i}\pi} = -1.$$

This compact and elegant identity is often cited as the ultimate example of mathematical beauty as it links the most basic mathematical numbers: π, e, -1, and i in an unexpected way.

Complex numbers are also immensely useful in applied mathematics. For instance, the solutions of the wave and diffusion equations as well as the

Fourier and Radon transforms, seen in the previous chapters, are all formulated compactly using complex analysis. What appeared at some point in the past as a figment of a mathematician's imagination turned out to be an essential pillar of all pure and applied mathematics.

One of the reasons complex numbers are so useful is that they are naturally connected to geometry in the plane. Each complex number $z = a + ib$ is associated with a point of coordinates (a, b) in the complex plane. Arithmetic operations on complex numbers correspond to operations on vectors in the plane. For instance, one can check that the sum of two complex numbers is equivalent to the sum of two vectors. More importantly, rotation in the plane can be expressed as a product of complex numbers as shown in Figure 30.



30. Complex numbers are naturally associated with geometry in the plane. Any complex number can be represented by a point on the plane by taking its real and imaginary parts as coordinates. A rotation in the plane of an angle θ is given by the multiplication with the complex exponential $e^{i\theta}$.

If complex numbers can be used to describe rotation in two dimensions, a natural question for a mathematician is: what is the natural structure or set of numbers to describe rotation in three dimensions? This is exactly the question that the Irish mathematician William Rowan Hamilton was pondering during a walk, near the Brougham Bridge in Dublin in 1843 when he had a flash of inspiration. He quickly carved with a knife his idea on the side of the bridge,

a prime example of mathematical vandalism. Hamilton's key realization was that complex numbers can be generalized. To do so, Hamilton introduced *quaternions*. Quaternions are numbers that are written with a quadruple (*a*, *b*, *c*, *d*) of real numbers such that

$$q = a + \mathrm{i}b + \mathrm{j}c + \mathrm{k}d,$$

where the symbols i, j, k satisfy rules inspired by the complex numbers:

$$\mathrm{i}\times\mathrm{i} = \mathrm{j}\times\mathrm{j} = \mathrm{k}\times\mathrm{k} = \mathrm{i}\times\mathrm{j}\times\mathrm{k} = -1.$$

A important new feature of quaternions is that the multiplication of different numbers depends on the order, that is, we have

$$\mathrm{i}\times\mathrm{j} = -\mathrm{j}\times\mathrm{i} = \mathrm{k}, \mathrm{j}\times\mathrm{k} = -\mathrm{k}\times\mathrm{j} = \mathrm{i}, \mathrm{k}\times\mathrm{i} = -\mathrm{i}\times\mathrm{k} = \mathrm{j}.$$

Based on these basic rules, quaternions can be added and multiplied. For instance, we can take the product of the two quaternions $q_1 = 2 + \mathrm{i}$ with $q_2 = 3 + 2\mathrm{i} + \mathrm{k}$:

$$\begin{aligned} q_1 \times q_2 &= (2+\mathrm{i})\times(3+2\mathrm{i}+\mathrm{k}) \\ &= 6 + 4\mathrm{i} + 2\mathrm{k} + 3\mathrm{i} + 2\mathrm{i}\times\mathrm{i} + \mathrm{i}\times\mathrm{k} \\ &= 6 + 4\mathrm{i} + 2\mathrm{k} + 3\mathrm{i} + 2(-1) - \mathrm{j} \\ &= 4 + 7\mathrm{i} - \mathrm{j} + 2\mathrm{k}. \end{aligned}$$

The multiplication of quaternions is said to be *non-commutative* since the order in which you perform the multiplication is important. For instance:

$$q_2 \times q_1 = 4 + 7\mathrm{i} + \mathrm{j} + 2\mathrm{k} \neq q_1 \times q_2.$$

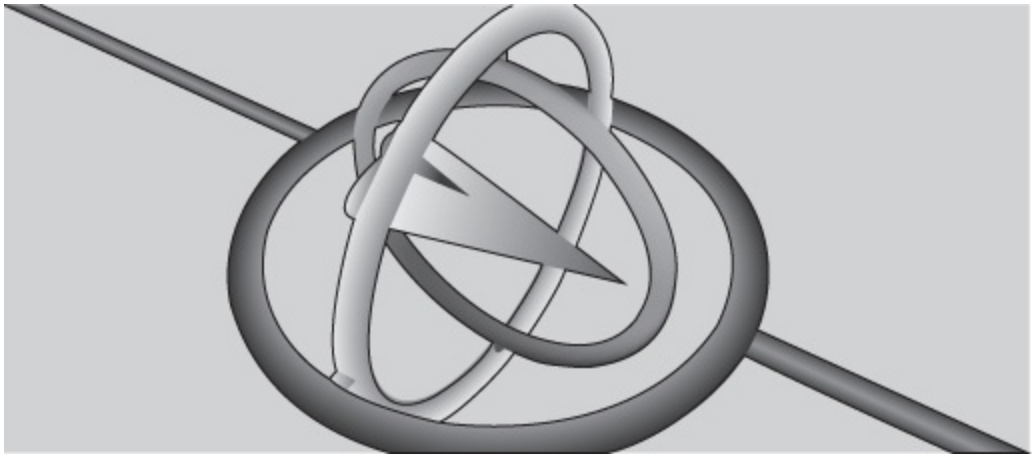
Hamilton created the first non-commutative algebra and, in the process, initiated the field of abstract algebra, dedicated to the study of such structures and their generalization. Quaternions are a true delight for mathematicians. Not only do they have an elegant algebraic structure, they also have a pleasing geometric interpretation. In particular, quaternions of norm one (i.e.

all quaternions q such that $q \times q = 1$) lie on a hypersphere in four dimensions (a hypersphere of radius $r > 0$ in n dimensions is the set of points in a space of n dimensions that are all at the same distance r from a given point. In two dimensions it is a circle, in three dimensions it is a sphere, and so on in higher dimensions). While Hamilton dedicated the rest of his career to the mathematical study of quaternions, they were not widely successful as a tool and they were eventually replaced by vectors and matrices.

Quaternions could be used as an example of mathematics going too far. Who needs a four-dimensional representation of numbers? Clearly these non-commutative algebraic structures have little or no bearing on everyday life? Yet, with the advent of computers it was realized that since quaternions have a simple geometric interpretation, they could be used in many applications where efficient methods to compute rotations are needed. Indeed, quaternions are now routinely used in computer graphics, robotics, missile and satellite guidance, as well as in orbital mechanics and control theory.

In these areas, quaternions have two main advantages over traditional matrix representations. First, only four numbers are required to code a rotation with quaternions rather than the nine elements of a rotation matrix, speeding up considerably the computation of repeated rotations. Second, they do not suffer from ‘gimbal lock’. To understand this idea, we can visualize a rotation in space by three angles. First, extend your arm. You can create a rotation by first moving it vertically (pitch), then horizontally (yaw), then by rotation your arm itself (roll). However, in doing so, you always have a position that is poorly defined (when your arm is straight up, there is no difference between a yaw and a roll). In the pitch/yaw/roll rotational system, shown in Figure 31, when the pitch is rotated 90° up or down, the yaw and roll correspond to the same motion and information about the rotation is lost. This virtual gimbal lock is a problem for actual gimbals where some angular zones are not accessible and can lead to catastrophic results when a plane is in a dive. During the first Moon landing mission in 1969, the Apollo 11 spacecraft was not equipped with a sufficient number of gimbals and experienced gimbal lock. After the Lunar Module landed, Mike Collins aboard the Command Module famously joked ‘How about sending me a fourth gimbal for Christmas?’ With quaternions, there is no virtual lock and rotations can be described smoothly by varying the four numbers

characterizing unit quaternions. This is why quaternions find their place in airplane technology, specifically for modern inertial navigation programs, in which they are used to represent both the rotation of the craft body frame (pitch, yaw, roll) and its geometric position (latitude and longitude).



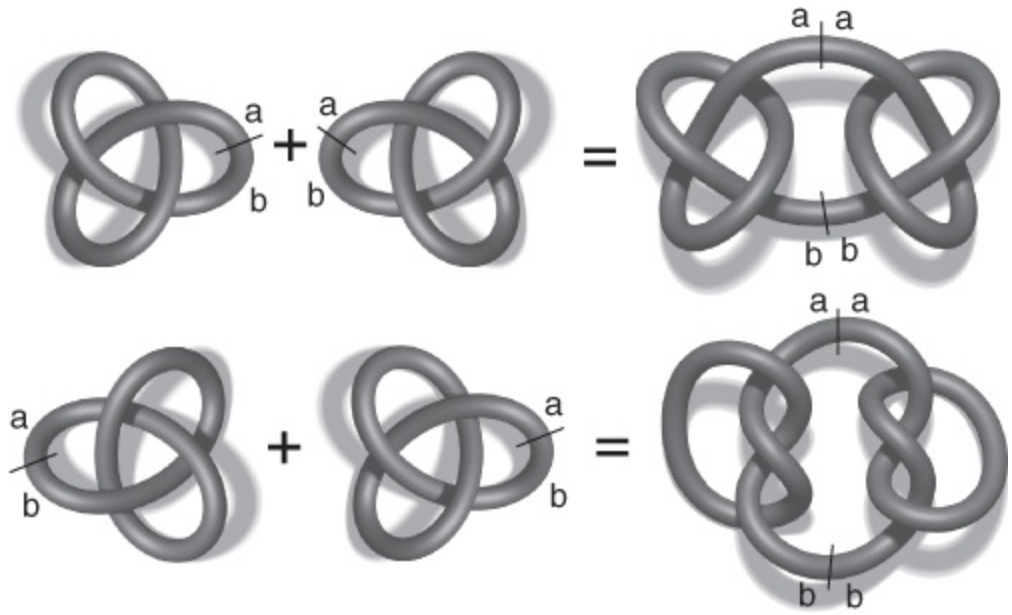
31. A gimbal is a system of multiple rings for rotation in space. In a set of three gimbals mounted together, each offers a degree of freedom: roll, pitch, and yaw.

In mathematics there is always a next question. Is there a natural way to generalize quaternions? An extension, the *octonions*, was already proposed in the 19th century when quaternions were fashionable. As the name suggests, the octonions are defined by eight components. Whereas complex numbers are pairs of real numbers and quaternions are pairs of complex numbers, octonions are pairs of quaternions. Octonionic multiplication is neither commutative nor associative (i.e. when you multiply three octonions together, the result depends, in general, on whether you first multiply the first two or the last two). Is this yet another useless mathematical construct? No, it turns out that following historic trends, this structure has found applications in various fields ranging from string theory to special relativity and quantum logic.

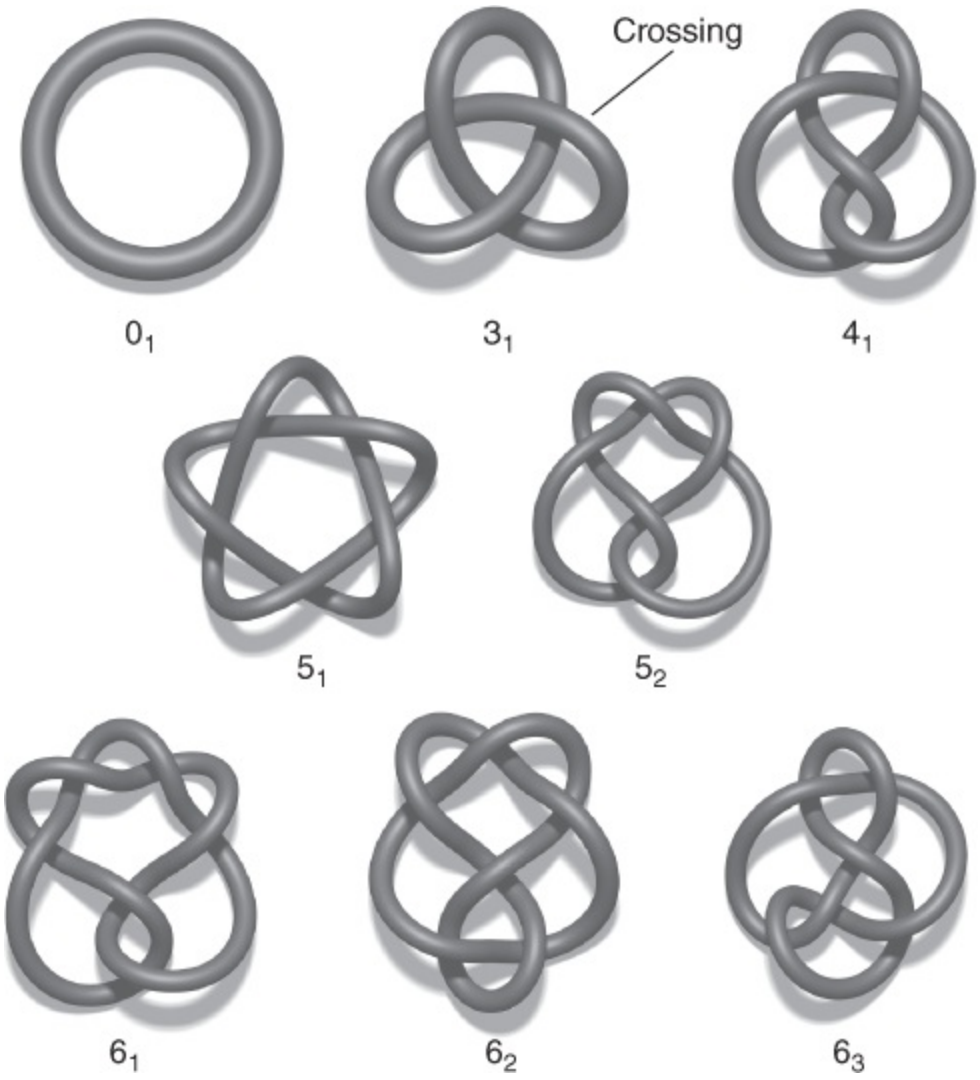
Knots and DNA

At the end of the 19th century, soon after Hamilton's death in 1865, William Thomson (who would eventually become Lord Kelvin) had a spark of inspiration. He had the magnificent idea that chemical elements could be classified through their internal structure. Based on the earlier work of Hermann von Helmholtz in fluid mechanics and the investigations of his friend, the Scottish natural philosopher Peter Tait, on smoke rings, Thomson believed that atoms are composed of tangled vortices. In Thomson's *vortex-atom theory*, atoms are knots of increasing complexity that can be deformed but can never lose their intrinsic shape: if a string is knotted and closed on itself, it can be deformed but its knottiness cannot be changed without cutting it. To each element there would be a corresponding knot. The problem was that, at the time, there was no classification of knots despite some early work by Carl Friedrich Gauss and Johann Benedict Listing on the topic.

At Thomson's suggestion, Tait and Little developed tools to study knots and they undertook a systematic enumeration of knots organized by their crossing numbers. Mathematically, knots are closed curves in space with no self-intersection (technically, a knot is an embedding of a circle in three-dimensional Euclidean space). A mathematical knot is very similar to the idea of a knot obtained by knotting a string and closing it. Two knots are the same if one can transform one into another by deforming and stretching the string but without letting the string go through itself or cutting it. Tait organized knots by their crossing numbers, the smallest possible number of intersections found in a *regular projection* (that is, by looking at all the possible shadows of the knot on a plane and disregarding the irregular shadows where three or more curves meet at a crossing point). The simplest knot, the unknot shown in Figure 32, has zero crossings. The trefoil knot 3_1 has three crossings, and so on. Tait's main classification is concerned with *prime knots*, which are simpler than the so-called *composite knots*. Composite knots are obtained as the sum of prime knots, such as the granny knots and the square knots shown in Figure 33. Prime knots can be used as the basic building blocks of knots, just as prime numbers are used to decompose any positive integer.



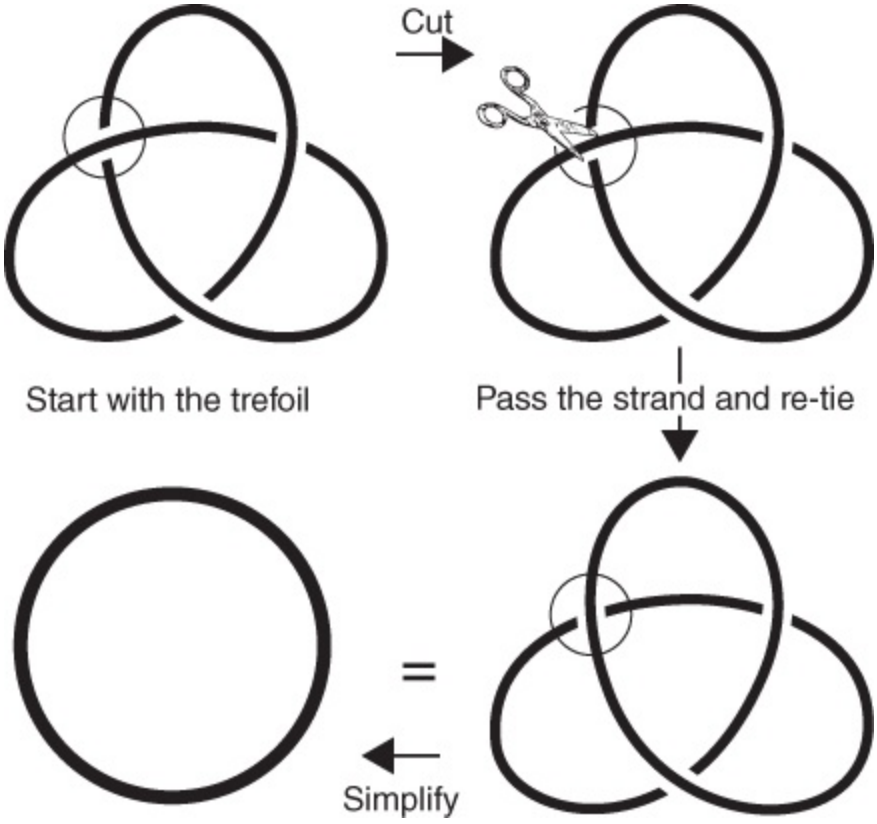
33. The square knot can be obtained as the sum of a trefoil knot and its mirror image. The granny knot is the sum of two identical trefoil knots. Composite knots are obtained by cutting prime knots and gluing together their segments, as shown for both knots.



32. A table of prime knots of up to six crossings. The first number defines the minimal number of crossing and the second number is used for classification: 6_2 designates the second knot with 6 crossings in the table of knots. The popular name for these knots are: the unknot 0_1, the trefoil 31, the figure-eight 4_1, the cinquefoil 5_1, the three-twist 5_2, and the stevedore 6_1.

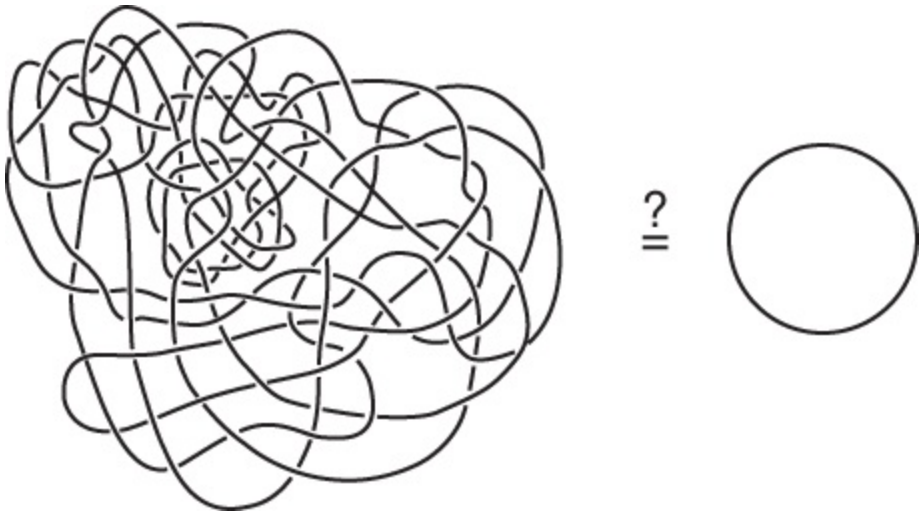
Unfortunately for Tait and Little, by the time they completed their tables of knots up to ten crossings, the Russian chemist Dmitri Mendeleev had solved the classification problem with his periodic table of elements. The vortex-atom theory was relegated to the dustbin of history. But their hard work was

not in vain because a new mathematical theory was born and, in the 20th century, the study of knots became a main focus of research in low-dimensional topology with surprising connections to other mathematical theories such as algebra and analysis. The topic is so relevant that one of the 1990 Fields Medals was awarded to Vaughan Jones for his contribution to knot theory and, in particular, for the invention of a function for each knot type, the *Jones polynomial*. In essence, a Jones polynomial $P(x)$ is defined for each knot (for instance, for the unknot $P = 1$ and for the trefoil $P = x + x^3 - x^4$). Operations on knots can be translated into operations on polynomials, henceforth linking geometry to algebra. This beautiful construction is intimately related to other mathematical theories and allows for the distinction between two knots (if two knots have different Jones polynomials, they are necessarily different). Thanks to such constructions, it is now possible to give precise meanings to surgical operations on knots such as inverting a crossing by cutting a strand, letting another strand pass through and reconnecting the cut strand afterwards, as shown in Figure 34. This operation may change the topology of the knot by changing its knot type.



34. Surgery on knots. The knot type can be changed by swapping the crossing (from an over-crossing to an under-crossing). In three dimensions, this operation can be performed by cutting, passing the strand, and re-tying the curve. After swapping a crossing, the knot has changed knot type from a trefoil to the unknot.

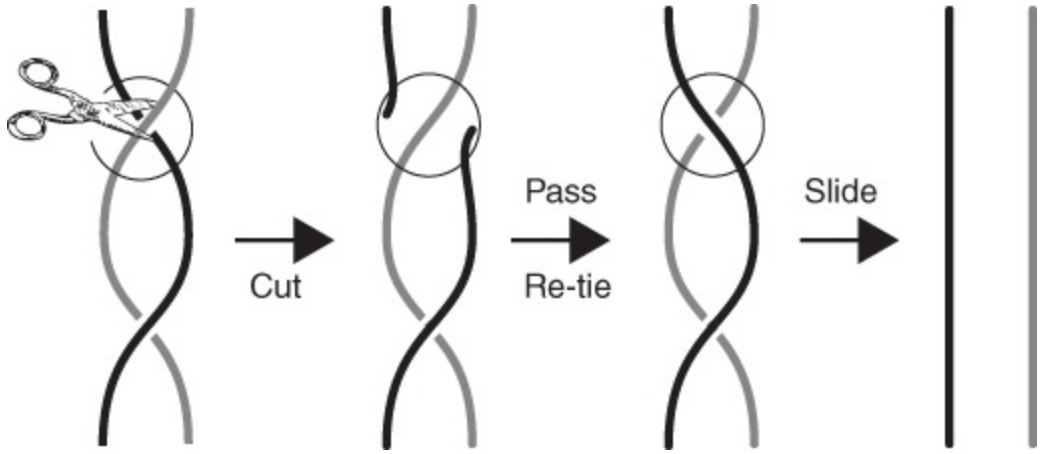
Nowadays, knot theory is still very much at the forefront of pure mathematics (see Figure 35 for a challenging problem). Yet, since the 1980s its results have found unexpected applications in the study of many physical and biological systems. Following the seminal discovery of DNA structure by Crick and Watson, as discussed in Chapter 6, it is now understood that DNA molecules carry the fundamental genetic information used by all living organisms for growth, development, proper functioning, and reproduction. Physically, the cellular DNA found in humans is an extremely long and thin chain. If straightened out, the largest human chromosome, chromosome number 1, would be 85 millimetres long and about 2 nanometres wide, a chain that is 42.5 million times longer than its diameter. These chains are tightly packed inside the nucleus and must be manipulated for key processes such as recombination, replication, and transcription to take place. DNA needs to solve not only a geometric packing problem but also many topological problems.



35. The unknotting problem consists in proving that a knot (left) is actually the unknot (right). Here is one of Wolfgang Haken's Gordian

unknots.

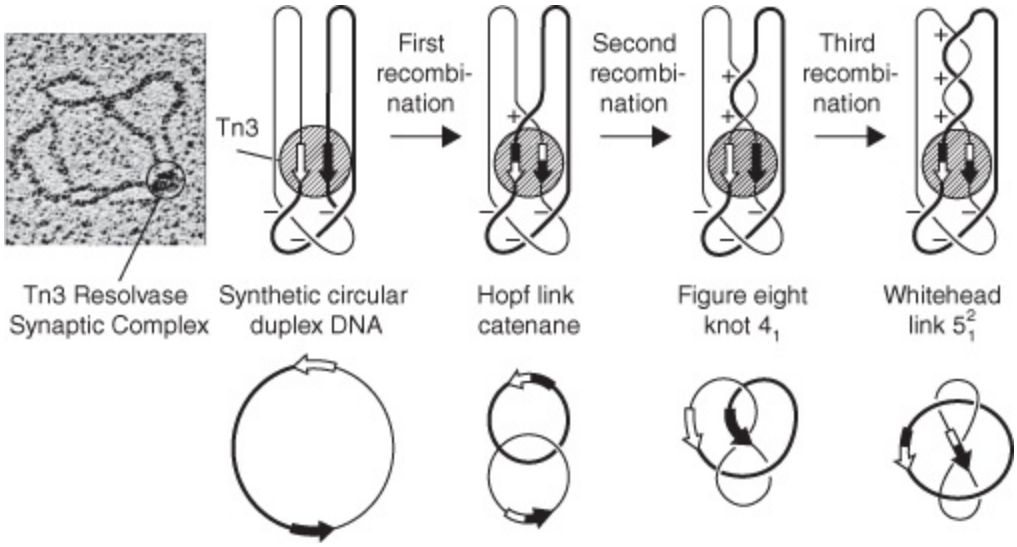
DNA is composed of two complementary twisted strands. During cell division, the strands must be completely separated. To do so requires completely unzipping the DNA and changing its knot type. Imagine that two twisted strings are attached at one end and you need to untwist them. If you just pull them apart, the twist in the two chains must be removed in order for them to be separated or the chain becomes overwound at the separation fork which prevents you from further separation. The solution to this geometric (and physical) impasse is to reduce the twist in the two strands of the DNA. This is accomplished in the cell by a special group of enzymes called *type I topoisomerases*, first discovered by James C. Wang in the 1970s (the prefix 'topo' stands for the topological action of the enzymes on the DNA, the ultimate mathematical enzyme!). Exactly like the surgical operation shown in Figure 34, these proteins cut one strand of a DNA double helix, let the other strand pass through, and then the cut strand is retied, reducing the overall twist, as shown in Figure 36.



36. Topoisomerase 1 action on DNA. One of the braids of DNA is cut, passed through, and re-tied. The overall action relaxes the twist of the two strands allowing for DNA replication.

Another important application of knot theory is in understanding the action of enzymes on DNA. For instance, *recombinases* are enzymes that manipulate

the structure of genomes to control gene expression. These enzymes are known to break the DNA apart at specific recombination sites and to recombine the ends by exchanging them. There is no direct observational method for enzyme action but recombinases are known to change the topology of DNA. Therefore, one can synthesize circular duplex DNA and study the enzyme mechanism by its action on DNA topology (see Figure 37). By looking theoretically and experimentally at the products created by recombination, mathematicians working with biologists are able to identify the mathematical operations that mirror the action of enzymes, hence uncovering key mechanisms taking place in the cell.



37. Action of the resolvase Tn3. The recombination sites are the two arrows bound by the resolvase (grey circle). During recombination, the resolvase transforms a circular duplex DNA into a Hopf link. Successive recombinations lead to the figure-eight knot and the Whitehead link. The next iteration would lead to the 62 knot (not shown).

The use of knot theory to study key processes in cellular biology was a pleasant surprise for many scientists. It also put new vigour into the theory of knots and transformed a rather esoteric subject into a mainstream applied topic. Nicholas Cozzarelli, a molecular biologist at the University of California Berkeley and one of the main proponents of topological enzymology once said:

> Before Jones, the math was incredibly arcane. The way the knots were classified had nothing to do with biology, but now you can calculate the things important to you.

Reflection

Quaternions and knots are only two examples of mathematical theories that have found unexpected applicability in sciences and engineering. Other examples include: the conic sections developed by Greek mathematicians in the 2nd century BC and used by Kepler to describe the motion of planets, the use of number theory in cryptography, and linear algebra and matrix theory with applications in physics and computer sciences. More recently, homology and the simplicial methods of algebraic topology and algebraic geometry are finding applications in data analysis.

This technology transfer from pure to applied mathematics was already understood in the 19th century by the mathematician Charles Babbage, credited with Ada Lovelace for the first conceptual discovery of the computer. In ‘Decline of science in England’ Babbage writes:

> In mathematical science, more than in all others, it happens that truths which are at one period the most abstract, and apparently the most remote from all useful application, become in the next age the bases of profound physical inquiries, and in the succeeding one, perhaps, by proper simplification and reduction to tables, furnish their ready and daily aid to the artist and the sailor.

The reason why some mathematical theories find applications in the real world even though they were built as purely abstract constructs is not clear. Maybe, unbeknownst to them, the brain of mathematicians is actually influenced by the world in which they live and the structures they create are inspired, consciously or not, by their surroundings. Or, maybe there are universal concepts to be found in nature that match the simplest and most elegant mathematical theories. In 1960, the physicist and Nobel Laureate Eugene Wigner wrote an essay entitled ‘The unreasonable effectiveness of mathematics in the natural sciences’ in which he concludes:

> The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve.

Governments are sometimes keen on funding successful future technologies. However, the future is particularly hard to predict. Which mathematical discoveries of today will change tomorrow's society? Nobody knows, but there will be plenty of surprises, surprises that would shock even the purest of mathematicians such as Godfrey Harold Hardy who famously said:

> No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world.

Little did he realize that his work in number theory would become the cornerstone of modern cryptography on which the modern digital economy rests.

Chapter 8

Where are we going? Networks and the brain

—

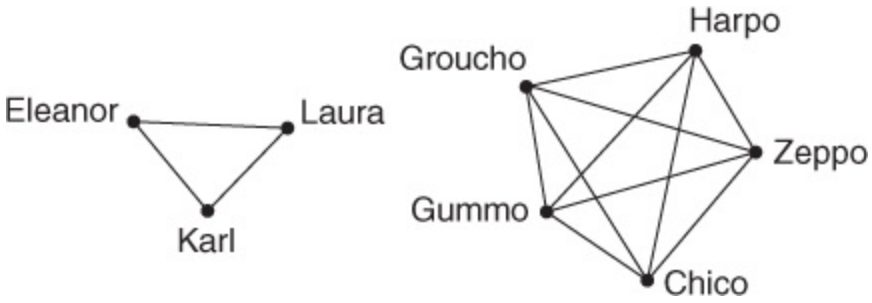
–Harpo

One of the striking traits of applied mathematics is that it has established itself by defying simple categorization. Applied mathematics, be it an art, a craft, or a discipline, is not bound by a particular scientific application, a particular mathematical universe, or a well-defined university department. Therefore, it can insert itself in all scientific disciplines and easily reinvent itself by moving fluidly from one field to the next, guided only by methods, theory, and applications. Many applied mathematicians see new challenges as an opportunity to expand their mathematical horizon, and in our rapidly changing society, such new challenges abound. Some of these challenges are related to traditional sciences, whereas others are generated by the world around us.

Networks

A relatively recent development in applied mathematics is the theory of

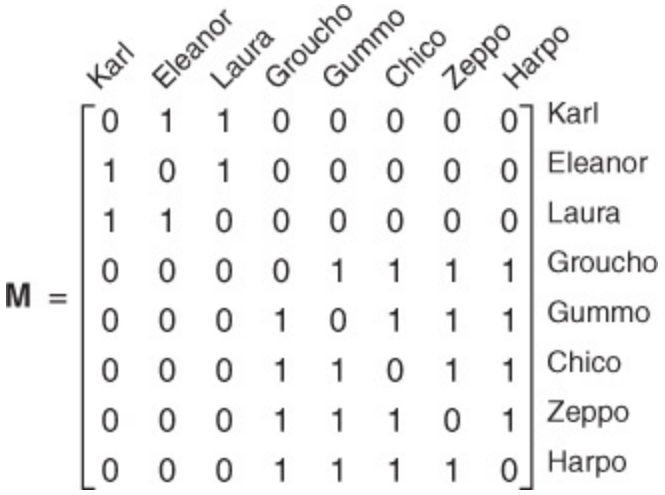
networks. Networks represent connections between multiple physical or virtual entities. In the simplest case, a network is a collection of nodes, representing the different elements of the system, connected by edges, representing the existence of a relationship between elements. An example of a network is given in Figure 38.



38. A network of family relationships for famous people with the surname Marx. Each person is represented by a node, and a family relationship is represented by an edge. The network is disconnected as Karl Marx and his two daughters have no family relationship with the five Marx brothers. This network has two components.

A network can be also represented by matrices. The elements of the matrix indicate whether pairs of nodes are connected or not in the network. For simple networks with no special orientation between nodes, the *adjacency matrix* is a matrix of zeros and ones. The element (i, j) and(j, i) is equal to one if there is an edge between node i and node j, and zero otherwise, as shown in Figure 39. Therefore, problems in network theory can be recast as problems on matrices and vectors, and most of network analysis is done by manipulations of (very) large matrices. Historically, the study of networks was initiated by Euler via the famous Königsberg bridge problem and further developed by mathematicians studying the mathematical structure of ordered and random graphs (the Königsberg problem is how to devise a walk through the city of Königsberg that would cross each of the bridges once and only once). In the 1950s, sociologists started to use networks to study social interactions and the formation of small interacting communities within larger social groups. But the subject really became prominent with the Internet and data revolution that led both to the creation of large networks (such as the World Wide Web and the social network of Facebook) and the availability of

large data sets.



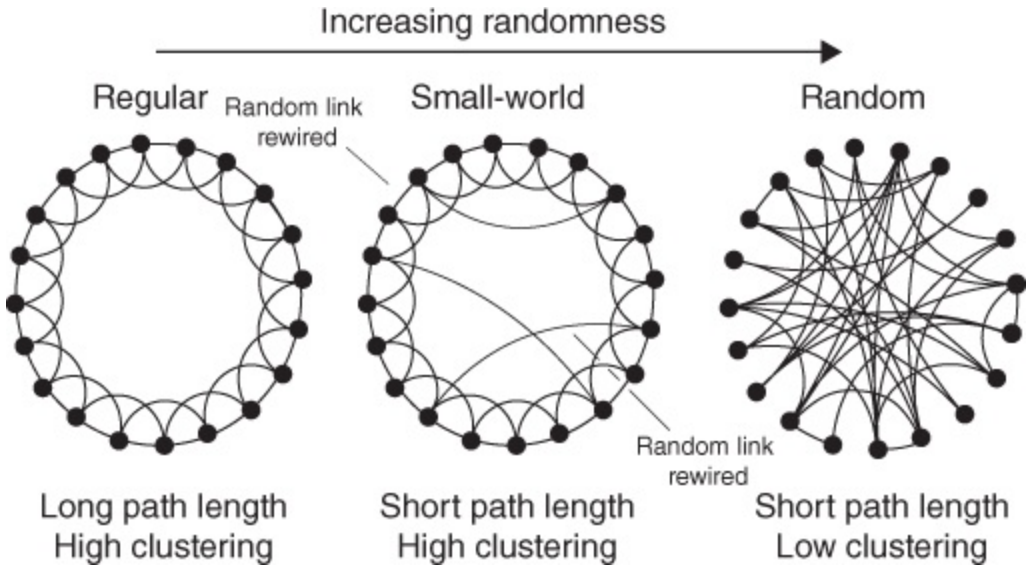
39. Matrices and networks. A network can be coded by a matrix representation. Each column and line of a matrix represents a node. A value of 1 at place (*i, j*) indicates that there is an edge between node *i* and node *j*.

Since a network is an abstraction of relationships between discrete elements, networks are ubiquitous in science and society. They are found in social systems (friendship networks, collaboration networks, human languages, …), biological systems (gene regulatory networks, metabolic networks, protein interaction maps, ecological interactions, evolutionary trees, …) and engineering systems (physical networks, power grids, metro stations, airline routes, web links, …).

Network theory really took off at the end of the 1990s when scientists from different communities realized the importance and ubiquity of networks. A key problem is to understand the structure of networks found in nature and society. Are they regular (e.g. every node is connected to its two neighbours) or random (nodes are randomly connected)? It turns out that most networks around us are neither regular nor random, as demonstrated by the *small-world* or *six-degrees-of-separation effect*. In social networks people seem to be connected to people they do not know directly through only a few connections (the popular idea is that it would only take a chain of six people

to connect you to Stephen Hawking through common acquaintances, or to connect the actor Kevin Bacon to any other actor through a chain of movies and cast members).

This small-world effect was formalized in 1998 by Steven Strogatz from Cornell University and his doctoral student at the time, Duncan Watts. In their seminal paper 'Collective dynamics of "small-world" networks', they explained that there are two important measures to quantify the connectedness of networks. The first one is the *average path length*, the average number of steps along the shortest paths for all possible pairs of network nodes. The second metric is *clustering*, measuring the degree to which nodes in a graph tend to cluster together. In a network of friends, high clustering means that there is a high probability that your friends also know each other. A regular network has a long path length but high clustering. A random network has a short path length and low clustering. Watts and Strogatz showed that real networks have both short average path lengths and high clustering: nearby members are closely connected but also connected to far-away members. In the finest tradition of applied mathematics, they considered an extremely simple toy model by building a family of networks ranging from regular to random as shown in Figure 40. They showed that by rewiring a few elements of a regular network to allow for random long distance connections, the average path length is dramatically reduced: a few long-distance shortcuts in a regular network brings all edges much closer, explaining the small-world effect.



40. The small-world effect. Starting from a regular network (left), each link is rewired with increasing probability (middle) up to a fully random model where each node is randomly connected to other nodes (right). If elements represent individuals and edges represent acquaintances: in the regular network, each element knows its two nearest neighbours on each side, whereas with added connections, some people know a few distant people.

Watts and Strogatz's work was immensely influential and marked the start of network science as a mainstream mathematical discipline. Since then, the field has exploded and there is considerable interest in better understanding, designing, and controlling networks. For instance, the Internet can be modelled as a network where each webpage is a node and an edge exists between links if there is a weblink between the two. It is particularly important for web users to be able to navigate this giant network and find the page they want. The page ranking algorithms, used by big search engine companies to present the most relevant weblinks to users, are based on the analysis of networks and the relative importance of different sites based on their connections. In their simplest form these algorithms can be understood as a mathematical problem involving rather large matrices (that can have billions of entries).

Another interesting application of networks is in epidemiology:

understanding how diseases spread on a network. The traditional models for the propagation of an infection take the form of a system of differential equations, similar to the ones seen in Chapter 4 for rabbits and wolves, but coupled to diffusion. These models describe the progression of a disease in a local population by computing the rate of change of infected people. The disease spreads spatially from direct contact, resulting in a propagating wave of infection. This approach is valid for the spreading of diseases from neighbour to neighbour. However, human mobility has completely changed, and we know that diseases can also spread by occasional contacts with people not living in close proximity. In trying to predict pandemics, we have to consider the possibility that an infected person could carry the disease to another continent through the air-transport network (see Figure 41). For instance, in 2003 the severe acute respiratory syndrome virus (SARS) was first observed in China, then moved to Hong Kong, then jumped to Europe, Canada, and the United States. These long-range connections change the dynamics by allowing for shortcuts in disease propagation. Network analysis shows that there is a critical number of these shortcuts that would change the progression from a localized epidemic to a global pandemic. These effects are now taken into account in order to pinpoint the origin of an outbreak and to formulate optimal public health policy in the case of future outbreaks.

41. The airline network. In this network, each airport is a node and an edge represents a flight between two airports. The study of disease

propagation on such networks led to unexpected behaviours that helped shape public health policy.

The mathematical brain

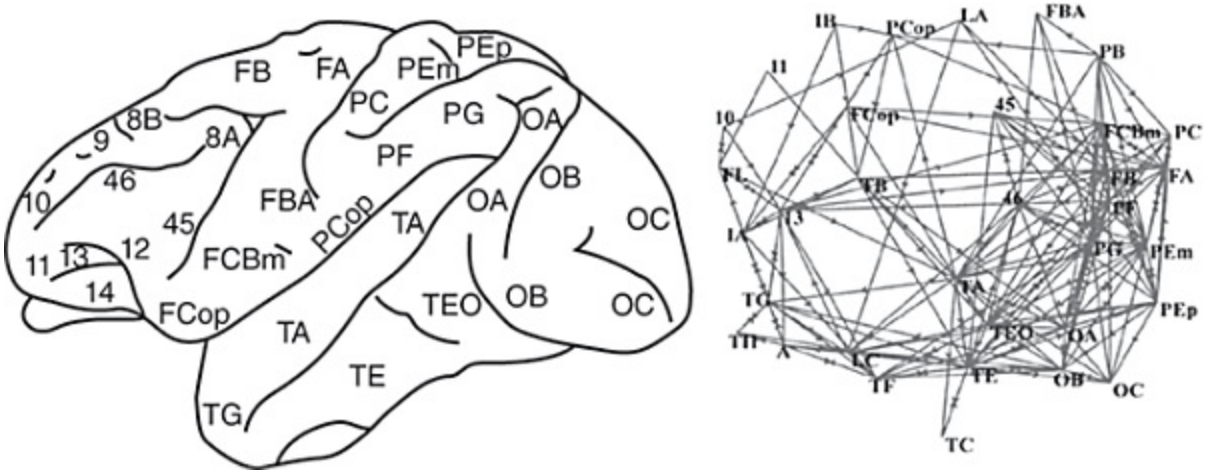
Major scientific efforts are required to address major societal challenges. These include fighting climate change, optimizing new renewable energy sources and developing new medical treatments. Traditionally, applied mathematicians involved with these collaborative efforts were considered a useful but small cog in a huge scientific machine. However, it is now appreciated that quality science requires clever modelling, state-of-the-art numerical methods, and fundamental theoretical insights from simplified models. This is the realm of applied mathematics.

For a final and perhaps most surprising example, I consider the challenges associated with understanding the brain, an organ of extreme complexity.

The human brain is an intricate composite of cells, fluids, and tissues which can broadly be distinguished into two regions: the cerebral cortex, made up of grey matter which contains the neuronal cell bodies in a thin layer of tissue close to the brain surface, and the white matter, the internal tissue composed of a glial cell scaffold as well as a number of axonal fibre bundles which facilitate signal transmission between neural regions.

At the basic functional level, the goal of neuroscientists is to understand the functions of the human brain resulting from the interaction of about eighty-six billion neurons with a hundred trillion connections. From this perspective, the problem consists of connecting the biochemical and electrophysiological behaviour of brain cells with the overall behaviour of networks of connected cells. The ultimate goal is to translate the resulting macroscopic electrophysiological behaviour into a functional dimension where direct relations can be established between neuronal response and, ultimately, human behaviour.

Different brain networks can be created by looking at the physical or functional connections between different parts of the brain achieved by axons. Physical maps are obtained non-invasively by *diffusion tensor imaging*, measuring the diffusion of water molecules within the cortex by MRI, and thus tracking the paths of white matter fibres which connect different parts of the brain (see Figure 42). Other networks can be obtained by looking at functional connections present during cognitive processes. These functional networks are obtained by functional magnetic resonance imaging (fMRI) measuring brain activity by detecting changes associated with blood flow.



42. The brain network. Left: a map of the parcellation scheme of the macaque cortex. Right: cortical connectivity maps in the macaque cortex constructed by placing nodes in close proximity if the connection between them is strong and far apart if the connection between them is weak.

These various networks can be studied by the methods we just discussed. In particular, repeated studies have shown that brain anatomical connectivity is sparse and has high clustering, with a few long-range connections mediating short path lengths between any pair of regions. That is, these connectivity maps exhibit small-worldness properties in much the same way as social networks do. It has been suggested that these features allow for dynamical complexity and are the result of an evolutionary pressure reducing wiring cost. This type of analysis has spawned a new field of study: *brain topology*,

which aims at understanding brain cognition through the study of connectivity maps. In particular, it has been shown that complex brain networks are sensitive to many features such as age, gender, and genetic background. Changes in the brain networks can also be found in diseases such as Alzheimer's, schizophrenia, epilepsy, and depression.

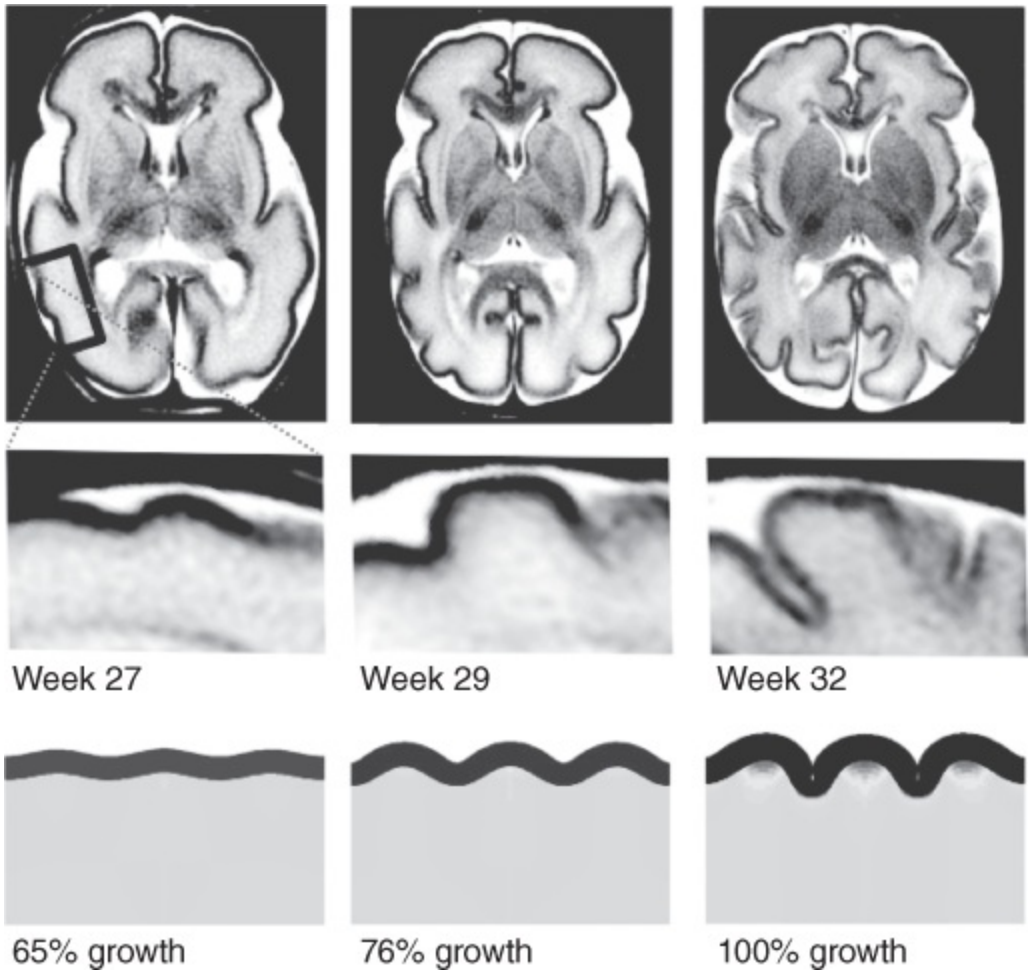
Whereas much is known about the brain at the functional level, comparatively little is known and understood at the geometric, physical, and mechanical levels. Recent findings have directly linked major brain development and diseases to the physical response of the brain both at the cellular and tissue levels. These include brain morphogenesis, axonal mechanics, brain tumours, brain surgery, traumatic brain injury, and neurodegenerative diseases. Physical studies of brain tissues are difficult. First, the brain is a fully enclosed organ that is particularly difficult to probe. Second, viewed as a solid, it is extremely soft and its mechanical response is heavily influenced by a fluid phase as well as multiple charged molecules found in its cells and in the extracellular matrix.

An interesting problem that has been tackled in recent years is the problem of *brain folding*: can we explain the intricate folded morphology of the human brain, both in terms of developmental origin and functional value?

The creation of sulci (valleys) and gyri (bumps) in the cortex allows for the rapid expansion of grey matter compared to white matter in the human brain. This *gyrification* process is a typical feature of large mammals. From approximately twenty-nine weeks from gestation onwards, the cerebral cortex develops a number of elaborate convolutions. These folds are associated with a significant increase in area of the brain's outer surface, resulting in a surface area approximately three times greater than the interior surface area of the skull. The folding patterns are attributed to a number of chemical and mechanical mechanisms including differential growth of the cortical layers and the activation and inhibition of chemical signalling on the cortical surface.

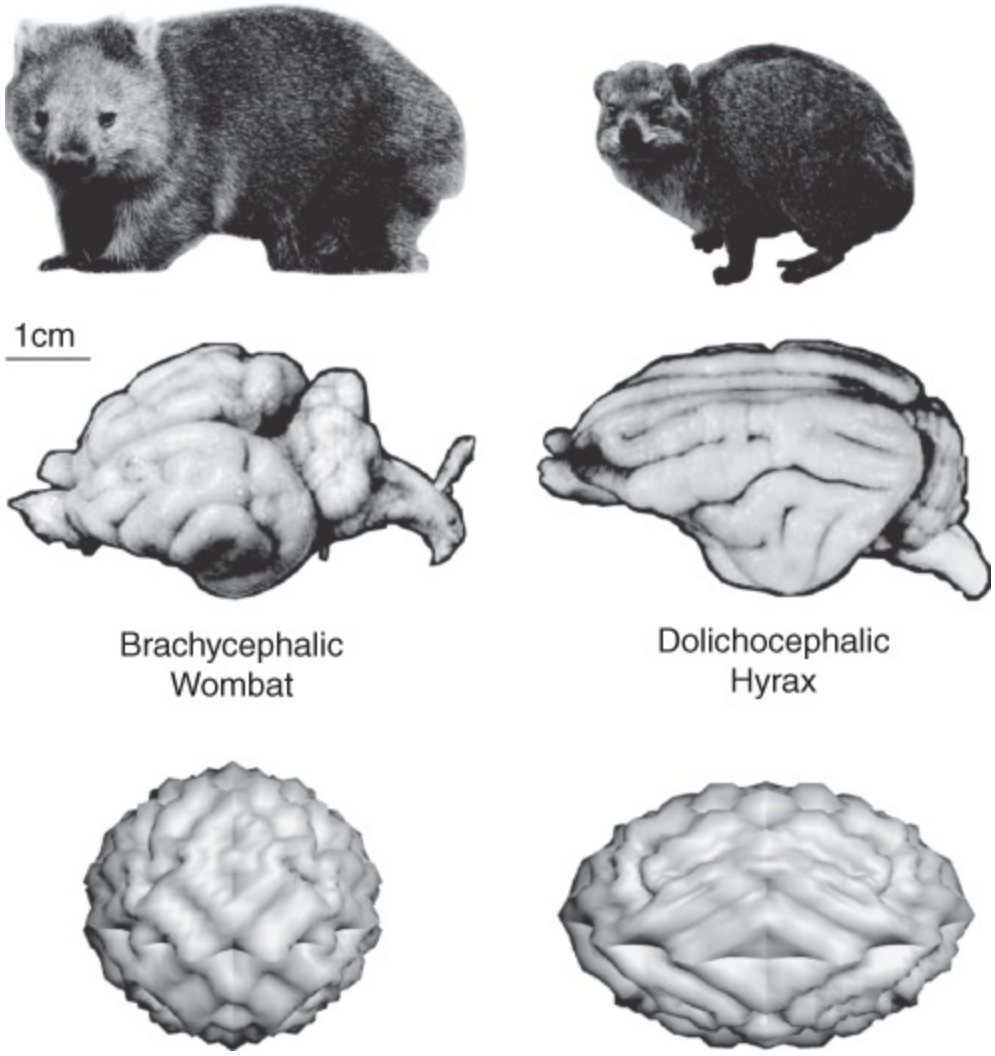
In recent years there has been mounting evidence that the onset of folding can be understood as a mechanical instability, similar to the folding pattern

achieved by the growth of a thin elastic layer attached to an elastic substrate, observed in swelling gel experiments. Simple toy models in two dimensions, using only the basic mechanical elements of the problem (see Figure 43), not only reproduce the known pattern but also explain some of the observed pathologies such as lissencephaly (poorly folded brain) and polymicrogyria (overly folded brain). The importance of geometry and physics in pattern selection becomes apparent when comparing the cortical folding patterns of brains with different shapes: elongated brains tend to be more longitudinally folded than spherical brains, as shown in Figure 44. The study of such patterns is particularly important as small perturbations during cortical development are closely associated with neurodevelopmental disorders such as autism spectrum disorders, schizophrenia, and epilepsy.



43. Brain folding. Magnetic resonance images of the developing human brain in preterm neonates at 27, 29, and 32 weeks of gestation. We see

the onset and development of the characteristic convoluted surface morphology. The bottom row shows the simulated morphologies of developing human brain section at 27, 29, and 32 weeks of gestation based on a simple mechanical model.



44. Mammalian brains vary greatly in ellipticity and gyrification. Simulations of a spheroid with growing cortical layer and varying ellipticity predict transverse folding in the brachycephalic skull and longitudinal folding in the dolichocephalic skull.

It seems clear that an understanding of the brain can only be achieved by coupling different mathematical methods and physical theories (fluid and solid mechanics, electrochemistry, signal propagation) at multiple scales

(from the sub-cellular processes taking places in the axons and glial cells all the way to the brain network as a whole). A mathematical theory of the brain must necessarily be multiscale (including different length and time scales) and multiphysics (including different physical and mechanical theories) and is therefore connected to some of the most outstanding challenges of modern applied mathematics. These aspects are crucial for the proper understanding of a host of conditions including neurodevelopment, cerebral oedema, hydrocephaly, craniectomy, spinal cord injury, tumour growth, traumatic brain injury, depression, and neurodegenerative diseases. With its efforts to understand the workings of the human brain, mathematics is now at the centre of research into what it is that makes us who we are.

Reflection

Unlike pure mathematics, applied mathematics is deeply influenced by society and scientific progress. The birth of applied mathematics as a formal discipline can be traced to World War II when major technological challenges required a systematic quantitative approach. At the time, the Allies faced tremendous logistic and technological challenges that needed radically new approaches. At the scientific level, this led to nuclear technology but also to the mathematical development of cryptography and modern continuum mechanics. At the organizational level, it initiated a new era of rational decision-making processes via the systematic use of mathematics through the development of *operations research*. The goal of operations research is to optimize decision-making tasks based on the application of advanced analytical methods such as mathematical modelling, statistical analysis, and mathematical optimization. These challenges required the unlikely collaboration of mathematicians and scientists from multiple disciplines thrown together in the war effort. In many ways, the structure and organization of our current scientific disciplines are still very much influenced by the war.

Our modern society is characterized by major challenges in the area of climate change, population increase and motion, energy resources, and medicine. Applied mathematics already plays a prominent role in these collaborative projects. It has served as an incubator for new ideas and

methods. For instance, for many decades climate modelling was deeply influenced by theories from applied mathematics such as dynamical systems, fluid mechanics, and numerical methods. Over the years, climate science has grown into its own discipline that does not require the same level of innovation and has naturally drifted away from mathematics despite remaining a fertile ground for new problems.

Our society is constantly producing vast amount of data. According to some estimates, more data was generated in the two years 2015–16 than in all of previous history. The creation of new data has become so overwhelming that it is now referred to as *big data*, a new major societal trend to generate and analyse very large chunks of data. The initial hope of big-data protagonists is that one could finally bypass theory and equations, go straight into data, and use empirical methods to unravel the mysteries of the universe. The data would be analysed by a black box that would provide answers to our most burning questions. This approach raises many interesting challenges. Yes, data is important, but understanding the processes that produce the data and the properties of the data may be equally important. The success of many methods has shed a bright light on the need to understand the underlying mathematical structure of both data and methods. Big data presents now a rich field of study that brings all mathematical sciences together, including statistics and computer science.

Applied mathematics sits at the convergence of collaborative modelling and data research. Its role is fundamental to how we understand our world and develop new technologies. As such it holds a very special role in science and society and will, undoubtedly, be at the centre of many future human endeavours in one way or another.

Epilogue

Quote me as saying, I was misquoted.

–Groucho (misquoted)

In an interview with Rolling Stone Magazine in 1965, Bob Dylan was pushed to define himself: ‘Do you think of yourself primarily as a singer or a poet?’ To which, Dylan famously replied: ‘Oh, I think of myself more as a song and dance man, y’know.’ Dylan’s attitude towards pigeonholing resonates with many applied mathematicians.

If I am ever invited again to a party and asked the question:

What is Applied Mathematics?

I will say:

It’s all aboutModels, Matrices, and Equations, y’know,

and maybe, just maybe, I’ll stay a little longer.

Further reading

Outside of a dog, a book is man's best friend.
Inside of a dog it's too dark to read.

–Groucho

References are rated from introductory (*) to advanced (***).

If you enjoy the *Very Short Introduction* series, I recommend the wonderful *Mathematics: A Very Short Introduction** by Tim Gowers (OUP, 2002).

At the other extreme of the spectrum, the prolific Tim Gowers is also the main editor of a very long introduction to mathematics: *Princeton Companion to Mathematics*** (Princeton University Press, 2008).

The applied companion to this last book is the *Princeton Companion to Applied Mathematics*** (Princeton University Press, 2015) edited by Nicholas J. Higham and Mark R. Dennis.

The description of the scaling and dynamics of a candle and the notion of power balance can be found in Alwyn Scott's *Encyclopaedia of Nonlinear Science* (Routledge, 2004; data for Figure 2).

Faraday's original lectures on the candle are available at <https://archive.org>.

Readers who want to learn more about the applications of reaction-diffusion equations in biology should have a look at James Murray's masterful treatise on *Mathematical Biology*** (Springer, 2007).

Thomas A. MacMahon's and John Tyler Bonner's *On Size and Life** (Scientific American, 1983) offers a wonderful collection of surprising scaling laws in nature and society, a must-read for all aspiring natural philosophers.

More fascinating examples of scaling laws in biology can be found in Knut Schmidt-Nielsen's *Scaling: Why is Animal Size So Important?** (Cambridge University Press, 1984).

Learn about the story of Poincaré's discoveries and mistakes in two excellent books: *Celestial Encounters: The Origins of Chaos and Stability** by Florin Diacu and Philip Holmes (Princeton Science Library, 1996); *Poincaré and the Three Body Problem*** by June Barrow-Green (American

Mathematical Society, 1996).

Learn about Sonya Kovalevskaya and her work in: *The Mathematics of Sonya Kovalevskaya*** by Roger Cooke (Springer-Verlag, 1984).

The development of chaos as a new science is told in: *Chaos: Making a New Science** by James Gleick (Open Road Media, 2011).

To know more about chaos theory itself, I recommend the excellent introduction *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering**** by Steven H. Strogatz (Westview Press, 2nd edition, 2014).

A very interesting account of the discovery process that led to the development of computed tomography can be found in the two lectures given by Hounsfield and Cormack when they received the Nobel prize. These can be found at <https://www.nobelprize.org>.

The early history of knot theory is nicely told in *Knots: Mathematics with a Twist** by Alexei Sossinsky (Harvard University Press, 2004).

For an introduction to the mathematics of knots: *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*** by Colin Adams (American Mathematical Society, 2004).

Learn more about DNA topology in *DNA Topology*** (Oxford Biosciences, 2005) by Andrew D. Bates and Anthony Maxwell.

An excellent mathematical introduction to the theory of networks: *Networks: An Introduction**** by Mark Newman (Oxford University Press, 2010).

The story of the discovery of small-world networks (and more) is found in a book by Strogatz himself: *Sync: The Emerging Science of Spontaneous Order** by Steven Strogatz (Penguin Press Science, 2004).

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z

SOCIAL MEDIA
Very Short Introduction

Join our community

www.oup.com/vsi

- Join us online at the official Very Short Introductions **Facebook** page.
- Access the thoughts and musings of our authors with our online **blog**.
- Sign up for our monthly **e-newsletter** to receive information on all new titles publishing that month.
- Browse the full range of Very Short Introductions online.
- Read **extracts** from the Introductions for free.
- If you are a teacher or lecturer you can order inspection copies quickly and simply via our website.