

Introduction to Applied Linear Algebra
Vectors, Matrices, and Least Squares

Introduction to Applied Linear Algebra

Vectors, Matrices, and Least Squares

Stephen Boyd

*Department of Electrical Engineering
Stanford University*

Lieven Vandenberghe

*Department of Electrical and Computer Engineering
University of California, Los Angeles*



CAMBRIDGE
UNIVERSITY PRESS



Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781316518960

DOI: 10.1017/9781108583664

© Cambridge University Press & Assessment 2018

This publication is in copyright. Subject to statutory exception and to the provisions
of relevant collective licensing agreements, no reproduction of any part may take
place without the written permission of Cambridge University Press & Assessment.

First published 2018 (version 4, January 2023)

Printed in the United Kingdom by TJ Books Limited, Padstow Cornwall, January 2023

A catalogue record for this publication is available from the British Library

ISBN 978-1-316-51896-0 Hardback

Additional resources for this publication at www.cambridge.org/IntroAppLinAlg

Cambridge University Press & Assessment has no responsibility for the persistence
or accuracy of URLs for external or third-party internet websites referred to in this
publication and does not guarantee that any content on such websites is, or will
remain, accurate or appropriate.

For

Anna, Nicholas, and Nora

Daniël and Margriet

Contents

Preface	xi
I Vectors	1
1 Vectors	3
1.1 Vectors	3
1.2 Vector addition	11
1.3 Scalar-vector multiplication	15
1.4 Inner product	19
1.5 Complexity of vector computations	22
Exercises	25
2 Linear functions	29
2.1 Linear functions	29
2.2 Taylor approximation	35
2.3 Regression model	38
Exercises	42
3 Norm and distance	45
3.1 Norm	45
3.2 Distance	48
3.3 Standard deviation	52
3.4 Angle	56
3.5 Complexity	63
Exercises	64
4 Clustering	69
4.1 Clustering	69
4.2 A clustering objective	72
4.3 The k -means algorithm	74
4.4 Examples	79
4.5 Applications	85
Exercises	87

viii	Contents
5	Linear independence 89
5.1	Linear dependence 89
5.2	Basis 91
5.3	Orthonormal vectors 95
5.4	Gram–Schmidt algorithm 97
	Exercises 103
II	Matrices 105
6	Matrices 107
6.1	Matrices 107
6.2	Zero and identity matrices 113
6.3	Transpose, addition, and norm 115
6.4	Matrix-vector multiplication 118
6.5	Complexity 122
	Exercises 124
7	Matrix examples 129
7.1	Geometric transformations 129
7.2	Selectors 131
7.3	Incidence matrix 132
7.4	Convolution 136
	Exercises 144
8	Linear equations 147
8.1	Linear and affine functions 147
8.2	Linear function models 150
8.3	Systems of linear equations 152
	Exercises 159
9	Linear dynamical systems 163
9.1	Linear dynamical systems 163
9.2	Population dynamics 164
9.3	Epidemic dynamics 168
9.4	Motion of a mass 169
9.5	Supply chain dynamics 171
	Exercises 174
10	Matrix multiplication 177
10.1	Matrix-matrix multiplication 177
10.2	Composition of linear functions 183
10.3	Matrix power 186
10.4	QR factorization 189
	Exercises 191

Contents	ix
11 Matrix inverses	199
11.1 Left and right inverses	199
11.2 Inverse	202
11.3 Solving linear equations	207
11.4 Examples	210
11.5 Pseudo-inverse	214
Exercises	217
 III Least squares	 223
12 Least squares	225
12.1 Least squares problem	225
12.2 Solution	227
12.3 Solving least squares problems	231
12.4 Examples	234
Exercises	239
 13 Least squares data fitting	 245
13.1 Least squares data fitting	245
13.2 Validation	260
13.3 Feature engineering	269
Exercises	279
 14 Least squares classification	 285
14.1 Classification	285
14.2 Least squares classifier	288
14.3 Multi-class classifiers	297
Exercises	305
 15 Multi-objective least squares	 309
15.1 Multi-objective least squares	309
15.2 Control	314
15.3 Estimation and inversion	316
15.4 Regularized data fitting	325
15.5 Complexity	330
Exercises	334
 16 Constrained least squares	 339
16.1 Constrained least squares problem	339
16.2 Solution	344
16.3 Solving constrained least squares problems	347
Exercises	352

x	Contents
17	Constrained least squares applications 357
17.1	Portfolio optimization 357
17.2	Linear quadratic control 366
17.3	Linear quadratic state estimation 372
	Exercises 378
18	Nonlinear least squares 381
18.1	Nonlinear equations and least squares 381
18.2	Gauss–Newton algorithm 386
18.3	Levenberg–Marquardt algorithm 391
18.4	Nonlinear model fitting 399
18.5	Nonlinear least squares classification 401
	Exercises 412
19	Constrained nonlinear least squares 419
19.1	Constrained nonlinear least squares 419
19.2	Penalty algorithm 421
19.3	Augmented Lagrangian algorithm 422
19.4	Nonlinear control 425
	Exercises 434
	Appendices 437
A	Notation 439
B	Complexity 441
C	Derivatives and optimization 443
C.1	Derivatives 443
C.2	Optimization 447
C.3	Lagrange multipliers 448
D	Further study 451
	Index 455

Preface

This book is meant to provide an introduction to vectors, matrices, and least squares methods, basic topics in applied linear algebra. Our goal is to give the beginning student, with little or no prior exposure to linear algebra, a good grounding in the basic ideas, as well as an appreciation for how they are used in many applications, including data fitting, machine learning and artificial intelligence, tomography, navigation, image processing, finance, and automatic control systems.

The background required of the reader is familiarity with basic mathematical notation. We use calculus in just a few places, but it does not play a critical role and is not a strict prerequisite. Even though the book covers many topics that are traditionally taught as part of probability and statistics, such as fitting mathematical models to data, no knowledge of or background in probability and statistics is needed.

The book covers less mathematics than a typical text on applied linear algebra. We use only one theoretical concept from linear algebra, linear independence, and only one computational tool, the QR factorization; our approach to most applications relies on only one method, least squares (or some extension). In this sense we aim for intellectual economy: With just a few basic mathematical ideas, concepts, and methods, we cover many applications. The mathematics we do present, however, is complete, in that we carefully justify every mathematical statement. In contrast to most introductory linear algebra texts, however, we describe many applications, including some that are typically considered advanced topics, like document classification, control, state estimation, and portfolio optimization.

The book does not require any knowledge of computer programming, and can be used as a conventional textbook, by reading the chapters and working the exercises that do not involve numerical computation. This approach however misses out on one of the most compelling reasons to learn the material: You can use the ideas and methods described in this book to do practical things like build a prediction model from data, enhance images, or optimize an investment portfolio. The growing power of computers, together with the development of high level computer languages and packages that support vector and matrix computation, have made it easy to use the methods described in this book for real applications. For this reason we hope that every student of this book will complement their study with computer programming exercises and projects, including some that involve real data. This book includes some generic exercises that require computation; additional ones, and the associated data files and language-specific resources, are available online.

If you read the whole book, work some of the exercises, and carry out computer exercises to implement or use the ideas and methods, you will learn a lot. While there will still be much for you to learn, you will have seen many of the basic ideas behind modern data science and other application areas. We hope you will be empowered to use the methods for your own applications.

The book is divided into three parts. Part I introduces the reader to vectors, and various vector operations and functions like addition, inner product, distance, and angle. We also describe how vectors are used in applications to represent word counts in a document, time series, attributes of a patient, sales of a product, an audio track, an image, or a portfolio of investments. Part II does the same for matrices, culminating with matrix inverses and methods for solving linear equations. Part III, on least squares, is the payoff, at least in terms of the applications. We show how the simple and natural idea of approximately solving a set of over-determined equations, and a few extensions of this basic idea, can be used to solve many practical problems.

The whole book can be covered in a 15 week (semester) course; a 10 week (quarter) course can cover most of the material, by skipping a few applications and perhaps the last two chapters on nonlinear least squares. The book can also be used for self-study, complemented with material available online. By design, the pace of the book accelerates a bit, with many details and simple examples in parts I and II, and more advanced examples and applications in part III. A course for students with little or no background in linear algebra can focus on parts I and II, and cover just a few of the more advanced applications in part III. A more advanced course on applied linear algebra can quickly cover parts I and II as review, and then focus on the applications in part III, as well as additional topics.

We are grateful to many of our colleagues, teaching assistants, and students for helpful suggestions and discussions during the development of this book and the associated courses. We especially thank our colleagues Trevor Hastie, Rob Tibshirani, and Sanjay Lall, as well as Nick Boyd, for discussions about data fitting and classification, and Jenny Hong, Ahmed Bou-Rabee, Keegan Go, David Zeng, and Jaehyun Park, Stanford undergraduates who helped create and teach the course EE103. We thank David Tse, Alex Lemon, Neal Parikh, and Julie Lancashire for carefully reading drafts of this book and making many good suggestions.

Stephen Boyd
Lieven Vandenbergh

Stanford, California
Los Angeles, California