Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics are generally aimed at third- and fourth-year undergraduate mathematics students at North American universities. These texts strive to provide students and teachers with new perspectives and novel approaches. The books include motivation that guides the reader to an appreciation of interrelations among different aspects of the subject. They feature examples that illustrate key concepts as well as exercises that strengthen understanding.

More information about this series at http://www.springer.com/series/666

Peter J. Olver • Chehrzad Shakiban

Applied Linear Algebra

Second Edition

Springer

Peter J. Olver
School of Mathematics
University of Minnesota
Minneapolis, MN
USA

Chehrzad Shakiban
Department of Mathematics
University of St. Thomas
St. Paul, MN
USA

ISSN 0172-6056 ISSN 2197-5604 (electronic)
Undergraduate Texts in Mathematics
ISBN 978-3-319-91040-6 ISBN 978-3-319-91041-3 (eBook)
https://doi.org/10.1007/978-3-319-91041-3

Library of Congress Control Number: 2018941541

Mathematics Subject Classification (2010): 15-01, 15AXX, 65FXX, 05C50, 34A30, 62H25, 65D05, 65D07, 65D18

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To our children and grandchildren.

You are the light of our life.

Preface

Applied mathematics rests on two central pillars: calculus and linear algebra. While calculus has its roots in the universal laws of Newtonian physics, linear algebra arises from a much more mundane issue: the need to solve simple systems of linear algebraic equations. Despite its humble origins, linear algebra ends up playing a comparably profound role in both applied and theoretical mathematics, as well as in all of science and engineering, including computer science, data analysis and machine learning, imaging and signal processing, probability and statistics, economics, numerical analysis, mathematical biology, and many other disciplines. Nowadays, a proper grounding in both calculus and linear algebra is an essential prerequisite for a successful career in science, technology, engineering, statistics, data science, and, of course, mathematics.

Since Newton, and, to an even greater extent following Einstein, modern science has been confronted with the inherent nonlinearity of the macroscopic universe. But most of our insight and progress is based on linear approximations. Moreover, at the atomic level, quantum mechanics remains an inherently linear theory. (The complete reconciliation of linear quantum theory with the nonlinear relativistic universe remains the holy grail of modern physics.) Only with the advent of large-scale computers have we been able to begin to investigate the full complexity of natural phenomena. But computers rely on numerical algorithms, and these in turn require manipulating and solving systems of algebraic equations. Now, rather than just a handful of equations, we may be confronted by gigantic systems containing thousands (or even millions) of unknowns. Without the discipline of linear algebra to formulate systematic, efficient solution algorithms, as well as the consequent insight into how to proceed when the numerical solution is insufficiently accurate, we would be unable to make progress in the linear regime, let alone make sense of the truly nonlinear physical universe.

Linear algebra can thus be viewed as the mathematical apparatus needed to solve potentially huge linear systems, to understand their underlying structure, and to apply what is learned in other contexts. The term “linear” is the key, and, in fact, it refers not just to linear algebraic equations, but also to linear differential equations, both ordinary and partial, linear boundary value problems, linear integral equations, linear iterative systems, linear control systems, and so on. It is a profound truth that, while outwardly different, all linear systems are remarkably similar at their core. Basic mathematical principles such as linear superposition, the interplay between homogeneous and inhomogeneous systems, the Fredholm alternative characterizing solvability, orthogonality, positive definiteness and minimization principles, eigenvalues and singular values, and linear iteration, to name but a few, reoccur in surprisingly many ostensibly unrelated contexts.

In the late nineteenth and early twentieth centuries, mathematicians came to the realization that all of these disparate techniques could be subsumed in the edifice now known as linear algebra. Understanding, and, more importantly, exploiting the apparent similarities between, say, algebraic equations and differential equations, requires us to become more sophisticated — that is, more abstract — in our mode of thinking. The abstraction process distills the essence of the problem away from all its distracting particularities, and, seen in this light, all linear systems rest on a common mathematical framework. Don't be afraid! Abstraction is not new in your mathematical education. In elementary algebra, you already learned to deal with variables, which are the abstraction of numbers. Later, the abstract concept of a function formalized particular relations between variables, say distance, velocity, and time, or mass, acceleration, and force. In linear algebra, the abstraction is raised to yet a further level, in that one views apparently different types of objects (vectors, matrices, functions, . . .) and systems (algebraic, differential, integral, . . .) in a common conceptual framework. (And this is by no means the end of the mathematical abstraction process; modern category theory, [**37**], abstractly unites different conceptual frameworks.)

In applied mathematics, we do not introduce abstraction for its intrinsic beauty. Our ultimate purpose is to develop effective methods and algorithms for applications in science, engineering, computing, statistics, data science, etc. For us, abstraction is driven by the need for understanding and insight, and is justified only if it aids in the solution to real world problems and the development of analytical and computational tools. Whereas to the beginning student the initial concepts may seem designed merely to bewilder and confuse, one must reserve judgment until genuine applications appear. Patience and perseverance are vital. Once we have acquired some familiarity with basic linear algebra, significant, interesting applications will be readily forthcoming. In this text, we encounter graph theory and networks, mechanical structures, electrical circuits, quantum mechanics, the geometry underlying computer graphics and animation, signal and image processing, interpolation and approximation, dynamical systems modeled by linear differential equations, vibrations, resonance, and damping, probability and stochastic processes, statistics, data analysis, splines and modern font design, and a range of powerful numerical solution algorithms, to name a few. Further applications of the material you learn here will appear throughout your mathematical and scientific career.

This textbook has two interrelated pedagogical goals. The first is to explain basic techniques that are used in modern, real-world problems. But we have not written a mere mathematical cookbook — a collection of linear algebraic recipes and algorithms. We believe that it is important for the applied mathematician, as well as the scientist and engineer, not just to learn mathematical techniques and how to apply them in a variety of settings, but, even more importantly, to understand why they work and how they are derived from first principles. In our approach, applications go hand in hand with theory, each reinforcing and inspiring the other. To this end, we try to lead the reader through the reasoning that leads to the important results. We do not shy away from stating theorems and writing out proofs, particularly when they lead to insight into the methods and their range of applicability. We hope to spark that eureka moment, when you realize "Yes, of course! I could have come up with that if I'd only sat down and thought it out." Most concepts in linear algebra are not all that difficult at their core, and, by grasping their essence, not only will you know how to apply them in routine contexts, you will understand what may be required to adapt to unusual or recalcitrant problems. And, the further you go on in your studies or work, the more you realize that very few real-world problems fit neatly into the idealized framework outlined in a textbook. So it is (applied) mathematical reasoning and not mere linear algebraic technique that is the core and raison d'être of this text!

Applied mathematics can be broadly divided into three mutually reinforcing components. The first is modeling — how one derives the governing equations from physical principles. The second is solution techniques and algorithms — methods for solving the model equations. The third, perhaps least appreciated but in many ways most important, are the frameworks that incorporate disparate analytical methods into a few broad themes. The key paradigms of applied linear algebra to be covered in this text include

- Gaussian Elimination and factorization of matrices;
- linearity and linear superposition;
- span, linear independence, basis, and dimension;
- inner products, norms, and inequalities;
- compatibility of linear systems via the Fredholm alternative;
- positive definiteness and minimization principles;
- orthonormality and the Gram–Schmidt process;
- least squares solutions, interpolation, and approximation;
- linear functions and linear and affine transformations;
- eigenvalues and eigenvectors/eigenfunctions;
- singular values and principal component analysis;
- linear iteration, including Markov processes and numerical solution schemes;
- linear systems of ordinary differential equations, stability, and matrix exponentials;
- vibrations, quasi-periodicity, damping, and resonance; .

These are all interconnected parts of a very general applied mathematical edifice of remarkable power and practicality. Understanding such broad themes of applied mathematics is our overarching objective. Indeed, this book began life as a part of a much larger work, whose goal is to similarly cover the full range of modern applied mathematics, both linear and nonlinear, at an advanced undergraduate level. The second installment is now in print, as the first author's text on partial differential equations, [**61**], which forms a natural extension of the linear analytical methods and theoretical framework developed here, now in the context of the equilibria and dynamics of continuous media, Fourier analysis, and so on. Our inspirational source was and continues to be the visionary texts of Gilbert Strang, [**79**, **80**]. Based on students' reactions, our goal has been to present a more linearly ordered and less ambitious development of the subject, while retaining the excitement and interconnectedness of theory and applications that is evident in Strang's works.

Syllabi and Prerequisites

This text is designed for three potential audiences:

- A beginning, in-depth course covering the fundamentals of linear algebra and its applications for highly motivated and mathematically mature students.
- A second undergraduate course in linear algebra, with an emphasis on those methods and concepts that are important in applications.
- A beginning graduate-level course in linear mathematics for students in engineering, physical science, computer science, numerical analysuis, statistics, and even mathematical biology, finance, economics, social sciences, and elsewhere, as well as master's students in applied mathematics.

Although most students reading this book will have already encountered some basic linear algebra — matrices, vectors, systems of linear equations, basic solution techniques, etc. — the text makes no such assumptions. Indeed, the first chapter starts at the very beginning by introducing linear algebraic systems, matrices, and vectors, followed by very basic Gaussian Elimination. We do assume that the reader has taken a standard two year calculus sequence. One-variable calculus — derivatives and integrals — will be used without comment; multivariable calculus will appear only fleetingly and in an inessential way. The ability to handle scalar, constant coefficient linear ordinary differential equations is also assumed, although we do briefly review elementary solution techniques in Chapter 7. Proofs by induction will be used on occasion. But the most essential prerequisite is a certain degree of mathematical maturity and willingness to handle the increased level of abstraction that lies at the heart of contemporary linear algebra.

Survey of Topics

In addition to introducing the fundamentals of matrices, vectors, and Gaussian Elimination from the beginning, the initial chapter delves into perhaps less familiar territory, such as the (permuted) LU and LDV decompositions, and the practical numerical issues underlying the solution algorithms, thereby highlighting the computational efficiency of Gaussian Elimination coupled with Back Substitution versus methods based on the inverse matrix or determinants, as well as the use of pivoting to mitigate possibly disastrous effects of numerical round-off errors. Because the goal is to learn practical algorithms employed in contemporary applications, matrix inverses and determinants are de-emphasized — indeed, the most efficient way to compute a determinant is via Gaussian Elimination, which remains *the* key algorithm throughout the initial chapters.

Chapter 2 is the heart of linear algebra, and a successful course rests on the students' ability to assimilate the absolutely essential concepts of vector space, subspace, span, linear independence, basis, and dimension. While these ideas may well have been encountered in an introductory ordinary differential equation course, it is rare, in our experience, that students at this level are at all comfortable with them. The underlying mathematics is not particularly difficult, but enabling the student to come to grips with a new level of abstraction remains the most challenging aspect of the course. To this end, we have included a wide range of illustrative examples. Students should start by making sure they understand how a concept applies to vectors in Euclidean space $\mathbb{R}^n$ before pressing on to less familiar territory. While one could design a course that completely avoids infinite-dimensional function spaces, we maintain that, at this level, they should be integrated into the subject right from the start. Indeed, linear analysis and applied mathematics, including Fourier methods, boundary value problems, partial differential equations, numerical solution techniques, signal processing, control theory, modern physics, especially quantum mechanics, and many, many other fields, both pure and applied, all rely on basic vector space constructions, and so learning to deal with the full range of examples is the secret to future success. Section 2.5 then introduces the fundamental subspaces associated with a matrix — kernel (null space), image (column space), coimage (row space), and cokernel (left null space) — leading to what is known as the Fundamental Theorem of Linear Algebra which highlights the remarkable interplay between a matrix and its transpose. The role of these spaces in the characterization of solutions to linear systems, e.g., the basic superposition principles, is emphasized. The final Section 2.6 covers a nice application to graph theory, in preparation for later developments.

Chapter 3 discusses general inner products and norms, using the familiar dot product and Euclidean distance as motivational examples. Again, we develop both the finite-dimensional and function space cases in tandem. The fundamental Cauchy–Schwarz inequality is easily derived in this abstract framework, and the more familiar triangle inequality, for norms derived from inner products, is a simple consequence. This leads to the definition of a general norm and the induced matrix norm, of fundamental importance in iteration, analysis, and numerical methods. The classification of inner products on Euclidean space leads to the important class of positive definite matrices. Gram matrices, constructed out of inner products of elements of inner product spaces, are a particularly fruitful source of positive definite and semi-definite matrices, and reappear throughout the text. Tests for positive definiteness rely on Gaussian Elimination and the connections between the LDL^T factorization of symmetric matrices and the process of completing the square in a quadratic form. We have deferred treating complex vector spaces until the final section of this chapter — only the definition of an inner product is not an evident adaptation of its real counterpart.

Chapter 4 exploits the many advantages of orthogonality. The use of orthogonal and orthonormal bases creates a dramatic speed-up in basic computational algorithms. Orthogonal matrices, constructed out of orthogonal bases, play a major role, both in geometry and graphics, where they represent rigid rotations and reflections, as well as in notable numerical algorithms. The orthogonality of the fundamental matrix subspaces leads to a linear algebraic version of the Fredholm alternative for compatibility of linear systems. We develop several versions of the basic Gram–Schmidt process for converting an arbitrary basis into an orthogonal basis, used in particular to construct orthogonal polynomials and functions. When implemented on bases of $\mathbb{R}^n$, the algorithm becomes the celebrated QR factorization of a nonsingular matrix. The final section surveys an important application to contemporary signal and image processing: the discrete Fourier representation of a sampled signal, culminating in the justly famous Fast Fourier Transform.

Chapter 5 is devoted to solving the most basic multivariable minimization problem: a quadratic function of several variables. The solution is reduced, by a purely algebraic computation, to a linear system, and then solved in practice by, for example, Gaussian Elimination. Applications include finding the closest element of a subspace to a given point, which is reinterpreted as the orthogonal projection of the element onto the subspace, and results in the least squares solution to an incompatible linear system. Interpolation of data points by polynomials, trigonometric function, splines, etc., and least squares approximation of discrete data and continuous functions are thereby handled in a common conceptual framework.

Chapter 6 covers some striking applications of the preceding developments in mechanics and electrical circuits. We introduce a general mathematical structure that governs a wide range of equilibrium problems. To illustrate, we start with simple mass–spring chains, followed by electrical networks, and finish by analyzing the equilibrium configurations and the stability properties of general structures. Extensions to continuous mechanical and electrical systems governed by boundary value problems for ordinary and partial differential equations can be found in the companion text [**61**].

Chapter 7 delves into the general abstract foundations of linear algebra, and includes significant applications to geometry. Matrices are now viewed as a particular instance of linear functions between vector spaces, which also include linear differential operators, linear integral operators, quantum mechanical operators, and so on. Basic facts about linear systems, such as linear superposition and the connections between the homogeneous and inhomogeneous systems, which were already established in the algebraic context, are shown to be of completely general applicability. Linear functions and slightly more general affine functions on Euclidean space represent basic geometrical transformations — rotations, shears, translations, screw motions, etc. — and so play an essential role in modern computer graphics, movies, animation, gaming, design, elasticity, crystallography, symmetry, etc. Further, the elementary transpose operation on matrices is viewed as a particular case of the adjoint operation on linear functions between inner product spaces, leading to a general theory of positive definiteness that characterizes solvable quadratic minimization problems, with far-reaching consequences for modern functional analysis, partial differential equations, and the calculus of variations, all fundamental in physics and mechanics.

Chapters 8–10 are concerned with eigenvalues and their many applications, including data analysis, numerical methods, and linear dynamical systems, both continuous and discrete. After motivating the fundamental definition of eigenvalue and eigenvector through the quest to solve linear systems of ordinary differential equations, the remainder of Chapter 8 develops the basic theory and a range of applications, including eigenvector bases, diagonalization, the Schur decomposition, and the Jordan canonical form. Practical computational schemes for determining eigenvalues and eigenvectors are postponed until Chapter 9. The final two sections cover the singular value decomposition and principal component analysis, of fundamental importance in modern statistical analysis and data science.

Chapter 9 employs eigenvalues to analyze discrete dynamics, as governed by linear iterative systems. The formulation of their stability properties leads us to define the spectral radius and further develop matrix norms. Section 9.3 contains applications to Markov chains arising in probabilistic and stochastic processes. We then discuss practical alternatives to Gaussian Elimination for solving linear systems, including the iterative Jacobi, Gauss–Seidel, and Successive Over–Relaxation (SOR) schemes, as well as methods for computing eigenvalues and eigenvectors including the Power Method and its variants, and the striking QR algorithm, including a new proof of its convergence. Section 9.6 introduces more recent semi-direct iterative methods based on Krylov subspaces that are increasingly employed to solve the large sparse linear systems arising in the numerical solution of partial differential equations and elsewhere: Arnoldi and Lanczos methods, Conjugate Gradients (CG), the Full Orthogonalization Method (FOM), and the Generalized Minimal Residual Method (GMRES). The chapter concludes with a short introduction to wavelets, a powerful modern alternative to classical Fourier analysis, now used extensively throughout signal processing and imaging science.

The final Chapter 10 applies eigenvalues to linear dynamical systems modeled by systems of ordinary differential equations. After developing basic solution techniques, the focus shifts to understanding the qualitative properties of solutions and particularly the role of eigenvalues in the stability of equilibria. The two-dimensional case is discussed in full detail, culminating in a complete classification of the possible phase portraits and stability properties. Matrix exponentials are introduced as an alternative route to solving first order homogeneous systems, and are also applied to solve the inhomogeneous version, as well as to geometry, symmetry, and group theory. Our final topic is second order linear systems, which model dynamical motions and vibrations in mechanical structures and electrical circuits. In the absence of frictional damping and instabilities, solutions are quasiperiodic combinations of the normal modes. We finish by briefly discussing the effects of damping and of periodic forcing, including its potentially catastrophic role in resonance.

Course Outlines

Our book includes far more material than can be comfortably covered in a single semester; a full year's course would be able to do it justice. If you do not have this luxury, several possible semester and quarter courses can be extracted from the wealth of material and applications.

First, the core of basic linear algebra that all students should know includes the following topics, which are indexed by the section numbers where they appear:

- Matrices, vectors, Gaussian Elimination, matrix factorizations, Forward and Back Substitution, inverses, determinants: 1.1–1.6, 1.8–1.9.
- Vector spaces, subspaces, linear independence, bases, dimension: 2.1–2.5.
- Inner products and their associated norms: 3.1–3.3.
- Orthogonal vectors, bases, matrices, and projections: 4.1–4.4.
- Positive definite matrices and minimization of quadratic functions: 3.4–3.5, 5.2
- Linear functions and linear and affine transformations: 7.1–7.3.
- Eigenvalues and eigenvectors: 8.2–8.3.
- Linear iterative systems: 9.1–9.2.

With these in hand, a variety of thematic threads can be extracted, including:

- Minimization, least squares, data fitting and interpolation: 4.5, 5.3–5.5.
- Dynamical systems: 8.4, 8.6 (Jordan canonical form), 10.1–10.4.
- Engineering applications: Chapter 6, 10.1–10.2, 10.5–10.6.
- Data analysis: 5.3–5.5, 8.5, 8.7–8.8.
- Numerical methods: 8.6 (Schur decomposition), 8.7, 9.1–9.2, 9.4–9.6.
- Signal processing: 3.6, 5.6, 9.7.
- Probabilistic and statistical applications: 8.7–8.8, 9.3.
- Theoretical foundations of linear algebra: Chapter 7.

For a first semester or quarter course, we recommend covering as much of the core as possible, and, if time permits, at least one of the threads, our own preference being the material on structures and circuits. One option for streamlining the syllabus is to concentrate on finite-dimensional vector spaces, bypassing the function space material, although this would deprive the students of important insight into the full scope of linear algebra.

For a second course in linear algebra, the students are typically familiar with elementary matrix methods, including the basics of matrix arithmetic, Gaussian Elimination, determinants, inverses, dot product and Euclidean norm, eigenvalues, and, often, first order systems of ordinary differential equations. Thus, much of Chapter 1 can be reviewed quickly. On the other hand, the more abstract fundamentals, including vector spaces, span, linear independence, basis, and dimension are, in our experience, still not fully mastered, and one should expect to spend a significant fraction of the early part of the course covering these essential topics from Chapter 2 in full detail. Beyond the core material, there should be time for a couple of the indicated threads depending on the audience and interest of the instructor.

Similar considerations hold for a beginning graduate level course for scientists and engineers. Here, the emphasis should be on applications required by the students, particularly numerical methods and data analysis, and function spaces should be firmly built into the class from the outset. As always, the students' mastery of the first five sections of Chapter 2 remains of paramount importance.

Comments on Individual Chapters

Chapter 1: On the assumption that the students have already seen matrices, vectors, Gaussian Elimination, inverses, and determinants, most of this material will be review and should be covered at a fairly rapid pace. On the other hand, the LU decomposition and the emphasis on solution techniques centered on Forward and Back Substitution, in contrast to impractical schemes involving matrix inverses and determinants, might be new. Sections 1.7, on the practical/numerical aspects of Gaussian Elimination, is optional.

Chapter 2: The crux of the course. A key decision is whether to incorporate infinite-dimensional vector spaces, as is recommended and done in the text, or to have an abbreviated syllabus that covers only finite-dimensional spaces, or, even more restrictively, only $\mathbb{R}^n$ and subspaces thereof. The last section, on graph theory, can be skipped unless you plan on covering Chapter 6 and (parts of) the final sections of Chapters 9 and 10.

Chapter 3: Inner products and positive definite matrices are essential, but, under time constraints, one can delay Section 3.3, on more general norms, as they begin to matter only in the later stages of Chapters 8 and 9. Section 3.6, on complex vector spaces, can be deferred until the discussions of complex eigenvalues, complex linear systems, and real and complex solutions to linear iterative and differential equations; on the other hand, it is required in Section 5.6, on discrete Fourier analysis.

Chapter 4: The basics of orthogonality, as covered in Sections 4.1–4.4, should be an essential part of the students' training, although one can certainly omit the final subsection in Sections 4.2 and 4.3. The final section, on orthogonal polynomials, is optional.

Chapter 5: We recommend covering the solution of quadratic minimization problems and at least the basics of least squares. The applications — approximation of data, interpolation and approximation by polynomials, trigonometric functions, more general functions, and splines, etc., are all optional, as is the final section on discrete Fourier methods and the Fast Fourier Transform.

Chapter 6 provides a welcome relief from the theory for the more applied students in the class, and is one of our favorite parts to teach. While it may well be skipped, the material is particularly appealing for a class with engineering students. One could specialize to just the material on mass/spring chains and structures, or, alternatively, on electrical circuits with the connections to spectral graph theory, based on Section 2.6, and further developed in Section 8.7.

Chapter 7: The first third of this chapter, on linear functions, linear and affine transformations, and geometry, is part of the core. This remainder of the chapter recasts many of the linear algebraic techniques already encountered in the context of matrices and vectors in Euclidean space in a more general abstract framework, and could be skimmed over or entirely omitted if time is an issue, with the relevant constructions introduced in the context of more concrete developments, as needed.

Chapter 8: Eigenvalues are absolutely essential. The motivational material based on solving systems of differential equations in Section 8.1 can be skipped over. Sections 8.2 and 8.3 are the heart of the matter. Of the remaining sections, the material on symmetric matrices should have the highest priority, leading to singular values and principal component analysis and a variety of numerical methods.

Chapter 9: If time permits, the first two sections are well worth covering. For a numerically oriented class, Sections 9.4–9.6 would be a priority, whereas Section 9.3 studies Markov processes — an appealing probabilistic/stochastic application. The chapter concludes with an optional introduction to wavelets, which is somewhat off-topic, but nevertheless serves to combine orthogonality and iterative methods in a compelling and important modern application.

Chapter 10 is devoted to linear systems of ordinary differential equations, their solutions, and their stability properties. The basic techniques will be a repeat to students who have already taken an introductory linear algebra and ordinary differential equations course, but the more advanced material will be new and of interest.

Changes from the First Edition

For the Second Edition, we have revised and edited the entire manuscript, correcting all known errors and typos, and, we hope, not introducing any new ones! Some of the existing material has been rearranged. The most significant change is having moved the chapter on orthogonality to before the minimization and least squares chapter, since orthogonal vectors, bases, and subspaces, as well as the Gram–Schmidt process and orthogonal projection play an absolutely fundamental role in much of the later material. In this way, it is easier to skip over Chapter 5 with minimal loss of continuity. Matrix norms now appear much earlier in Section 3.3, since they are employed in several other locations. The second major reordering is to switch the chapters on iteration and dynamics, in that the former is more attuned to linear algebra, while the latter is oriented towards analysis. In the same vein, space constraints compelled us to delete the last chapter of the first edition, which was on boundary value problems. Although this material serves to emphasize the importance of the abstract linear algebraic techniques developed throughout the text, now extended to infinite-dimensional function spaces, the material contained therein can now all be found in the first author's Springer Undergraduate Text in Mathematics, *Introduction to Partial Differential Equations*, [**61**], with the exception of the subsection on splines, which now appears at the end of Section 5.5.

There are several significant additions:

- In recognition of their increasingly essential role in modern data analysis and statistics, Section 8.7, on singular values, has been expanded, continuing into the new Section 8.8, on Principal Component Analysis, which includes a brief introduction to basic statistical data analysis.
- We have added a new Section 9.6, on Krylov subspace methods, which are increasingly employed to devise effective and efficient numerical solution schemes for sparse linear systems and eigenvalue calculations.
- Section 8.4 introduces and characterizes invariant subspaces, in recognition of their importance to dynamical systems, both finite- and infinite-dimensional, as well as linear iterative systems, and linear control systems. (Much as we would have liked also to add material on linear control theory, space constraints ultimately interfered.)
- We included some basics of spectral graph theory, of importance in contemporary theoretical computer science, data analysis, networks, imaging, etc., starting in Section 2.6 and continuing to the graph Laplacian, introduced, in the context of electrical networks, in Section 6.2, along with its spectrum — eigenvalues and singular values — in Section 8.7.

- We decided to include a short Section 9.7, on wavelets. While this perhaps fits more naturally with Section 5.6, on discrete Fourier analysis, the convergence proofs rely on the solution to an iterative linear system and hence on preceding developments in Chapter 9.
- A number of new exercises have been added, in the new sections and also scattered throughout the text.

Following the advice of friends, colleagues, and reviewers, we have also revised some of the less standard terminology used in the first edition to bring it closer to the more commonly accepted practices. Thus "range" is now "image" and "target space" is now "codomain". The terms "special lower/upper triangular matrix" are now "lower/upper unitriangular matrix", thus drawing attention to their unipotence. On the other hand, the term "regular" for a square matrix admitting an LU factorization has been kept, since there is really no suitable alternative appearing in the literature. Finally, we decided to retain our term "complete" for a matrix that admits a complex eigenvector basis, in lieu of "diagonalizable" (which depends upon whether one deals in the real or complex domain), "semi-simple", or "perfect". This choice permits us to refer to a "complete eigenvalue", independent of the underlying status of the matrix.

Exercises and Software

Exercises appear at the end of almost every subsection, and come in a medley of flavors. Each exercise set starts with some straightforward computational problems to test students' comprehension and reinforce the new techniques and ideas. Ability to solve these basic problems should be thought of as a minimal requirement for learning the material. More advanced and theoretical exercises tend to appear later on in the set. Some are routine, but others are challenging computational problems, computer-based exercises and projects, details of proofs that were not given in the text, additional practical and theoretical results of interest, further developments in the subject, etc. Some will challenge even the most advanced student.

As a guide, some of the exercises are marked with special signs:

♢ indicates an exercise that is used at some point in the text, or is important for further development of the subject.

♡ indicates a project — usually an exercise with multiple interdependent parts.

♠ indicates an exercise that requires (or at least strongly recommends) use of a computer. The student could either be asked to write their own computer code in, say, MATLAB, MATHEMATICA, MAPLE, etc., or make use of pre-existing software packages.

♣ = ♠ + ♡ indicates a computer project.

Advice to instructors: Don't be afraid to assign only a couple of parts of a multi-part exercise. We have found the True/False exercises to be a particularly useful indicator of a student's level of understanding. Emphasize to the students that a full answer is not merely a T or F, but must include a detailed explanation of the reason, e.g., a proof, or a counterexample, or a reference to a result in the text, etc.

Conventions and Notations

Note: A full symbol and notation index can be found at the end of the book.

Equations are numbered consecutively within chapters, so that, for example, (3.12) refers to the 12^{th} equation in Chapter 3. Theorems, Lemmas, Propositions, Definitions, and Examples are also numbered consecutively within each chapter, using a common index. Thus, in Chapter 1, Lemma 1.2 follows Definition 1.1, and precedes Theorem 1.3 and Example 1.4. We find this numbering system to be the most conducive for navigating through the book.

References to books, papers, etc., are listed alphabetically at the end of the text, and are referred to by number. Thus, [**61**] indicates the 61^{st} listed reference, which happens to be the first author's partial differential equations text.

Q.E.D. is placed at the end of a proof, being the abbreviation of the classical Latin phrase *quod erat demonstrandum*, which can be translated as "what was to be demonstrated".

$\mathbb{R}, \mathbb{C}, \mathbb{Z}, \mathbb{Q}$ denote, respectively, the real numbers, the complex numbers, the integers, and the rational numbers. We use $e \approx 2.71828182845904\ldots$ to denote the base of the natural logarithm, $\pi = 3.14159265358979\ldots$ for the area of a circle of unit radius, and i to denote the imaginary unit, i.e., one of the two square roots of -1, the other being $-\mathrm{i}$. The absolute value of a real number x is denoted by $|\,x\,|$; more generally, $|\,z\,|$ denotes the modulus of the complex number z.

We consistently use boldface lowercase letters, e.g., $\mathbf{v}, \mathbf{x}, \mathbf{a}$, to denote vectors (almost always column vectors), whose entries are the corresponding non-bold subscripted letter: v_1, x_i, a_n, etc. Matrices are denoted by ordinary capital letters, e.g., A, C, K, M — but not all such letters refer to matrices; for instance, V often refers to a vector space, L to a linear function, etc. The entries of a matrix, say A, are indicated by the corresponding subscripted lowercase letters, a_{ij} being the entry in its i^{th} row and j^{th} column.

We use the standard notations

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n, \qquad \prod_{i=1}^{n} a_i = a_1 a_2 \cdots a_n,$$

for the sum and product of the quantities $a_1, \ldots, a_n$. We use max and min to denote maximum and minimum, respectively, of a closed subset of $\mathbb{R}$. Modular arithmetic is indicated by $j = k \bmod n$, for $j, k, n \in \mathbb{Z}$ with $n > 0$, to mean $j - k$ is divisible by n.

We use $S = \{\, f \mid C \,\}$ to denote a set, where f is a formula for the members of the set and C is a list of conditions, which may be empty, in which case it is omitted. For example, $\{\, x \mid 0 \le x \le 1 \,\}$ means the closed unit interval from 0 to 1, also denoted $[0,1]$, while $\{\, a x^2 + b x + c \mid a, b, c \in \mathbb{R} \,\}$ is the set of real quadratic polynomials, and $\{0\}$ is the set consisting only of the number 0. We write $x \in S$ to indicate that x is an element of the set S, while $y \notin S$ says that y is not an element. The cardinality, or number of elements, in the set A, which may be infinite, is denoted by $\#A$. The union and intersection of the sets A, B are respectively denoted by $A \cup B$ and $A \cap B$. The subset notation $A \subset B$ includes the possibility that the sets might be equal, although for emphasis we sometimes write $A \subseteq B$, while $A \subsetneq B$ specifically implies that $A \neq B$. We can also write $A \subset B$ as $B \supset A$. We use $B \setminus A = \{\, x \mid x \in B, x \notin A \,\}$ to denote the set-theoretic difference, meaning all elements of B that do not belong to A.

An arrow $\rightarrow$ is used in two senses: first, to indicate convergence of a sequence: $x_n \rightarrow x^\star$ as $n \rightarrow \infty$; second, to indicate a function, so $f\colon X \rightarrow Y$ means that f defines a function from the domain set X to the codomain set Y, written $y = f(x) \in Y$ for $x \in X$. We use $\equiv$ to emphasize when two functions agree everywhere, so $f(x) \equiv 1$ means that f is the constant function, equal to 1 at all values of x. Composition of functions is denoted $f \circ g$.

Angles are always measured in radians (although occasionally degrees will be mentioned in descriptive sentences). All trigonometric functions, $\cos, \sin, \tan, \sec$, etc., are evaluated on radians. (Make sure your calculator is locked in radian mode!)

As usual, we denote the natural exponential function by e^x. We always use $\log x$ for its inverse — the natural (base e) logarithm (never the ugly modern version $\ln x$), while $\log_a x = \log x / \log a$ is used for logarithms with base a.

We follow the reference tome [**59**] (whose mathematical editor is the first author's father) and use $\operatorname{ph} z$ for the phase of a complex number. We prefer this to the more common term "argument", which is also used to refer to the argument of a function $f(z)$, while "phase" is completely unambiguous and hence to be preferred.

We will employ a variety of standard notations for derivatives. In the case of ordinary derivatives, the most basic is the Leibnizian notation $\dfrac{du}{dx}$ for the derivative of u with respect to x; an alternative is the Lagrangian prime notation u'. Higher order derivatives are similar, with u'' denoting $\dfrac{d^2u}{dx^2}$, while $u^{(n)}$ denotes the n^{th} order derivative $\dfrac{d^n u}{dx^n}$. If the function depends on time, t, instead of space, x, then we use the Newtonian dot notation, $\dot{u} = \dfrac{du}{dt}$, $\ddot{u} = \dfrac{d^2u}{dt^2}$. We use the full Leibniz notation $\dfrac{\partial u}{\partial x}, \dfrac{\partial u}{\partial t}, \dfrac{\partial^2 u}{\partial x^2}, \dfrac{\partial^2 u}{\partial x\, \partial t}$, for partial derivatives of functions of several variables. All functions are assumed to be sufficiently smooth that any indicated derivatives exist and mixed partial derivatives are equal, cf. [**2**].

Definite integrals are denoted by $\displaystyle\int_a^b f(x)\,dx$, while $\displaystyle\int f(x)\,dx$ is the corresponding indefinite integral or anti-derivative. In general, limits are denoted by $\displaystyle\lim_{x \to y}$, while $\displaystyle\lim_{x \to y^+}$ and $\displaystyle\lim_{x \to y^-}$ are used to denote the two one-sided limits in $\mathbb{R}$.

History and Biography

Mathematics is both a historical and a social activity, and many of the algorithms, theorems, and formulas are named after famous (and, on occasion, not-so-famous) mathematicians, scientists, engineers, etc. — usually, but not necessarily, the one(s) who first came up with the idea. We try to indicate first names, approximate dates, and geographic locations of most of the named contributors. Readers who are interested in additional historical details, complete biographies, and, when available, portraits or photos, are urged to consult the wonderful University of St. Andrews MacTutor History of Mathematics archive:

http://www-history.mcs.st-and.ac.uk

Some Final Remarks

To the student: You are about to learn modern applied linear algebra. We hope you enjoy the experience and profit from it in your future studies and career. (Indeed, we recommended holding onto this book to use for future reference.) Please send us your comments, suggestions for improvement, along with any errors you might spot. Did you find our explanations helpful or confusing? Were enough examples included in the text? Were the exercises of sufficient variety and at an appropriate level to enable you to learn the material?

To the instructor: Thank you for adopting our text! We hope you enjoy teaching from it as much as we enjoyed writing it. Whatever your experience, we want to hear from you. Let us know which parts you liked and which you didn't. Which sections worked and which were less successful. Which parts your students enjoyed, which parts they struggled with, and which parts they disliked. How can we improve it?

Like every author, we sincerely hope that we have written an error-free text. Indeed, all known errors in the first edition have been corrected here. On the other hand, judging from experience, we know that, no matter how many times you proofread, mistakes still manage to sneak through. So we ask your indulgence to correct the few (we hope) that remain. Even better, email us with your questions, typos, mathematical errors and obscurities, comments, suggestions, etc.

The second edition's dedicated web site

http://www.math.umn.edu/∼olver/ala2.html

will contain a list of known errors, commentary, feedback, and resources, as well as a number of illustrative Matlab programs that we've used when teaching the course. Links to the *Selected Solutions Manual* will also be posted there.

Acknowledgments

First, let us express our profound gratitude to Gil Strang for his continued encouragement from the very beginning of this undertaking. Readers familiar with his groundbreaking texts and remarkable insight can readily find his influence throughout our book. We thank Pavel Belik, Tim Garoni, Donald Kahn, Markus Keel, Cristina Santa Marta, Nilima Nigam, Greg Pierce, Fadil Santosa, Wayne Schmaedeke, Jackie Shen, Peter Shook, Thomas Scofield, and Richard Varga, as well as our classes and students, particularly Taiala Carvalho, Colleen Duffy, and Ryan Lloyd, and last, but certainly not least, our late father/father-in-law Frank W.J. Olver and son Sheehan Olver, for proofreading, corrections, remarks, and useful suggestions that helped us create the first edition. We acknowledge Mikhail Shvartsman's contributions to the arduous task of writing out the solutions manual. We also acknowledge the helpful feedback from the reviewers of the original manuscript: Augustin Banyaga, Robert Cramer, James Curry, Jerome Dancis, Bruno Harris, Norman Johnson, Cerry Klein, Doron Lubinsky, Juan Manfredi, Fabio Augusto Milner, Tzuong-Tsieng Moh, Paul S. Muhly, Juan Carlos Álvarez Paiva, John F. Rossi, Brian Shader, Shagi-Di Shih, Tamas Wiandt, and two anonymous reviewers.

We thank many readers and students for their strongly encouraging remarks, that cumulatively helped inspire us to contemplate making this new edition. We would particularly like to thank Nihat Bayhan, Joe Benson, James Broomfield, Juan Cockburn, Richard Cook, Stephen DeSalvo, Anne Dougherty, Ken Driessel, Kathleen Fuller, Mary Halloran, Stuart Hastings, David Hiebeler, Jeffrey Humpherys, Roberta Jaskolski, Tian-Jun Li, James Meiss, Willard Miller, Jr., Sean Rostami, Arnd Scheel, Timo Schürg, David Tieri, Peter Webb, Timothy Welle, and an anonymous reviewer for their comments on, suggestions for, and corrections to the three printings of the first edition that have led to this improved second edition. We particularly want to thank Linda Ness for extensive help with the sections on SVD and PCA, including suggestions for some of the exercises. We also thank David Kramer for his meticulous proofreading of the text.

And of course, we owe an immense debt to Loretta Bartolini and Achi Dosanjh at Springer, first for encouraging us to take on a second edition, and then for their willingness to work with us to produce the book you now have in hand — especially Loretta's unwavering support, patience, and advice during the preparation of the manuscript, including encouraging us to adopt and helping perfect the full-color layout, which we hope you enjoy.

Peter J. Olver
University of Minnesota
olver@umn.edu

Cheri Shakiban
University of St. Thomas
cshakiban@stthomas.edu

Minnesota, March 2018

Table of Contents