
Preface

The purpose of the book

This book was designed as a textbook for a junior-senior level second course in linear algebra at the University of Texas at Austin. In that course, and in this book, I try to show math, physics, and engineering majors the incredible power of linear algebra in the real world. The hope is that, when faced with a linear system (or a nonlinear system that can be reasonably linearized), future engineers will think to decompose the system into modes that they can understand. Usually this is done by diagonalization. Sometimes this is done by decomposing into a convenient orthonormal basis, such as Fourier series. Sometimes a continuous decomposition, into δ functions or by Fourier transforms, is called for. The underlying ideas of breaking a vector into modes (the Superposition Principle) and of decoupling a complicated system by a suitable choice of linear coordinates (the Decoupling Principle) appear throughout physics and engineering. My goal is to impress upon students the importance of these principles, while giving them enough tools to use them effectively.

There are many existing types of second linear algebra courses, and many books to match, but few if any make this goal a priority. Some courses are theoretical, going in the direction of functional analysis, Lie Groups or abstract algebra. “Applied” second courses tend to be heavily numerical, teaching efficient and robust algorithms for factorizing or diagonalizing matrices. Some courses split the difference, developing matrix theory in depth, proving classification theorems (e.g., Jordan form) and estimates (e.g., Gershgorin’s Theorem). While each of these courses is well-suited for its chosen

audience, none give a prospective physicist or engineer substantial insight into how or why to apply linear algebra at all.

Notes to the instructor. The readers of this book are assumed to have taken an introductory linear algebra class, and hence to be familiar with basic matrix operations such as row reduction, matrix multiplication and inversion, and taking determinants. The reader is also assumed to have had some exposure to vector spaces and linear transformations. This material is reviewed in Chapters 2 and 3, pretty much from the beginning, but a student who has never seen an abstract vector space will have trouble keeping up. The subject of Chapter 4, eigenvalues, is typically covered quite hastily at the end of a first course (if at all), so I work under the assumption that readers do not have any prior knowledge of eigenvalues.

The key concept of these introductory chapters is that a basis makes a vector space look like \mathbb{R}^n (or sometimes \mathbb{C}^n) and makes linear transformations look like matrices. Some bases make the conversion process simple, while others make the end results simple. The standard basis in \mathbb{R}^n makes coordinates easy to find, but may result in an operator being represented by an ugly matrix. A basis of eigenvectors, on the other hand, makes the operator appear simple but makes finding the coordinates of a vector difficult. To handle problems in linear algebra, one must be adept in coordinatization and in performing change-of-basis operations, both for vectors and for operators.

One premise of this book is that standard software packages (e.g., MATLAB, Maple or Mathematica) make it easy to diagonalize matrices without any knowledge of sophisticated numerical algorithms. This frees us to consider the *use* of diagonalization, and some general features of important classes of operators (e.g., Hermitian or unitary operators). Diagonalization, by computer or by hand, gives a set of coordinates in which a problem, even a problem with an infinite number of degrees of freedom, decouples into a collection of independent scalar equations. This is what I call the Decoupling Principle.

(Strictly speaking this is only true for diagonalizable operators. However, a matrix or operator coming from the real world is almost certainly diagonalizable, especially since Hermitian and unitary matrices are *always* diagonalizable. For completeness, I have included sections in the book about nondiagonalizable matrices, power vectors, and Jordan form, but these issues are not stressed, and these sections can be skipped with little loss of continuity.)

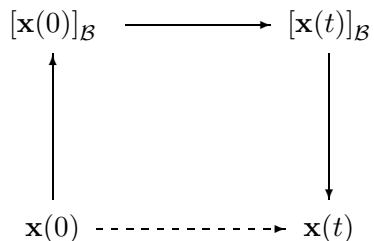


Figure 0.1. The decoupling strategy

The Decoupling Principle is first applied systematically in Chapter 5, where we consider a variety of coupled linear differential equations or difference equations. Students may have seen some of these problems in previous courses on differential equations, probability or classical mechanics, but typically have not understood that the right choice of coordinates (achieved by using a basis of eigenvectors) is independent of the type of problem. By presenting a sequence of problems solved via the Decoupling Principle, this point is driven home. It is also hoped that the examples are of interest in their own right, and provide an applied counterweight to the fairly theoretical introductory chapters.

In Chapter 5, students are also exposed to questions of linear and nonlinear stability. They learn to linearize nonlinear equations near fixed points, and to use their stability calculations to determine for how long linearized equations can adequately model an underlying nonlinear problem. These are questions of crucial importance to physicists and engineers.

Up through Chapter 5, our calculational model is as follows (see Figure 0.1). To solve a time evolution problem (say, $d\mathbf{x}/dt = L\mathbf{x}$) we find a basis \mathcal{B} of eigenvectors of L . We then convert the initial vector $\mathbf{x}(0)$ into coordinates $[\mathbf{x}(0)]_{\mathcal{B}}$, compute the coordinates $[\mathbf{x}(t)]_{\mathcal{B}}$ of the vector at a later time, and from that reconstitute the vector $\mathbf{x}(t)$. The basis of eigenvectors makes the middle (horizontal) step easy, but the vertical steps, especially finding the coordinates $[\mathbf{x}(0)]_{\mathcal{B}}$, can be difficult.

In Chapter 6 we introduce inner products and see how coordinatization becomes easy if our basis is orthogonal. Fourier series on L^2 of an interval is then a natural consequence. Chapter 6 also contains several subjects that, while interesting, may not fit into a course syllabus.

In Section 6.1, the subsection on nonstandard inner products may be skipped if desired. In Section 6.3, the general discussion of dual spaces is included mostly for reference, and may also be skipped, especially as many students find this material to be quite difficult. However, the beginning of the section should be covered thoroughly. It is important for students to

understand that $|\mathbf{v}\rangle$ is a vector, while $\langle\mathbf{w}|$ is an operation, namely “take the inner product with \mathbf{w} ”. They should also understand the representation of bras and kets as rows and columns, respectively.

Section 6.7 on least squares also deserves comment. This topic is off the theme of the Decoupling Principle, but is far too useful to leave out. Instructors should feel free to spend as much or as little time on this digression as they see fit.

Chapter 5 demonstrates the utility of bases of eigenvectors, Chapter 6 demonstrates the utility of orthogonal bases, and Chapter 7 reconciles the two approaches, showing how several classes of important operators are diagonalizable with orthogonal eigenvectors. Fourier series, introduced in Chapter 6 as an expansion in an orthogonal basis, can then be reconsidered as an eigenfunction expansion for the Laplacian.

Another important premise of this book is that infinite dimensional systems are important, but that a full treatment of Banach spaces (or even just Hilbert spaces) would only distract students from the Decoupling Principle. The last third of the book is devoted to infinite dimensional problems (e.g., the wave equation in $1+1$ dimensions), with the idea of transferring intuition from finite to infinite dimensions. My attitude is summarized in the advice to the student at the end of Section 3.4, where infinite dimensional spaces are first introduced:

In short, infinite dimensional spaces and infinite dimensional operations are neither totally bizarre nor totally tame, but somewhere in between. If an argument or technique works in finite dimensions, it is probable, but by no means certain, that it will work in infinite dimensions. As a first approximation, applying your finite dimensional intuition to infinite dimensions is a very good idea. However, you should be prepared for an occasional surprise, almost always due to a lack of convergence of some sum.

Chapter 8 is the infinite dimensional sequel to Chapter 5, using the wave equation to demonstrate the Decoupling Principle for partial differential equations. The key idea is to think of a scalar-valued partial differential equation as an ordinary differential equation on an infinite dimensional vector space. The results of Chapter 5 then carry over directly to give the general solution to the vibrating string problem in terms of standing waves. The wave equation can also be attacked in different ways, each demonstrating a different linear algebraic principle. Solving the wave equation on the whole line in terms of forward and backward traveling waves involves both the Superposition Principle and the properties of commuting operators. The wave equation on the half line gives us the method of images. Comparing the standing wave and traveling wave solutions to the vibrating string leads us naturally to consider two kinds of Fourier series on an interval $[0, L]$; the

first in terms of $\sin(n\pi x/L)$, the second in terms of $\exp(2\pi i n x/L)$. Both are eigenfunction expansions for Hermitian operators, the first for the Laplacian with Dirichlet boundary conditions, the second for $-id/dx$ with periodic boundary conditions.

In Chapter 9 we make the transition from discrete to continuous spectra, introducing the Dirac δ function and expansions that involve integrating over generalized eigenfunctions. Fourier transforms (Chapter 10) then naturally appear as generalized eigenfunction expansions for the “momentum” operator $-id/dx$.

Finally, once we have a generalized basis of δ functions, we can decompose with respect to that basis to get integral kernels (a.k.a. Green’s functions) for linear operators. Like the earlier discussion of least squares, this is a departure from the central theme of the book, but is much too useful to leave out.

Possible course outlines. There are three recommended courses that can be built from this book. The first 8 chapters, with an occasional section skipped, forms a coherent one semester course on diagonalization and on infinite dimensional problems with discrete spectra. This is essentially the course I have taught at the University of Texas at Austin. For such a course Chapters 2 and 3 should be presented quickly, as only the last section or two of each chapter is likely to be new material.

For universities on the quarter system, the entire book can be used for the second and third quarters of a year-long linear algebra course. In that case I recommend that the first quarter concentrate on matrix manipulations, solutions to linear equations, and vector space properties of \mathbb{R}^n and its subspaces. (E.g., the first four chapters of David Lay’s excellent text.) There is no need to discuss eigenvalues or inner products at all in the first quarter, as they are covered from scratch in Chapters 4 and 6, respectively. In such a sequence, Chapters 2 and 3 would be treated as new material and presented slowly, not as review material to be skimmed.

There are several sections (2.5, 3.4, 4.7, 4.9, 5.6, 5.7, 6.7, 6.8, 6.9) that can be skipped without too much loss of continuity. Some of these sections (especially 5.7: Linearization of nonlinear problems, 6.7: Least squares, and 6.9: Fourier series) are of tremendous importance in their own right and should be learned at some point, but it is certainly possible to construct a course without them. Which of these to include and which to skip is largely a matter of course pace and instructor taste.

As a third option, the first seven chapters of this book can make a substantial first course in linear algebra for strong students who have already learned about row reduction and matrix algebra in high school. For

such a course I recommend emphasizing finite dimensional applications (e.g., Markov chains and least squares) and de-emphasizing infinite dimensional extensions.

Finally, this book can be used for self-study by advanced undergraduate or beginning graduate students who need more linear algebra than is typically taught in a first course. Chapters 6, 7, 9, and 10 are of particular interest to physics students struggling with the formalism of quantum mechanics, Chapter 11 to physics and engineering students studying electromagnetism, and Chapters 7–11 to students of applied math and functional analysis.

To serve the needs of such students, the last three chapters are written at a more sophisticated level than the earlier chapters. They are logically self-contained, treating each subject from scratch, but assume a significant background in general mathematics. For example, to appreciate Fourier transforms, it helps to be adept at computing them, and that often means doing contour integrals. These chapters are probably most useful to students who have been exposed to Fourier transforms and/or Green's functions in their physics and engineering coursework, but who lack a conceptual framework for these subjects.

Notes to the student. You will probably find the beginning of the book to be largely review. You may have seen much of Chapters 2 (Vector spaces) and 3 (Linear transformations) and some of Chapter 4 (Eigenvalues) in a first linear algebra course, but I do not assume that you have mastered these concepts. As befits review material, most of the concepts are presented quickly from the beginning. If you thoroughly understood your first course, you should be able to skim these chapters, concentrating on the last section or two of each. On the other hand, if your first course was not enough preparation, you should take the time to go through Chapters 2 and 3 carefully, and work out many of the problems. It's worth the extra effort, as the entire book depends strongly on the ideas of Chapters 2 and 3.

I typically present each major concept in three settings. The first setting is in \mathbb{R}^n , where the problem is essentially a (frequently familiar) matrix computation. The second setting is in a general n -dimensional vector space, where a choice of basis reduces the problem to one on \mathbb{R}^n . The key is to choose the right basis, and I put considerable emphasis on understanding what stays the same and what changes when you change basis. The third setting is in an infinite dimensional vector space. The goal is not to develop a general theory, but for you to see enough examples to start building up intuition. Infinite dimensional spaces appear more and more often later in the book.

While (almost) all results in finite dimensions are proven, most infinite dimensional theorems (such as the spectral theorem for bounded self-adjoint operators on a Hilbert space) are merely stated, and in some cases I just argue formally, by analogy to finite dimensions. Although such analogies can sometimes fail (I give examples), they are a very good intuitive starting point.

This book is aimed at a mixture of math, physics, computer science, engineering, and economics majors. The only absolute prerequisite is familiarity with matrix manipulation (Gaussian elimination, inverses, determinants, etc.). These topics are invariably covered in a first linear algebra course, but can also be picked up elsewhere. Basic vector space concepts are covered from the beginning in Chapters 2 and 3. However, if you have never seen a vector space (not even \mathbb{R}^n), you may have difficulty keeping up. These chapters are aimed primarily at students who once were exposed to vector spaces but may be rusty. In Chapters 4–8, no prior knowledge is expected or required. Chapters 9–11 are generally more sophisticated, but are self-contained, with no prior knowledge of the subject assumed.

Finally, software such as MATLAB, Mathematica and Maple can make many linear algebra computations much easier, and I encourage you to use them. With technology, you can avoid the drudgery of, say, computing the eigenvalues and eigenvectors of a large matrix by hand. However, technology should enhance your thinking, not replace it! A computer can tell you what the eigenvalues of a matrix are, but it's up to you to figure out what to do with them.

Notation and terminology. Many objects in linear algebra are referred to by different names by mathematicians, physicists and engineers. What's worse, the same terms are often used by the different communities to mean different things. As Winston Churchill said of Americans and Englishmen, *We are divided by a common language.*

Here are some key differences. Others are noted in the text. In each case I have adopted the physics conventions and terminology, especially the terminology of quantum mechanics.

- Mathematicians usually denote their complex inner products (\mathbf{x}, \mathbf{y}) , linear in \mathbf{x} and conjugate-linear in \mathbf{y} . Physicists use Dirac's bra-ket notation $\langle \mathbf{x} | \mathbf{y} \rangle$, linear in \mathbf{y} and conjugate-linear in \mathbf{x} , and refer to the individual factors $\langle \mathbf{x} |$ and $|\mathbf{y} \rangle$ as “bras” and “kets”, respectively.
- An eigenvalue whose geometric multiplicity is greater than one is called “repeated” or “multiple” by mathematicians and “degenerate” by physicists.

- An eigenvalue whose algebraic multiplicity is greater than its geometric multiplicity is called “degenerate” by mathematicians and “deficient” by physicists.
- If A is an operator, ξ is a vector and $(A - \lambda I)^p \xi = 0$ for some exponent p , then ξ is often called a “generalized eigenvector” corresponding to the eigenvalue λ . I prefer Sheldon Axler’s term “power vector”, and use “generalized eigenvector” in the context of continuous spectrum to mean a non-normalizable eigenfunction. (E.g., on the real line, e^{ikx} is a generalized eigenfunction of $-id/dx$ with generalized eigenvalue k .)

Acknowledgments. This book grew out of lecture notes for a new class in Applied Linear Algebra at the University of Texas at Austin. I thank Karen Uhlenbeck and Jerry Bona for suggesting that I develop this class and to Martin Speight for helping to design it. The students in this class, number M375 in Fall 1997 and M346 in Fall 1998, provided invaluable help, both as test subjects and as editors. In particular, their feedback led to a much needed overhaul of the early chapters.

Parts of this book were written at the Park City Math Institute in 1997 and 1998. In both summers, the participants in the Undergraduate Faculty Program were generous with their expertise and insight into what would or would not work in an actual classroom. I especially thank Jane Day, Daniel Goroff, Dan Kalman and John Polking. This book was completed at the Physics Department of the Israel Institute of Technology (Technion).

Finally, I am especially grateful to Margaret Combs, whose artistry, \TeX technical expertise and many hours of work are responsible for most of the figures.

Lorenzo Sadun
sadun@math.utexas.edu
June, 2000