Preface

A reasonable question for any author is the seemingly innocuous "Why did you write it?" This is especially relevant for a mathematical text. After all, there aren't any new ground-breaking results here — the results in this book are all "well-known." (See for example Lax [**24**], Meckes and Meckes [**28**], or Garcia and Horn's [**12**].) Why did I write it? The simple answer is, that it is a book that I wished I had had when I finished my undergraduate degree. I knew that I liked analysis and analytic methods, but I didn't know about the wide range of useful applications of analysis. It was only after I began to teach analysis that I learned about many of the useful results that can be proved by analytic methods. What do I mean by "analytic methods"? To me, an analytic method is any method that uses tools from analysis: convergence, inequalities, and compactness being very common ones. That means that, from my perspective, using the triangle inequality or the Cauchy-Schwarz-Bunyakovsky inequality means applying analytic methods. (As an aside, in grad school, my advisor referred to himself as a "card-carrying analyst", and so I too am an analyst.)

A much harder question to address is: what does "useful" mean? This is somewhat related to the following: when you hear a new result, what is your first reaction? Is it *"Why is it true?"* or *"What can I do with it?"* I definitely have the first thought, but many will have the second thought. For example, I think the Banach Fixed Point Theorem is useful, since it can be used to prove lots of other results (an existence and uniqueness theorem for initial value problems and the inverse function theorem). But many of those results require yet more machinery, and so students have to wait to see why the Banach Fixed Point Theorem is useful until we have that machinery. On the other hand, after having been told that math is useful for several years, students can be understandably dubious when being told that what they're learning is useful.

For the student: What should you get out of this book? First, a better appreciation of the "applicability" of the analytic tools you have, as well as a sense of how many of the basic ideas you know can be generalized. On a more itemized level, you will see how linear algebra and analysis can be used in several "data science" type problems: determining how close a given set of data is to a given subspace (the "best" subspace problem), how to solve least squares problems (the Moore-Penrose pseudo-inverse), how to best approximate a high rank object with a lower rank one (low rank approximation and the Eckart-Young-Mirsky Theorem), and how to find the best transformation that preserves angles and distances to compare a given data set to a reference one (the orthogonal Procrustes problem). As you read the text, you will find exercises — you should do them as you come to them, since they are intended to help strengthen and reinforce your understanding, and many of them will be helpful later on!

For the student and instructor: What is the topic here? The extraordinary utility of linear algebra and analysis. And there are many, many examples of that usefulness. One of the most "obvious" examples of the utility of linear algebra comes from Google's PageRank Algorithm, which has been covered extremely well by Langville and Meyer, in [**22**] (see also [**3**]). Our main topic is the Singular Value Decomposition (SVD). To quote from Golub and Van Loan [**13**], Section 2.4, "[t]he practical and theoretical importance of the SVD is hard to overestimate." There is a colossal number of examples of SVD's usefulness. (See for example the Netflix Challenge, which offered a million dollar prize for improving Netflix's recommendations by 10% and was won by a team which used the SVD.) What then justifies (at least in this author's mind) another book? Most of the application oriented books do not provide proofs (see my interest in "why is that true?") of the foundational parts, commonly saying "…as is well known…" Books that go deeply into the proofs tend more to the numerical linear algebra side of things, which are usually oriented to the (**incredibly important**) questions of how to efficiently and accurately calculate the SVD of a given matrix. Here, the emphasis is on the proof of the existence of the SVD, inequalities for singular values, and a few applications. For applications, I have chosen four: determining the "best" approximating subspace to a given collection of points, compression/approximation by low-rank matrices (for the operator and Frobenius norms), the Moore-Penrose pseudo-inverse, and a Procrustes-type problem asking for the orthogonal transformation that most closely transforms a given configuration to a reference configuration (as well as the closely related problem that adds the requirement of preserving orientation). Proofs are provided for the solutions of these problems, and each one uses analytic ideas (broadly construed). So, what is this book? A showcase of the utility of analytic methods in linear algebra, with an emphasis on the SVD.

What is it not? You will not find algorithms for calculating the SVD of a given matrix, nor any discussion of efficiency of such algorithms. Those questions are very difficult, and beyond the scope of this book. A standard reference for those questions is Golub and Van Loan's book [**13**]. Another reference which discusses the history of and the current (as of 2020) state of the art for algorithms computing the SVD is [**9**]. Dan Kalman's article [**20**] provides an excellent overview of the general idea of the SVD, as well as references to applications. For a deeper look into the history of the SVD, we suggest G. W. Stewart's article [**36**]. In addition, while we do consider four applications, we do not go into tremendous depth and cover all of the possible applications of the SVD. One major application that we do not discuss is Principal Component Analysis (PCA). PCA is a standard tool in statistics and is covered in [**15**] (among many other places, see also the references in [**16**]). SVD is also useful in actuarial science, where it is used in the Lee - Carter method [**25**] to make forecasts of life expectancy. One entertaining application is in analyzing cryptograms, see Moler and Morrison's article [**30**]. A few more fascinating applications (as well as references to many, many more) may be found in Martin and Porter's article [**26**]. My first exposure to the SVD was in Browder's analysis text [**6**]. There, the SVD was used to give a particularly slick proof that if $T : \mathbb{R}^n \to \mathbb{R}^n$ is linear, then $m\big(T(\Omega)\big) = |\det T| m(\Omega)$, where m is Lebesgue measure and Ω is any measurable set. The SVD can also be used for information retrieval, see for example [**3**] or [**42**]. For more applications of linear algebra (not "just" the SVD), we suggest Elden's book [**11**], or Gil Strang's book [**37**].

Another book that shows some clever applications of linear algebra to a variety of mathematical topics is Matousek's book [**27**]. Finally, note that I make no claim that the list of references is in any way complete, and I apologize to the many experts whose works I have missed. I have tried to reference surveys whose references will hopefully be useful for those who wish to dig deeper.

What is in this book? Chapter 1 starts with a quick review of the linear algebra pre-requisites (vectors, vector spaces, bases, dimension, and subspaces). We then move on to a discussion of some applications of linear algebra that may not be familiar to a student with only a single linear algebra course. Here we discuss how matrices can be used to encode information, and how the structure provided by matrices allows us to find information with some simple matrix operations. We then discuss the four applications mentioned above: the approximating subspace problem, compression/approximation by low-rank matrices (for the operator and Frobenius norms), the Moore-Penrose pseudo-inverse, and a Procrustes-type problem asking for the orthogonal transformation that most closely transforms a given configuration to a reference configuration, as well as the orientation preserving orthogonal transformation that most closely transforms a given configuration to a given reference configuration.

Chapter 2 covers the background material necessary for the subsequent chapters. We begin with a discussion of the sum of subspaces, the formula $\dim(\mathcal{U}_1 + \mathcal{U}_2) = \dim \mathcal{U}_1 + \dim \mathcal{U}_2 - \dim(\mathcal{U}_1 \cap \mathcal{U}_2)$, as well as the Fundamental Theorem of Linear Algebra. We then turn to analytic tools: norms and inner products. We give important examples of norms and inner products on matrices. We then turn to associated analytic and topological concepts: continuity, open, closed, completeness, the Bolzano-Weierstrass Theorem, and sequential compactness.

Chapter 3 uses the tools from Chapter 2 to cover some of the fundamental ideas (orthonormality, projections, adjoints, orthogonal complements, etc.) involved in the four applications. We also cover the separation by a linear functional of two disjoint closed convex sets when one is also assumed to be bounded (in an inner-product space). We finish Chapter 3 with a short discussion of the Singular Value Decomposition and how it can be used to solve the four basic problems. The proofs that the solutions are what we claim are postponed to Chapter 6.

Chapter 4 is devoted to a proof of the Spectral Theorem, as well as the minimax and maximin characterizations of the eigenvalues. We also prove Weyl's inequalities about eigenvalues and an interlacing theorem. These are the basic tools of spectral graph theory, see for example [**7**] and [**8**].

Chapter 5 provides a proof of the Singular Value Decomposition, and gives two additional characterizations of the singular values. Then, we prove Weyl's inequalities for singular values.

Chapter 6 is devoted to proving the statements made at the end of Chapter 3 about the solutions to the four fundamental problems.

Finally, Chapter 7 takes a short glimpse towards changes in infinite dimensions, and provides examples where the infinite-dimensional behavior is different.

Pre-Requisites

It is assumed that readers have had a standard course in linear algebra and are familiar with the ideas of vector spaces (over $\mathbb{R}$), subspaces, bases, dimension, linear independence, matrices as linear transformations, rank of a linear transformation, and nullity of a linear transformation. We also assume that students are familiar with determinants, as well as eigenvalues and how to calculate them. Some familiarity with linear algebra software is useful, but not essential.

In addition, it is assumed that readers have had a course in basic analysis. (There is some debate as to what such a course should be called, with two common titles being "advanced calculus" or "real analysis.") To be more specific, students should know the definition of infimum and supremum for a non-empty set of real numbers, the basic facts about convergence of sequences, the Bolzano-Weierstrass Theorem (in the form that a bounded sequence of real numbers has a convergent subsequence), and the basic facts about continuous functions. (For a much more specific background, the first three chapters of [**31**] are sufficient.) Any reader familiar with metric spaces at the level of Rudin [**32**] is definitely prepared, although exposure to metric space topology is not necessary. We will work with a very particular type of metric space: normed vector spaces, and Chapter 2 provides a background for students who may not have seen it. (Even students familiar with metric spaces may benefit by reading the sections in Chapter 2 about convexity and coercivity.)

Notation

If A is an $m \times n$ real matrix, a common way to write the Singular Value Decomposition is $A = U\Sigma V^T$, where U and V are orthogonal (so their columns form orthonormal bases), and the only non-zero entries in Σ are on the main diagonal. (And V^T is the transpose of V.) With this notation, if u_i are the columns of U, v_j are the columns of V, and the diagonal entries of Σ are σ_k, we will have $Av_i = \sigma_i u_i$ and $A^T u_i = \sigma_i v_i$. Thus, A maps the v to the u, which means that A is mapping vector space $\mathcal{V}$ into a vector space $\mathcal{U}$. However, I prefer to preserve alphabetical order when writing domain and co-domain, which means $A : \mathcal{V} \to \mathcal{U}$ feels awkward to me. One solution would be to simply reverse the role of u and v and write the Singular Value Decomposition as $A = V\Sigma U^T$, which would be at odds with just about every single reference and software out there and make it extraordinarily difficult to compare to other sources (or software). On the other hand, it is very common to think of x as the inputs and y as outputs for a function (and indeed it is common to write $f(x) = y$ or $Ax = y$ in linear algebra), and so I have chosen to write the Singular Value Decomposition as $A = Y\Sigma X^T$. From this point of view, the columns of X will form an orthonormal basis for the domain of A, which makes Ax_i fairly natural. Similarly, the columns of Y will form an orthonormal basis for the codomain of A, which hopefully makes $Ax_i = \sigma_i y_i$ feel natural.

Elements of $\mathbb{R}^n$ will be written as $\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T$, where the superscript T indicates the transpose. Recall that if A is an $m \times n$ matrix with ijth entry given by a_{ij}, then A^T is the transpose of A, which means A^T is $n \times m$ and the ij^{th} entry of A^T is a_{ji}. In particular, this means that elements of $\mathbb{R}^n$ should be thought of as *column* vectors. This means that $x = [x_1\ x_2\ \dots\ x_n]^T$ is equivalent to

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Functions may be referred to without an explicit name by writing "the function $x \mapsto$ [appropriate formula]". Thus, the identity function would be $x \mapsto x$, and the exponential function would be $x \mapsto e^x$. Similarly, a function may be defined by

$$f : x \mapsto \text{[appropriate formula]}.$$

For example, given a matrix A, the linear operator defined by multiplying by A is written $L : x \mapsto Ax$. (We will often abuse notation and identify A with the operator $x \mapsto Ax$.) My use of this notation is to remind us that functions are not equations or expressions. We may also use $:=$ to mean "is defined to equal".

Acknowledgements

I would like to first thank CWU's Department of Mathematics and all my colleagues there for making it possible for me to go on sabbatical, during which I wrote most of this book. Without that opportunity, this book would certainly not exist. Next, my thanks to all of the students who made it through my year long sequence in "applicable analysis": Amanda Boseck, John Cox, Raven Dean, John-Paul Mann, Nate Minor, Kerry Olivier, Christopher Pearce, Ben Squire, and Derek Wheel. I'm sorry that it has taken me so long to finally get around to writing the book. I'm even more sorry that I didn't know any of this material when I taught that course. There is also a large group of people who read various parts of the early drafts: Dan Curtis, Ben Freeman, James Harper, Ralf Hoffmann, Adrian Jenkins, Mary Kastning, Dominic Klyve, Brianne and George Kreppein, Mike Lundin, Aaron Montgomery, James Morrow, Ben Squire, Mike Smith, Jernej Tonejc, and Derek Wheel. Their feedback was very useful, and they caught many typos and mistakes. In particular, Brianne Kreppein, Aaron Montgomery, and Jernej Tonejc deserve special thanks for reading the entire draft. Any mistakes and typos that remain are solely my fault! Next, Ina Mette at AMS was fantastic about guiding me through the process. On my sabbatical, there were many places I worked. I remember working out several important details in the Suzallo Reading Room at the University of Washington, and a few more were worked out at Uli's Bierstube in Pike Place Market. My family was also a source of inspiration. My parents Carl and

Ann Bisgard, as well as my siblings Anders Bisgard and Sarah Bisgard-Chaudhari, went out of their way to encourage me to finally finish this thing. Finally, my wonderful wife Kathryn Temple has offered constant encouragement to me through the process, and without her, I doubt I would ever have finished.

James Bisgard