STUDENT MATHEMATICAL LIBRARY Volume 94

# Analysis and Linear Algebra: The Singular Value Decomposition and Applications 

James Bisgard

# Analysis and Linear Algebra: The Singular Value Decomposition and Applications 

# Analysis and Linear Algebra: The Singular Value Decomposition and Applications 

James Bisgard

# EDITORIAL COMMITTEE 

John McCleary<br>Kavita Ramanan<br>Rosa C. Orellana<br>John Stillwell (Chair)

2020 Mathematics Subject Classification. Primary 15-01, 15A18, 26-01, 26Bxx, 49Rxx.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-94

## Library of Congress Cataloging-in-Publication Data

Names: Bisgard, James, 1976- author.
Title: Analysis and linear algebra : the singular value decomposition and applications / James Bisgard.
Description: Providence, Rhode Island : American Mathematical Society, [2021] | Series: Student mathematical library, 1520-9121 ; volume 94 | Includes bibliographical references and indexes.
Identifiers: LCCN 2020055011 | (paperback) ISBN 9781470463328 | (ebook) 9781470465131
Subjects: LCSH: Algebras, Linear-Textbooks. | Mathematical analysis-Textbooks. | Singular value decomposition-Textbooks. | AMS: Linear and multilinear algebra; matrix theory - Instructional exposition (textbooks, tutorial papers, etc.). | Real functions - Instructional exposition (textbooks, tutorial papers, etc.). | Real functions - Functions of several variables. | Calculus of variations and optimal control; optimization - Variational methods for eigenvalues of operators.
Classification: LCC QA184.2 .B57 2021 | DDC 512/.5-dc23
LC record available at https://lccn.loc.gov/2020055011
Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy select pages for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Requests for permission to reuse portions of AMS publication content are handled by the Copyright Clearance Center. For more information, please visit www.ams.org/ publications/pubpermissions.

Send requests for translation rights and licensed reprints to reprint-permission @ams.org.
(c) 2021 by the American Mathematical Society. All rights reserved.

The American Mathematical Society retains all rights
except those granted to the United States Government.
Printed in the United States of America.
The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at https://www.ams.org/
$10987654321 \quad 262524232221$

To my family, especially my loving wife Kathryn, and to the cats, especially Smack and Buck.

## Contents

Preface ..... xi
Pre-Requisites ..... XV
Notation ..... xvi
Acknowledgements ..... Xvii
Chapter 1. Introduction ..... 1
§1.1. Why Does Everybody Say Linear Algebra is "Useful"? ..... 1
§1.2. Graphs and Matrices ..... 4
§1.3. Images ..... 7
§1.4. Data ..... 9
§1.5. Four "Useful" Applications ..... 9
Chapter 2. Linear Algebra and Normed Vector Spaces ..... 13
§2.1. Linear Algebra ..... 14
§2.2. Norms and Inner Products on a Vector Space ..... 20
§2.3. Topology on a Normed Vector Space ..... 30
§2.4. Continuity ..... 38
§2.5. Arbitrary Norms on $\mathbb{R}^{d}$ ..... 44
§2.6. Finite-Dimensional Normed Vector Spaces ..... 48
§2.7. Minimization: Coercivity and Continuity ..... 52
§2.8. Uniqueness of Minimizers: Convexity ..... 54
§2.9. Continuity of Linear Mappings ..... 56
Chapter 3. Main Tools ..... 61
§3.1. Orthogonal Sets ..... 61
§3.2. Projection onto (Closed) Subspaces ..... 67
§3.3. Separation of Convex Sets ..... 73
§3.4. Orthogonal Complements ..... 77
§3.5. The Riesz Representation Theorem and Adjoint Operators ..... 79
§3.6. Range and Null Spaces of $L$ and $L^{*}$ ..... 84
§3.7. Four Problems, Revisited ..... 85
Chapter 4. The Spectral Theorem ..... 99
§4.1. The Spectral Theorem ..... 99
§4.2. Courant-Fischer-Weyl Min-Max Theorem for Eigenvalues 111
§4.3. Weyl's Inequalities for Eigenvalues ..... 117
§4.4. Eigenvalue Interlacing ..... 119
§4.5. Summary ..... 121
Chapter 5. The Singular Value Decomposition ..... 123
§5.1. The Singular Value Decomposition ..... 124
§5.2. Alternative Characterizations of Singular Values ..... 147
§5.3. Inequalities for Singular Values ..... 161
§5.4. Some Applications to the Topology of Matrices ..... 166
§5.5. Summary ..... 170
Chapter 6. Applications Revisited ..... 171
§6.1. The "Best" Subspace for Given Data ..... 171
§6.2. Least Squares and Moore-Penrose Pseudo-Inverse ..... 179
§6.3. Eckart-Young-Mirsky for the Operator Norm ..... 182
§6.4. Eckart-Young-Mirsky for the Frobenius Norm and Image Compression ..... 185
§6.5. The Orthogonal Procrustes Problem ..... 188
§6.6. Summary ..... 198
Contents ..... ix
Chapter 7. A Glimpse Towards Infinite Dimensions ..... 201
Bibliography ..... 209
Index of Notation ..... 213
Index ..... 215

## 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 $T: \mathbb{R}^{n} \rightarrow \mathbb{R}^{n}$ is linear, then $m(T(\Omega))=|\operatorname{det} T| m(\Omega)$, where $m$ is Lebesgue measure and $\Omega$ 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 $\operatorname{dim}\left(\mathcal{U}_{1}+\mathcal{U}_{2}\right)=\operatorname{dim} \mathcal{U}_{1}+\operatorname{dim} \mathcal{U}_{2}-\operatorname{dim}\left(\mathcal{U}_{1} \cap \mathcal{U}_{2}\right)$, 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 $\Sigma$ 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 $\Sigma$ are $\sigma_{k}$, we will have $A v_{i}=\sigma_{i} u_{i}$ 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 $A: \mathcal{V} \rightarrow \mathcal{U}$ 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 $A x=y$ 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 $A x_{i}$ fairly natural. Similarly, the columns of $Y$ will form an orthonormal basis for the codomain of $A$, which hopefully makes $A x_{i}=\sigma_{i} y_{i}$ feel natural.

Elements of $\mathbb{R}^{n}$ will be written as $\left[\begin{array}{llll}x_{1} & x_{2} & \ldots & x_{n}\end{array}\right]^{T}$, where the superscript $T$ indicates the transpose. Recall that if $A$ is an $m \times n$ matrix with $i j$ th entry given by $a_{i j}$, then $A^{T}$ is the transpose of $A$, which means $A^{T}$ is $n \times m$ and the $i j^{\text {th }}$ entry of $A^{T}$ is $a_{j i}$. In particular, this means that elements of $\mathbb{R}^{n}$ should be thought of as column vectors. This means that $x=\left[\begin{array}{llll}x_{1} & x_{2} & \ldots & x_{n}\end{array}\right]^{T}$ is equivalent to

$$
x=\left[\begin{array}{c}
x_{1} \\
x_{2} \\
\vdots \\
x_{n}
\end{array}\right] .
$$

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 $L: x \mapsto A x$. (We will often abuse notation and identify $A$ with the operator $x \mapsto A x$.) 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 BisgardChaudhari, 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.

## Bibliography

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, Princeton, NJ, 2008. With a foreword by Paul Van Dooren. MR2364186
[2] Javier Bernal and Jim Lawrence, Characterization and computation of matrices of maximal trace over rotations, J. Geom. Symmetry Phys. 53 (2019), 21-53, DOI 10.7546/jgsp-53-2019-21-53. MR3971648
[3] Michael W. Berry, Susan T. Dumais, and Gavin W. O'Brien, Using linear algebra for intelligent information retrieval, SIAM Rev. 37 (1995), no. 4, 573-595, DOI 10.1137/1037127. MR1368388
[4] Béla Bollobás, Linear analysis, 2nd ed., Cambridge University Press, Cambridge, 1999. An introductory course. MR1711398
[5] Haim Brezis, Functional analysis, Sobolev spaces and partial differential equations, Universitext, Springer, New York, 2011. MR2759829
[6] Andrew Browder, Mathematical analysis, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1996. An introduction. MR1411675
[7] Fan R. K. Chung, Spectral graph theory, CBMS Regional Conference Series in Mathematics, vol. 92, Published for the Conference Board of the Mathematical Sciences, Washington, DC;
by the American Mathematical Society, Providence, RI, 1997. MR1421568
[8] Dragoš Cvetković, Peter Rowlinson, and Slobodan Simić, An introduction to the theory of graph spectra, London Mathematical Society Student Texts, vol. 75, Cambridge University Press, Cambridge, 2010. MR2571608
[9] Jack Dongarra, Mark Gates, Azzam Haidar, Jakub Kurzak, Piotr Luszczek, Stanimire Tomov, and Ichitaro Yamazaki, The singular value decomposition: anatomy of optimizing an algorithm for extreme scale, SIAM Rev. 60 (2018), no. 4, 808-865, DOI 10.1137/17M1117732. MR3873018
[10] G. Eckart and G. Young, The approximation of one matrix by another of lower rank, Psychometrika 1 (1936), 211-218.
[11] Lars Eldén, Matrix methods in data mining and pattern recognition, Fundamentals of Algorithms, vol. 15, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2019. Second edition of [ MR2314399]. MR3999331
[12] Stephan Ramon Garcia and Roger A. Horn, A second course in linear algebra, Cambridge University Press, Cambridge, 2017.
[13] Gene H. Golub and Charles F. Van Loan, Matrix computations, 4th ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 2013. MR3024913
[14] Anne Greenbaum, Ren-Cang Li, and Michael L. Overton, Firstorder perturbation theory for eigenvalues and eigenvectors, SIAM Rev. 62 (2020), no. 2, 463-482, DOI 10.1137/19M124784X. MR4094478
[15] I. T. Jolliffe, Principal component analysis, 2nd ed., Springer Series in Statistics, Springer-Verlag, New York, 2002. MR2036084
[16] Ian T. Jolliffe and Jorge Cadima, Principal component analysis: a review and recent developments, Philos. Trans. Roy. Soc. A 374 (2016), no. 2065, 20150202, 16, DOI 10.1098/rsta.2015.0202. MR3479904
[17] Jürgen Jost, Postmodern analysis, 3rd ed., Universitext, SpringerVerlag, Berlin, 2005. MR2166001
[18] Wolfang Kabsch, $A$ solution for the best rotation to relate two sets of vectors, Acta Crystallographica A32 (1976), 922-923.
[19] Wolfgang Kabsch, $A$ discussion of the solution for the best rotation to relate two sets of vectors, Acta Crystallographica A34 (1978), 827828.
[20] Dan Kalman, A singularly valuable decomposition: The SVD of a matrix, College Math. J. 27 (1996), no. 1, 2-23.
[21] Tosio Kato, A short introduction to perturbation theory for linear operators, Springer-Verlag, New York-Berlin, 1982. MR678094
[22] Amy N. Langville and Carl D. Meyer, Google's PageRank and beyond: the science of search engine rankings, Princeton University Press, Princeton, NJ, 2006. MR2262054
[23] J. Lawrence, J. Bernal, and C. Witzgall, A purely algebraic justification of the Kabsch-Umeyama algorithm, Journal of Research of the National Institute of Standards and Technology 124 (2019), 1-6.
[24] Peter D. Lax, Linear algebra and its applications, 2nd ed., Pure and Applied Mathematics (Hoboken), Wiley-Interscience [John Wiley \& Sons], Hoboken, NJ, 2007. MR2356919
[25] R. Lee and L. Carter, Modeling and forecasting US mortality, J. American Statistical Assoc. 87 (1992), 659-671.
[26] Carla D. Martin and Mason A. Porter, The extraordinary SVD, Amer. Math. Monthly 119 (2012), no. 10, 838-851, DOI 10.4169/amer.math.monthly.119.10.838. MR2999587
[27] Jiří Matoušek, Thirty-three miniatures, Student Mathematical Library, vol. 53, American Mathematical Society, Providence, RI, 2010. Mathematical and algorithmic applications of linear algebra. MR2656313
[28] Elizabeth S. Meckes and Mark W. Meckes, Linear algebra, Cambridge University Press, Cambridge, 2018.
[29] L. Mirsky, Symmetric gauge functions and unitarily invariant norms, Quart. J. Math. Oxford Ser. (2) 11 (1960), 50-59, DOI 10.1093/qmath/11.1.50. MR114821
[30] Cleve Moler and Donald Morrison, Singular value analysis of cryptograms, Amer. Math. Monthly 90 (1983), no. 2, 78-87, DOI 10.2307/2975804. MR691178
[31] Kenneth A. Ross, Elementary analysis, 2nd ed., Undergraduate Texts in Mathematics, Springer, New York, 2013. The theory of calculus; In collaboration with Jorge M. López. MR3076698
[32] Walter Rudin, Principles of mathematical analysis, 3rd ed., McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976. International Series in Pure and Applied Mathematics. MR0385023
[33] Bryan P. Rynne and Martin A. Youngson, Linear functional analysis, 2nd ed., Springer Undergraduate Mathematics Series, SpringerVerlag London, Ltd., London, 2008. MR2370216
[34] Amol Sasane, A friendly approach to functional analysis, Essential Textbooks in Mathematics, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2017. MR3752188
[35] Karen Saxe, Beginning functional analysis, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2002. MR1871419
[36] G. W. Stewart, On the early history of the singular value decomposition, SIAM Rev. 35 (1993), no. 4, 551-566, DOI 10.1137/1035134. MR 1247916
[37] G. Strang, Linear algebra and learning from data, WellesleyCambridge Press, 2019.
[38] Gilbert Strang, The fundamental theorem of linear algebra, Amer. Math. Monthly 100 (1993), no. 9, 848-855, DOI 10.2307/2324660. MR1247531
[39] Michael J. Todd, Minimum-volume ellipsoids, MOS-SIAM Series on Optimization, vol. 23, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2016. Theory and algorithms. MR3522166
[40] Madeleine Udell and Alex Townsend, Why are big data matrices approximately low rank?, SIAM J. Math. Data Sci. 1 (2019), no. 1, 144-160, DOI 10.1137/18M1183480. MR3949704
[41] S. Umeyama, Least-squares estimation of transformation parameters between two point patterns, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991), 376-380.
[42] Eugene Vecharynski and Yousef Saad, Fast updating algorithms for latent semantic indexing, SIAM J. Matrix Anal. Appl. 35 (2014), no. 3, 1105-1131, DOI 10.1137/130940414. MR3249365

## Index of Notation

:=, defined to be, xvii
$A^{T}$, the transpose of $A, \mathrm{XVI}^{\mathrm{XV}}$
$A^{\dagger}$, Moore-Penrose pseudo-inverse, 92
$B_{r}(x)$, ball of radius $r>0$ around $x, 31$
$L^{*}, 81$
$R_{L}$, Rayleigh quotient, 100
$U_{1} \oplus U_{2}$, direct sum of subspaces, 17
$V^{T}$, transpose of a matrix $V$, 区vi
$[x]_{U}$, coordinates of $x$ with respect to the basis $U, 48$
$\langle A, B\rangle_{F}$, the Frobenius inner product of $A$ and $B, 23$
$\langle\cdot, \cdot\rangle$, an inner product, 22
$\lambda_{k}^{\downarrow}(L)$, eigenvalues of $L$ in decreasing order, 110
$\lambda_{k}^{\uparrow}(L), 110$
$\mathcal{L}(\mathcal{V}, \mathcal{W})$, set of linear mappings from $\mathcal{v}$ to $\mathcal{W}, 29$
$\mapsto$, function mapping, xvii
$\mathbf{e}_{j}$, the $j$ th standard basis element, 45
$\mathcal{B} \mathcal{L}(\mathcal{V}, \mathcal{W}), 29$
$\mathcal{U}^{+}$, non-zero elements of the subspace U, 127
$u_{1}+u_{2}$, sum of subspaces, 15
$\mathcal{V} \times \mathcal{W}$, product of vector spaces, 125
$\mathcal{N}(L)$, nullspace of $L, 18$
$\mathcal{U}^{\perp}$, the orthogonal complement of $\mathcal{U}$, 77
$u^{\perp}$, the orthogonal complement of $\operatorname{span}\{u\}, 77$
$\mathcal{R}(L)$, the range of $L$, 18
$\operatorname{tr} C$, trace of a matrix $C, 19$
$d(a, \mathcal{U})$, distance from $a$ to a subspace u, 92
$f^{-1}(B)$, the pre-image of $B$ under $f, 40$

## Index

adjacency matrix, 5
adjoint, 99
of a matrix, 84
operator, 81
Axiom of Choice, 203
balls, 31
open, 32
Banach space, 202
best subspace problem, 9, 91, 171
Bolzano-Weierstrass Theorem, 37, 44, 46, 50, 54

Cauchy-Schwarz-Bunyakovsky
Inequality, 24
CFW Theorem, 111
closed, 31, 32
relatively, 38
subspace, 67
closest point, 67
calculation, 71
for a convex set, 73
coercive, 53, 54
compact, 31, 37, 43, 46, 50, 75
completeness, 31, 38, 50
component functions, 43
continuous, 39, 41, 43
$\varepsilon-\delta, 39$
component-wise, 43
eigenvalues, 116
matrix multiplication, 42
norm, 46
sequentially, 39
topologically, 39
convergence, 31, 32
component-wise, 36
convexity, 55
for a set, 55
for a function, 55
coordinates, 56
in an orthonormal basis, 63
Courant-Fischer-Weyl Min-Max
Theorem
for singular values, 163
Courant-Fischer-Weyl Min-max
Theorem, 111
CSB inequality, 24

## degree, 5

density
of full rank operators, 166
of invertible matrices, 167
of symmetric matrices with simple eigenvalues, 168
direct sum of subspaces, 17
Eckart-Young-Mirsky, 93
Eckart-Young-Mirsky Theorem, 11
for Frobenius norm, 185, 186
for operator norm, 182, 184
eigenvalues
continuity, 116
interlacing, 120
min-max characterization, 111
relation to singular values, 147, 159
Weyl's inequality, 118
equivalence
for continuity, 40
for norms, 34, 45, 46, 52
Fundamental Theorem of Linear Algebra, 18, 84, 85

Gram-Schmidt Process, 65
graph, $\square^{4}$
graph Laplacian, 7
gray-scale matrix, 7
Hilbert space, 204
induced norm, 25
inner product, 22, 61
dot product, 22
for product space, 126
Frobenius, 23
invariant subspace, 104
Kabsch-Umeyama Algorithm, 197
kernel, 18
length of a path, 4
low rank approximation, 11, 93
for Frobenius norm, 185, 186
for matrices and Frobenius norm, 187
for matrices and operator norm, 184
for operator norm, 182, 184
minimization, 43
minimizers, 53
minimizing sequences, 53
Moore-Penrose Pseudo-Inverse, 10
Moore-Penrose pseudo-inverse, 92, 179
for matrices, 181
norm, 13, 49
continuity of,46
definition, 20
equivalence, 34, 45, 46, 52
Euclidean, 21

Frobenius, 26,161
induced by inner product, 25, 99, 126
max, 21
on $\mathbb{R}^{d}, 21$
operator, 29
sub-multiplicative matrix norm, 28
taxi-cab, 21
normed vector space, 20
nullity, 18
nullspace, 18
open, 31
balls, 32
relatively, 38
Orthogonal
Procrustes Problem, 11
orthogonal, 61
decomposition, 78
complement, 77, 104
Procrustes Problem, 95, 188, 190
orthogonal matrix, 26
orthonormal, 61
outer product, 24
path, 4
Principal Component Analysis, xiii
Procrustes Problem
orientation preserving, 97, 197
orthogonal, 95, 188, 190
product space, 125
projection, 71
protractors, 29
Pythagorean Theorem, 64
range, 18
rank, 18
lower semi-continuity, 168
Rayleigh quotient, 100, 127
reduced SVD, 88, 145
reverse triangle inequality, 21, 40, 45
Riesz Representation Theorem, 79, 80
rulers, 29
self-adjoint, 99
Separation of convex sets, 73
sequences
Cauchy, 31
convergence, 31, 32
minimizing, 53
sequentially compact, 31, 37, 43, 46, 50,

## 75

singular triples, 85, 124, 170
Singular Value Decomposition for matrices, 143
Singular Value Decomposition, 85
Singular Value Decomposition, 125 by norm, 152 by Spectral Theorem, 159 reduced, 88
singular values, 170 by norm, 152 continuity, 165 Courant-Fischer-Weyl
characterization, 163
Frobenius norm, 161
in terms of eigenvalues, 159
Weyl's inequality, 164
singular vectors
left, 85
right, 85
Spectral Theorem, 99
for matrices, 108
rank one decomposition, 109
standard norm, 20
sum of subspaces, 15
SVD, 85, 125
reduced, 88
by norm, 152
by Spectral Theorem, 159
for matrices, 143
rank one decomposition, 146
reduced, 145
topology
normed vector space, 32
trace of a matrix, 19
transpose, xvi
Weyl's inequality, 118
for singular values, 164, 184, 187
zero mapping, 29

## Selected Published Titles in This Series

94 James Bisgard, Analysis and Linear Algebra: The Singular Value Decomposition and Applications, 2021
93 Iva Stavrov, Curvature of Space and Time, with an Introduction to Geometric Analysis, 2020
92 Roger Plymen, The Great Prime Number Race, 2020
91 Eric S. Egge, An Introduction to Symmetric Functions and Their Combinatorics, 2019
90 Nicholas A. Scoville, Discrete Morse Theory, 2019
89 Martin Hils and François Loeser, A First Journey through Logic, 2019
88 M. Ram Murty and Brandon Fodden, Hilbert's Tenth Problem, 2019
87 Matthew Katz and Jan Reimann, An Introduction to Ramsey Theory, 2018
86 Peter Frankl and Norihide Tokushige, Extremal Problems for Finite Sets, 2018
85 Joel H. Shapiro, Volterra Adventures, 2018
84 Paul Pollack, A Conversational Introduction to Algebraic Number Theory, 2017
83 Thomas R. Shemanske, Modern Cryptography and Elliptic Curves, 2017
82 A. R. Wadsworth, Problems in Abstract Algebra, 2017
81 Vaughn Climenhaga and Anatole Katok, From Groups to Geometry and Back, 2017
80 Matt DeVos and Deborah A. Kent, Game Theory, 2016
79 Kristopher Tapp, Matrix Groups for Undergraduates, Second Edition, 2016
78 Gail S. Nelson, A User-Friendly Introduction to Lebesgue Measure and Integration, 2015
77 Wolfgang Kühnel, Differential Geometry: Curves - Surfaces Manifolds, Third Edition, 2015
76 John Roe, Winding Around, 2015
75 Ida Kantor, Jiří Matoušek, and Robert Šámal, Mathematics++, 2015
74 Mohamed Elhamdadi and Sam Nelson, Quandles, 2015
73 Bruce M. Landman and Aaron Robertson, Ramsey Theory on the Integers, Second Edition, 2014
72 Mark Kot, A First Course in the Calculus of Variations, 2014
71 Joel Spencer, Asymptopia, 2014

For a complete list of titles in this series, visit the AMS Bookstore at www.ams.org/bookstore/stmlseries/.


This book provides an elementary analytically inclined journey to a fundamental result of linear algebra: the Singular Value Decomposition (SVD). SVD is a workhorse in many applications of linear algebra to data science. Four important applications relevant to data science are considered throughout the book: determining the subspace that "best" approximates a given set (dimension reduction of a data set); finding the "best" lower rank approximation of a given matrix (compression and general approximation problems); the Moore-Penrose pseudoinverse (relevant to solving least squares problems); and the orthogonal Procrustes problem (finding the orthogonal transformation that most closely transforms a given collection to a given configuration), as well as its orientation-preserving version.
The point of view throughout is analytic. Readers are assumed to have had a rigorous introduction to sequences and continuity. These are generalized and applied to linear algebraic ideas. Along the way to the SVD, several important results relevant to a wide variety of fields (including random matrices and spectral graph theory) are explored: the Spectral Theorem; minimax characterizations of eigenvalues; and eigenvalue inequalities. By combining analytic and linear algebraic ideas, readers see seemingly disparate areas interacting in beautiful and applicable ways.

