LINEAR ALGEBRA FOR EVERYONE

GILBERT STRANG

Massachusetts Institute of Technology

WELLESLEY - CAMBRIDGE PRESS

Box 812060 Wellesley MA 02482

Linear Algebra for Everyone

ISBN 978-1-7331466-3-0

LaTeX typesetting by Ashley C. Fernandes (info@problemsolvingpathway.com)
Printed in the United States of America

9 8 7 6 5 4 3 2 1

QA184.2 .S773 2020 | DDC 512/.5–dc23

Other texts from Wellesley - Cambridge Press

Linear Algebra and Learning from Data, 2019, Gilbert Strang ISBN 978-0-6921963-8-0

Introduction to Linear Algebra, 5th Ed., 2016, Gilbert Strang ISBN 978-0-9802327-7-6

Computational Science and Engineering, Gilbert Strang ISBN 978-0-9614088-1-7

Wavelets and Filter Banks, Gilbert Strang and Truong Nguyen ISBN 978-0-9614088-7-9

Introduction to Applied Mathematics, Gilbert Strang ISBN 978-0-9614088-0-0

Calculus Third Edition, Gilbert Strang ISBN 978-0-9802327-5-2

Algorithms for Global Positioning, Kai Borre & Gilbert Strang ISBN 978-0-9802327-3-8

Essays in Linear Algebra, Gilbert Strang ISBN 978-0-9802327-6-9

Differential Equations and Linear Algebra, Gilbert Strang ISBN 978-0-9802327-9-0

An Analysis of the Finite Element Method, 2008 edition, Gilbert Strang and George Fix ISBN 978-0-9802327-0-7

Wellesley - Cambridge Press
Box 812060, Wellesley MA 02482 USA
www.wellesleycambridge.com
LAFEeveryone@gmail.com

Gilbert Strang's page : **math.mit.edu/∼gs**
For orders : **math.mit.edu/weborder.php**
Outside US/Canada : **www.cambridge.org**
Select books, India : **www.wellesleypublishers.com**

The website for this book (with Solution Manual) is **math.mit.edu/everyone**
2019 book : **Linear Algebra and Learning from Data** (**math.mit.edu/learningfromdata**)
2016 book : **Introduction to Linear Algebra, 5th Edition** (**math.mit.edu/linearalgebra**)
2014 book : **Differential Equations and Linear Algebra** (**math.mit.edu/dela**)

Linear Algebra is included in MIT's OpenCourseWare **ocw.mit.edu/courses/mathematics**
Those videos (including 18.06SC and 18.065) are also on **www.youtube.com/mitocw**
18.06 Linear Algebra 18.06SC with problem solutions 18.065 Learning from Data

The cover design was created by Gail Corbett and Lois Sellers : **lsellersdesign.com**

Table of Contents

Preface v

1 Vectors and Matrices 1
- 1.1 Linear Combinations of Vectors . . . 2
- 1.2 Lengths and Angles from Dot Products . . . 11
- 1.3 Matrices and Column Spaces . . . 20
- 1.4 Matrix Multiplication and $\boldsymbol{A} = \boldsymbol{CR}$ . . . 29

2 Solving Linear Equations $\boldsymbol{Ax} = \boldsymbol{b}$ 39
- 2.1 The Idea of Elimination . . . 40
- 2.2 Elimination Matrices and Inverse Matrices . . . 49
- 2.3 Matrix Computations and $\boldsymbol{A} = \boldsymbol{LU}$ . . . 57
- 2.4 Permutations and Transposes . . . 64

3 The Four Fundamental Subspaces 74
- 3.1 Vector Spaces and Subspaces . . . 75
- 3.2 The Nullspace of $\boldsymbol{A}$: Solving $\boldsymbol{Ax} = \boldsymbol{0}$ . . . 83
- 3.3 The Complete Solution to $A\boldsymbol{x} = \boldsymbol{b}$ . . . 96
- 3.4 Independence, Basis, and Dimension . . . 107
- 3.5 Dimensions of the Four Subspaces . . . 121

4 Orthogonality 134
- 4.1 Orthogonality of the Four Subspaces . . . 135
- 4.2 Projections onto Subspaces . . . 143
- 4.3 Least Squares Approximations . . . 153
- 4.4 Orthogonal Matrices and Gram-Schmidt . . . 165

5 Determinants and Linear Transformations 177
- 5.1 $\boldsymbol{3}$ by $\boldsymbol{3}$ Determinants . . . 178
- 5.2 Properties and Applications of Determinants . . . 184
- 5.3 Linear Transformations . . . 192

6 Eigenvalues and Eigenvectors 201
- 6.1 Introduction to Eigenvalues . . . 202
- 6.2 Diagonalizing a Matrix . . . 215
- 6.3 Symmetric Positive Definite Matrices . . . 227
- 6.4 Systems of Differential Equations . . . 243

7 The Singular Value Decomposition (SVD) **258**
7.1 Singular Values and Singular Vectors 259
7.2 Compressing Images by the SVD 269
7.3 Principal Component Analysis 274
7.4 The Victory of Orthogonality (and a Revolution) 280

8 Learning from Data **286**
8.1 Piecewise Linear Learning Functions 289
8.2 Convolutional Neural Nets 299
8.3 Minimizing Loss by Gradient Descent 306
8.4 Mean, Variance, and Covariance 321

Appendix 1 The Ranks of AB and $A+B$ **334**

Appendix 2 Eigenvalues and Singular Values : Rank One **335**

Appendix 3 Counting Parameters in the Basic Factorizations **336**

Appendix 4 Codes and Algorithms for Numerical Linear Algebra **337**

Appendix 5 Matrix Factorizations **338**

Appendix 6 The Column-Row Factorization of a Matrix **340**

Appendix 7 The Jordan Form of a Square Matrix **343**

Appendix 8 Tensors **344**

Appendix 9 The Condition Number **345**

Appendix 10 Markov Matrices and Perron-Frobenius **346**

Index **348**

Index of Symbols **354**

Six Great Theorems / Linear Algebra in a Nutshell **356**

Preface

This is a linear algebra textbook with a new start. Chapter 1 begins as usual with vectors. We see their linear combinations and dot products. Then the new ideas come with matrices. Let me illustrate those ideas right away by an example.

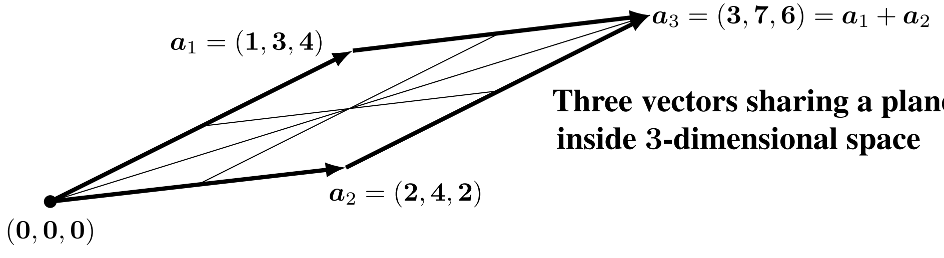
Suppose we are given a 3 by 3 matrix A with columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 7 \\ 4 & 2 & 6 \end{bmatrix}.$$

Those columns are three-dimensional vectors. The first vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ connect the center point $(0,0,0)$ to the points $(1,3,4)$ and $(2,4,2)$. The picture shows those points in 3-dimensional space (xyz space). The key to this matrix is the third vector $\boldsymbol{a}_3$ going to the point $(3,7,6)$.

When I look at those vectors, I see something exceptional. Adding columns 1 and 2 produces column 3. In other words $\boldsymbol{a}_1 + \boldsymbol{a}_2 = \boldsymbol{a}_3$. In a 3-dimensional picture, $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ go from the center point $(0,0,0)$ to the points $(1,3,4)$ and $(2,4,2)$. The picture shows how to add those vectors. It is normal that all combinations of two vectors will fill up a plane. (The plane is actually infinite, we just drew the part between the vectors.) What is really exceptional is that **the third point $\boldsymbol{a_3}$ =$(\mathbf{3}, \mathbf{7}, \mathbf{6})$ lies on this plane of $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$.**

Most points don't lie on that plane. Most vectors $\boldsymbol{a}_3$ are *not* combinations of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. Most 3 by 3 matrices have *independent* columns. Then the matrix will be invertible. But these three columns are *dependent* because they lie on the same plane : $\boldsymbol{a}_1 + \boldsymbol{a}_2 = \boldsymbol{a}_3$.



Three vectors sharing a plane inside 3-dimensional space

That picture reveals the most important fact about those three vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$. But we need the right language to describe it. Our goal is to see the idea, and to get better and better at expressing it. Here are three steps in a good direction.

1 Idea in words Column 3 is the sum of column 1 and column 2

2 Idea in symbols $\boldsymbol{a}_3 = \boldsymbol{a}_1 + \boldsymbol{a}_2$

3 Matrix times vector $\begin{bmatrix} \\ \boldsymbol{a}_3 \\ \\ \end{bmatrix} = \begin{bmatrix} & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 \\ & \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}$

Step 3 shows how a matrix C multiplies a vector $\boldsymbol{x}$. The columns $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in C multiply the numbers $x_1 = 1$ and $x_2 = 1$ in $\boldsymbol{x}$. *The output $C\boldsymbol{x}$ is a combination of the columns.* Here that column combination is $\boldsymbol{a}_1 + \boldsymbol{a}_2$.

One more crucial step allows several combinations at once. This is the way forward. We cannot take this step with pictures or words. **A matrix multiplies a matrix**.

4 Matrix times matrix $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{R}$ is $\begin{bmatrix} & & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \\ & & \end{bmatrix} = \begin{bmatrix} & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 \\ & \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \end{bmatrix}$

Those three columns of $\boldsymbol{R}$ give three combinations in $\boldsymbol{A}$ of the two columns in $\boldsymbol{C}$

Column 1 : $\mathbf{1}\boldsymbol{a}_1 + \mathbf{0}\boldsymbol{a}_2 = \boldsymbol{a}_1$ **Column 2** : $\mathbf{0}\boldsymbol{a}_1 + \mathbf{1}\boldsymbol{a}_2 = \boldsymbol{a}_2$

Column 3 : $\mathbf{1}\boldsymbol{a}_1 + \mathbf{1}\boldsymbol{a}_2 = (\mathbf{1}, \mathbf{3}, \mathbf{4}) + (\mathbf{2}, \mathbf{4}, \mathbf{2}) = (\mathbf{3}, \mathbf{7}, \mathbf{6}) = \boldsymbol{a}_3$

That matrix multiplication $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{R}$ displays major information :

A has dependent columns : the combination of columns $1 + 2 - 3$ gives $(0, 0, 0)$

A also has dependent rows : some combination of its rows will give $(0, 0, 0)$!

The "*column space*" of this A is only a plane and not the whole 3D space

The "*row space*" of this A is also a plane and not the whole 3D space

The square matrix A has no inverse. Its determinant is zero. It is unusual.

Subspaces of Vector Spaces

In Chapters 1 to 4, the organizing ideas are **vector spaces**. The columns of A are in m-dimensional space $\mathbf{R}^m$. The rows are in n-dimensional space $\mathbf{R}^n$. But the action is in **four subspaces inside $\mathbf{R}^m$ and $\mathbf{R}^n$**. The matrix A multiplies $\boldsymbol{x}_{\text{row}}$ in its *row space* and $\boldsymbol{x}_{\text{null}}$ in its *nullspace*. The output is $A\boldsymbol{x}_{\text{row}} = \boldsymbol{b}$ in its *column space* because $A\boldsymbol{x}_{\text{null}} =$ *zero*. Then the complete solution to $A\boldsymbol{x} = \boldsymbol{b}$ has a row space part and a nullspace part :

$$\boxed{\boldsymbol{x} = \boldsymbol{x}_{\textbf{row}} + \text{ any } \boldsymbol{x}_{\textbf{null}} \qquad A\boldsymbol{x} = A\boldsymbol{x}_{\textbf{row}} + A\boldsymbol{x}_{\textbf{null}} = \boldsymbol{b} + \mathbf{0} = \boldsymbol{b}}$$

Every linear equation $A\boldsymbol{x} = \boldsymbol{b}$ is solved this way : particular $\boldsymbol{x}_{\text{row}}$+ any homogeneous $\boldsymbol{x}_{\text{null}}$.

The Plan for the Book

That example is part of a new start. I believe it is a better start (for the reader and the course). By working with specific matrices to introduce the algebra, the subject unfolds. Chapter 1 develops the matrix equation $A = C$ times R. C takes independent vectors like $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ from the column space of A. R takes independent vectors from the row space of A. Those two "vector spaces" are at the center of linear algebra. We meet them properly in Chapter 3. But you will know about independence from examples in Chapter 1.

The big step is to factor A into C times R. Matrix multiplication is a crucial operation, and Chapter 1 ends with four different ways to do it—seen on the back cover of the book. This sets the path to all the great factorizations of linear algebra:

$\boldsymbol{A = LU}$	Chapter 2 solves n equations $A\boldsymbol{x} = \boldsymbol{b}$ in n unknowns: A is square
$\boldsymbol{A = CR}$	Chapter 3 reduces to r independent columns and r independent rows
$\boldsymbol{A = QR}$	Chapter 4 changes the columns of A into perpendicular columns of Q
$\boldsymbol{S = Q\Lambda Q^{\mathrm{T}}}$	Chapter 6 has **eigenvectors** in $\boldsymbol{Q}$. The **eigenvalues** of $\boldsymbol{S}$ are in $\boldsymbol{\Lambda}$
$\boldsymbol{A = U\Sigma V^{\mathrm{T}}}$	Chapter 7 has **singular vectors** in $\boldsymbol{U}$ and $\boldsymbol{V}$ and **singular values** in $\boldsymbol{\Sigma}$

The columns of A are in m-dimensional space $\mathbf{R}^m$, the rows are in $\mathbf{R}^n$. The m by n matrix multiplies vectors $\boldsymbol{x}$ in $\mathbf{R}^n$ to produce $A\boldsymbol{x}$ in $\mathbf{R}^m$. But the real action of A is seen in the four fundamental subspaces.

Chapter 2 only allows one solution $\boldsymbol{x}_{\textbf{row}}$. The matrix A is square and invertible. Chapter 3 finds every solution to $A\boldsymbol{x} = \boldsymbol{b}$ by adding every $\boldsymbol{x}_{\textbf{null}}$. Chapter 4 deals with equations that don't have a solution (because $\boldsymbol{b}$ has a piece from the mysterious fourth subspace). I hope you will like the "big picture of linear algebra" on page 124: **all four subspaces**.

Those five factorizations are a perfect way to organize and remember linear algebra. The eigenvalues in the matrix $\boldsymbol{\Lambda}$ and the singular values in $\boldsymbol{\Sigma}$ come from S and A in a beautiful way—but not a simple way. Those numbers in $S\boldsymbol{x} = \lambda\boldsymbol{x}$ and $A\boldsymbol{v} = \sigma\boldsymbol{u}$ see deeply into the symmetric matrix S and the m by n matrix A. Often S appears in engineering and physics. Often A is a matrix of data. And data is now coming from everywhere.

Please don't miss Tim Baumann's page 272 on compressing photographs by the SVD.

Chapter 5 explains the amazing formulas produced by determinants. Amazing but unfortunately difficult to compute! We solve equations $A\boldsymbol{x} = \boldsymbol{b}$ before (not after) we find the determinant of A. Those equations ask us to produce $\boldsymbol{b}$ from the columns of A.

$$\boldsymbol{Ax} = \boldsymbol{x_1}\,(\textbf{column 1}) + \boldsymbol{x_2}\,(\textbf{column 2}) + \cdots + \boldsymbol{x_n}\,(\textbf{column } \boldsymbol{n}) = \textbf{ right side } \boldsymbol{b}.$$

In principle, determinants lead to eigenvalues (Chapter 6) and singular values (Chapter 7). In practice, we look for other ways to find and use those important numbers. And yet determinants tell us about geometry too—like the volume of a tilted box in n dimensions.

A short course can go directly from dimensions in 3.5 to eigenvalues for 2×2 matrices.

Final Chapter : Learning from Data

We added Chapter 8 on **data science**. The data often comes in a rectangular matrix A. Each row measures n properties of a sample. That number n is large, and the number of samples is often very large. Out of this big matrix, applied linear algebra has to find out what is important. The goal is to produce understanding that leads to a decision.

In machine learning the output is a nonlinear function of the input. Deep learning aims to find that function from the training data. It produces (often with giant computations) a learning function $\boldsymbol{F(x,v)}$. The input vectors $\boldsymbol{v}$ contain the features of each training sample (many $\boldsymbol{v}$'s from many samples). The vector $\boldsymbol{x}$ contains the weights assigned to those features. And hidden inside F is a nonlinear function that mixes the inputs and the weights. A favorite is the simple function ReLU $(y) =$ maximum of 0 and y : a *ramp function*.

Optimizing the weights to learn from the training data $\boldsymbol{v}$ is the big computation. When that is well done, we can input new samples that the system has never seen. The success of deep learning is that $F(\boldsymbol{x},\boldsymbol{v})$ is often close to the correct output. The system has identified an image, or translated a sentence, or chosen a winning move.

This is *piecewise linear algebra*, a mixture of weight matrices and ReLU. It is included with no expectation of testing students. This is a chapter to use later, in any way you want. You could experiment with the website **playground.tensorflow.org**.

Machine learning has become important and powerful—based on linear algebra and calculus (optimizing the weights) and on statistics (controlling the mean and variance). *But it is not required or even expected to be part of the course. What I hope is that the faster start allows you to reach eigenvalues and singular values*—those are true highlights of this subject. Chapters 2 to 7 confirm the jump of intuition near the end of Chapter 1 :

> If all columns of A lie on an r-dimensional plane, then all rows of A also lie on a (usually different) r-dimensional plane. That fact has far-reaching consequences.

This is a textbook for a normal linear algebra class—to explain the key ideas of this beautiful subject to *everyone*. As clearly as I can. Thank you.

Gilbert Strang

After writing this introduction, I looked back at the opening example. That 3 by 3 matrix A has dependent columns (column $1 +$ column $2 =$ column 3). By linear algebra, A must also have dependent rows. The numbers to show this are $\boldsymbol{-5}$ and $\boldsymbol{3}$:

$$\boldsymbol{-5}\,(\text{row } 1) + \boldsymbol{3}\,(\text{row } 2) = \text{row } 3 \text{ of } A \qquad \boldsymbol{-5}\,(\boldsymbol{1},\boldsymbol{2},\boldsymbol{3}) + \boldsymbol{3}\,(\boldsymbol{3},\boldsymbol{4},\boldsymbol{7}) = (\boldsymbol{4},\boldsymbol{2},\boldsymbol{6}).$$

The number of independent rows equals the number of independent columns. *Wonderful* !