Preface

These Notes are like a sparse textbook. The essential ideas are here and the unnecessary words are gone. The goal is to present ideas and examples in good order.

Section 7.1 may serve as an example. It explains the "Singular Value Decomposition" of a matrix. This is one way to factor the matrix A into simple pieces—and it has the great merit of succeeding for every matrix : no exceptions. Here are the three pieces that multiply to give A :

(**orthogonal** matrix) (**diagonal** matrix) (**orthogonal** matrix) = (*rotation*) (*stretch*) (*rotation*)

This idea has become important because those matrices are so special. Section 7.1 begins with this example :

$$\boldsymbol{A}=\begin{bmatrix}\mathbf{3}&\mathbf{0}\\\mathbf{4}&\mathbf{5}\end{bmatrix}=\begin{bmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{bmatrix}\begin{bmatrix}\sigma_1&\\&\sigma_2\end{bmatrix}\begin{bmatrix}\cos\alpha&\sin\alpha\\-\sin\alpha&\cos\alpha\end{bmatrix}$$

We are starting with four numbers $\mathbf{3,0,4,5}$. We are ending with four different numbers—two rotation angles θ and α, and two stretching factors σ_1 and σ_2 (those are the singular values of A). The numbers and their meaning are revealed at the start of Chapter 7, after preparing the way in Chapters 1 to 6.

Most linear algebra courses end with the eigenvalues of a matrix. (Eigenvalues are explained in Chapter 6.) My hope is to encourage that next step to singular values. All we need are the eigenvalues of $A^{\mathrm{T}}A$, the "transpose matrix" A^{T} times the original matrix.

$$\boldsymbol{A}^{\mathbf{T}}\boldsymbol{A}=\begin{bmatrix}3&4\\0&5\end{bmatrix}\begin{bmatrix}3&0\\4&5\end{bmatrix}=\begin{bmatrix}\mathbf{25}&\mathbf{20}\\\mathbf{20}&\mathbf{25}\end{bmatrix}\text{ has eigenvalues }\ \lambda_1=45\ \text{ and }\ \lambda_2=5$$

Then $\sigma_1=\sqrt{45}$ and $\sigma_2=\sqrt{5}$ are the singular values (stretching factors) of this A.

Now I need to start again. There are five full chapters before an eigenvalue appears. The basic problem in those chapters is to understand and solve linear equations $\boldsymbol{Ax}=\boldsymbol{b}$. The matrix A (square or rectangular) has n column vectors $\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n$. Multiplying A times $\boldsymbol{x}$ gives a combination $x_1\boldsymbol{a}_1+\cdots+x_n\boldsymbol{a}_n$ of the columns of A.

We are asked to find a combination of those columns that produces the vector $\boldsymbol{b}$:

$$\boxed{\boldsymbol{Ax} = \boldsymbol{b} \text{ means } x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}}$$

Depending on A and $\boldsymbol{b}$, a solution $x_1, x_2, \ldots, x_n$ may or may not be possible. Our equations may have one solution $\boldsymbol{x}$, or no solutions, or infinitely many solutions. Chapters 1 to 3 explain the steps that look for $\boldsymbol{x}$.

Question Do the equations $A\boldsymbol{x} = \boldsymbol{b}$ have a solution vector $\boldsymbol{x}$?

Is the vector $\boldsymbol{b}$ a combination of the columns of A?

Answers (one of three)

1 There is no solution $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$

2 The only solution is $\boldsymbol{x} = (\_\_, \_\_, \ldots, \_\_)$

3 There are nonzero solutions to $A\boldsymbol{y} = \boldsymbol{0}$. Those $\boldsymbol{y}$'s can be added to a particular solution $\boldsymbol{x}$ to give more solutions $A(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{b}$.

You are seeing that linear algebra is constantly working with **combinations of vectors**: combinations of columns of A, combinations of rows of A, combinations of solutions to $A\boldsymbol{y} = \boldsymbol{0}$. To solve $A\boldsymbol{x} = \boldsymbol{b}$, we look for a particular combination that matches $\boldsymbol{b}$.

A happy part of linear algebra is the wonderful variety of matrices: diagonal, triangular, symmetric, orthogonal, and many more. The organizing principles have become matrix factorizations like $\boldsymbol{A} = \boldsymbol{LU}$ (lower triangular times upper triangular). The idea of elimination—to simplify the equations $\boldsymbol{Ax} = \boldsymbol{b}$ by introducing zeros in the matrix—appears early as it must. Please don't spend forever on those computations. Linear algebra has so many more good ideas.

These pages were created in 2020 and 2021. Their goal is to help instructors and students see linear algebra in an organized way. The notes go from vectors to matrices to subspaces to bases. "*Linear independence*" is a crucial idea, so it comes early—for vectors of integers. We test independence in Chapter 3, but the idea is already in Chapter 1.

The reader may know my video lectures for **Mathematics 18.06** and **18.065** on MIT's OpenCourseWare **ocw.mit.edu** and **www.youtube.com/mitocw**. I am so grateful that those have been helpful. Now these lecture notes can help in a different way. You will quickly gain a picture of the whole course—the structure of the subject, the key topics in a natural order, and the connecting ideas that make linear algebra so beautiful.

This structure is the basis for two textbooks from Wellesley-Cambridge Press :

Introduction to Linear Algebra **Linear Algebra for Everyone**

I don't try to teach every topic in those books. I do teach eigenvalues and eigenvectors. A basis of eigenvectors for square matrices—and a basis of singular vectors for all matrices—those take you to the heart of a matrix in a way that elimination cannot do.

The last chapters of these notes extend to a third textbook and a second MIT math course $\mathbf{18.065}$. The subject is *data science*. The viewpoint is again linear algebra. The videos on OpenCourseWare and YouTube explain the ideas in the book. The first lecture develops $\boldsymbol{A} = \boldsymbol{CR}$, the factorization that reveals the rank and the column space of A—that is a highly recommended way to start a linear algebra course.

Linear Algebra and Learning from Data (Wellesley-Cambridge Press 2019)

This is "Deep Learning" and it is not entirely linear. A learning function $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v})$ is created from training data $\boldsymbol{v}$ (like images of handwritten numbers) and matrix weights $\boldsymbol{x}$. The piecewise linear "ReLU function" plays a mysterious but crucial part in $\boldsymbol{F}$. With ReLU and good weights, the output $F(\boldsymbol{x}, \boldsymbol{v})$ is correct for the training data. Then $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v}_{\text{new}})$ can come close to perfection for new data that the system has never seen.

The learning function $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v})$ grows out of linear algebra and optimization and statistics and high performance computing. Our aim in Chapters 10–11 is to understand (in part) why it succeeds.

Above all, I hope these Notes help you to teach linear algebra and learn linear algebra. Each chapter ends with suggested problems (solutions online). This subject is used in so many valuable ways. And it rests on ideas that everyone can understand.

Thank you ! **Gilbert Strang**

Triangular basis for 4-dimensional space (one example : columns of U)
The four columns of U are independent
The inverse matrix U^{-1} is also triangular : $\boldsymbol{U^{-1}U = I}$

$$\boldsymbol{U} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{1} & \mathbf{1} \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix} \qquad \boldsymbol{U^{-1}} = \begin{bmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{1} & \mathbf{-1} & \mathbf{0} \\ 0 & 0 & \mathbf{1} & \mathbf{-1} \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix}$$

Orthogonal basis for 4-dimensional space (one example : columns of Q)
The four columns of Q are perpendicular
The inverse matrix Q^{-1} is the transpose matrix Q^{T} : $\boldsymbol{Q^{\mathrm{T}}Q = I}$

$$\boldsymbol{Q} = \frac{1}{2}\begin{bmatrix} \mathbf{1} & \mathbf{1} & 1 & 1 \\ \mathbf{1} & \mathbf{-1} & -1 & 1 \\ \mathbf{1} & \mathbf{1} & -1 & -1 \\ \mathbf{1} & \mathbf{-1} & 1 & -1 \end{bmatrix} \qquad \boldsymbol{Q^{-1}} = \frac{1}{2}\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{-1} & \mathbf{1} & \mathbf{-1} \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{bmatrix} = \boldsymbol{Q^{\mathrm{T}}}$$

Four Ways to Multiply $AB = C$

$$\begin{bmatrix} \boldsymbol{X} & \boldsymbol{X} \\ \boldsymbol{x} & \boldsymbol{x} \\ \boldsymbol{x} & \boldsymbol{x} \end{bmatrix}\begin{bmatrix} \boldsymbol{X} & \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} \\ \boldsymbol{X} & \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} \end{bmatrix}$$

(**Row** $\boldsymbol{i}$ of A) $\cdot$ (**Column** $\boldsymbol{k}$ of B) = **Number** $\boldsymbol{C_{ik}}$
$i = 1$ to 3 $\quad k = 1$ to 4 $\quad$ **12 numbers**

$$\begin{bmatrix} \boldsymbol{X} & \boldsymbol{X} \\ \boldsymbol{X} & \boldsymbol{X} \\ \boldsymbol{X} & \boldsymbol{X} \end{bmatrix}\begin{bmatrix} \boldsymbol{X} & \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} \\ \boldsymbol{X} & \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} \end{bmatrix}$$

$\boldsymbol{A}$ times (**Column** $\boldsymbol{k}$ of B) = **Column** $\boldsymbol{k}$ **of** $\boldsymbol{C}$
$k = 1$ to 4 $\quad$ **4 columns**

$$\begin{bmatrix} \boldsymbol{X} & \boldsymbol{X} \\ \boldsymbol{x} & \boldsymbol{x} \\ \boldsymbol{x} & \boldsymbol{x} \end{bmatrix}\begin{bmatrix} \boldsymbol{X} & \boldsymbol{X} & \boldsymbol{X} & \boldsymbol{X} \\ \boldsymbol{X} & \boldsymbol{X} & \boldsymbol{X} & \boldsymbol{X} \end{bmatrix}$$

(**Row** $\boldsymbol{i}$ of A) times $\boldsymbol{B}$ = **Row** $\boldsymbol{i}$ **of** $\boldsymbol{C}$
$i = 1$ to 3 $\quad$ **3 rows**

$$\begin{bmatrix} \boldsymbol{X} & \boldsymbol{x} \\ \boldsymbol{X} & \boldsymbol{x} \\ \boldsymbol{X} & \boldsymbol{x} \end{bmatrix}\begin{bmatrix} \boldsymbol{X} & \boldsymbol{X} & \boldsymbol{X} & \boldsymbol{X} \\ \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} \end{bmatrix}$$

(**Column** $\boldsymbol{j}$ of A) (**Row** $\boldsymbol{j}$ of B) = **Rank 1 Matrix**
$j = 1$ to 2 $\quad$ **2 matrices**

Those four ways use the same multiplications in different orders.

Row times Column is the usual way for hand computation.

For understanding : $A\boldsymbol{x}$ is a combination of the columns of A.