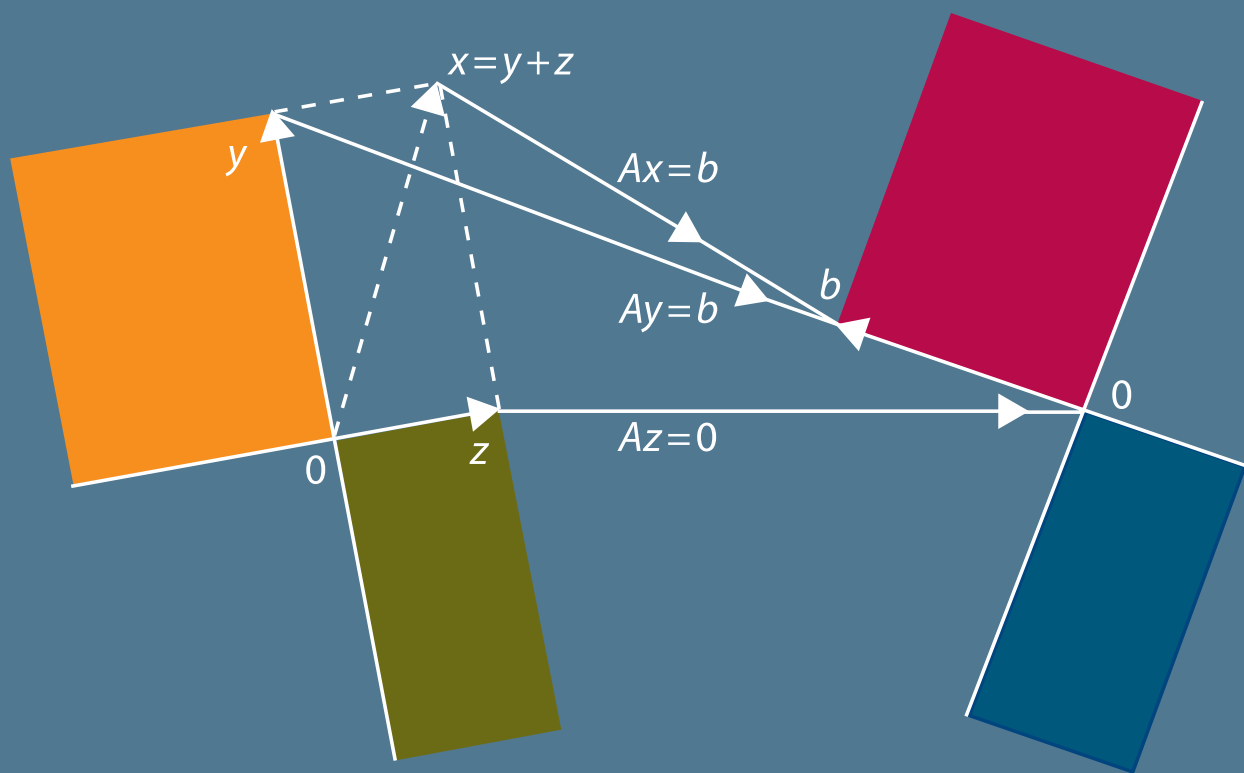


Introduction to

# LINEAR ALGEBRA

SIXTH EDITION



GILBERT STRANG

Wellesley-Cambridge Press  
 Box 812060 Wellesley MA 02482  
 gs @ math.mit.edu

# Introduction to Linear Algebra, Sixth Edition

Gilbert Strang      Wellesley-Cambridge Press

One goal of this Preface can be achieved right away. You need to know about the video lectures for MIT's Linear Algebra course **Math 18.06**. Those videos go with this book, and they are part of MIT's OpenCourseWare. The direct links to linear algebra are

<https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/>

<https://ocw.mit.edu/courses/18-06sc-linear-algebra-fall-2011/>

On YouTube those lectures are at <https://ocw.mit.edu/1806videos> and [/1806scvideos](https://ocw.mit.edu/1806scvideos)

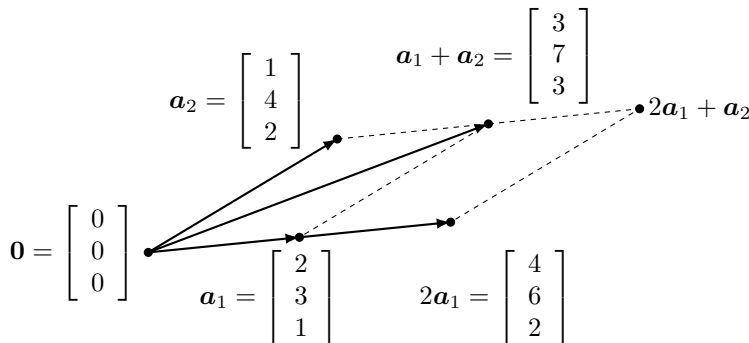
The first link brings the original lectures from the dawn of OpenCourseWare. Problem solutions by graduate students (really good) and also a short introduction to linear algebra were added to the new 2011 lectures. On both websites, the left column is a link to the contents (click on +). And the course today has a new start—the crucial ideas of linear independence and the column space of a matrix have moved near the front.

I would like to tell you about those ideas in this Preface.

Start with two column vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$ . They can have three components each, so they correspond to points in 3-dimensional space. The picture needs a center point which locates the zero vector:

$$\mathbf{a}_1 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} \quad \text{zero vector} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

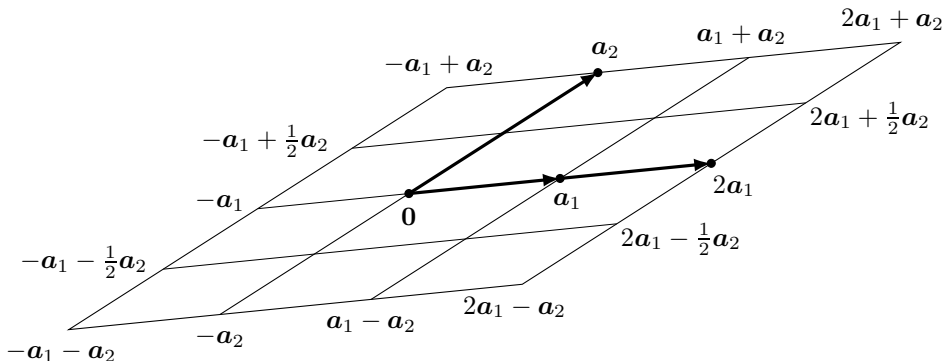
The vectors are drawn on this 2-dimensional page. But we all have practice in visualizing three-dimensional pictures. Here are  $\mathbf{a}_1$ ,  $\mathbf{a}_2$ ,  $2\mathbf{a}_1$ , and the vector sum  $\mathbf{a}_1 + \mathbf{a}_2$ .



That picture illustrated two basic operations—adding vectors  $a_1 + a_2$  and multiplying a vector by 2. Combining those operations produced a “**linear combination**”  $2a_1 + a_2$ :

$$\text{Linear combination} = ca_1 + da_2 \text{ for any numbers } c \text{ and } d$$

Those numbers  $c$  and  $d$  can be negative. In that case  $ca_1$  and  $da_2$  will reverse their directions: they go right to left. Also very important,  $c$  and  $d$  can involve fractions. Here is a picture with a lot more linear combinations. **Eventually we want all vectors  $ca_1 + da_2$ .**



Here is the key! **The combinations  $ca_1 + da_2$  fill a whole plane.** It is an infinite plane in 3-dimensional space. By using more and more fractions and decimals  $c$  and  $d$ , we fill in a complete plane. *Every point on the plane is a combination of  $a_1$  and  $a_2$ .*

Now comes a fundamental idea in linear algebra: **a matrix.** The matrix  $A$  holds  $n$  **column vectors**  $a_1, a_2, \dots, a_n$ . At this point our matrix has two columns  $a_1$  and  $a_2$ , and those are vectors in 3-dimensional space. So the matrix has **three rows and two columns.**

<b>3 by 2 matrix</b> $m = 3$ rows $n = 2$ columns	$A = \begin{bmatrix} & a_1 & a_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 2 \end{bmatrix}$
---	---

The combinations of those two columns produced a plane in three-dimensional space. There is a natural name for that plane. **It is the column space of the matrix.** For any  $A$ , **the column space of  $A$  contains all combinations of the columns.**

Here are the four ideas introduced so far. You will see them all in Chapter 1.

1. **Column vectors**  $a_1$  and  $a_2$  in three dimensions
2. **Linear combinations**  $ca_1 + da_2$  of those vectors
3. **The matrix  $A$**  contains the columns  $a_1$  and  $a_2$
4. **Column space of the matrix** = all linear combinations of the columns = plane

Now we include 2 more columns in  $A$

The 4 columns are in 3-dimensional space

$$A = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 3 & 4 & 7 & 0 \\ 1 & 2 & 3 & -1 \end{bmatrix}$$

Linear algebra aims for an understanding of every column space. Let me try this one.

Columns 1 and 2 produce the same plane as before (same  $a_1$  and  $a_2$ )

Column 3 contributes nothing new **because  $a_3$  is on that plane**:  $a_3 = a_1 + a_2$

Column 4 is **not on the plane**: Adding in  $c_4 a_4$  raises or lowers the plane

The column space of this matrix  $A$  is the **whole 3-dimensional space**: all points!

You see how we go a column at a time, left to right. Each column can be **independent** of the previous columns or it can be a **combination** of those columns. To produce every point in 3-dimensional space, you need three independent columns.

### Matrix Multiplication $A = CR$

Using the words “linear combination” and “independent columns” gives a good picture of that 3 by 4 matrix  $A$ . Column 3 is a linear combination: **column 1 + column 2**. **Columns 1, 2, 4 are independent**. The only way to produce the zero vector as a combination of the independent columns 1, 2, 4 is to multiply all those columns by **zero**.

We are so close to a key idea of Chapter 1 that I have to go on. Matrix multiplication is the perfect way to write down what we know. From the 4 columns of  $A$  we pick out the independent columns  $a_1, a_2, a_4$  in the column matrix  $C$ . **Every column of  $R$  tells us the combination of  $a_1, a_2, a_4$  in  $C$  that produces a column of  $A$ .  $A$  equals  $C$  times  $R$** :

$$A = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 3 & 4 & 7 & 0 \\ 1 & 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 4 & 0 \\ 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = CR$$

Column 3 of  $A$  is dependent on columns 1 and 2 of  $A$ , and column 3 of  $R$  shows how. Add the independent columns 1 and 2 of  $C$  to get column  $a_3 = a_1 + a_2 = (3, 7, 3)$  of  $A$ .

### Matrix multiplication: Each column $j$ of $CR$ is $C$ times column $j$ of $R$

Section 1.3 of the book will *multiply a matrix times a vector* (two ways). Then Section 1.4 will *multiply a matrix times a matrix*. This is the key operation of linear algebra. It is important that there is more than one good way to do this multiplication.

I am going to stop here. The normal purpose of the Preface is to tell you about the big picture. The next pages will give you two ways to organize this subject—especially the first seven chapters that more than fill up most linear algebra courses. Then come optional chapters, leading to the most active topic in applications today: **deep learning**.

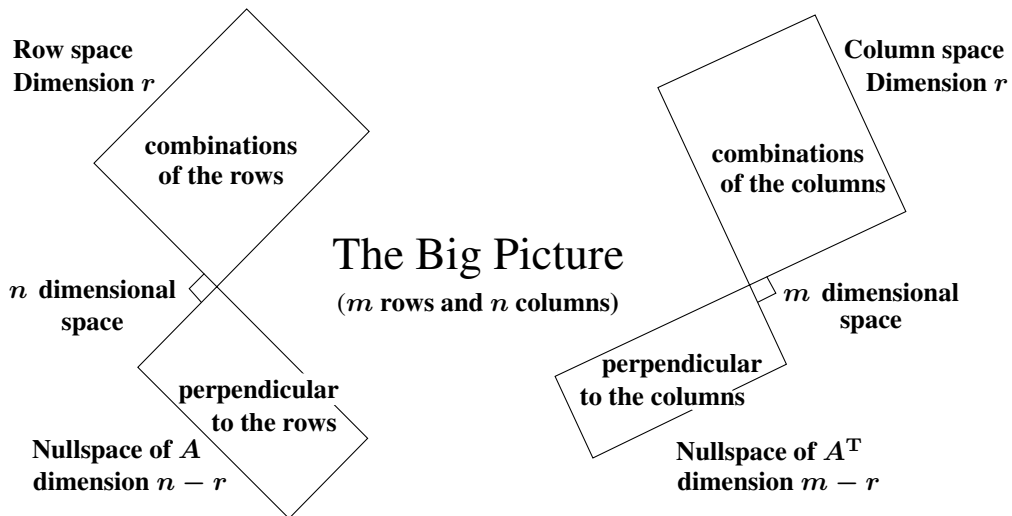
## The Four Fundamental Subspaces

You have just seen how the course begins—with the columns of a matrix  $A$ . There were two key steps. One step was to take all combinations  $ca_1 + da_2 + ea_3 + fa_4$  of the columns. This led to the **column space of  $A$** . The other step was to **factor the matrix into  $C$  times  $R$** . That matrix  $C$  holds a full set of **independent columns**.

I fully recognize that this is only the Preface to the book. You have had zero practice with the column space of a matrix (and even less practice with  $C$  and  $R$ ). But the good thing is: Those are the right directions to start. Eventually, *every matrix will lead to four fundamental spaces*. Together with the column space of  $A$  comes the **row space—all combinations of the rows**. When we take all combinations of the  $n$  columns and all combinations of the  $m$  rows—those combinations fill up “spaces” of vectors.

The other two subspaces complete the picture. Suppose the row space is a plane in three dimensions. Then there is one special direction in the 3D picture—**that direction is perpendicular to the row space**. That perpendicular line is the **nullspace** of the matrix. We will see that the vectors in the nullspace (perpendicular to all the rows) solve  $Ax = \mathbf{0}$ : the most basic of linear equations.

And if vectors perpendicular to all the rows are important, so are the vectors perpendicular to all the columns. Here is the picture of the **Four Fundamental Subspaces**.



**The Four Fundamental Subspaces : An  $m$  by  $n$  matrix with  $r$  independent columns.**

This picture of four subspaces comes in Chapter 3. The idea of perpendicular spaces is developed in Chapter 4. And special “basis vectors” for all four subspaces are discovered in Chapter 7. That step is the final piece in the **Fundamental Theorem of Linear Algebra**. The theorem includes an amazing fact about any matrix, square or rectangular:

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

## Five Factorizations of a Matrix

Here are the organizing principles of linear algebra. When our matrix has a special property, these factorizations will show it. Chapter after chapter, they express the key idea in a direct and useful way.

The usefulness increases as you go down the list. **Orthogonal matrices** are the winners in the end, because their columns are perpendicular unit vectors. That is perfection.

$$\text{2 by 2 Orthogonal Matrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} = \text{Rotation by Angle } \theta$$

**Here are the five factorizations from Chapters 1, 2, 4, 6, 7 :**

- 1  $A = CR$  =  $R$  combines independent columns in  $C$  to give all columns of  $A$
- 2  $A = LU$  = Lower triangular  $L$  times Upper triangular  $U$
- 4  $A = QR$  = Orthogonal matrix  $Q$  times Upper triangular  $R$
- 6  $S = Q\Lambda Q^T$  = (Orthogonal  $Q$ ) (Eigenvalues in  $\Lambda$ ) (Orthogonal  $Q^T$ )
- 7  $A = U\Sigma V^T$  = (Orthogonal  $U$ ) (Singular values in  $\Sigma$ ) (Orthogonal  $V^T$ )

May I call your attention to the last one? It is the **Singular Value Decomposition (SVD)**. It applies to every matrix  $A$ . Those factors  $U$  and  $V$  have perpendicular columns—all of length one. Multiplying any vector by  $U$  or  $V$  leaves a vector of the same length—so computations don't blow up or down. And  $\Sigma$  is a positive diagonal matrix of “singular values”. If you learn about eigenvalues and eigenvectors in Chapter 6, *please* continue a few pages to singular values in Section 7.1.

## Deep Learning

For a true picture of linear algebra, applications have to be included. Completeness is totally impossible. At this moment, the dominating direction of applied mathematics has one special requirement: *It cannot be entirely linear!*

One name for that direction is “deep learning”. It is an extremely successful approach to a fundamental scientific problem: **Learning from data**. In many cases the data comes in a matrix. Our goal is to look inside the matrix for the connections between variables. Instead of solving matrix equations or differential equations that express known input-output rules, *we have to find those rules*. The success of deep learning is to build a function  $F(x, v)$  with inputs  $x$  and  $v$  of two kinds:

The vectors  $v$  describes the features of the training data.

The matrices  $x$  assign weights to those features.

The function  $F(x, v)$  is close to the correct output for that training data  $v$ .

**When  $v$  changes to unseen test data,  $F(x, v)$  stays close to correct.**

This success comes partly from the form of the learning function  $F$ , which allows it to include vast amounts of data. In the end, a linear function  $F$  would be totally inadequate. The favorite choice for  $F$  is **piecewise linear**. This combines simplicity with generality.

## Applications in the Book and on the Website

I hope this book will be useful to you long after the linear algebra course is complete. It is all the applications of linear algebra that make this possible. Matrices carry data, and other matrices *operate on that data*. The goal is to “see into a matrix” by understanding its eigenvalues and eigenvectors and singular values and singular vectors. And each application has special matrices—here are four examples:

<b>Markov matrices <math>M</math></b>	Each column is a set of probabilities adding to 1.
<b>Incidence matrices <math>A</math></b>	Graphs and networks start with a set of nodes. The matrix $A$ tells the <i>connections</i> (edges) between those nodes.
<b>Transform matrices <math>F</math></b>	The Fourier matrix uncovers the <i>frequencies</i> in the data.
<b>Covariance matrices <math>C</math></b>	The <b>variance</b> is key information about a random variable. <i>The covariance explains dependence between variables.</i>

We included those applications and more in this Sixth Edition. For the crucial computation of matrix weights in deep learning, Chapter 9 presents the ideas of **optimization**. This is where linear algebra meets calculus: **derivative = zero** becomes a matrix equation at the minimum point because  $F(x)$  has many variables.

---

Several topics from the Fifth Edition gave up their places but not their importance. Those sections simply moved onto the Web. The website for this new Sixth Edition is

**[math.mit.edu/linearalgebra](http://math.mit.edu/linearalgebra)**

That website includes sample sections from this new edition and solutions to all Problem Sets. These sections (and more) are saved online from the Fifth Edition:

<b>Fourier Series</b>	<b>Norms and Condition Numbers</b>
<b>Iterative Methods and Preconditioners</b>	<b>Linear Algebra for Cryptography</b>

---

Here is a small touch of linear algebra—three questions before this course gets serious:

1. Suppose you draw three straight line segments of lengths  $r$  and  $s$  and  $t$  on this page. What are the conditions on those three lengths to allow you to make the segments into a triangle? In this question you can choose the directions of the three lines.
2. Now suppose the directions of three straight lines  $u, v, w$  are fixed and different. But you could stretch those lines to  $au, bv, cw$  with any numbers  $a, b, c$ . Can you always make a closed triangle out of the three vectors  $au, bv, cw$ ?
3. Linear algebra doesn't stay in a plane! Suppose you have **four lines**  $u, v, w, z$  in different directions in 3-dimensional space. Can you always choose the numbers  $a, b, c, d$  (zeros not allowed) so that  $au + bv + cw + dz = \mathbf{0}$ ?

---

For typesetting this book, maintaining its website, offering quality textbooks to Indian fans, I am grateful to Ashley C. Fernandes of Wellesley Publishers ([www.wellesleypublishers.com](http://www.wellesleypublishers.com))

# Introduction to Linear Algebra, Sixth Edition




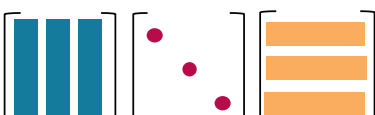

Gilbert Strang Wellesley-Cambridge Press

<b>1</b>	<b>Vectors and Matrices</b>	<b>1</b>
1.1	Vectors and Linear Combinations . . . . .	2
1.2	Lengths and Angles from Dot Products . . . . .	9
1.3	Matrices and Their Column Spaces . . . . .	18
1.4	Matrix Multiplication $AB$ and $CR$ . . . . .	27
<b>2</b>	<b>Solving Linear Equations <math>Ax = b</math></b>	<b>39</b>
2.1	Elimination and Back Substitution . . . . .	40
2.2	Elimination Matrices and Inverse Matrices . . . . .	49
2.3	Matrix Computations and $A = LU$ . . . . .	57
2.4	Permutations and Transposes . . . . .	64
<b>3</b>	<b>The Four Fundamental Subspaces</b>	<b>75</b>
3.1	Vector Spaces and Subspaces . . . . .	76
3.2	Computing the Nullspace by Elimination: $A = CR$ . . . . .	84
3.3	The Complete Solution to $Ax = b$ . . . . .	95
3.4	Independence, Basis, and Dimension . . . . .	106
3.5	Dimensions of the Four Subspaces . . . . .	120
<b>4</b>	<b>Orthogonality</b>	<b>135</b>
4.1	Orthogonality of Vectors and Subspaces . . . . .	136
4.2	Projections onto Lines and Subspaces . . . . .	143
4.3	Least Squares Approximations . . . . .	155
4.4	Orthonormal Bases and Gram-Schmidt . . . . .	168
4.5	The Pseudoinverse of a Matrix . . . . .	182
<b>5</b>	<b>Determinants</b>	<b>191</b>
5.1	<b>3</b> by <b>3</b> Determinants and Cofactors . . . . .	192
5.2	Computing and Using Determinants . . . . .	198
5.3	Areas and Volumes by Determinants . . . . .	204
<b>6</b>	<b>Eigenvalues and Eigenvectors</b>	<b>209</b>
6.1	Introduction to Eigenvalues: $Ax = \lambda x$ . . . . .	210
6.2	Diagonalizing a Matrix . . . . .	225
6.3	Symmetric Positive Definite Matrices . . . . .	239
6.4	Complex Numbers and Vectors and Matrices . . . . .	255
6.5	Solving Linear Differential Equations . . . . .	263



<b>7</b>	<b>The Singular Value Decomposition (SVD)</b>	<b>286</b>
7.1	Singular Values and Singular Vectors . . . . .	287
7.2	Image Processing by Linear Algebra . . . . .	297
7.3	Principal Component Analysis (PCA by the SVD) . . . . .	302
<b>8</b>	<b>Linear Transformations</b>	<b>308</b>
8.1	The Idea of a Linear Transformation . . . . .	309
8.2	The Matrix of a Linear Transformation . . . . .	318
8.3	The Search for a Good Basis . . . . .	327
<b>9</b>	<b>Linear Algebra in Optimization</b>	<b>335</b>
9.1	Minimizing a Multivariable Function . . . . .	336
9.2	Backpropagation and Stochastic Gradient Descent . . . . .	346
9.3	Constraints, Lagrange Multipliers, Minimum Norms . . . . .	355
9.4	Linear Programming, Game Theory, and Duality . . . . .	364
<b>10</b>	<b>Learning from Data</b>	<b>370</b>
10.1	Piecewise Linear Learning Functions . . . . .	372
10.2	Creating and Experimenting . . . . .	381
10.3	Mean, Variance, and Covariance . . . . .	386
<b>Appendix 1</b>	<b>The Ranks of <math>AB</math> and <math>A + B</math></b>	<b>400</b>
<b>Appendix 2</b>	<b>Matrix Factorizations</b>	<b>401</b>
<b>Appendix 3</b>	<b>Counting Parameters in the Basic Factorizations</b>	<b>403</b>
<b>Appendix 4</b>	<b>Codes and Algorithms for Numerical Linear Algebra</b>	<b>404</b>
<b>Appendix 5</b>	<b>The Jordan Form of a Square Matrix</b>	<b>405</b>
<b>Appendix 6</b>	<b>Tensors</b>	<b>406</b>
<b>Appendix 7</b>	<b>The Condition Number of a Matrix Problem</b>	<b>407</b>
<b>Appendix 8</b>	<b>Markov Matrices and Perron-Frobenius</b>	<b>408</b>
<b>Appendix 9</b>	<b>Elimination and Factorization</b>	<b>410</b>
<b>Appendix 10</b>	<b>Computer Graphics</b>	<b>413</b>
	<b>Index of Equations</b>	<b>419</b>
	<b>Index of Notations</b>	<b>422</b>
	<b>Index</b>	<b>423</b>

## The Five Factorizations of a Matrix

$A=CR$		<p><b>C</b> First <math>r</math> independent columns of <math>A</math></p> <p><b>R</b> Combines the columns in <math>C</math> to produce all columns in <math>A</math></p>
$A=LU$		<p><b>L</b> Lower triangular matrix/all ones on the diagonal</p> <p><b>U</b> Upper triangular matrix/no zeros on the diagonal</p>
$A=QR$		<p><b>Q</b> Columns are orthogonal unit vectors</p> <p><b>R</b> Triangular <math>R</math> combines those orthonormal columns of <math>Q</math> to produce the columns of <math>A</math></p>
$S=Q\Lambda Q^T$ $SQ = Q\Lambda$		<p><b>Q</b> Columns of <math>Q</math> are orthonormal eigenvectors of <math>S</math></p> <p><b>Λ</b> Diagonal matrix: Real eigenvalues of <math>S</math></p>
$A=U\Sigma V^T$ $AV=U\Sigma$		<p><b>U</b> Orthonormal singular vectors (outputs from <math>A</math>)</p> <p><b>Σ</b> Diagonal matrix: Positive singular values of <math>A</math></p> <p><b>V</b> Orthonormal singular vectors (inputs to <math>A</math>)</p>

This is the textbook that goes with Gilbert Strang's video lectures on [ocw.mit.edu](http://ocw.mit.edu).

ISBN 978-1-7331466-7-8



9781733146678

**Wellesley-Cambridge Press**  
**Box 812060 Wellesley MA 02482**  
**gs @ math.mit.edu**