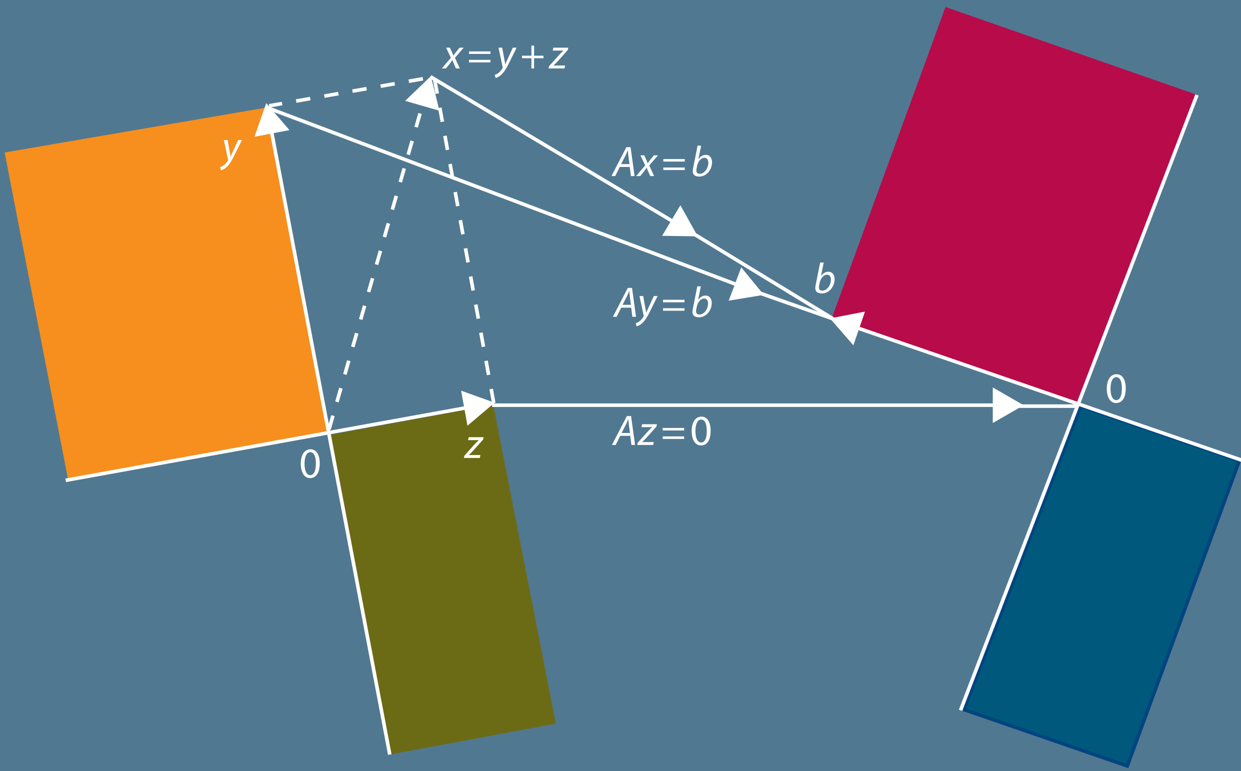
Introduction to

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**

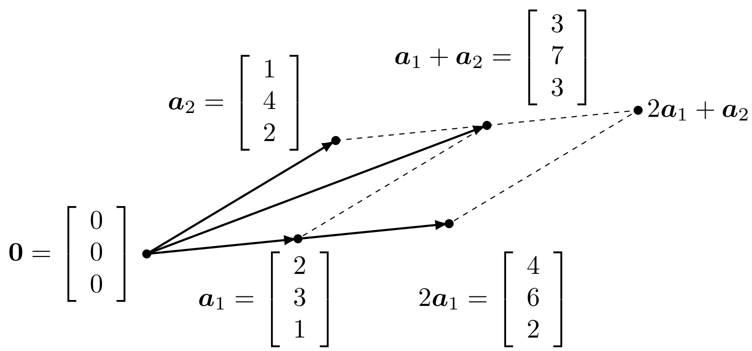
The first link brings the original lectures from the dawn of OpenCourseWare. Problem solutions by graduate students (really good) and also a short introduction to linear algebra were added to the new 2011 lectures. On both websites, the left column is a link to the contents (click on $+$). And the course 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 $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. 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:

$$\boldsymbol{a}_1 = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \qquad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} \qquad \textbf{zero vector} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

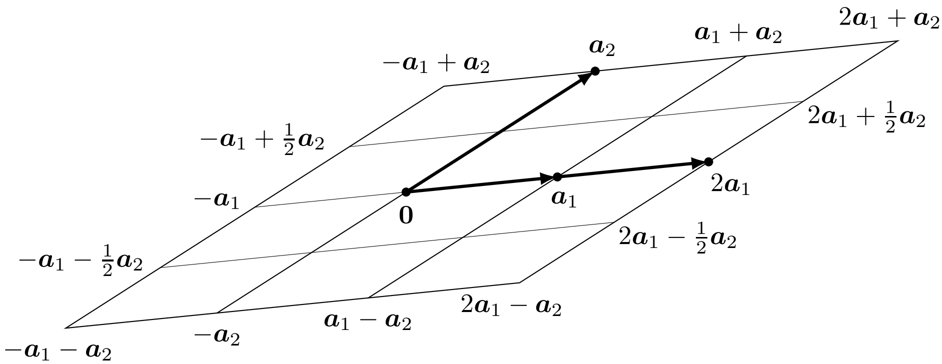
The vectors are drawn on this 2-dimensional page. But we all have practice in visualizing three-dimensional pictures. Here are $\boldsymbol{a}_1, \boldsymbol{a}_2, 2\boldsymbol{a}_1$, and the vector sum $\boldsymbol{a}_1 + \boldsymbol{a}_2$.



That picture illustrated two basic operations—adding vectors $\boldsymbol{a}_1 + \boldsymbol{a}_2$ and multiplying a vector by 2. Combining those operations produced a **"linear combination"** $2\boldsymbol{a}_1 + \boldsymbol{a}_2$:

$$\boxed{\textbf{Linear combination} = c\boldsymbol{a_1} + d\boldsymbol{a_2} \ \textbf{ for any numbers } c \textbf{ and } d}$$

Those numbers c and d can be negative. In that case $c\boldsymbol{a}_1$ and $d\boldsymbol{a}_2$ 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 $c\boldsymbol{a}_1 + d\boldsymbol{a}_2$.**



Here is the key ! **The combinations $c\boldsymbol{a_1} + d\boldsymbol{a_2}$ fill a whole plane**. It is an infinite plane in 3-dimensional space. By using more and more fractions and decimals $\boldsymbol{c}$ and $\boldsymbol{d}$, we fill in a complete plane. *Every point on the plane is a combination of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$.*

Now comes a fundamental idea in linear algebra: **a matrix**. The matrix A holds **n column vectors** $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. At this point our matrix has two columns $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, and those are vectors in 3-dimensional space. So the matrix has **three rows and two columns**.

$$\boxed{\begin{array}{ll} \textbf{3 by 2 matrix} & \\ \boldsymbol{m} = \textbf{3 rows} & \boldsymbol{A} = \begin{bmatrix} & & \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 \\ & & \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 2 \end{bmatrix} \\ \boldsymbol{n} \ = \textbf{2 columns} & \end{array}}$$

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 $\boldsymbol{A}$ 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** $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in three dimensions

2. **Linear combinations** $\boldsymbol{c}\boldsymbol{a}_1 + \boldsymbol{d}\boldsymbol{a}_2$ of those vectors

3. **The matrix $\boldsymbol{A}$** contains the columns $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$

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

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

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 $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$)

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

Column 4 is **not on the plane** : Adding in $\boldsymbol{c_4 a_4}$ 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 $\boldsymbol{a_1}, \boldsymbol{a_2}, \boldsymbol{a_4}$ in the column matrix $\boldsymbol{C}$. **Every column of $\boldsymbol{R}$ tells us the combination of $\boldsymbol{a_1, a_2, a_4}$ in $\boldsymbol{C}$ that produces a column of $\boldsymbol{A}$. $\boldsymbol{A}$ equals $\boldsymbol{C}$ times $\boldsymbol{R}$** :

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

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 $\boldsymbol{a}_3 = \boldsymbol{a}_1 + \boldsymbol{a}_2 = (\mathbf{3}, \mathbf{7}, \mathbf{3})$ 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**.

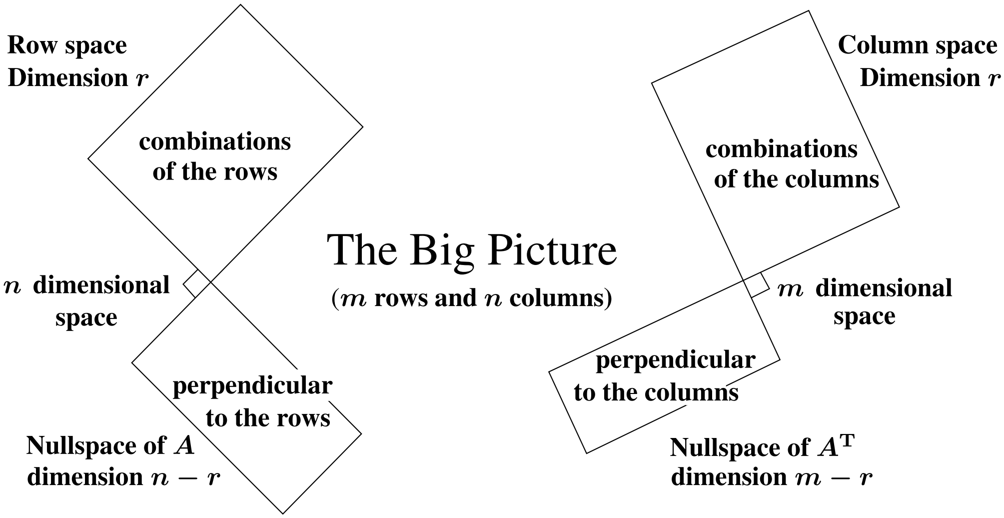
The Four Fundamental Subspaces

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 $c\boldsymbol{a}_1 + d\boldsymbol{a}_2 + e\boldsymbol{a}_3 + f\boldsymbol{a}_4$ 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 $A\boldsymbol{x} = \mathbf{0}$: 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.

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

Here are the five factorizations from Chapters $\mathbf{1, 2, 4, 6, 7}$:

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

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 Σ 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 $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v})$ with inputs $\boldsymbol{x}$ and $\boldsymbol{v}$ of two kinds :

The vectors $\boldsymbol{v}$ describes the features of the training data.

The matrices $\boldsymbol{x}$ assign weights to those features.

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

When $\boldsymbol{v}$ changes to unseen test data, $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v})$ stays close to correct.

This success comes partly from the form of the learning function $\boldsymbol{F}$, which allows it to include vast amounts of data. In the end, a linear function $\boldsymbol{F}$ would be totally inadequate. The favorite choice for $\boldsymbol{F}$ 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:

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

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

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:

Fourier Series	**Norms and Condition Numbers**
Iterative Methods and Preconditioners	**Linear Algebra for Cryptography**

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 $au + bv + cw + dz = 0$?

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)

gilstrang@gmail.com Gilbert Strang

Introduction to Linear Algebra, Sixth Edition

Gilbert Strang **Wellesley-Cambridge Press**

Elimination and Factorization $A = CR$

Gilbert Strang

Abstract If a matrix A has rank r, then its row echelon form (from elimination) contains the identity matrix in its first r independent columns. How do we *interpret the matrix* F that appears in the remaining columns of that echelon form? F multiplies those first r independent columns of A to give its $n-r$ dependent columns. Then F reveals bases for the row space and the nullspace of the original matrix A. And F is the key to the column-row factorization $\boldsymbol{A} = \boldsymbol{CR}$.

1. Elimination must be just about the oldest algorithm in linear algebra. By systematically producing zeros in a matrix, it simplifies the solution of m equations $A\boldsymbol{x} = \boldsymbol{b}$. We take as example this 3 by 4 matrix A, with row $1+$ row $2 =$ row 3. Then its rank is $r = 2$, and we execute the familiar elimination steps to find its *reduced row echelon form* Z:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 11 & 17 \\ 3 & 7 & 37 & 57 \\ 4 & 9 & 48 & 74 \end{bmatrix} \to \boldsymbol{Z} = \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

At this point, we pause the algorithm to ask a question: **How is $\boldsymbol{Z}$ related to $\boldsymbol{A}$?** One answer comes from the fundamental subspaces associated with A:

1) The two nonzero rows of $\boldsymbol{Z}$ (call them $\boldsymbol{R}$) are a basis for the row space of A.

2) The first two columns of $\boldsymbol{A}$ (call them $\boldsymbol{C}$) are a basis for the column space of A.

3) The nullspace of $\boldsymbol{Z}$ equals the nullspace of $\boldsymbol{A}$ (orthogonal to the same row space).

Those were our reasons for elimination in the first place. "Simplify the matrix A without losing the information it contains." By applying the same steps to the right hand side of $A\boldsymbol{x} = \boldsymbol{b}$, we reach an equation $Z\boldsymbol{x} = \boldsymbol{d}$—with the same solutions $\boldsymbol{x}$ and the simpler matrix Z.

The object of this short note is to express the result of elimination in a different way. This factorization cannot be new, but it deserves new emphasis.

Elimination factors $\boldsymbol{A}$ into $\boldsymbol{C}$ times $\boldsymbol{R} = (\boldsymbol{m} \times \boldsymbol{r})$ times $(\boldsymbol{r} \times \boldsymbol{n})$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 11 & 17 \\ 3 & 7 & 37 & 57 \\ 4 & 9 & 48 & 74 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \end{bmatrix} = \boldsymbol{CR}$$

C has full column rank $r = 2$ and R has full row rank $r = 2$. When we establish that $A = CR$ is true for every matrix A, this factorization brings with it a proof of the first great theorem in linear algebra: **Column rank equals row rank**.

2. Here is a description of C and R that is independent of the algorithm (row operations) that computes them.

Suppose the first $\boldsymbol{r}$ independent columns of $\boldsymbol{A}$ go into $\boldsymbol{C}$. Then the other $\boldsymbol{n-r}$ columns of $\boldsymbol{A}$ must be combinations $\boldsymbol{CF}$ of those independent columns in $\boldsymbol{C}$. That key matrix $\boldsymbol{F}$ is part of the row factor $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix} \boldsymbol{P}$, with $\boldsymbol{r}$ independent rows. Then right away we have $A = CR$:

$$\boldsymbol{A} = \boldsymbol{CR} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{CF} \end{bmatrix} \boldsymbol{P} = \begin{bmatrix} \textbf{Independent cols} & \textbf{Dependent cols} \end{bmatrix} \textbf{Permute cols} \tag{1}$$

If the r independent columns come first in A, that permutation matrix will be $P = I$. Otherwise we need P to permute the n columns of C and CF into correct position in A.

Here is an example of $A = \boldsymbol{C}\begin{bmatrix} I & F \end{bmatrix} P$ in which P exchanges columns 2 and 3:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} P = \boldsymbol{CR}. \tag{2}$$

The essential information in $Z = \textbf{rref}(A)$ is the list of r independent columns of A, and the matrix F (r by $n-r$) that combines those independent columns to give the $n-r$ dependent columns CF in A. This uniquely defines Z in equation (3).

3. The factorization $A = CR$ is confirmed. But how do we determine the first r independent columns in A (going into C) and the dependencies CF of the remaining $n-r$ columns? This is the moment for **row operations** on A. Three operations are allowed, to put A into its reduced row echelon form $Z = \textbf{rref}(A)$:

(a) Subtract a multiple of one row from another row (below or above)

(b) Exchange two rows

(c) Divide a row by its first nonzero entry

All linear algebra teachers and a positive fraction of students know those steps and their outcome $\textbf{rref}(A)$. It contains an r by r identity matrix I (only zeros can precede those 1's). The position of I reveals the first r independent columns of A. And equation (1) above reveals the meaning of F! It tells us how the $n-r$ *dependent* columns CF of A come from the independent columns in C. The remaining $m-r$ dependent rows of A must become zero rows in Z:

$$\textbf{Elimination reduces } \boldsymbol{A} \textbf{ to } \boldsymbol{Z} = \textbf{rref}(A) = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{P} \tag{3}$$

All our row operations (a)(b)(c) are invertible. (This is Gauss-Jordan elimination: operating on rows above the pivot row as well as below.) But the matrix that reduces A to this echelon form is less important than the **factorization $\boldsymbol{A = CR}$** that it uncovers in equation (1).

4. Before we apply $A = CR$ to solving $A\boldsymbol{x} = \mathbf{0}$, we want to give a column by column (left to right) construction of **rref**(A) from A. After elimination on k columns, that part of the matrix is in its own **rref** form. We are ready for elimination on the current column $k+1$. This new column has an upper part $\boldsymbol{u}$ and a lower part $\boldsymbol{\ell}$:

$$\textbf{First } \boldsymbol{k+1} \textbf{ columns} \qquad \begin{bmatrix} I_k & F_k \\ 0 & 0 \end{bmatrix} P_k \text{ followed by } \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\ell} \end{bmatrix}. \tag{4}$$

The big question is: **Does this new column $\boldsymbol{k+1}$ join with $\boldsymbol{I_k}$ or $\boldsymbol{F_k}$?**

If $\boldsymbol{\ell}$ is all zeros, the new column is **dependent** on the first k columns. Then $\boldsymbol{u}$ joins with F_k to produce F_{k+1} in the next step to column $k+2$.

If $\boldsymbol{\ell}$ is not all zero, the new column is **independent** of the first k columns. Pick any nonzero in $\boldsymbol{\ell}$ (preferably the largest) as the pivot. Move that row of A to the top of $\boldsymbol{\ell}$. Then use that pivot row to zero out (by standard elimination) all the rest of column $k+1$. (That step is expected to change the columns after $k+1$.) Column $k+1$ joins with I_k to produce I_{k+1}. We adjust P_k and we are ready for column $k+2$.

At the end of elimination, we have a most desirable list of column numbers. They tell us the **first $\boldsymbol{r}$ independent columns of $\boldsymbol{A}$**. Those are the columns of C. They led to the identity matrix $I_{r \text{ by } r}$ in the row factor R of $A = CR$.

5. What is achieved by reducing A to **rref**(A)? The row space is not changed! Then its orthogonal complement (**the nullspace of $\boldsymbol{A}$**) is not changed. Each column of CF tells us how a dependent column of A is a combination of the independent columns in C. Effectively, **the columns of $\boldsymbol{F}$ are telling us $\boldsymbol{n-r}$ solutions to $\boldsymbol{Ax} = \mathbf{0}$**. This is easiest to see by example.

$$\begin{matrix} x_1 + 2x_2 + 11x_3 + 17x_4 = 0 \\ 3x_1 + 7x_2 + 37x_3 + 57x_4 = 0 \end{matrix} \quad \text{reduces to} \quad \begin{bmatrix} I & F \end{bmatrix} \boldsymbol{x} = \begin{matrix} x_1 + 3x_3 + 5x_4 = 0 \\ x_2 + 4x_3 + 6x_4 = 0 \end{matrix}$$

The solution with $x_3 = 1$ and $x_4 = 0$ is $\boldsymbol{x} = \begin{bmatrix} -3 & -4 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$. Notice 3 and 4 from F. The second solution with $x_3 = 0$ and $x_4 = 1$ is $\boldsymbol{x} = \begin{bmatrix} -5 & -6 & 0 & 1 \end{bmatrix}^{\mathrm{T}}$. Those solutions are the two columns of $\boldsymbol{X}$ in $\boldsymbol{AX} = \mathbf{0}$. (This example has $P = I$.) **The $\boldsymbol{n-r}$ columns of $\boldsymbol{X}$ are a natural basis for the nullspace of $\boldsymbol{A}$:**

$$\boldsymbol{A} = \boldsymbol{C}\begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix}\boldsymbol{P} \text{ multiplies } \boldsymbol{X} = \boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} -\boldsymbol{F} \\ \boldsymbol{I_{n-r}} \end{bmatrix} \text{ to give } \boldsymbol{AX} = -\boldsymbol{CF} + \boldsymbol{CF} = \mathbf{0}. \tag{5}$$

With $PP^{\mathrm{T}} = I$ for the permutation, each column of X solves $A\boldsymbol{x} = \mathbf{0}$. Those $n-r$ solutions in X tell us what we know: Each dependent column of A is a combination of the independent columns in C. In the example, column 3 = 3 (column 1) + 4 (column 2).

Gauss-Jordan elimination leading to $A = CR$ is less efficient than the Gauss process that directly solves $A\boldsymbol{x} = \boldsymbol{b}$. The latter stops at a triangular system $U\boldsymbol{x} = \boldsymbol{c}$: back substitution creates $\boldsymbol{x}$. Gauss-Jordan has the extra cost of eliminating upwards. If we only want to solve equations, stopping at a triangular factorization is faster.

6. Block elimination The row operations that reduce A to its echelon form produce one zero at a time. A key part of that echelon form is an r by r identity matrix. If we think on a larger scale—instead of one row at a time—**that output $\boldsymbol{I}$ tells us that some $\boldsymbol{r}$ by $\boldsymbol{r}$ matrix has been inverted**. Following that lead brings a "matrix understanding" of elimination.

Suppose that the matrix $\boldsymbol{W}$ in the first r rows and columns of A is invertible. Then *elimination takes all its instructions from W!* One entry at a time—or all at once by "block elimination"—**$\boldsymbol{W}$ will change to $\boldsymbol{I}$**. In other words, the first r rows of A will yield I and F. **This identifies $\boldsymbol{F}$ as $\boldsymbol{W^{-1}H}$**. And the last $m - r$ rows will become zero rows.

$$\textbf{Block elimination} \quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \text{ reduces to } \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \textbf{rref}(\boldsymbol{A}). \tag{6}$$

That just expresses the facts of linear algebra: If A begins with r independent rows and its rank is r, then the remaining $m - r$ rows are combinations of those first r rows : $\begin{bmatrix} J & K \end{bmatrix} = JW^{-1} \begin{bmatrix} W & H \end{bmatrix}$. Those $m - r$ rows become zero rows in elimination.

In general that first r by r block might not be invertible. But elimination will find W. We can move W to the upper left corner by row and column permutations $\boldsymbol{P_r}$ and $\boldsymbol{P_c}$.Then the full expression of block elimination to reduced row echelon form is

$$\boldsymbol{P_r A P_c} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{7}$$

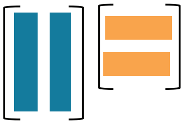
7. This raises an interesting point. Since A has rank r, we know that it has r independent rows and r independent columns. Suppose those rows are in a submatrix B and those columns are in a submatrix C. Is it always true that the r by r "intersection" W of those rows with those columns will be **invertible**? (Everyone agrees that somewhere in A there is an r by r invertible submatrix. The question is whether $B \cap C$ can be counted on to provide such a submatrix.) Is W automatically full rank?

The answer is *yes*. **The intersection of $\boldsymbol{r}$ independent rows of $\boldsymbol{A}$ with $\boldsymbol{r}$ independent columns does produce a matrix $\boldsymbol{W}$ of rank $\boldsymbol{r}$. $\boldsymbol{W}$ is invertible.**

Proof: Every column of A is a combination of the r columns of C.
Every column of the r by n matrix B is the same combination of the r columns of W.
Since B has rank r, its column space is all of $\mathbf{R}^{\boldsymbol{r}}$.
Then the column space of W is also $\mathbf{R}^{\boldsymbol{r}}$ and **the square submatrix $\boldsymbol{W}$ has rank $\boldsymbol{r}$**.

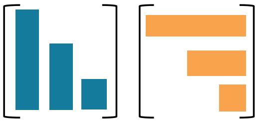
The Five Factorizations of a Matrix

$A = CR$



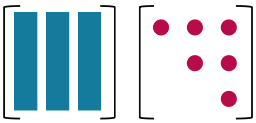
C First r independent columns of A
R Combines the columns in C to produce all columns in A

$A = LU$



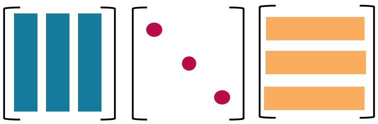
L Lower triangular matrix/all ones on the diagonal
U Upper triangular matrix/no zeros on the diagonal

$A = QR$



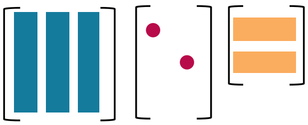
Q Columns are orthogonal unit vectors
R Triangular R combines those orthonormal columns of Q to produce the columns of A

$S = Q\Lambda Q^{T}$
$SQ = Q\Lambda$



Q Columns of Q are orthonormal eigenvectors of S
Λ Diagonal matrix: Real eigenvalues of S

$A = U\Sigma V^{T}$
$AV = U\Sigma$



U Orthonormal singular vectors (outputs from A)
Σ Diagonal matrix: Positive singular values of A
V Orthonormal singular vectors (inputs to A)

This is the textbook that goes with Gilbert Strang's video lectures on *ocw.mit.edu*.

ISBN 978-1-7331466-7-8
90000
9781733146678

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