LINEAR ALGEBRA NOTES

ADRIAN PĂCURAR

Contents

1. Matrices, EF, RREF

Identification of matrix echelon form (EF) and reduced row echelon form (RREF) is a crucial component of this course. Make sure you know the various forms well.

For EF, we don't need the pivot entries to be 1. We also don't require the entries above the pivots to be zero. This means there are multiple representations of a matrix in EF.

EF is useful for identifying the number of pivots, or determining if a system is consistent or not. EF is NOT suitable for finding the solution space of a matrix associated to a linear system of equations.

Example 1. A few EF matrices:

$$a)\ \begin{pmatrix} 1 & -3 & 0 & 5 & 1 & 4 \\ 0 & 0 & 1 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad b)\ \begin{pmatrix} 1 & -2 & 1 & -1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad c)\ \begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix}$$

The matrix in (a) is almost RREF, except for the $(1,5)$ entry (row 1, column 5) which is nonzero. The same is true for (b), which has a nonzero entry above the second pivot. Matrix (c) is the most common EF – it tells us all the pivot positions (it is not necessary to reduce it further if that's all we need).

Example 2. The following matrix in EF corresponds to a linear system. The system is clearly inconsistent (consider row 3). We do not need to reduce any further to see this.

$$\left(\begin{array}{ccccc|c} 1 & 2 & -5 & -6 & 0 & -5 \\ 0 & 7 & -6 & -3 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

RREF requires a bit more work, and is needed when trying to determine the solution(s) of a system of equations. All pivot positions (leading nonzero entries in each row) must be equal to 1, and there are zeroes above and below each pivot. RREF is unique. Regardless of the steps taken in the row-reduction process, you will always end up with the same RREF matrix.

Often times we are interested in finding the solution space for a matrix. The preferred representation for this is the *parametric/vector form*, which looks like this:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \underbrace{\begin{pmatrix} * \\ * \\ \vdots \\ * \end{pmatrix} s + \begin{pmatrix} * \\ * \\ \vdots \\ * \end{pmatrix} t + \cdots + \begin{pmatrix} * \\ * \\ \vdots \\ * \end{pmatrix} w}_{\text{col vectors associated to free variables}} + \underbrace{\begin{pmatrix} * \\ * \\ \vdots \\ * \end{pmatrix}}_{\text{constant vector}}$$

The constant part is not always present. In tutorial I taught a shortcut to finding the parametric solution straight from the matrix, without having to go through the hassle of

Example 3. The RREF matrix

$$\left(\begin{array}{ccc|c} \mathbf{1} & -2 & 0 & -1 \\ 0 & 0 & \mathbf{1} & -2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

implies that we have one free variable x_2. To find the parametric form we perform the following steps:

1) (Alignment step) Row-interchange until all the pivots are on the main diagonal. If necessary, you may need to introduce rows of zero until you have as many rows as there are variables. For this example, we have enough rows. We get:

$$\left(\begin{array}{ccc|c} \mathbf{1} & -2 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & -2 \end{array}\right)$$

2) Identify the location of missing pivots:

$$\left(\begin{array}{ccc|c} \mathbf{1} & -2 & 0 & -1 \\ 0 & \boxed{0} & 0 & 0 \\ 0 & 0 & \mathbf{1} & -2 \end{array}\right)$$

3) The solution will involve using the constant (augmented) part of the matrix, and the columns associated to free variables. In this case, x_2 is the free variable, so we will use the second column, where we replace the highlighted zero (missing pivot) by -1. The solution is:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 \\ \boxed{-1} \\ 0 \end{pmatrix} s + \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix}$$

Example 4. Here is another example for finding parametric solutions of the following matrix:

$$\left(\begin{array}{ccc|c} 1 & -3 & -5 & 0 \\ 0 & 3 & -3 & -6 \end{array}\right)$$

Notice that this is not RREF, so we need to do that first. The existing pivots are already aligned as they should be, but we need to introduce a row of zeroes (so we have 3 rows total, which is the number of unknowns).

$$\left(\begin{array}{ccc|c} 1 & -3 & -5 & 0 \\ 0 & 3 & -3 & -6 \end{array}\right) \xrightarrow{rref} \left(\begin{array}{ccc|c} 1 & 0 & -8 & -6 \\ 0 & 1 & -1 & -2 \end{array}\right) \xrightarrow{align} \left(\begin{array}{ccc|c} 1 & 0 & -8 & -6 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & \boxed{0} & 0 \end{array}\right)$$

Now we are ready to give the solution:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -8 \\ -1 \\ \boxed{-1} \end{pmatrix} s + \begin{pmatrix} -6 \\ -2 \\ 0 \end{pmatrix}$$

Example 5. Here is an example that has more than one free variable:

$$\left(\begin{array}{ccccc|c} 1 & 2 & -5 & -6 & 0 & -5 \\ 0 & 1 & -6 & -3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{rref} \left(\begin{array}{ccccc|c} 1 & 0 & 7 & 0 & 0 & -9 \\ 0 & 1 & -6 & -3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{align} \left(\begin{array}{ccccc|c} \mathbf{1} & 0 & 7 & 0 & 0 & -9 \\ 0 & \mathbf{1} & -6 & -3 & 0 & 2 \\ 0 & 0 & \boxed{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{0} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 \end{array}\right)$$

so we have 5 unknowns, 2 free variables, and the solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 7 \\ -6 \\ \boxed{-1} \\ 0 \\ 0 \end{pmatrix} s + \begin{pmatrix} 0 \\ -3 \\ 0 \\ \boxed{-1} \\ 0 \end{pmatrix} t + \begin{pmatrix} -9 \\ 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

2. Linear Combinations, Span, Independence

Linear algebra has at its core the concept of a linear combination. In multivariable calculus we used this very heavily. Every time we gave a vector in coordinate form like $\mathbf{v} = \langle a, b, c\rangle$ for real numbers a, b, c, we are really saying that $\mathbf{v}$ is a linear combination of the unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ (the standard basis for $\mathbb{R}^3$) as follows:

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

What makes this *linear* is the fact that everything is raised to the first power, and there are no mixing of variables. For example, a surface like $f = x + 2y + z$ is linear, while $g = x - xy + 5z + y^2$ is not (we have xy and y^2 which are nonlinear terms; they introduce curvature to the surface).

Definition 1. In general, for k given vectors $\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_k} \in \mathbb{R}^n$, a **linear combination** is

$$c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_k\mathbf{v_k}$$

where $c_1, c_2, \dots, c_k \in \mathbb{R}$ are scalars. Technically we can talk about linear combinations where the scalars (and the vector entries/coordinates) are complex numbers, but that is beyond what we will study in this course. Notice that a linear combination is another vector with the same dimension as the original $\mathbf{v_i}$'s.

Definition 2. The **span** of a set of vectors $\{\mathbf{v_1}, \dots, \mathbf{v_n}\}$ is the set of all possible linear combinations that can be formed using the given vectors:

$$span(\mathbf{v_1}, \dots, \mathbf{v_n}) = \{c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_k\mathbf{v_n} : c_i \in \mathbb{R}\}$$

Geometrically the span represents a linear surface (hyperplane) that is determined by the given vectors. We did this over and over again in multivariable calc, when we talked about how 2 vectors are needed to determine a plane. The determined plane is the span of the 2 given vectors (plus an offset if the plane does not go through the origin).

Linear algebra just takes that concept and generalizes it to higher dimensions. Notice how the solution of Example 5 involves a span:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \underbrace{\begin{pmatrix} 7 \\ -6 \\ \boxed{-1} \\ 0 \\ 0 \end{pmatrix} s + \begin{pmatrix} 0 \\ -3 \\ 0 \\ \boxed{-1} \\ 0 \end{pmatrix} t}_{\text{free variable column span}} + \underbrace{\begin{pmatrix} -9 \\ 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{\text{constant offset from origin}}$$

The span of a set of vectors always contains the origin (set all constants in the linear combo to zero). In the parametric form of a solution set, we have two pieces:

- The span of vectors associated to the columns of the free variables. Geometrically this is a hyperplane that goes through the origin.
- A constant offset that shifts the above hyperplane from the origin. This is sometimes zero. The offset is always zero for solution spaces of homogeneous systems.

Sometimes when we consider the span of a set of vectors, we have redundancies in that set: the object spanned by the vectors can be obtained by using less vectors than what is given. This leads us to the idea of **linear independence**. There are several ways to characterize when a set of vectors is linearly independent:

- If there is no way to reduce the size of the vector set (by removing vectors from it) while generating the same spanned object, the set of vectors is linearly independent. If we have redundancies, the set is linearly dependent.
- If we construct the matrix whose columns are the vectors $\{\mathbf{v_1}, \dots, \mathbf{v_n}\}$, and perform RREF, the vectors are linearly independent precisely when we have as many pivots as there are vectors. Otherwise, they are linearly dependent. The number of pivots tell us the dimesion of the space spanned by the given vectors (so naturally if we have less pivots than vectors, we must have redundancies in our set).

If you are given a set of vectors and you want to determine precisely which of them to keep in order to generate their span, the row-reduction process helps here.

Example 6. Consider the set of vectors

$$B = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ -7 \\ 1 \\ 0 \end{pmatrix} \right\}$$

What is the minimal set that generates the same span as $span(B)$? First create a matrix whose columns are the given vectors, and get it to echelon form. I happened to have picked the vectors so that we already have echelon form:

$$\begin{pmatrix} 1 & -3 & 3 & 5 & 1 & 4 \\ 0 & 0 & 1 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Look at which columns have pivots (in this case, columns 1, 3, 5). Then the minimal set of generators is the set containing the first, third, and fifth vector from the original set B.

3. Homogeneous Systems

There is a slightly more abstract way to phrase linear independence. If we can write the zero vector as a nontrivial (not all coefficients = zero) linear combination of the vectors, the set is linearly dependent. If we can only write $\mathbf{0}$ as the trivial linear combination, the set is independent.

This idea of writing zero as a linear combination of columns of a matrix ties to **homogeneous systems** of linear equations. This is simply a system of linear equations where the RHS of every equation is zero, like:

$$\begin{cases} ax_1 + bx_2 + cx_3 = 0 \\ dx_1 + ex_2 + fx_3 = 0 \\ gx_1 + hx_2 + lx_3 = 0 \end{cases}$$

That means that instead of working with an augmented of the form

$$\left(\begin{array}{ccc|c} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \end{array}\right)$$

we simply omit the zero column and work with the matrix containing only the coefficients of our variables:

$$\begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}$$

(This works because any row operations we perform leave the zero column unchanged.)

A homogeneous system is ALWAYS solvable – we can always plug in zero for all the variables, and the equations hold. We call this the **trivial solution**. In math, a good way to think of the work "trivial" is that it means "easy" or "obvious" – something that you can spot right away.

If the columns of the matrix above are independent, we will only ever have the trivial solution. If they are dependent, we will find **nontrivial solutions** (remember dependence means that we can write $\mathbf{0}$ as a nontrivial linear combination). These solutions are given by the same process we use for nonhomogeneous systems (row-reduce and find the parametric solution).

4. The Matrix Equation $A\mathbf{x} = \mathbf{b}$

While it may seem intimidating at first, the equation $A\mathbf{x} = \mathbf{b}$ is nothing but just a way to rephrase the following: *can we write* $\mathbf{b}$ *as a linear combination of the columns of matrix* A*? If so, what are the coefficients?* Suppose that the columns of A are given by

$$A = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{a_1} & \mathbf{a_2} & \dots & \mathbf{a_n} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

and write the vector $\mathbf{x} = \langle x_1, x_2, \dots, x_n \rangle$ (as a column vector). Then $A\mathbf{x}$ expands as

$$A\mathbf{x} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{a_1} & \mathbf{a_2} & \dots & \mathbf{a_n} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{a_1} + x_2\mathbf{a_2} + \dots + x_n\mathbf{a_n}$$

which is just a linear combination of the columns of A. Hence answering the question

Can we write $\mathbf{b}$ *as a linear combination of the columns of* A*?*

is equivalent to solving $A\mathbf{x} = \mathbf{b}$, which in turn is equivalent to performing RREF on the augmented matrix $(A|\mathbf{b})$:

$$\left(\begin{array}{cccc|c} \uparrow & \uparrow & & \uparrow & b_1 \\ \mathbf{a_1} & \mathbf{a_2} & \dots & \mathbf{a_n} & \vdots \\ \downarrow & \downarrow & & \downarrow & b_m \end{array}\right) \xrightarrow{rref} \left(\begin{array}{c|c} & \uparrow \\ * & \mathbf{x} \\ & \downarrow \end{array}\right)$$

The result of the RREF process tells us exactly what coefficients (entries of $\mathbf{x}$) we need in order to write $\mathbf{b}$ as a linear combination of the columns. So we have the following theorem:

Theorem 1. If A is an $m \times n$ matrix (m rows and n columns), with column vectors $\mathbf{a_1}, \dots, \mathbf{a_n}$ and $\mathbf{b} \in \mathbb{R}^m$, the matrix equation

$$A\mathbf{x} = \mathbf{b}$$

has the same solution set as the vector equation

$$x_1\mathbf{a_1} + x_2\mathbf{a_2} + \cdots + x_n\mathbf{a_n} = \mathbf{b}$$

which has the same solution set as the system of linear equations with augmented matrix

$$\begin{pmatrix}\mathbf{a_1} & \mathbf{a_2} & \dots & \mathbf{a_n} \mid \mathbf{b}\end{pmatrix} = \begin{pmatrix}A \mid \mathbf{b}\end{pmatrix}$$

To summarize, we have different interpretations/uses for matrices and the row-reduction process:

- Solving a linear system. Usually these will be augmented matrices, but not always (one exception would be homogeneous systems).
- Determining linear dependence/independence for a set of vectors.

Theorem 2. Suppose A is $m \times n$. The following are equivalent (all true or all false):

a) For each $\mathbf{b} \in \mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.
b) Each $b \in \mathbb{R}^m$ is a linear combination of the columns of A.
c) The columns of A span all of $\mathbb{R}^m$.
d) A has a pivot in every row.

Notice that this is not the same as saying the columns of A are independent! If that were the case, we would have a pivot in each column as well. In fact, if we want to achieve both column vector independence AND have the above statements be true, A would need to be a square matrix with $m = n$ (a necessary but NOT sufficient condition, we have seen square matrices that don't have all the pivots).

5. Linear Transformations And Matrices

Another key concept in linear algebra is linear transformations.

Definition 3. A **linear transformation** is a map (or function) $T : \mathbb{R}^n \to \mathbb{R}^m$ whose inputs are vectors from $\mathbb{R}^n$, and whose outputs are vectors from $\mathbb{R}^n$ (i.e. $\mathbf{u}_{n\times 1} \mapsto \mathbf{v}_{m\times 1}$), satisfying the following properties:

a) $T(\mathbf{0}) = \mathbf{0}$ (explicitly $\mathbf{0} \in \mathbb{R}^m$ maps to $\mathbf{0} \in \mathbb{R}^n$)
b) For scalars $a, b \in \mathbb{R}$, we have $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$.

In other words, T maps linear combinations from the domain space to linear combinations in the codomain space; it preserves linear combos.

Example 7. From calculus, given functions f, g and constants c, we have

$$\frac{d}{dx}(cf(x) + g(x)) = c\frac{d}{dx}f(x) + \frac{d}{dx}g(x)$$

and

$$\int [cf(x) + g(x)]dx = c\int f(x)dx + \int g(x)dx$$

In fact, both the derivative and integral operators are linear transformations acting on functions. Since in the definition of linear transformations, we require them to have vectors as inputs/outputs, indeed we will be thinking of functions f and g as vectors later in this course. We will also discuss and define the definition of linear independence of functions.

Example 8. A polynomial of degree 2 or less has the general form $a + bx + cx^2$. Notice that it is completely determined by its coefficients $a, b, c \in \mathbb{R}$. Indeed, we can identify every polynomial $P(x)$ with degree $\deg(P) \leq 2$ by its vector of coefficients $\langle a, b, c\rangle \in \mathbb{R}^3$, by simply mapping

$$\langle a, b, c\rangle \mapsto a + bx + cx^2$$

(the rule is: the first coordinate of the vector becomes the constant coefficient, the second coordinate is the coeff of x^1, and so on). In some sense, we will learn that the *space of polynomials with degree at most 2* is equivalent to $\mathbb{R}^3$. That's all I can say for now, but great things are coming!

The usefulness of linear algebra in studying these kind of transformations comes from the following fact:

Theorem 3. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A of dimension $m \times n$ such that

$$T(\mathbf{x}) = A\mathbf{x} \qquad \text{(so } T \text{ can be computed via matrix product)}$$

for all $\mathbf{x} \in \mathbb{R}^n$. Furthermore, this matrix A is entirely determined by how T acts on the standard basis vectors of $\mathbb{R}^n$:

$$\mathbf{e_1} = \begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}, \quad \mathbf{e_2} = \begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix}, \quad \mathbf{e_3} = \begin{pmatrix}0\\0\\1\\\vdots\\0\end{pmatrix}, \quad \dots \quad \mathbf{e_n} = \begin{pmatrix}0\\0\\0\\\vdots\\1\end{pmatrix}$$

To get the matrix, we simply take the images of the basis vectors under T, and place them as the column vectors of A:

$$A = \begin{pmatrix} \uparrow & \uparrow & \uparrow & & \uparrow \\ T(\mathbf{e_1}) & T(\mathbf{e_2}) & T(\mathbf{e_3}) & \dots & T(\mathbf{e_n}) \\ \downarrow & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

We call this the *standard matrix of* T.

6. Matrix Algebra

Definition 4. Suppose $A = (a_{i,j})_{m\times n}$ and $B = (b_{i,j})_{m\times n}$ are two matrices of the same dimensions. Then the **matrix sum** $A + B$ is simply entry-wise addition:

$$A + B = (a_{i,j}) + (b_{i,j}) = (a_{i,j} + b_{i,j})_{m\times n}$$

Theorem 4. Let A, B, C be matrices of the same dimension $m \times n$, and $r, s \in \mathbb{R}$. Then matrix addition has the following properties:

a) $A + B = B + A$ (commutativity)
b) $(A + B) + C = A + (B + C)$ (associativity)
c) $A + 0_{m\times n} = A$ (where $0_{m\times n}$ is the zero matrix)
d) $r(A + B) = rA + rB$
e) $(r + s)A = rA + sA$
f) $r(sA) = (rs)A$

Matrix addition satisfies the same properties as vector addition from multivariable calculus.

Definition 5. Suppose $A = (a_{i,j})_{m\times n}$ and $B = (b_{i,j})_{n\times p}$ are two matrices, where the number of columns of A matches the number of rows of B. Then the **matrix product** AB is a matrix C of dimensions $m \times p$, where the entry in row i and column j is given by

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} \cdot b_{k,j}$$

The easiest way to think about matrix product: the entry $c_{i,j}$ (in row i, col j) is simply the dot product of the i-th row of A with the j-th column of B.

Here is another slightly more cumbersome way is to think about this. The book mentions it and everyone's brain works differently, so maybe it will make more sense to you. If $\mathbf{a_{1,*}} \ldots \mathbf{a_{m,*}}$ are the m row vectors of matrix A, and $\mathbf{b_{*,1}} \ldots \mathbf{b_{*,p}}$ are the p column vectors of matrix B, then the matrix product becomes

$$AB = \begin{pmatrix} \mathbf{a_{1,*}} \\ \mathbf{a_{2,*}} \\ \vdots \\ \mathbf{a_{m,*}} \end{pmatrix}_{m\times n} \begin{pmatrix} \mathbf{b_{*,1}} & \mathbf{b_{*,2}} & \ldots & \mathbf{b_{*,p}} \end{pmatrix}_{n\times p} = \left(\begin{pmatrix} \mathbf{a_{1,*}} \\ \mathbf{a_{2,*}} \\ \vdots \\ \mathbf{a_{m,*}} \end{pmatrix} \mathbf{b_{*,1}} \quad \ldots \quad \begin{pmatrix} \mathbf{a_{1,*}} \\ \mathbf{a_{2,*}} \\ \vdots \\ \mathbf{a_{m,*}} \end{pmatrix} \mathbf{b_{*,p}} \right)$$

In a sense, we are distributing the matrix A into the "vector" B whose "entries" are, well, the column vectors of B. Written more compactly, this looks like:

$$AB = A \begin{pmatrix} \mathbf{b_{*,1}} & \mathbf{b_{*,2}} & \ldots & \mathbf{b_{*,p}} \end{pmatrix} = \begin{pmatrix} A\mathbf{b_{*,1}} & A\mathbf{b_{*,2}} & \ldots & A\mathbf{b_{*,p}} \end{pmatrix}$$

If you really sit down and think about what is going on, you will notice that *each column of AB is a linear combination of the columns of A, using weights from the corresponding column of B.*

If you didn't follow this, don't worry too much about it. Move on – you aren't missing anything. I've certainly never thought about it this way, and I prefer the easier method of using dot products between rows of A and columns of B.

Definition 6. The identity matrix I_n is the $n \times n$ (square) matrix with 1's along the main diagonal, and zeroes everywhere else:

$$I_n = \begin{pmatrix} 1 & & & O \\ & 1 & & \\ & & \ddots & \\ O & & & 1 \end{pmatrix}_{n\times n}$$

It serves as the multiplicative identity for matrices in the same way as the number 1 is the multiplicative identity for real numbers.

Theorem 5. Let A be an $m \times n$ matrix, and B, C be matrices with dimensions so that the indicated operations below are defined. Then matrix multiplication has the following properties:

a) $A(BC) = (AB)C$ (associativity)
b) $A(B + C) = AB + AC$
c) $(B + C)A = BA + CA$
d) $r(AB) = (rA)B$ for any scalar $r \in \mathbb{R}$
e) $I_m A = A = AI_n$

f) We usually do NOT have $AB = BA$ for matrices!
g) If $AB = 0$, it is not always true that $A = 0$ or $B = 0$!
h) When we write A^k for some positive integer k, we mean $\underbrace{A \cdot A \cdots A}_{k \text{ times}}$

Definition 7. Suppose $A = (a_{i,j})_{m\times n}$ is a matrix. The **transpose of** A is defined as another matrix $A^\intercal = (a_{j,i})_{n\times m}$. In other words, row 1 of A becomes column 1 of $A^\intercal$, row 2 of A becomes column 2 of $A^\intercal$, and so on.

We can also think about this the other way around: column 1 of A becomes row 1 of $A^\intercal$, column 2 of A becomes row 2 of $A^\intercal$, and so on. Pick whichever way is easiest for you to think about, they are equivalent.

Theorem 6. Let A and B be matrices with dimensions compatible for the operations indicated below. The transpose operation satisfies the following properties:

a) $(A^\intercal)^\intercal = A$
b) $(A + B)^\intercal = A^\intercal + B^\intercal$
c) $(rA)^\intercal = r(A^\intercal)$ for any $r \in \mathbb{R}$
d) $(AB)^\intercal = B^\intercal A^\intercal$ (unexpectedly!)
e) More generally, for matrices $A_1, A_2, \cdots A_r$, we have $(A_1 A_2 \dots A_r)^\intercal = {A_r}^\intercal {A_{r-1}}^\intercal \cdots {A_2}^\intercal {A_1}^\intercal$

Example 9. Going back to linear transformations, in order to see why replacing $T(\mathbf{x})$ by matrix multiplication $A\mathbf{x}$ would work, recall the dimension restrictions that are needed for general matrix multiplication to be compatible:

$$\begin{pmatrix} \boxed{\qquad\qquad} \\ \\ \\ \end{pmatrix}_{\mathbf{m}\times\underline{n}} \cdot \begin{pmatrix} \quad \\ \\ \\ \end{pmatrix}_{\underline{n}\times\mathbf{k}} = \begin{pmatrix} \qquad\qquad \\ \\ \end{pmatrix}_{\mathbf{m}\times\mathbf{k}}$$

When applying A to a column vector $\mathbf{x} \in \mathbb{R}^n$, this has dimension $n \times 1$, and the output will have dimension $m \times 1$ (a vector from $\mathbb{R}^m$), so at least dimensionally speaking this product makes sense. Now to see the equivalence, first write $\mathbf{x} = \langle x_1, x_2, \dots, x_n\rangle$ in terms of the standard basis as:

$$\mathbf{x} = x_1\mathbf{e_1} + x_2\mathbf{e_2} + \cdots + x_n\mathbf{e_n}$$

Then use the linearity of T to get

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e_1} + x_2\mathbf{e_2} + \cdots + x_n\mathbf{e_n}) \\ &= x_1T(\mathbf{e_1}) + x_2T(\mathbf{e_2}) + \cdots + x_nT(\mathbf{e_n}) \\ &= A\mathbf{x} \end{aligned}$$

where the last equality follows from the fact that $T(\mathbf{e_j})$ are the columns of A.

There are several types of problems you may solve by using this equivalence between A and T. Here are two of the simplest:

a) Given how T acts on some given vectors $\mathbf{u}, \mathbf{v}$, compute T for a linear combination of $\mathbf{u}$ and $\mathbf{v}$. See WKSH4, problem 1b.
b) Given a formula for T, find its matrix. To do this, apply the formula to the standard basis vectors, and use the results to construct the columns of A. See WKSH4, problem 2b.

c) (hard) Given T at nonbasis vectors, find the standard matrix A. This will come later when we learn about change of basis.

7. Inverse Of A Matrix

We saw that linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ map vectors $\mathbf{u} \in \mathbb{R}^n$ to other vectors $\mathbf{v} \in \mathbb{R}^m$. Sometimes we are interested in reversing the operation $T(\mathbf{u}) = \mathbf{v}$. Given $\mathbf{v} \in \mathbb{R}^m$, we would like to find its *preimage* $\mathbf{u} \in \mathbb{R}^n$ under the transformation T.

In general this is not possible. However, if the dimensions of the domain and codomain spaces are equal, we may reverse the operation. Knowing that $T(\mathbf{u})$ is equivalent to the matrix product $A\mathbf{u}$, reversing the transformation boils down to finding the **matrix inverse** A^{-1} of A:

$$A\mathbf{u} = \mathbf{v} \quad \iff \quad \mathbf{u} = A^{-1}\mathbf{v}$$

This is only possible if A is a square matrix, but not all square matrices are invertible. An invertible matrix is sometimes also called **nonsingular**, while a non-invertible matrix is called **singular**. If you want to go all SciFi, a singular matrix is one where the dimension of the image space decreases – a collapse in dimension aka a singularity. Cool, huh? This is why they are not invertible.

So when exactly are square matrices (and their corresponding linear transformations) invertible? The next theorem gives the full description:

Theorem 7. Suppose A is a square matrix of dimension n. The following are equivalent:

a) A is invertible
b) $A \xrightarrow{rref} I_n$ (so A has n pivots)
c) The columns of A are independent (so they span all of $\mathbb{R}^n$)
d) For every $\mathbf{b} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a **unique** solution $\mathbf{x} = A^{-1}\mathbf{b}$ (this means consistent system, and no free variables; it is not enough to have solutions, there must be a unique one)
e) The homogeneous $A\mathbf{x} = 0$ has only the trivial solution (again, no free variables)
f) The transformation $\mathbf{u} \mapsto A\mathbf{u}$ (whose standard matrix is A) is one-to-one (so you can think of invertible transformations as you did for one-to-one functions from calculus: they had an inverse)
g) The transformation $\mathbf{u} \mapsto A\mathbf{u}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$ (so no loss in dimension occurs). As a side note, this condition only works because the input and output spaces have the same dimension – it is NOT true in general that an onto transformation is invertible when the dimension of the spaces are different.
h) There exist matrices C and D such that $CA = I$ and $AD = I$ (in fact, C will equal D, and they equal A^{-1})
i) $\det A \neq 0$

The last condition has an explicit expression when A is 2×2, as we have a general formula for the 2×2 determinant.

Corollary 1. Suppose A is a 2×2 matris given by

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then A is invertible if and only if $ad - bc \neq 0$, and the inverse is given by

$$A^{-1} = \frac{a}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If $ad - bc = 0$, then A is not invertible.

Notice that the fraction has exactly $\det A$ in the denominator. Later on we will learn something called *Cramer's Rule*, which is a generalization of the above corollary, and it too will involve the fraction $1/\det A$, multiplied to some "special matrix" for computing inverses.

If you want to memorize this, that's fine. I've honestly never used it, and don't bother to remember it. Instead, I use row reduction to compute matrix inverses:

Theorem 8. Suppose A is an invertible matrix of dimension n.

$$\left(A \mid I_n\right) \xrightarrow{rref} \left(I_n \mid A^{-1}\right)$$

(OK that's a lie, I actually just use a calculator haha, but *do as I say, not as I do*).

Theorem 9. Suppose A and B are invertible matrices of the same dimension. Then the matrix inverse operation satisfies the following properties:

a) $(A^{-1})^{-1} = A$
b) $(AB)^{-1} = B^{-1}A^{-1}$ (same as the transpose) and more generally, for invertible matrices $A_1, \ldots A_n$, we have $(A_1 A_2 \cdots A_n)^{-1} = {A_n}^{-1} {A_{n-1}}^{-1} \cdots {A_2}^{-1} {A_1}^{-1}$
c) $(A^\intercal)^{-1} = (A^{-1})^\intercal$

8. Span, Subspaces, Basis, Column And Null Spaces

Consider the set of p distinct vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \ldots, \mathbf{v}_p\}$ (of the same dimension). We've talked about the **span** of S, which is the space generated by all possible linear combinations of the p vectors in our set:

$$span\{\mathbf{v}_1, \ldots \mathbf{v}_p\} = \{\mathbf{u} : \mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p \text{ for any } c_i \in \mathbb{R}\}$$

Example 10. From multivariable calculus, the space generated by a single vector $\mathbf{v}_1$ in $\mathbb{R}^2$ (or $\mathbb{R}^3$) is a line $L = \{c_1\mathbf{v}_1 : c \in \mathbb{R}\}$, which goes through the origin whose direction is given by v_1.

Example 11. If you have two nonparallel vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$, they will span (generate) a plane $P = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2\}$ which goes through the origin.

In both of these examples, how do we know that the objects spanned go through the origin? Simple: picking the constants c_i to be all equal to zero will result in the linear combination $0\mathbf{v}_1 + \cdots + 0\mathbf{v}_p = \mathbf{0}$, the zero vector.

We have a special name for these generated spaces. They are called **subspaces**. Any subspace $H \subseteq \mathbb{R}^n$ satisfies the following properties:

a) The zero vector $\mathbf{0} \in H$ (since $\mathbf{0}$ can always be achieved as a linear combination, by setting all coefficients to 0).
b) For each $\mathbf{u}, \mathbf{v} \in H$, their sum is also in H.
c) For any $\mathbf{u} \in H$ and $c \in \mathbb{R}$, the vector $c\mathbf{u}$ is in H.

Every subspace can be generated by some set of vectors. In other words, *every subspace can be thought of as a span of vectors.* The properties (b) and (c) can be combined by saying: once you generate the subspace H, you cannot get any new linear combinations from the existing ones, no matter how you combine vectors in H (addition of two vectors in H is just another lin. combo. of the generators, and so is multiplication by a scalar).

Example 12. The **column space** of a $n \times p$ matrix

$$A = \begin{pmatrix} \uparrow & \uparrow & \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_2 & \mathbf{v}_4 & \dots & \mathbf{v}_p \\ \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow \end{pmatrix}_{n \times p}$$

(which I will denote by $col(A)$) is the *subspace spanned/generated by the column vectors of A*. Since the matrix has n rows, each column vector is of dimension n, so $col(A)$ is a subspace of $\mathbb{R}^n$.

The maximum dimension will be $col(A)$ is the smallest of p, n, since the dimension of any subspace cannot be larger than the number of generators which span that subspace (so it can be at most p), nor can it be larger than the dimension of the vectors themselves (so it can be at most n).

How do we actually determine the dimension of $col(A)$? This is easy:

- reduce A down to EF
- # of pivots $= \dim colsp(A)$

Notice that to determine $\dim col(A)$, it is not necessary to reduce all the way to RREF.

Question: suppose we have less pivots than the p columns of A. That means we have redundancies, and $col(A)$ can be generated by less vectors. How do we know what we can throw out from among $\mathbf{v}_1, \dots \mathbf{v}_p$?

This is also easy: if in the echelon form of A, you have 4 pivots, and they are located (for example) in columns 1, 3, 4, and p of the EF, then the corresponding column vectors *of the original matrix* (namely $\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_p$) are the only vectors you need to generate all of $colsp(A)$! We have:

- $\dim col(A) = 4$ (since you had 4 pivots)
- $col(A) = \underbrace{span\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}}_{\text{with redundancies}} = \underbrace{span\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_p\}}_{\text{without redundancies}}$

The original set $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ was linearly dependent (less pivots than columns). However we were able to reduce this to a **linearly independent** set $B = \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_p\}$, and we can NOT reduce it any further.

The set $B = \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_p\}$ forms a **basis for the column space of** A. It is a **minimal set of generators** that spans the column space of A.

In general, a basis B for any subspace $H \subseteq \mathbb{R}^n$ is a minimal set of vectors that span H. It has no "redundancies." This is equivalent to the book definition that B is a linearly independent set that spans H.

The **null space** of a $n \times p$ matrix

$$A = \begin{pmatrix} \uparrow & \uparrow & \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_2 & \mathbf{v}_4 & \dots & \mathbf{v}_p \\ \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow \end{pmatrix}_{n \times p}$$

(which I will denote by $null(A)$) is the **solution space to the homogeneous system** $A\mathbf{x} = \mathbf{0}$. Unlike $col(A)$, this is NOT spanned by columns of A. In order to determine which vectors span $null(A)$, we do the following:

- reduce A all the way to RREF (remember, RREF is needed when trying to pinpoint ANY solution, whereas EF is used if we only care about pivots)
- use the free variable column vectors (with the modification I discussed in my shortcut from Section 1) to write the solution vector of the homogeneous system. The resulting modified vectors will span the null space, and they WILL be independent! (this is because of the placement of the special -1 when modifying the free columns – they end up in different coordinate positions and will prevent any dependence from happening)

Example 13. Consider the system $A\mathbf{x} = \mathbf{0}$ where the matrix A is given by

$$\begin{pmatrix} 1 & 2 & -5 & -6 & 0 \\ 0 & 1 & -6 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Our goal is to reduce this to RREF, and then write the parametric solution:

$$\begin{pmatrix} 1 & 2 & -5 & -6 & 0 \\ 0 & 1 & -6 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{rref} \begin{pmatrix} 1 & 0 & 7 & 0 & 0 \\ 0 & 1 & -6 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{align} \begin{pmatrix} \mathbf{1} & 0 & 7 & 0 & 0 \\ 0 & \mathbf{1} & -6 & -3 & 0 \\ 0 & 0 & \boxed{0} & 0 & 0 \\ 0 & 0 & 0 & \boxed{0} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

so we have 5 unknowns, 2 free variables, and the parametric solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 7 \\ -6 \\ \boxed{-1} \\ 0 \\ 0 \end{pmatrix} s + \begin{pmatrix} 0 \\ -3 \\ 0 \\ \boxed{-1} \\ 0 \end{pmatrix} t$$

This solution space to the homogeneous system IS the null space of A. It is the span of the basis set

$$B = \left\{ \begin{pmatrix} 7 \\ -6 \\ \boxed{-1} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \\ 0 \\ \boxed{-1} \\ 0 \end{pmatrix} \right\}$$

and it has dimension 2. Notice that independence is forced by the positions of the -1 entries. In general, we will ALWAYS have

$$\dim null(A) = \# \text{ of free variables}$$

Example 14. Referring back to Example 5, we looked at a matrix corresponding to a nonhomogeneous system $A\mathbf{x} = \mathbf{b}$ (where $\mathbf{b}$ was nonzero), namely

$$\left(\begin{array}{ccccc|c} 1 & 2 & -5 & -6 & 0 & -5 \\ 0 & 1 & -6 & -3 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

and we saw that the solution set was given by a span of the same two vectors that appeared in Example 13, plus a constant offset. This is not a coincidence! Here is the connection:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \underbrace{\underbrace{\begin{pmatrix} 7 \\ -6 \\ \boxed{-1} \\ 0 \\ 0 \end{pmatrix} s + \begin{pmatrix} 0 \\ -3 \\ 0 \\ \boxed{-1} \\ 0 \end{pmatrix} t}_{\text{free column span = null space}} + \underbrace{\begin{pmatrix} -9 \\ 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{\text{offset}}}_{\text{solution set of } A\mathbf{x} = \mathbf{b}\text{. This is NOT a subspace!}}$$

In general, when trying to solve $A\mathbf{x} = \mathbf{b}$, the parametric solution will be some offset of the homogeneous solution (null space), plus an offset. However, because of the offset, the solution set to $A\mathbf{x} = \mathbf{b}$ will not contain the zero vector, so it does not form a subspace (unless of course $\mathbf{b}$ is zero, but that's just solving the homogeneous system). So

$$(\text{solution to } A\mathbf{x} = \mathbf{b}) = (\text{homogeneous solution}) + \text{offset}$$

$$(\text{solution to } A\mathbf{x} = \mathbf{b}) = (\text{null space of } A) + \text{offset}$$

While it may not be a subspace, the solution to $A\mathbf{x} = \mathbf{b}$ is still an object sitting in $\mathbb{R}^n$, with the same dimension as the null space of A. (think back to multivariable calc – translating a plane that goes through the origin still gives a plane of the same dimension, but it no longer goes through the origin)

9. Dimension, Rank

We have talked about **dimension** in the case of column spaces and null spaces, but this concept applies to any subspace $H \subseteq \mathbb{R}^n$.

Definition 8. Given a subspace H generated by a basis set of vectors B, the dimension of H is just the number of basis vectors. In other words,

$$\dim H = |B|$$

A special case is when $B = \{\mathbf{0}\}$, so only a single point can be generated (the origin). This point is the zero subspace, and the convention is that it has dimension zero.

Notice how I never used "the" when referring to basis B. That is because **a basis is not unique**. There are infinitely many possible basis sets B, but they will ALWAYS have the same number of vectors, which is the dimension of the generated subspace.

Example 15. For $\mathbb{R}^2$, we used the *standard basis* $\mathbf{e}_1, \mathbf{e}_2$, but another perfectly valid basis for $\mathbb{R}^2$ is $B = \{\langle 1, 2\rangle, \langle 3, -5\rangle\}$. In fact, any two nonparallel vectors $\{\langle a, b\rangle, \langle c, d\rangle\}$ will form a basis for $\mathbb{R}^2$. It is just not the standard basis.

This is true in higher generality. In fact we have the following result:

Theorem 10. For any subspace H of $\mathbb{R}^n$ of dimension p, any collection $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of p linearly independent vectors from H will automatically form a basis for H.

Going back to a matrix A and its reduced form, we saw over and over again that

$$(\#\text{ columns of } A) = (\#\text{ of pivots}) + (\#\text{ of free variables})$$

The number of pivots has a special name. It is called the **rank** of A, and it is equal to the dimension of the column space of A. Using this terminology, we have the following theorem:

Theorem 11. (Rank Equation) For any matrix A with p columns, we have

$$n = rank(A) + \dim null(A)$$

This is nothing more than a rephrasing of what I wrote above. Lastly, there is a close connection between rank and invertible matrices:

Theorem 12. For any $n \times n$ matrix A, the following are equivalent to invertiblity of A:

a) The columns of A form a basis for $\mathbb{R}^n$
b) $col(A) = \mathbb{R}^n$
c) $\dim col(A) = rank(A) = n$ (columns of A are independent)
d) $null(A) = \{\mathbf{0}\}$ ($A\mathbf{x} = \mathbf{0}$ has only the trivial solution)
e) $\dim null(A) = 0$ (no free variables)

10. Determinants

Definition 9. For a $n \times n$ matrix $A = (a_{ij})_{1 \leq i,j \leq n}$, let $A_{i,j}$ denote the **submatrix** of A obtained by deleting row i and column j. The **determinant of** A is then given by

$$\begin{aligned} \det A &= \sum_{j=1}^{n} a_{k,j}(-1)^{k+j} \det A_{k,j} \quad \text{(expand horizontally along row } k\text{)} \\ &= \sum_{i=1}^{n} a_{i,k}(-1)^{i+k} \det A_{i,k} \quad \text{(expand vertically along column } k\text{)} \end{aligned}$$

Sometimes we write $C_{i,j} = (-1)^{i+j} \det A_{i,j}$ – this is called the (i,j)**-cofactor**.

Theorem 13. If A is a lower/upper triangular matrix, then $\det A$ is the product of the entries on the main diagonal.

Instead of using the above, one may use the row-reduction process to compute determinants (especially for larger matrices). The following theorem tells us how to do this:

Theorem 14. Suppose A is a square matrix, and consider the row operations indicated below:

(a) If $A \xrightarrow{R_i = R_i + kR_j} B$, then $\det B = \det A$.
(b) If $A \xrightarrow{R_i = kR_i} B$, then $\det B = k \det A$.
(c) If $A \xrightarrow{R_i \leftrightarrow R_j} B$, then $\det B = -\det A$.
(d) $A \xrightarrow{R_i = kR_i + R_j} B$, then $\det B = k \det A$. The determinant will change, since you stored the result back in R_i. You can think of the row operation $R_i = kR_i + R_j$ as a combination of (a) and (b) above, so the determinant will change. This is not considered an elementary row operation.

Suppose that matrix B is the *echelon* form of A, and in the process you performed exactly r row interchanges. Then a combination of theorems 13 and 14 tells us that

$$\det A = \begin{cases} (-1)^r \cdot (\text{product of pivots in } U) & \text{if } A \text{ is invertible} \\ 0 & \text{if } A \text{ is not invertible} \end{cases}$$

Theorem 15. Let A, B be square matrices of dimension n, and k a scalar. Then
(a) $\det(AB) = \det A \cdot \det B$
(b) $\det(kA) = k^n \det A$
(c) $\det(A^{-1}) = 1/\det A$
(d) $\det(A^\intercal) = \det A$
The last property tells us theorem 14 applies to column operations as well.

Theorem 16. Matrix A is invertible if and only if $\det A \neq 0$.

I mentioned in tutorial the **geometric interpretation** of the determinant. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation with matrix A. Suppose R is a region of dimension n inside $\mathbb{R}^n$, and $T(R)$ is the image of R under T (also inside $\mathbb{R}^n$). Applying T to R may distort the region, by stretching/compressing, therefore changing its area. The determinant $\det A$ can be thought of as the **magnification factor** of T:

$$\det A = \frac{\text{area/volume/"measure" of the image } T(R)}{\text{area/volume/"measure" of the original } R}$$

The word "measure" is a generalization of the concept of length (in $\mathbb{R}$), area (in $\mathbb{R}^2$), volume (in $\mathbb{R}^3$) to higher dimensions. We don't really have specific names for this in higher dimensions, so we just call it "measure".

With this geometric interpretation, we can easily see why we have $\det(AB) = \det A \cdot \det B$, and also why $\det(A^{-1}) = 1/\det A$. In the case where $\det A = 0$, so A is singular, you can think of the non-invertibility of T to mean that there is a "loss in dimension" – the image is smaller than the original.

Example 16. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the projection onto the xy-plane, so $T(x, y, z) = (x, y, 0)$. One can easily check that this is given by the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Clearly this is noninvertible, and $\det A = 0$. Notice that T maps the unit ball onto the unit circle in the xy-plane. The ball has volume $4\pi/3$, but the volume of the circle is zero. This is another way to see why $\det A = 0$.

Example 17. We've been using this since multivariable calculus, when doing change of variable (a transformation) and computing the Jacobian (a determinant). For 2D integrals, this looked like

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix}$$

It represents the magnification factor of the differential $dA = dxdy$ under T, which is why

$$dxdy = \frac{\partial(x, y)}{\partial(u, v)} dudv$$

John Hannah wrote a very accessible paper on the geometry of determinants. It is available at `https://www.jstor.org/stable/2974931`

Example 18. For a 2×2 matrix A, $|\det A|$ represents the area of the parallelogram spanned by the columns of A. For a 3×3 matrix A, then $|\det A|$ represents the volume of the parallelipiped spanned by the columns of A.

11. Cramer's Rule

This has two applications: solving a system of linear equations, or finding the inverse of a matrix.

Theorem 17. (Cramer for solving systems) For an invertible $n \times n$ square matrix A and a column vector $\mathbf{b}$ of dimension n, let $A_j(\mathbf{b})$ be the matrix obtained by replacing column j of A by b. Then the solution $\mathbf{x} = (x_1, \ldots x_n)$ to the matrix equation $A\mathbf{x} = \mathbf{b}$ is given by

$$x_j = \frac{\det A_j(\mathbf{b})}{\det A}$$

Definition 10. For a matrix A, the **adjoint/adjugate of** A is given by

$$adj(A) = \begin{pmatrix} C_{11} & C_{12} & \ldots & C_{1n} \\ C_{21} & C_{22} & \ldots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \ldots & C_{nn} \end{pmatrix}^T$$

where $C_{i,j}$ is the (i,j)-cofactor. (the transpose is important!)

Theorem 18. (Cramer for finding inverses) For an invertible matrix A, let $adj(A)$ be its adjoint. Then A^{-1} is given by

$$A^{-1} = \frac{1}{\det A} adj(A)$$

12. Vector Spaces

In multivariable calculus, we spend a great deal of time working with vectors. They had some nice algebraic properties when working with vector addition and scalar multiplication. In linear algebra we are developing even more tools for working with vectors.

The idea of the vector space is meant to extend these algebraic properties to other more general objects. If we can show that some collection of objects (eg. functions) enjoys the same algebraic properties of vectors, this means we can use the framework of linear algebra and matrices to study them! The properties we'd like to have are given below:

Definition 11. A set V of objects (called vectors) is called a **vector space** if, for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $c, d \in \mathbb{R}$, we have the following axioms:

(1) $\mathbf{u} + \mathbf{v} \in V$ ("closure" under addition)
(2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
(3) $(\mathbf{u} + \mathbf{v}) + \mathbf{w}) = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
(4) we have a zero vector $\mathbf{0}$ satisfying $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for all $\mathbf{u}$
(5) every $\mathbf{u}$ has an additive inverse $-\mathbf{u}$ satisfying $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
(6) $c\mathbf{u} \in V$ ("closure" under scalar multiplication)
(7) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(8) $(c+d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
(9) $c(d\mathbf{u}) = (cd)\mathbf{u}$
(10) $1\mathbf{u} = \mathbf{u}$ (scalar multiplication by our unit $1 \in \mathbf{R}$ doesn't change a vector)

Notice how these are exactly the same properties we are used to seeing from multivariable calculus and vector algebra.

The idea of "closure" under some operation simply means that if I take two objects from my set V, and perform the binary operation on them, the result also lies the set V. By binary operation I mean an operation taking two inputs (like addition or multiplication). For example, the integers $\mathbb{Z}$ are closed under addition and multiplication (the result is always an integer). However $\mathbb{Z}$ is not closed under taking inverses.

Property (10) may seem strange and obvious, but in higher math we don't restrict ourselves to working with $\mathbb{R}$. We can easily work with vector spaces over complex numbers instead, or any other *fields* ($\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are all examples of something called *fields*, but we won't worry about that in this class).

Example 19. The following are all vector spaces:
(a) $\mathbb{R}^n$
(b) any span of a set of vectors $\{\mathbf{v}_1, \ldots \mathbf{v}_n\}$. If these vectors are independent, they also form a basis for the spanned vector space.
(c) $\mathbb{P}_n$, the set of polynomials of degree at most n
(d) $C[a, b]$, the set of real-valued continuous function $f : [a, b] \to \mathbb{R}$
(e) $C^\infty(\mathbb{R})$, the set of all infinitely differentiable functions $f : \mathbb{R} \to \mathbb{R}$
(f) Formal power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ with real coefficients. Here "formal" means we are not concerned with the question of convergence. Instead we just look at it as a polynomial with infinitely many terms. While these are "infinitely differentiable" in some sense, this is not the same as (e) – think about the geometric series, its corresponding function, and the domain of convergence.

13. Subspaces

We've worked with vector spaces enough to see that sometimes we have smaller "spaces" inside a bigger space. For instance, $\mathbb{R}^2$ sits inside $\mathbb{R}^3$. We can work with vectors of the form $(x, y, 0) \in \mathbb{R}^3$ in the same way that we would work with vectors $(x, y) \in \mathbb{R}^2$ – no matter how I add or scalar-multiply objects of the form $(x, y, 0)$, I can never leave the xy-plane. There is some "closure" at play here. Objects of the form $(x, y, 0)$ are a vector space, and at the same time a subspace of (x, y, z).

Definition 12. A subset H of a vector space V is called a **subspace** if
(a) $\mathbf{0} \in H$
(b) H is closed under addition
(c) H is closed under scalar multiplication
I want to emphasize that this definition ensures *every subspace H of V is also vector space*, just "smaller" (in the dimensional sense). Another way to think about it is that **subspaces are closed/stable under linear combinations**.

It turns out that *every span is a subspace*, and likewise *every subspace (of finite dimension) can be represented as a span of finitely many vectors*:

Theorem 19. Let V be a vector space, and $\mathbf{v}_1, \ldots \mathbf{v}_r \in V$. Then $sp\{\mathbf{v}_1, \ldots \mathbf{v}_r\}$ is a subspace of V.

Example 20. The set $\{\mathbf{0}\}$ consisting of only the zero vector of a vector space V is always a subspace. This is the **zero subspace**, and has dimension zero.

Example 21. Technically $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$, since objects $(x, y) \in \mathbb{R}^2$ have only 2 coordinates, and therefore lack the correct form $(x, y, z) \in \mathbb{R}^3$.

Example 22. $H = \{(x, y, 0) : x, y \in R\}$ and $G = \{(x, 0, z) : x, z \in \mathbb{R}\}$ are both examples of 2-dimensional subspaces of $\mathbb{R}^3$. They "act" like $\mathbb{R}^2$.

Example 23. Any plane in $\mathbb{R}^3$ that passes through the origin is a 2-dimensional subspace of $\mathbb{R}^3$. Similarly, any line that passes through the origin is a 1-dimensional subspace of $\mathbb{R}^3$.

Example 24. A plane or a line NOT passing through the origin are not subspaces (they lack the zero vector).

14. Null, Column, Row Space

Let A be a $m \times n$ matrix, with its corresponding linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. When we try to solve $A\mathbf{x} = \mathbf{0}$ (the homogeneous solution), we are basically asking the question *which vectors in* $\mathbb{R}^n$ *get sent to* $\mathbf{0} \in \mathbb{R}^m$ *via the transformation* T*?* We already saw that the homogeneous solution is expressible as a span, and therefore is a subspace:

Definition 13. The **null space** of A, denoted $Nul(A)$, is the solution space of $A\mathbf{x} = \mathbf{0}$. This is a subspace of $\mathbb{R}^n$.

Definition 14. The **column space** of A, denoted $Col(A)$, is the set of all linear combinations of the columns of A, and a subspace of $\mathbb{R}^m$.

Definition 15. The **row space** of A, denoted $Row(A)$, is the set of all linear combinations of the rows of A, and a subspace of $\mathbb{R}^n$. The row space of A is the column space of $A^\intercal$.

I mentioned earlier that subspaces are stable under linear combinations. Since row-reduction operations are simply performing linear combinations with the rows (addition and scalar multiplication), then for any two matrices A and B that are row equivalent, $Row(A) = Row(B)$. If B is in EF, then the nonzero rows form a basis for both $Row(A)$ and $Row(B)$.

Notice the following difference: $Nul(A)$ is a subspace in the domain $\mathbb{R}^n$ of T, while $Col(A)$ is a subspace of the codomain $\mathbb{R}^m$ of T. $Col(A)$ is also the *image of* T. The dimension of the entire image object $T(\mathbb{R}^n)$ is the dimension of the column space, and equal to the number of pivots of A. If the nullity $\dim Null(A)$ is not zero, then there is some dimension loss in T. This can be calculated using the rank-nullity theorem:

$$(\#\text{ of columns}) = (\text{dim col space}) + (\text{dim null space})$$

We can talk about these concepts more generally. Instead of working with $\mathbb{R}^n$ and $\mathbb{R}^m$, consider any two vector spaces V and W.

Definition 16. A **linear transformation** $T : V \to W$ is a map/function that preserves linear combinations:

(a) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$
(b) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and $\mathbf{u} \in V$

Definition 17. The **kernel** (or **null space**) of a linear transformation $T : V \to W$ is the set of all $\mathbf{u} \in V$ with $T(\mathbf{u}) = 0$.

Example 25. Consider $\mathbb{P}_3 = \{a + bx + cx^2 + dx^3 : a, b, c \in \mathbb{R}\}$, with standard basis $\{1, x, x^2, x^3\}$. This is a 4-dimensional vector space. To every such polynomial we associate vector $\langle a, b, c, d\rangle$. The derivative operator d/dx is a linear transformation (we know this from calculus), which means it is representable as a matrix. To get this matrix, we first see how it acts on the standard basis:

$$\frac{d}{dx}(1) = 0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \frac{d}{dx}(x) = 1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \frac{d}{dx}(x^2) = 2x = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 0 \end{pmatrix} \qquad \frac{d}{dx}(x^3) = 3x^2 = \begin{pmatrix} 0 \\ 0 \\ 3 \\ 0 \end{pmatrix}$$

Then the derivative matrix is

$$D = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

You can check that multiplication by this matrix gives the correct derivative of any polynomial of degree at most 3. Notice that the kernel of D is the set of all constant functions (a subspace of dimension 1). The image can have at most degree 2, and so it forms a subspace of dimension 3. The dimension drops from 4 to 3 since the kernel has dimension 1.

Another way to think about it is this: by taking a derivative, we lose all information about constants. This is why the D matrix is not invertible – we know from calculus that computing the indefinite integral $\int f(x)dx$ gives us $F(x) + C$, a family of functions all of which map to $f(x)$ under the transformation D.

This is true in general for any linear transformation $T : V \to W$. For any vector $\mathbf{w} \in W$, the preimage will have some kernel component involved. If you remember how we solved $A\mathbf{x} = \mathbf{w}$ for a particular $\mathbf{w}$, the solution looked like

$$\mathbf{x} = (\text{something concrete/constant}) + (\text{an element from the kernel})$$

The kernel component was the solution to the homogeneous system $A\mathbf{x} = \mathbf{0}$, involving the free variables.

15. Linear Independence, Bases, Coordinates

Linear independence is easy to tell in $\mathbb{R}^n$, but that method doesn't work for general vector spaces. Generally vectors are not tuples, like the formal power series I mentioned in example 19, or even continuous functions. For general vector spaces, we use the following:

Theorem 20. A set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ of two or more vectors is linearly independent if no vector in the set (with the given ordering) can be expressed as a linear combination of the preceeding vectors.

Definition 18. Let H be a subspace of V. Then $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_r\}$ is a **basis** for H if

(a) $\mathcal{B}$ is linearly independent and
(b) $span(\mathcal{B}) = H$

Then the **dimension** of H is just $\dim H = |\mathcal{B}|$, as before.

Since I mentioned that every subspace H of V (including V itself) can be written as a span of vectors from V, it is always possible to find a basis. If you have a nonindependent set S that spans H, you can check Theorem 20 and remove elements which fail the test (i.e. remove elements that are expressible as linear combinations of preceeding vectors). What you are left with forms a basis.

Example 26. Given a matrix A, a basis for $Col(A)$ is obtained by taking the pivoting columns of A. A basis for $Row(A)$ is obtained by taking the pivoting rows from the EF/RREF form of A. A basis for $Nul(A)$ is obtained by solving the homogeneous system – the vectors involved in the parametric form of the solution form a basis.

Basis and coordinates are very closely related. Every time we wrote a vector $\mathbf{x} \in \mathbb{R}^3$ as $\langle x_1, x_2, x_3 \rangle$ back in multivariable calculus, what we were really saying is that

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$$

so the coordinates in $\langle x_1, x_2, x_3 \rangle$ are the coefficients of the above linear combination. This same idea applies to any other basis.

Definition 19. Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis for a vector space V (of dimension n). Any vector $\mathbf{x} \in V$ can be expressed as some linear combination of the basis vectors in $\mathcal{B}$, say

$$\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n$$

Then the **coordinate vector of x relative to the basis** $\mathcal{B}$, denoted by $[\mathbf{x}]_{\mathcal{B}}$, is given by the coefficients of the linear combination above:

$$[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

When $\mathcal{B}$ is not specified, we usually assume we are dealing with the standard basis.

Given a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, let B be the matrix whose columns are the vectors of B. For a vector $\mathbf{x}$ in terms of the standard basis, we have the following:

$$B[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$$

which is equivalent to expressing $\mathbf{x}$ as a linear combination of the $\mathbf{b}_i$'s. I showed how this works in tutorial. The matrix B is called a **change of basis** or **change of coordinate matrix**. Here are a few types of problems you could try to solve:

Example 27. Suppose you are given $\mathcal{B}$, and $[\mathbf{x}]_{\mathcal{B}}$. To express $\mathbf{x}$ in terms of the standard basis, simply do the matrix product $B[\mathbf{x}]_{\mathcal{B}}$. An example of this is on Quiz 5, Problem 2.

Example 28. You may be given $\mathcal{B}$ and $\mathbf{x}$ in standard coordinates. To find $[\mathbf{x}]$, augment the matrix B with column vector x, and solve. In other words, you are solving for the unknown $[\mathbf{x}]_{\mathcal{B}}$ in $B[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$. An example of this is on Quiz 5, Problem 1.

Example 29. If you have to find $[\mathbf{x}]_{\mathcal{B}}$ for a number of vectors $\mathbf{x}$, you may instead invert the matrix B (always possible since $\mathcal{B}$ is an independent set). Then the equation $B[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$ becomes

$$[\mathbf{x}]_{\mathcal{B}} = B^{-1}\mathbf{x}$$

Notice how coordinate mapping from a vector $\mathbf{x} \in V$ to its coodinate vector $[\mathbf{x}]_{\mathcal{B}}$ (even with respect to the standard basis) is basically a linear transformation transformation $V \to \mathbb{R}^n$. An instance of this was when we mapped polynomials in $\mathbb{P}_n$ to vectors in $\mathbb{R}^{n+1}$. This is a one-to-one transformation, meaning that **such a representation is unique**.

In Example 25, we see that $\mathbb{P}_3$ and $\mathbb{R}^4$ are both vector spaces of dimension 4, and adding polynomials of degree ≤ 3 is very "similar" in some sense to adding vectors in $\mathbb{R}^4$. While not at all obvious at first, the coordinate mapping between polynomials and vectors allowed us to see the structural similarities between these vector spaces. In math, we say that they are **isomorphic**, and the one-to-one coordinate map between them is a **isomorphism**. More generally, coordinate mapping is an isomorphism from P_n to $\mathbb{R}^{n+1}$.

As another example, $\mathbb{R}^2$ is isomorphic to any 2-dimensional subspace of $\mathbb{R}^3$.

16. Dimension For General Vector Spaces

The idea of dimension is very simple to define for "finitely generated vector spaces" (i.e. with a finite set as their basis). The dimension of such a space is just the size of the basis set.

Definition 20. Let $\mathcal{B}$ be a basis for V. If $\mathcal{B}$ contains finitely many elements, then we say that V is **finite-dimensional** and $\dim V = |\mathcal{B}|$. In the case where V is not spanned by a finite set, then V is **infinite-dimensional**.

Example 30. We know that $\dim \mathbb{R}^n = n$, and we are very familiar with the standard basis for this space. However, if you consider the space of formal power series

$$\left\{ \sum_{i \geq 0} a_i x^i : a_i \in \mathbb{R} \right\}$$

it is impossible to come up with a finite basis for this set. It would have to look like $\mathcal{B} = \{1, x, x^2, x^3, \ldots\}$ and it contains infinitely many elements. This is an example of an infinite-dimensional vector space.

When referring to subspaces H of V, I talk about H "sitting inside" V. I also described H as being "smaller" than V. What I really mean is the following:

Theorem 21. If H is a subspace of a finite-dimensional vector space V, then

$$\dim H \leq \dim V$$

In particular, H is also finite-dimensional. Equality in the above implies that $H = V$ (it's perfectly fine to say that a vector space is a subspace of itself).

Example 31. Theorem 21 fails for infinite-dimensional vector spaces V. It's entirely possible to have subspaces H of infinite dimension, but $H \neq V$. Consider the following

$$H = \left\{ \sum_{i \geq 0} b_i x^{2i} : b_i \in \mathbb{R} \right\} \qquad V = \left\{ \sum_{i \geq 0} a_i x^i : a_i \in \mathbb{R} \right\}$$

where $\dim V = \dim H = \infty$, but $H \neq V$.

17. Change Of Basis

Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots \mathbf{c}_n\}$ are two bases for a vector space V, and let B, C be the matrices whose columns are the vectors in $\mathcal{B}$, $\mathcal{C}$, respectively.

For any $\mathbf{x} \in V$, let's say coordinate vectors are $[\mathbf{x}]_{\mathcal{B}}$ and $[\mathbf{x}]_{\mathcal{C}}$. We already know how to relate this to the standard basis, via the relations

$$B[\mathbf{x}]_{\mathcal{B}} = \mathbf{x} \qquad C[\mathbf{x}]_{\mathcal{C}} = \mathbf{x}$$

One thing we may want to do is to change coordinates/basis from $\mathcal{B}$ to $\mathcal{C}$. This is possible, and it can be done in the following steps:

(1) convert from $\mathcal{B}$ to standard (easy, do the multiplication $B[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$)
(2) convert standard to $\mathcal{C}$

We already know how to do both of these steps, see Section 15 on page 21. Step 2 takes a bit longer.

Theorem 22. There is a unique matrix $P_{\mathcal{B}\to\mathcal{C}}$ called the **change of coordiantes matrix from $\mathcal{B}$ to $\mathcal{C}$**, so that

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{B}\to\mathcal{C}}[\mathbf{x}]_{\mathcal{B}}$$

The columns of $P_{\mathcal{B}\to\mathcal{C}}$ are the $\mathcal{C}$-coordinates of the $\mathbf{b}_i$'s:

$$P_{\mathcal{B}\to\mathcal{C}} = \begin{pmatrix}[\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \dots & [\mathbf{b}_n]_{\mathcal{C}}\end{pmatrix}$$

Since we have to compute coordinates relative to $\mathcal{C}$ for multiple vectors, we won't use row-reduction to find these. It is more efficient to find C^{-1} and apply it to the $\mathbf{b}_i$'s.

While this is what the book defines as the change of basis, and it's nice theoretically, it's not very efficient. In the steps (1) and (2) I describe above, the first step corresponds to multiplying $B[\mathbf{x}]_{\mathcal{B}}$ to get $\mathbf{x}$ (standard). The second step corresponds to multiplying $C^{-1}x$ to get $[\mathbf{x}]_{\mathcal{C}}$. The net result is that we multiplied $[\mathbf{x}]_{\mathcal{B}}$ by $C^{-1}B$ (order is right to left). This means that

$$P_{\mathcal{B}\to\mathcal{C}} = C^{-1}B$$

which is an improvement.

There's an even faster way, which the book goes over. Instead of inverting C, then multiplying the result to B, you can do it all in one go by performing RREF on the augmented matrix $(C|B)$:

$$\begin{pmatrix}\mathbf{c}_1 & \dots & \mathbf{c}_n \mid \mathbf{b}_1 & \dots & \mathbf{b}_n\end{pmatrix} \xrightarrow{rref} \begin{pmatrix}I \mid P_{\mathcal{B}\to C}\end{pmatrix}$$

Doing this in one step has the effect of taking the inverse of C, which would normally involve performing RREF on $(C|I) \to (I|C^{-1})$, and then doing the product $C^{-1}B$.

Suppose you wanted to go the other way, i.e. you wanted to find $P_{\mathcal{C}\to\mathcal{B}}$. You simply augment in the reverse order:

$$\begin{pmatrix}\mathbf{b}_1 & \dots & \mathbf{b}_n \mid \mathbf{c}_1 & \dots & \mathbf{c}_n\end{pmatrix} \xrightarrow{rref} \begin{pmatrix}I \mid P_{\mathcal{C}\to B}\end{pmatrix}$$

The connection between changing basis back and forth is as you would expect:

$$(P_{\mathcal{B}\to C})^{-1} = P_{\mathcal{C}\to B}$$

18. Eigenvalues And Eigenvectors: Basic Definitions

A linear transformation $T : V \to V$ maps vectors in V to other vectors in V. If we think about what this does to all of V, T *morphs/reshapes* the original space V into itself (if T is invertible), or a subspace of itself (if T is not invertible). Examples include rotations, reflections, dilations, contractions, etc. (Note: translations are not linear transformations as they do not send the zero vector back to zero)

Even though T reshapes the space V (and usually involves a change in direction when applied to a vector), it is possible that T acts on certain special vectors by only rescaling the vector, without modifying its direction.

Definition 21. An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. In other words, multiplication by A has the same effect on $\mathbf{x}$ as scalar multiplication.

To determine if a given vector $\mathbf{x}$ is an eigenvector of A, simply multiply $A\mathbf{x}$ and check that the result is a scalar multiple of $\mathbf{x}$.

Definition 22. The scalar λ is an **eigenvalue** of A if there is some nontrivial solution $\mathbf{x}$ to the equation $A\mathbf{x} = \lambda\mathbf{x}$ (we call this solution $\mathbf{x}$ the **eigenvector corresponding to** λ).

To determine if a specified scalar λ is an eigenvalue of A, notice that the equation $A\mathbf{x} = \lambda\mathbf{x}$ in the definition can be rewritten as

$$A\mathbf{x} = \lambda I\mathbf{x}$$

where I is the identity matrix. Move everything to one side, and "factor out" $\mathbf{x}$ to get

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

If this homogeneous system has a nontrivial solution for a given λ, then that λ is an eigenvalue of the matrix A.

The question of finding eigenvectors for a specified λ uses the same process. Any nontrivial solution to this homogeneous system will be an eigenvector corresponding to the given λ. Of course, there are infinitely many such nontrivial solutions. The entire solution set is the null space of the matrix

$$A - \lambda I$$

and is called the **eigenspace** of A corresponding to λ.

What's interesting is, if $\mathbf{v}_1, \dots, \mathbf{v}_r$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_r$ (so they come from different eigenspaces), then they are linearly independent.

Notice this allows us to find eigenvectors for a specific λ, but how do we know which λ to use in the first place? The following theorem describes a special case.

Theorem 23. The eigenvalues of an upper/lower triangular matrix are the entries on the main diagonal.

I will describe the general method in the next section.

19. Eigenvalues: The Characteristic Equation and Diagonalization

Definition 23. Given a square matrix A of dimension n, the **characteristic polynomial** of A is the determinant of the matrix $A - \lambda I$:

$$P(\lambda) = \det(A - \lambda I)$$

Notice this is a polynomial in the variable λ. The next theorem allows us to find eigenvalues for the matrix A:

Theorem 24. The scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

In other words, the eigenvalues of A are the *roots of the characteristic polynomial* $P(\lambda)$ *of* A. If A is $n \times n$, then $P(\lambda)$ has degree n, and the fundamental theorem of algebra tells us that we have exactly n roots (some of them can be repeated, or even complex, but we will have at most n real roots). If the polynomial has repeated roots, then the **multiplicity** of the root is the multiplicitiy of that eigenvalue. The equation $P(\lambda) = 0$ is also called the **characteristic equation** of A.

Definition 24. If A and B are $n \times n$ matrices, we say that A **is similar to** B if we can write $B = P^{-1}AP$ for some invertible matrix P. This process if multiplying by P^{-1} and P is called a **similarity transformation** or **conjugation** by P.

If A is similar to B, it is also true that B is similar to A, so the order doesn't really matter. We just say A and B are similar. The next thereom tells us why we care about similarity:

Theorem 25. If A and B are similar, then $P_A(\lambda) = P_B(\lambda)$. In other words, similar matrices will have the same characteristic polynomial, and the same eigenvalues (with the same multiplicities). This also means A and B have the same determinant and trace. In general, if A has eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct), then:

$$\det(A) = \prod_{i=1}^{n} \lambda_i \qquad \text{and} \qquad tr(A) = \sum_{i=1}^{n} \lambda_i$$

where $tr(A)$ is the **trace of** A (sum of the entries on the main diagonal).

Note: while conjugation doesn't change eigenvalues, row operations does! Just being *row equivalent* is not the same as being *similar*.

The advantage is that, if we are able to somehow say A is similar to a diagonal matrix D, then we automatically know the eigenvalues of A (they are the diagonal entries of D). In this case, we say that A is **diagonalizable**. Sadly, not all matrices are diagonalizable:

Theorem 26. A matrix A of dimension n is diagonalizable if and only if A has n linearly independent eigenvectors.

For a diagonalizable A, we should be able to write it as $A = PDP^{-1}$ for some P. This is possible if and only if the columns of P are n linearly independent eigenvectors of A (these form an **eigenvector basis** for $\mathbb{R}^n$). Notice how this gives us a way to construct P! The diagonal entries of D will be the eigenvalues corresponding to the columns of P (with the respective order that they appear in).

We see that diagonalizing an $n \times n$ matrix A involves the following steps:

(1) Find the eigenvalues of A with their corresponding multiplicities
(2) Find n independent eigenvectors. You can just take the generators/basis for the solution of the homogeneous system here, as they will be independent and nontrivial eigenvectors.
(3) Construct P, the invertible matrix whose columns form the eigenvector basis.
(4) Construct D from the corresponding eigenvalues.

Repeated Eigenvalues

Since the eigenvectors belonging to distint eigenvalues are always independent, we have the following result:

Theorem 27. If A has n distinct eigenvalues, then A is diagonalizable.

The converse is NOT true. Notice that in the above steps, there is nothing preventing us from finding n independent eigenvectors even if we have repeated eigenvalues with multiplicity > 1. It is entirely possible for a diagonalizable matrix to have repeated eigenvalues. However, the multiplicities of the eigenvalues does give us some information about the eigenspaces of A. The following theorem fully describes this:

Theorem 28. Suppose that A is a matrix of dimension n, with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$.

(1) The dimension of the eigenspace for λ_i is at most the multiplicity of λ_i.
(2) A is diagonalizable if and only if the sum of the dimensions of all eigenspaces is n. In order for this to occur, $P_A(\lambda)$ must factor completely into linear factors (assuming we want real eigenvalues), AND the dimension of each eigenspace is exactly equal to the multiplicity of its eigenvalue.

(3) If A is diagonalizable, and $\mathcal{E}_k$ forms a basis for the eigenspace associated to λ_k, then

$$\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \cdots \cup \mathcal{E}_r$$

forms an eigenvector basis for $\mathbb{R}^n$.

Putting Theorems 27 and 28 together, if all eigenvalues are distinct, then each eigenspace has dimension 1. We are guaranteed that, for any eigenvalue regardless of multiplicity, the eigenspace has dimension at least one (we have at least one free variable for $A - \lambda_i I$), and at most that multiplicity. If all eigenvalues are distinct, they all have multiplicity 1, and we get one eigenvector from each.

Remark: Looking at the formula $A = PDP^{-1}$, and knowing that the columns of P form the eigenbasis $\mathcal{E}$, we can break down the meaning of PDP^{-1} as a series of 3 operations. The matrix P changes basis from $\mathcal{E}$ to the standard basis S, and thus P^{-1} changes basis from S to $\mathcal{E}$. Instead of applying the transformation A to a vector $\mathbf{v}$ in standard coordinates, we are doing $PDP^{-1}\mathbf{v}$, which means we are first changing basis from S to $\mathcal{E}$ (via P^{-1}), then applying D (the matrix representation of our transformation in the eigenbasis), and switching coordinates back from eigen to standard (via P).

While this seems like more work, if one needs to apply the transformation a large number of times, computationally this is more desirable as applying D in $\mathcal{E}$ is very fast – multiplication by a diagonal matrix to $\mathbf{v}$ is as fast as scalar multiplication. It requires n multiplication operations, whereas multiplication by A needs as much as n^2 multiplications, and $n(n-1)$ additions to complete the dot products.

20. Linear Transformations In Various Bases

Suppose we are looking at $T : V \to W$, where V and W are abstract vector spaces of dimensions n and m, respectively. In order to talk about the matrix of T, we need a basis $\mathcal{B}$ for V and $\mathcal{C}$ for W. Then the action $\mathbf{x} \mapsto T(\mathbf{x})$ can be studied in the familiar spaces $\mathbb{R}^n \to \mathbb{R}^m$:

$$[\mathbf{x}]_{\mathcal{B}} \mapsto [T(\mathbf{x})]_{\mathcal{C}}$$

The columns of the **matrix for T relative to $\mathcal{B}, \mathcal{C}$** is given (as usual) by how it acts on the basis $\mathcal{B}$:

$$A = \begin{pmatrix} \uparrow & & \uparrow \\ [T(\mathbf{b}_1)]_{\mathcal{C}} & \dots & [T(\mathbf{b}_n)]_{\mathcal{C}} \\ \downarrow & & \downarrow \end{pmatrix}$$

This is what the book talks about in section 5.4. There's really nothing special about it, and they use it to describe $A = PDP^{-1}$, which I mention in the remark above.

This is useful in higher generality, not just when D is the diagonal eigenvalue matrix. The book has a useful shortcut in these cases. Suppose you are given any matrix A (relative to standard) and a new basis $\mathcal{B}$. To compute D, the representation of this transformation in $\mathcal{B}$, do the following:

(1) Construct P, whose columns are the basis vectors from $\mathcal{B}$.
(2) Compute AP.
(3) Row-reduce the following augmented matrix to get $D = P^{-1}AP$:

$$\big(P \mid AP\big) \xrightarrow{rref} \big(I \mid P^{-1}AP\big)$$

21. Complex Eigenvalues

When we found eigenvalues for an $n \times n$ matrix A, we ended up finding the roots of a polynomial of degree n. Usually those roots were real, but in general they don't need to be. In fact, we have:

Theorem 29. A polynomial $P(x)$ of degree n with complex coefficients will have exactly n complex roots (not necessarily distinct). Furthermore, if $\alpha = a + bi$ is a complex root, then its complex conjugate $\overline{\alpha} = a - bi$ is also a root (this is always the case).

Note: A consequence is that, for a matrix A with real entries, if $\lambda, \mathbf{v}$ is a complex eigenvalue-eigenvector pair, then $\overline{\lambda}, \overline{\mathbf{v}}$ is another "eigenpair".

Since $\mathbb{R} \subset \mathbb{C}$, all the polynomials we've been studying have exactly n roots, but some of them may be complex. Finding them is generally difficult, so for our class we restrict ourselves to $n = 2$. The quadratic formula is the most effective way of finding these roots (factoring in the complex numbers is difficult). I showed in tutorial how to find an eigenvector once you have a complex eigenvalue – in general, DO NOT try to use row-reduction (this would involve dividing by complex numbers at some point, which is hard).

For matrices of the special form, we have the following result:

Theorem 30. Let A be a 2×2 matrix with real entries, and suppose A has a complex eigenvalue $\lambda = a - bi$ with corresponding complex eigenvector $\mathbf{v}$. Then

$$A = PCP^{-1}$$

where

$$P = \begin{pmatrix} Re(\mathbf{v}) & Im(\mathbf{v}) \end{pmatrix} \qquad \text{and} \qquad C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

22. Inner Product, Length, Orthogonality

When dealing with $\mathbb{R}^3$, we had a good geometric grasp of length and orthogonality. We used the dot product to determine these things, but the dot product is defined for higher dimensions as well. It can be used to generalize these concepts to higher dimensions, and even abstract vector spaces, where we don't have a geometric intuition of what a *vector* looks like (e.g. polynomials).

Definition 25. The **length/norm** of a vector $\mathbf{v}$ is defined as

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2}$$

and the **Euclidean distance** between two vectors $\mathbf{v}$ and $\mathbf{w}$ is

$$dist(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}||$$

Definition 26. We say that two vectors $\mathbf{v}$ and $\mathbf{w}$ are **orthogonal** when

$$\mathbf{v} \cdot \mathbf{w} = 0$$

The idea comes from $\mathbf{v} \cdot \mathbf{w} = ||\mathbf{v}|| \cdot ||\mathbf{w}|| \cos(\theta)$, which is zero when θ is a right angle. This also allows us to talk about angles between vectors in higher dimensions.

With these definitions, geometric properties we are used to seeing in 2 or 3 dimensions extend to $\mathbb{R}^n$:

Theorem 31. (Pythagorean Theorem) Two vectors $\mathbf{v}$ and $\mathbf{w}$ are orthogonal if and only if $||\mathbf{v}+\mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2$. (the sum of the vectors is another vector with the same length as the hypotenuse of the right triangle determined by $\mathbf{v}$ and $\mathbf{w}$).

While the concept of orthogonality is fairly straightforward for vectors, we expand this further to subspaces, and the concept of orthogonal complements.

Definition 27. Suppose W is a subspace of $\mathbb{R}^n$ (or more generally any vector space V).

(1) A vector $\mathbf{x} \in \mathbb{R}^n$ is **orthogonal to** W if it is orthogonal to every $\mathbf{w} \in W$, i.e $\mathbf{x}\cdot\mathbf{w} = 0$ for every $\mathbf{w} \in W$.
(2) The **orthogonal complement of** W, denoted by $W^\perp$, is the set of all vectors $\mathbf{x} \in \mathbb{R}^n$ that are orthogonal to W.
(3) $W^\perp$ is also a subspace of $\mathbb{R}^n$.

Theorem 32. Let A be an $m\times n$ matrix. The orthogonal complement of $\mathit{Row}(A)$ is $Null(A)$, and the orthogonal complement of $Col(A)$ is $Null(A^\intercal)$.

$$(\mathit{Row}\ A)^\perp = Null(A) \quad \text{and} \quad (Col\ A)^\perp = Null(A^\intercal)$$

Proof. (idea) Think about what it means for $\mathbf{x} \in Null(A)$. It means $A\mathbf{x} = \mathbf{0}$, so the dot product of $\mathbf{x}$ to every row in A is 0. This means that $\mathbf{x}$ is orthogonal to $\mathit{Row}(A)$, and so is every vector from the null space. For the second part of the theorem, we know $Col(A) = \mathit{Row}(A^\intercal)$, so just apply the same argument to $A^\intercal$. □

23. Orthogonal Sets/Bases, Projections, Gram-Schmidt

In Calc 3, we talked about scalar and vector projections of $\mathbf{v}$ onto $\mathbf{b}$. These were:

$$comp_{\mathbf{b}}(\mathbf{v}) = \frac{\mathbf{v}\cdot\mathbf{b}}{||\mathbf{b}||} \qquad \text{and} \qquad \widehat{\mathbf{v}} = proj_{\mathbf{b}}(\mathbf{v}) = \left(\frac{\mathbf{v}\cdot\mathbf{b}}{||\mathbf{b}||}\right)\frac{\mathbf{b}}{||\mathbf{b}||} = \left(\frac{\mathbf{v}\cdot\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}}\right)\mathbf{b}$$

At the same time, the perpendicular/normal component of $\mathbf{v}$ onto $\mathbf{b}$ was just $\mathbf{v} - \widehat{\mathbf{v}}$. These are good to keep in mind for this entire section, but we are particularly interested in the vector projection.

Definition 28. A set $\{\mathbf{b}_1, \dots, \mathbf{b}_r\} \subset \mathbb{R}^n$ is an **orthogonal set** if $\mathbf{b}_i \cdot \mathbf{b}_j = 0$ for every $i \neq j$ (the dot product of any pair of vectors form the set is zero).

Theorem 33. Every orthogonal set $\mathcal{B}$ of (nonzero) vectors is linearly independent, and forms a basis for the subspace W spanned by $\mathcal{B}$. This is called an **orthogonal basis** for W.

There is a close connection between the coordinates of a vector $[\mathbf{v}]_{\mathcal{B}} = \langle c_1, \dots, c_r\rangle$ relative to an orthogonal basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_r\}$. Since we can express $\mathbf{v}$ as the linear combination

$$\mathbf{v} = c_1\mathbf{b}_1 + \cdots + c_r\mathbf{b}_r$$

by orthogonalilty we see that j-th coordinate of $[\mathbf{v}]_{\mathcal{B}}$ is given by the formula:

$$c_j = \frac{\mathbf{v}\cdot\mathbf{b}_j}{\mathbf{b}_j\cdot\mathbf{b}_j}$$

This expression is just another way of stating the decompositon theorem from the book.

Note that the formula for c_j is slightly different from the scalar projection we are used to, which gives the component relative to the *unit vector* in the direction of $\mathbf{b}_j$. It is off by a factor of $|\mathbf{b}_j|$, as shown below:

$$comp_{b_j}(\mathbf{v}) = \frac{\mathbf{v}\cdot\mathbf{b}_j}{||\mathbf{b}_j||} = \frac{\mathbf{v}\cdot\mathbf{b}_j}{\mathbf{b}_j\cdot\mathbf{b}_j}\cdot||\mathbf{b}_j|| = c_j||\mathbf{b}_j||$$

Definition 29. If the vectors in $\mathcal{B}$ are not only orthogonal, but also unit vectors, then we say $\mathcal{B}$ forms an *orthonormal basis.* These are essentially rotations of the standard basis.

We have a quick way to determine if a set is orthonormal. This is given by the following:

Theorem 34. An $m \times n$ matrix U has orthonormal columns if and only if $U^\intercal U = I$. If U has orthogonal columns (not normalized), then $U^\intercal U$ will be a diagonal matrix.

Proof. To see why this is true, consider the matrix product $U^\intercal U = M$. The (i,j)-entry in M is the dot product of row i of $U^\intercal$ and col j of U, but this is the same as doing the dot product of $\mathbf{U}_i$ (col i) and $\mathbf{U}_j$ (col j) of the matrix U, which will be:

$$M_{ij} = \mathbf{U}_i\cdot\mathbf{U}_j = \begin{cases} 1 & \text{if } i=j \text{ (or some constant for the orthogonal case)} \\ 0 & \text{if } i\neq j \text{ (due to orthogonality)} \end{cases}$$

The result is the identity matrix (or a diagonal matrix for the orthogonal case). □

In practice, we construct a matrix whose columns are the vectors form $\mathcal{B}$, and test the above condition. If U has orthonormal columns, , then the following properties hold for $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$:

(1) $||U\mathbf{x}|| = ||\mathbf{x}||$ (multiplication by U preserves length)
(2) $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$
(3) The expression in (2) is zero if and only if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal to begin with.

This is another way of saying that multiplication by such U preserves length and orthogonality.

Gram-Schmidt

Sometimes we start with some basis $B = \{\mathbf{x}_1,\ldots,\mathbf{x}_p\}$ for a subspace W of $\mathbb{R}^n$ (or the entire $\mathbb{R}^n$), and we are interested in turning that into an orthogonal frame. We do this as follows:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_1\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p\cdot\mathbf{v}_{p-1}}{\mathbf{v}_{p-1}\cdot\mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

so we iteratively remove the parallel components from each $\mathbf{x}_i$ relative to all the preceeding $\mathbf{v}_i$. At the end, the set $V = \{\mathbf{v}_1,\ldots,\mathbf{v}_p\}$ is an orthogonal basis for the same subspace. Furthermore, each initial subset of V spans the same subspace as each initial subset of the original B (so if you truncate the sets at the same position, you get the same subspace).

To turn this into an orthonormal basis, only one last step is required: divide everything in V by its length.

24. The Method Of Least-Squares And QR Decomposition

Looking at the equation $A\mathbf{x} = \mathbf{b}$, sometimes when trying to solve the system, we found it to be inconsistent. This begs the question, can we find some $\hat{\mathbf{x}}$ so that the difference between $A\mathbf{x}$ and $\mathbf{b}$ is minimal?

The book gives the reasoning behind the process, but to find such an $\hat{\mathbf{x}}$, you do the following steps:

(1) Consider the system $A^\intercal A\mathbf{x} = A^\intercal\mathbf{b}$.
(2) Row-reduce the matrix
$$\left(A^\intercal A \mid A^\intercal\mathbf{b}\right) \xrightarrow{rref} \left(I \mid \hat{\mathbf{x}}\right)$$
(3) Alternately, one can do $\hat{\mathbf{x}} = (A^\intercal A)^{-1}(A^\intercal\mathbf{b})$ (if $A^\intercal A$ is invertible).

In some cases, the new system $A^\intercal A\mathbf{x} = A^\intercal\mathbf{b}$ has a nontrivial null space. If that happens, the row reduction above will give infinitely many solutions. The following theorem describes when we have a unique solution:

Theorem 35. Let A be an $m \times n$ matrix. The following are equivalent:

(a) The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b} \in \mathbb{R}^m$.
(b) The columns of A are independent.
(c) $A^\intercal A$ is invertible, in which case step (3) in the process above is possible.

QR Factorization

Theorem 36. Suppose A is an $m \times n$ matrix with independent columns. Then we can write $A = QR$, where $Q_{m\times n}$ is a matrix whose columns form an **orthonormal** basis for $Col(A)$, and R is an upper triangular matrix.

In fact, the columns of R are related to expressing the columns of A in terms of the columns of Q, but there is a simpler method for finding R.

(1) Use Gram-Schmidt on the columns of A to get an orthogonal set.
(2) Normalize this set to get the columns of Q
(3) In $A = QR$, the matrix Q (and A) is not always square, so we cannot just multiply by Q^{-1} to get R. Instead, consider $(Q^\intercal A) = (Q^\intercal Q)R$. Now since Q is an orthonormal matrix, $Q^\intercal Q = I$ and this reduces to $R = Q^\intercal A$.

QR factorization can be used as an alternate way to get the least-square solution to $A\mathbf{x} = \mathbf{b}$.

Theorem 37. Given a QR factorization $A = QR$ of an $m \times n$ matrix A with independent columns (as in Theorem 35), the equation $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution given by
$$\hat{\mathbf{x}} = R^{-1}Q^\intercal\mathbf{b}$$

Note: this is not always convenient, as it involves taking inverses. Instead, one can solve the system
$$R\mathbf{x} = Q^\intercal\mathbf{b}$$
using row-reduction to speed things up.