Math 291-1: Intensive Linear Algebra & Multivariable Calculus

Northwestern University, Lecture Notes

Written by Santiago Cañez

These are notes which provide a basic summary of each lecture for Math 291-1, the first quarter of "MENU: Intensive Linear Algebra & Multivariable Calculus", taught by the author at Northwestern University. The book used as a reference is the 5th edition of *Linear Algebra with Applications* by Bretscher. Watch out for typos! Comments and suggestions are welcome.

These notes will focus on material covered in Math 291 which is not normally covered in Math 290, and should thus be used in conjunction with my notes for Math 290, which are available at http://www.math.northwestern.edu/∼scanez/courses/290/notes.php.

Contents

Lecture 1: Introduction

Linear algebra is the study of "linear" things. In the simplest setting, "linear" things are those which can be described using equations where variables appear to only the first power; for instance, lines and planes both fall into this category. This does not seem all that interesting on its own, but this course is all about convincing you that it is incredibly interesting, mainly due to the fact that there are so many different types of questions which can be phrased in terms of "linear" things.

One of the key facts we outlined on the first day is that any system of linear equations, no matter the number of variables nor the number of equations, has either zero, exactly one, or infinitely many solutions. This makes sense geometrically when we interpret solutions of linear systems in terms of intersecting lines, planes, or analogous higher-dimensional objects (which we can't really picture), but we can show it algebraically too. Here was the argument for the simple case of a system of two equations in two unknowns:

Claim. If the system of linear equations:

$$\begin{aligned} ax + by &= c \\ dx + fy &= g \end{aligned}$$

where x and y are the variables and a, b, c, d, f, g are fixed numbers, has at least two solutions, then it has infinitely many.

Proof. Suppose that (x_1, y_1) and (x_2, y_2) are two distinct solutions of the given system, so that

$$\begin{aligned} ax_1 + by_1 &= c \\ dx_1 + fy_1 &= g \end{aligned} \quad \text{and} \quad \begin{aligned} ax_2 + by_2 &= c \\ dx_2 + fy_2 &= g. \end{aligned}$$

We claim that then $(x, y) = \left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}y_1 + \frac{1}{2}y_2\right)$ is also a solution and that it is different from both (x_1, y_1) and (x_2, y_2). First, to verify that it is a solution, we substitute in for x and y:

$$a\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) + b\left(\frac{1}{2}y_1 + \frac{1}{2}y_2\right) = \frac{1}{2}(ax_1 + by_1) + \frac{1}{2}(ax_2 + by_2) = \frac{1}{2}c + \frac{1}{2}c = c,$$

so the first equation is satisfied, and

$$d\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) + f\left(\frac{1}{2}y_1 + \frac{1}{2}y_2\right) = \frac{1}{2}(dx_1 + fy_1) + \frac{1}{2}(dx_2 + fy_2) = \frac{1}{2}g + \frac{1}{2}g = g,$$

so the second is as well. Hence $(x, y) = \left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}y_1 + \frac{1}{2}y_2\right)$ is a solution of the given linear system as claimed.

To see that this solution is different from both (x_1, y_1) and (x_2, y_2), suppose instead that $\left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}y_1 + \frac{1}{2}y_2\right)$ was the same as (x_1, y_1). Then we would have

$$\frac{1}{2}x_1 + \frac{1}{2}x_2 = x_1 \quad \text{and} \quad \frac{1}{2}y_1 + \frac{1}{2}y_2 = y_1.$$

But the first equation gives $\frac{1}{2}x_1 = \frac{1}{2}x_2$, so $x_1 = x_2$, and the second gives $\frac{1}{2}y_1 = \frac{1}{2}y_2$, so $y_1 = y_2$. Thus, in order for $\left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}y_1 + \frac{1}{2}y_2\right)$ to be the same as (x_1, y_1), it would have to be true that $(x_1, y_1) = (x_2, y_2)$, which is not true since we are assuming that (x_1, y_1) and (x_2, y_2) are distinct solutions of the given system. A similar computation shows that $\left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}y_1 + \frac{1}{2}y_2\right)$ is different from (x_2, y_2), so $\left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}y_1 + \frac{1}{2}y_2\right)$ is indeed a solution of the given system distinct from the two we started with.

Now, repeating the same argument to the solutions (x_1, y_1) and $\left(\frac{1}{2}x_1 + \frac{1}{2}x_2, \frac{1}{2}y_1 + \frac{1}{2}y_2\right)$ will produce a fourth solution, and repeating the argument with (x_1, y_1) and this fourth solution will give a fifth, and so on. Thus we end up with infinitely many solutions. □

Exercise. More generally, with the same setup as above, what types of coefficients m and n will result in

$$(mx_1 + nx_2, my_1 + ny_2)$$

being a solution of the same system?

A similar idea works with larger linear systems, but the algebra gets messier to write out in detail. Later we will interpret this question in terms of so-called *linear transformations*, where we can give a simpler, unified answer as to why this works in general.

We also saw that something similar happens for certain so-called *systems of differential equations*, which (hopefully) suggests that there is some connection between these topics. In particular, say we have the following system of differential equations:

$$\begin{aligned} x_1' &= 10x_1 + 8x_2 - 2 \\ x_2' &= x_1 - 3x_2 + 4. \end{aligned}$$

Recall that these equations characterize functions $x_1(t)$ and $x_2(t)$ with the property that the derivative of x_1 is 10 times x_1 itself plus 8 times x_2 minutes 2, and similarly the derivative of x_2 satisfies the requirement of the second equation. To solve this system means to find all pairs of functions $x_1(t), x_2(t)$ satisfying both equations simultaneously. (Of course, it is not assumed coming into this course that you know what differential equations are nor how to solve them—this example is only meant to illustrate a connection between questions concerning solutions of differential equations and questions concerning solutions of linear systems. We'll learn more about differential equations at some point.)

Claim. If $x_1(t) = f_1(t), x_2(t) = g_1(t)$ and $x_1(t) = f_2(t), x_2(t) = g_2(t)$ are both solutions of the above system of differential equations, then

$$x_1(t) = \frac{1}{2}f_1(t) + \frac{1}{2}f_2(t), x_2(t) = \frac{1}{2}g_1(t) + \frac{1}{2}g_2(t)$$

is also a solution, and it is distinct from the two we started with.

Proof. To say that f_1, g_1 is a solution of the given system means that their derivatives satisfy

$$\begin{aligned} f_1' &= 10f_1 + 8g_1 - 2 \\ g_1' &= f_1 - 3g_1 + 4 \end{aligned}$$

and similarly saying that f_2, g_2 is a solution means

$$f_2' = 10f_2 + 8g_2 - 2$$

$$g_2' = f_2 - 3g_2 + 4.$$

To check that

$$x_1 = \frac{1}{2}f_1 + \frac{1}{2}f_2, x_2 = \frac{1}{2}g_1 + \frac{1}{2}g_2$$

is also a solution, we simply verify that substituting in the first function in for x_1 and the second in for x_2 in our system results in true statements. We have:

$$\begin{aligned} 10\left(\frac{1}{2}f_1 + \frac{1}{2}f_2\right) + 8\left(\frac{1}{2}g_1 + \frac{1}{2}g_2\right) - 2 &= \frac{1}{2}(10f_1 + 8g_1 - 2) + \frac{1}{2}(10f_2 + 8g_2 - 2) \\ &= \frac{1}{2}f_1' + \frac{1}{2}f_2' \\ &= \left(\frac{1}{2}f_1 + \frac{1}{2}f_2\right)', \end{aligned}$$

so $x_1 = \frac{1}{2}f_1 + \frac{1}{2}f_2, x_2 = \frac{1}{2}g_1 + \frac{1}{2}g_2$ satisfies the first of our differential equations. Similarly:

$$\begin{aligned} \left(\frac{1}{2}f_1 + \frac{1}{2}f_2\right) - 3\left(\frac{1}{2}g_1 + \frac{1}{2}g_2\right) + 4 &= \frac{1}{2}(f_1 - 3g_1 + 4) + \frac{1}{2}(f_2 - 3g_2 + 4) \\ &= \frac{1}{2}g_1' + \frac{1}{2}g_2' \\ &= \left(\frac{1}{2}g_1 + \frac{1}{2}g_2\right)', \end{aligned}$$

so the second differential equation is satisfied as well. Thus

$$x_1 = \frac{1}{2}f_1 + \frac{1}{2}f_2, x_2 = \frac{1}{2}g_1 + \frac{1}{2}g_2$$

is indeed also a solution of the given system of differential equations.

The same reasoning used in the previous claim to show that the newly obtained solution was different from the previous ones also applies here to show that this newly obtained solution is different from the ones with which we started. □

As in the previous claim about intersecting lines, it follows that if the given system of differential equations has at least two solutions, it must have infinitely many.

Is it a coincidence that solutions of this and other systems of differential equations exhibit the same type of behavior as do solutions of linear systems? Of course not, or else we wouldn't be mentioning it here. Indeed, when phrased in the right terms, we will see that we are really asking the *same* question here. The setting of *vector spaces* and *linear transformations* provides the unified framework from which to approach the topics we've looked at here and so many more. It will be a glorious journey for us all!

Lecture 2: Mathematical Induction

Warm-Up. Consider the equation of a line

$$ax + by = c$$

where a, b, c are fixed numbers. Suppose that $\left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right]$ are two points on this line. We show that if $c \neq 0$, then $\left[\begin{smallmatrix} mx_1+nx_2 \\ my_1+ny_2 \end{smallmatrix}\right]$ is also on this line if and only if $n = 1 - m$; while if $c = 0$, then $\left[\begin{smallmatrix} mx_1+nx_2 \\ my_1+ny_2 \end{smallmatrix}\right]$ is on this line for any values of m and n.

To say that $\left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right]$ are on the given line means that

$$ax_1 + by_1 = c \quad \text{and} \quad ax_2 + by_2 = c.$$

We compute:

$$a(mx_1 + nx_2) + b(my_1 + ny_2) = m(ax_1 + by_1) + n(ax_2 + by_2) = mc + nc.$$

Thus, $\left[\begin{smallmatrix} mx_1+nx_2 \\ my_1+ny_2 \end{smallmatrix}\right]$ satisfies the equation of the same line if and only if

$$mc + nc = c, \text{ or equivalently } nc = (1 - m)c.$$

If $c \neq 0$, after dividing this equation through by c we can see that it holds if and only if $n = 1 - m$ as claimed, while if $c = 0$ this equation holds for any values of m and n, as claimed.

Question to think about. In the above scenario, why is it that in the $c = 0$ case, *any* values of m and n work, while in the $c \neq 0$ case only those values of m and n for which $m + n = 1$ work? What property of lines with $c = 0$ vs lines with $c \neq 0$ does this reflect?

Exercise. In the case where $c \neq 0$, is it true that *every* point on the line $ax + by = c$ above is of the form $\left[\begin{smallmatrix} mx_1+(1-m)x_2 \\ my_1+(1-m)y_2 \end{smallmatrix}\right]$, where $\left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right]$ are the two points we started with?

Induction. Induction is a technique for proving statements of the form: "For any positive integer n past some value, $P(n)$ is true", where $P(n)$ is whatever formula/inequality/other depending on n we're interested in. For us, induction is something we'll make good use of when justifying many things in linear algebra, but it also serves to set the tone of the course in that it is a good introduction to proof-writing in general. The underlying core behind induction is the idea that we can use known results for certain values of n to justify those same results for larger values of n, thereby deriving new information from old.

To prove a statement of the form "For any positive integer $n \geq 1, P(n)$ is true", we proceed as follows:

- (Base case) First we verify that the claim is true in the base case $n = 1$; that is, we verify that $P(1)$ is true,
- (Induction step) Next we assume that the claim we're looking at is true for *some* $n = k \geq 1$, and use this to show that the claim is then also true for $n = k + 1$; that is, we show that if $P(k)$ is true for some $k \geq 1$, then $P(k + 1)$ is also true. (The statement $P(k)$ which we assume to be true in this step is usually referred to as the *induction hypothesis*.)

If we can carry out these two steps, the conclusion is that $P(n)$ is in fact true for all $n \geq 1$ as claimed. This is the basic format, but we'll see that there are modifications we can make, such as when the base case is something other than $n = 1$.

Why does induction work? How we do know that carrying out the two steps in the induction procedure indeed justifies that the claim we're looking at is true for all n? Here is the basic idea, with the key point being that the induction step guarantees that if we know $P(k)$ to be true, then we automatically get that $P(k + 1)$ is true.

By the base case, we know that $P(1)$ is true. Applying the induction step to this then implies that $P(2)$ is true. But now that we know $P(2)$ is true, applying the induction step to *this* instead (i.e. applying the induction step to $k = 2$) implies that $P(3)$ is true. Applying the induction step

to $P(3)$ now implies that $P(4)$ is true, and so on. This is often described as a "domino-like" effect, where the induction step "knocks down" the next domino, one at a time. In the end all dominos will be knocked down, so $P(n)$ is true for all n.

Exercise. If the above description as to why induction works is a bit too "hand-wavy" for you (and perhaps it should be), rest assured that we can actually *prove* that induction works! That is, suppose that for each $n \geq 1$, $P(n)$ is some statement. Assume further that we know $P(1)$ to be true and we know that if $P(k)$ is true for some $k \geq 1$, then $P(k+1)$ is true. Prove that then $P(n)$ must be true for all $n \geq 1$. Here is a hint: if we assume instead that $P(n)$ wasn't true for all $n \geq 1$, then there would have to be a smallest value of n for which $P(n)$ wasn't true; if we call this value M, what can we say about $P(M-1)$?

Example 1. Let's see induction in action. We show that if $x \neq 1$ and $n \geq 1$, then

$$1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x}.$$

This is a well-known formula, and in some sense we don't really need induction to justify it, since it can also be justified by expanding the product

$$(1 - x)(1 + x + x^2 + \cdots + x^n).$$

However, we'll approach it using induction here. The statement $P(n)$ in this case is the equality stated above.

First, since

$$1 + x = \frac{1 - x^2}{1 - x} \text{ because } 1 - x^2 = (1 - x)(1 + x),$$

the claimed equality is true for the base case $n = 1$. Now, suppose that

$$1 + x + x^2 + \cdots + x^k = \frac{1 - x^{k+1}}{1 - x} \text{ for some } k \geq 1.$$

(In other words, assume that $P(k)$ is true for some k—this is our induction hypothesis.) We then want to show that the claimed equality holds for $n = k + 1$ as well; that is, we want to show that

$$1 + x + x^2 + \cdots + x^{k+1} = \frac{1 - x^{k+2}}{1 - x}.$$

The key point, as with any induction procedure, is to figure out how to use the induction hypothesis that the formula is true for $n = k$ to justify that it holds for the next larger value $n = k+1$. In this case, the observation needed is that the left-hand side of the equality we're trying to justify in fact *includes* among its terms the left hand side of the equality we're assuming to be true in our induction hypothesis. Indeed, the term right before x^{k+1} is x^k, so if we think of the given sum as

$$1 + x + x^2 + \cdots + x^{k+1} = (1 + x + \cdots + x^k) + x^{k+1},$$

we are able to use our induction hypothesis to replace the term in parentheses by $\frac{1-x^{k+1}}{1-x}$. All that will remain then is some algebra.

Here are the final details. We have:

$$1 + x + x^2 + \cdots + x^{k+1} = (1 + x + \cdots + x^k) + x^{k+1}$$

$$
\begin{aligned}
&= \frac{1-x^{k+1}}{1-x} + x^{k+1} \\
&= \frac{1-x^{k+1}+x^{k+1}(1-x)}{1-x} \\
&= \frac{1-x^{k+2}}{1-x},
\end{aligned}
$$

where in the second step we used the induction hypothesis and in the third we added fractions using a common denominator. This final equality is precisely the formula we're looking at when $n = k+1$, so we have shown that $P(k)$ being true implies that $P(k+1)$ is true. By induction, we conclude that our claimed equality in fact holds for all $n \geq 1$.

Example 2. Justifying equalities involving positive integers is a common use of induction, but the technique of induction extends beyond that. Here we justify an inequality:

$$\text{for all } n \geq 4,\ n^2 > 2n+3.$$

Note that the given inequality is not true for $n = 1, 2, 3$, which is why we are only considering values of n which are 4 or larger. The only change to the induction procedure is that our base case changes from $n = 1$ to $n = 4$, and so in the end our conclusion will only hold for $n \geq 4$.

Since $4^2 > 2(4)+3$, the inequality holds for the base case $n = 4$. Suppose for the induction hypothesis that $k^2 > 2k+3$ for some $k \geq 4$. We want to show that this implies $(k+1)^2 > 2(k+1)+3$. Since $k^2 > 2k+3$, we have

$$k^2 + 2k + 1 > 2k + 3 + 2k + 1$$

since adding $2k+1$ to both sides does not alter the inequality. The left-hand side is now $(k+1)^2$, so we have $(k+1)^2 > 4k+4$. Since $k \geq 4$, $2k \geq 1$ so

$$4k + 4 = 2k + 2k + 4 \geq 2k + 5,$$

and hence

$$(k+1)^2 > 4k + 4 \geq 2k + 5 = 2(k+1) + 3,$$

as we wanted. We conclude by induction that $n^2 > 2n+3$ for all $n \geq 4$.

Complex numbers. We'll be using complex numbers from time to time, so here is a very quick crash course. We'll talk about other properties of complex numbers as they're needed.

A *complex number* is an expression of the form $a+ib$, where a and b are both real numbers and the symbol i satisfies $i^2 = -1$. (The point is that you should think of i as something like $\sqrt{-1}$: $\sqrt{-1}$ does not exist as a real number—since no real number x satisfies $x^2 = -1$—so we introduce a new symbol which plays this role.) Addition of complex numbers work as you would expect it to:

$$(a+ib) + (c+id) = (a+c) + i(b+d)$$

where we group together the ib and id terms and factor out i, and multiplication of complex numbers does as well, as long as you remember that $i^2 = -1$:

$$(a+ib)(c+id) = ac + iad + ibc + bdi^2 = (ac-bd) + i(ad+bc).$$

For a complex number $a+ib$, its *complex conjugate* is the complex number $a-ib$ obtained by changing the sign of the i term, and is denoted by $\overline{a+ib}$:

$$\overline{a+ib} = a - ib.$$

As stated above, other properties of complex numbers will be introduced as needed.

Example 3. Here is one final induction example, this time justifying a property of complex numbers. The claim is that for any complex numbers $z_1, \dots, z_n$, the conjugate of their sum is the sum of their conjugates:

$$\overline{z_1 + \cdots + z_n} = \overline{z_1} + \cdots + \overline{z_n}.$$

The induction here occurs on n, the number of complex numbers we are considering. Note that for $n = 1$ this says $\overline{z_1} = \overline{z_1}$, which isn't really saying much, so the base case should really be $n = 2$.

Suppose that $z_1 = a + ib$ and $z_2 = c + id$ with a, b, c, d all real. Then

$$z_1 + z_2 = (a + ib) + (c + id) = (a + c) + i(b + d),$$

so

$$\overline{z_1 + z_2} = (a + c) - i(b + d) = (a - ib) + (c - id) = \overline{z_1} + \overline{z_2}.$$

Hence the claimed equality holds for the base case $n = 2$. Suppose now that the claimed equality holds for any k complex numbers, where $k \geq 2$, and suppose that $z_1, \dots, z_{k+1}$ is a list of $k + 1$ complex numbers. By thinking of the sum $z_1 + \cdots + z_{k+1}$ as $(z_1 + \cdots + z_k) + z_{k+1}$, the base case gives

$$\overline{z_1 + \cdots + z_{k+1}} = \overline{(z_1 + \cdots + z_k) + z_{k+1}} = \overline{z_1 + \cdots + z_k} + \overline{z_{k+1}}.$$

To be clear, we are applying the base case to the two complex numbers $z_1 + \cdots + z_k$ and z_{k+1}. Now, by the induction hypothesis

$$\overline{z_1 + \cdots + z_k} = \overline{z_1} + \cdots + \overline{z_k},$$

so we get

$$\overline{z_1 + \cdots + z_{k+1}} = \overline{z_1 + \cdots + z_k} + \overline{z_{k+1}} = \overline{z_1} + \cdots + \overline{z_k} + \overline{z_{k+1}}$$

as desired. We conclude by induction that the claimed equality holds for all $n \geq 2$.

Remark. Note the subtle difference between this final example and, say, the first example. In this example, the induction step requires applying both the induction hypothesis *and* the base case, whereas in the first example the base case was no where to be found in the induction step. It will often be the case that carrying out an induction successfully requires using the base case and/or other special cases together with the induction hypothesis as part of the induction step.

Lecture 3: Vectors

Warm-Up. We show that for any n complex numbers $z_1, \dots, z_n$,

$$|z_1 z_2 \cdots z_n| = |z_1||z_2| \cdots |z_n|$$

where the *absolute value* $|a + ib|$ of a complex number is defined to be $|a + ib| := \sqrt{a^2 + b^2}$. (The notation ":=" is used to indicate that the left-hand side is being defined via the right-hand side.) Geometrically, if you associate to a complex numbers $z = a + ib$ the point (a, b) in $\mathbb{R}^2$, the absolute value of z gives the distance from (a, b) to the origin.

We proceed by induction. First suppose that $z_1 = a + ib$ and $z_2 = c + id$ are two complex numbers. Then

$$z_1 z_2 = (a + ib)(c + id) = (ac - bd) + (ad + bc)i,$$

so

$$\begin{aligned}|z_1 z_2| &= \sqrt{(ac-bd)^2+(ad+bc)^2} \\ &= \sqrt{a^2c^2-2acbd+b^2d^2+a^2d^2+2adbc+b^2c^2} \\ &= \sqrt{a^2c^2+b^2d^2+a^2d^2+b^2c^2}.\end{aligned}$$

But also,

$$|z_1||z_2| = \sqrt{a^2+b^2}\sqrt{c^2+d^2} = \sqrt{(a^2+b^2)(c^2+d^2)} = \sqrt{a^2c^2+a^2d^2+b^2c^2+b^2d^2}.$$

Thus we see that $|z_1 z_2| = |z_1||z_2|$, so the claimed equality holds for the base case $n = 2$.

Suppose now that the claimed equality holds for any k complex numbers, and let $z_1, \ldots, z_{k+1}$ denote $k+1$ complex numbers. Then

$$\begin{aligned}|z_1 \cdots z_{k+1}| &= |(z_1 \cdots z_k) z_{k+1}| \\ &= |z_1 \cdots z_k||z_{k+1}| \\ &= |z_1| \cdots |z_k||z_{k+1}|,\end{aligned}$$

where the second line follows by applying the base case to the complex numbers $z_1 \cdots z_k$ and z_{k+1}, and the third line by the induction hypothesis that $|z_1 \cdots z_k| = |z_1| \cdots |z_k|$. Hence, assuming that the claimed equality holds for k complex numbers implies it holds for $k+1$ complex numbers, so by induction we conclude that it holds for any n complex numbers for any $n \geq 2$.

Remark. Of course, if we write express each complex number above concretely as

$$z_1 = a_1 + ib_1, \ldots, z_n = a_n + ib_n,$$

we can try to compute out the expressions

$$|z_1 z_2 \cdots z_n| = |(a_1 + ib_1)(a_2 + ib_2) \cdots (a_n + ib_n)|$$

and

$$|z_1||z_2| \cdots |z_n| = \sqrt{a_1^2 + b_1^2}\sqrt{a_2^2 + b_2^2} \cdots \sqrt{a_n^2 + b_n^2}$$

concretely to see if they are equal, but the point is that this will involve some incredibly messy algebra. Instead, it is much cleaner to work out this algebra only in the simplest case where we have two terms, and then let induction take care of the general scenario.

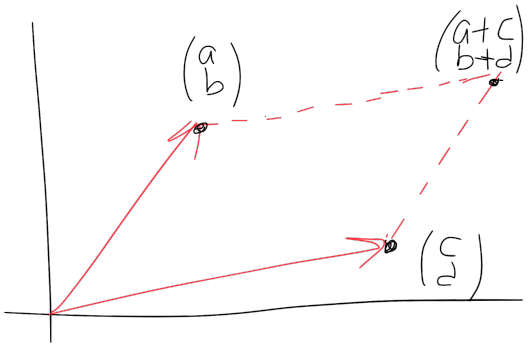
Exercise. To avoid the impression that induction only applies to proving concrete formulas or inequalities—as in all examples we've seen so far—here is a fun induction problem we looked at in class, which focuses more on using a procedure which works in a smaller setting to build up a procedure which works in a larger setting. For $n \geq 1$, take a $2^n \times 2^n$ sized chessboard and remove one square. Show that what remains can be covered using "L"-shaped pieces consisting of three squares, so that you get when you glue three squares together into an "L" shape.

Exercise. Here is another fun induction problem, which we didn't look at in class, but also focuses on using induction to "do something" rather than using it to prove some formula. A polygon is called *convex* if it has the property that for any two points inside of it, the entire line segment connecting these two points remains inside the polygon. Show that for any $n \geq 3$, any convex polygon with n vertices can be dividing up into $n-2$ triangles.

Vectors. Rather than spell out all of the material about vectors which we discussed in class, I'll point out that it (and more) can all be found in Appendix A of the book, which you should look at if vectors are an unfamiliar concept to you. The key points are understanding the geometric and algebric descriptions of vectors, vector addition, and scalar multiplication. Next time we'll focus more on understanding why various properties of vectors are actually true, as a way to further introduce ourselves to the notion of "proving" things.

Adding vectors geometrically. There is, however, one fact not clarified in the book, which we did look at in class. The question is why the geometric interpretation of vector addition via parallelograms gives the same answer as the algebraic interpretation in terms of adding together corresponding components.

Here is the setup. Let $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$ be two vectors (or points) in $\mathbb{R}^2$. Algebraically their sum is the vector $\left[\begin{smallmatrix} a+c \\ b+d \end{smallmatrix}\right]$, and the claim is that this sum is indeed the vector obtained by viewing $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$ (now thought of as arrows) as the sides of a parallelogram and then drawing the arrow which extends from $(0,0)$ to the opposite corner of this parallelogram:



To guarantee this we have to know that the quadrilateral with vertices at

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} a \\ b \end{bmatrix}, \quad \begin{bmatrix} c \\ d \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} a+c \\ b+d \end{bmatrix}$$

is in fact a parallelogram. Now, one side of this quadrilateral is the line segment between $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$, which has length $\sqrt{a^2+b^2}$, and the opposite side is the line segment between $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} a+c \\ b+d \end{smallmatrix}\right]$, which has length

$$\sqrt{(a+c-c)^2+(b+d-d)^2} = \sqrt{a^2+b^2},$$

so these two opposite sides have length. Similarly, the remaining sides are the line segment between $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$ and the line segment between $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} a+c \\ b+d \end{smallmatrix}\right]$, which have lengths

$$\sqrt{c^2+d^2} \quad \text{and} \quad \sqrt{(a+c-a)^2+(b+d-b)^2} = \sqrt{c^2+d^2}$$

respectively. Since opposite sides of this quadrilateral have the same length, it is actually a parallelogram, so the algebraic addition of vectors indeed corresponds to the geometric addition as claimed.

Lecture 4: More on Vectors

Warm-Up 1. We verify the *associative* property of vector addition:

$$(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z}) \text{ for any } \mathbf{x},\mathbf{y},\mathbf{z} \in \mathbb{R}^n.$$

(Recall that $\mathbb{R}^n$ denotes the space of vectors with n coordinates, and the symbol "$\in$" means "in", so $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ means that $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are vectors in $\mathbb{R}^n$.)

First we justify this algebraically. Suppose that concretely

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \ \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

are the coordinate expressions for $\mathbf{x}, \mathbf{y}, \mathbf{z}$. Then

$$\begin{aligned} (\mathbf{x} + \mathbf{y}) + \mathbf{z} &= \left[\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \right] + \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \\ &= \begin{bmatrix} (x_1 + y_1) + z_1 \\ \vdots \\ (x_n + y_n) + z_n \end{bmatrix}. \end{aligned}$$

On the other hand,

$$\begin{aligned} (\mathbf{x} + \mathbf{y}) + \mathbf{z} &= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \left[\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \right] \\ &= \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 + z_1 \\ \vdots \\ y_n + z_n \end{bmatrix} \\ &= \begin{bmatrix} x_1 + (y_1 + z_1) \\ \vdots \\ x_n + (y_n + z_n) \end{bmatrix}. \end{aligned}$$

By the associativity property of additional of real numbers, we have that

$$(x_i + y_i) + z_i = x_i + (y_i + z_i) \text{ for each } i = 1, \ldots, n,$$

so

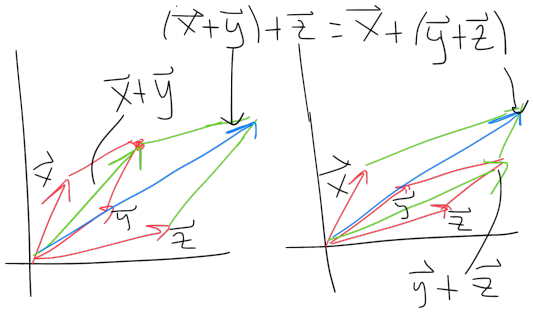
$$\begin{bmatrix} (x_1 + y_1) + z_1 \\ \vdots \\ (x_n + y_n) + z_n \end{bmatrix} = \begin{bmatrix} x_1 + (y_1 + z_1) \\ \vdots \\ x_n + (y_n + z_n) \end{bmatrix}$$

since saying that two vectors are equal means that their corresponding components are equal. Thus

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$$

as claimed.

Now we justify this geometrically, which requires drawing a good picture. The point is that, geometrically, $(\mathbf{x}+\mathbf{y})+\mathbf{z}$ means we first take the parallelogram with sides $\mathbf{x}+\mathbf{y}$ and draw the diagonal representing $\mathbf{x}+\mathbf{y}$, and then we take this diagonal and draw the parallelogram with it as one side and $\mathbf{z}$ as the other, whereas $\mathbf{x}+(\mathbf{y}+\mathbf{z})$ means we start with the parallelogram having $\mathbf{y}$ and $\mathbf{z}$ as sides instead. We only draw the picture for vectors in $\mathbb{R}^2$, which is a little easier to visualize than $\mathbb{R}^3$—which we can still picture—or $\mathbb{R}^n$ for $n \geq 4$, which we can't picture.



Remark. What if we had more than three vectors? With four we have more than two possible ways of grouping terms:

$$(\mathbf{x}+\mathbf{y})+(\mathbf{z}+\mathbf{w}),\ ((\mathbf{x}+\mathbf{y})+\mathbf{z})+\mathbf{w},\ \mathbf{x}+((\mathbf{y}+\mathbf{z})+\mathbf{w}),\ \textit{etc.}$$

Is it true that all such expressions are equal, so that we can write $\mathbf{x}+\mathbf{y}+\mathbf{z}+\mathbf{w}$ and unambiguously know what this means? Working this out algebraically as before shows that this comes down the analogous property of addition of real numbers, but how do we know this property holds for real numbers instead?

Of course, all such expressions are equal, and they remain so even if we have n vectors in general. Rather than doing a lot of messy computations to verify this, this is the type of thing which is taylor-made for induction. In particular, all we need is the fact that this associativity property holds for three vectors at time, and from this we can derive the analogous property for any number of vectors. (Technically, in the proof for the case of three vectors we did assume that $(x+y)+z = x+(y+z)$ holds for real numbers, which is one thing we will take for granted, but no other extra assumptions are necessary.) The following exercise is about using induction to prove the most general associativity property.

Exercise. (Inspired by John Alongi) We will use the following notation: for any number of vectors, $\mathbf{v}_1+\cdots+\mathbf{v}_k$ denotes specifically the sum obtained by first adding $\mathbf{v}_1$ and $\mathbf{v}_2$, then adding $\mathbf{v}_3$ to the result, then $\mathbf{v}_4$, then $\mathbf{v}_5$, and so on. So, for instance:

$$\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3+\mathbf{v}_4 := ((\mathbf{v}_1+\mathbf{v}_2)+\mathbf{v}_3)+\mathbf{v}_4.$$

The goal is to show that this specific grouping of terms is the same as any other grouping.

Fix a positive integer $\ell \geq 1$. Then for any $m \geq 1$ and any $\ell+m$ vectors $\mathbf{v}_1,\ldots,\mathbf{v}_{\ell+m}$, show that

$$(\mathbf{v}_1+\cdots+\mathbf{v}_\ell)+(\mathbf{v}_{\ell+1}+\cdots+\mathbf{v}_{\ell+m}) = \mathbf{v}_1+\cdots+\mathbf{v}_{\ell+m}.$$

Proceeding by induction on m, the induction step will use the base case of three vectors in addition to the induction hypothesis.

Warm-Up 2. We verify that for any vector $\mathbf{x} \in \mathbb{R}^n$ and scalar a, the length of $a\mathbf{x}$ is given by $|a|$ times the length of $\mathbf{x}$. This justifies the geometric picture of scalar multiplication, that multiplying

a vector by a scalar scales its length by that amount. (Of course, when $a < 0$, the vector also flips direction, but here we are only interested in what happens to the length.) Before doing so, we must define what we mean by the length of a vector in a higher-dimensional space $\mathbb{R}^n$: for

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n, \text{ the } \textit{length} \text{ of } \mathbf{x} \text{ is } \sqrt{x_1^2 + \cdots + x_n^2}.$$

The motivation is that, in the case of $\mathbb{R}^2$ or $\mathbb{R}^3$, this value indeed gives the usual notion of length we're accustomed to as the distance from $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ or $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ to the origin. We thus use this special case to define what "length" should mean in spaces we can't easily visualize.

Let $\mathbf{x}$ be as above and let $a \in \mathbb{R}$ be a scalar. Then

$$a\mathbf{x} = \begin{bmatrix} ax_1 \\ \vdots \\ ax_n \end{bmatrix},$$

so the length of $a\mathbf{x}$ is

$$\sqrt{(ax_1)^2 + \cdots + (ax_n)^2} = \sqrt{a^2(x_1^2 + \cdots + x_n^2)} = |a|\sqrt{x_1^2 + \cdots + x_n^2},$$

which is indeed $|a|$ times the length of $\mathbf{x}$ as claimed.

Properties of vector addition. Vector addition has some basic properties, none of which should seem very surprising. Here we will focus on two of them, where the point is to understand them from a "higher" point of view:

- (existence of an additive identity): there exists a unique vector $\mathbf{0} \in \mathbb{R}^n$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, and
- (existence of additive inverses): for any $\mathbf{x} \in \mathbb{R}^n$, there exists a unique vector $-\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.

The *zero vector* $\mathbf{0}$ is thus the "additive identity" for vector addition, meaning that it has the property that adding it to anything else gives back that other vector, and for a given $\mathbf{x}$ its *negative* $-\mathbf{x}$ is its "additive inverse", meaning that it has the property that adding it to $\mathbf{x}$ gives the zero vector, so in effect $-\mathbf{x}$ "cancels out" $\mathbf{x}$.

Concretely, $\mathbf{0}$ is the vector in coordinates which has all entires equal to 0, and it should not be hard to see that this indeed satisfies the "additive identity" property. Perhaps the more important point here is that $\mathbf{0}$ is the *only* vector satisfying the additive identity property, which where the "unique" in the definition comes from. To show this is true, we argue using only the additive identity property alone and nothing about coordinate expressions for vectors. A general strategy when proving something is unique is to suppose there are *two* things having the required property, and then show that they must actually be the same.

Proof that additive identities are unique. Suppose that $\mathbf{0}$ and $\mathbf{0}'$ are both additive identities for vector addition and consider the sum $\mathbf{0} + \mathbf{0}'$. On the one hand, since $\mathbf{0}$ is an additive identity this sum should equal $\mathbf{0}'$, but on the other hand since $\mathbf{0}'$ is an additive identity, this sum should also equal $\mathbf{0}$. Thus

$$\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}',$$

so $\mathbf{0} = \mathbf{0}'$ and hence there is only one additive identity. □

Why go through this trouble? The point of proving something like this—which might seem "obvious"—is that the same argument then applies to *any* other type of situation where we have a notion of "addition" defined. That is, later we will be seeing other types of "spaces" where we have an addition operation, and rather than having to check that additive identities are unique in each instance, we will know for sure that this is true since the proof we gave above works in any such settings. Similarly, after proving (on the homework) that additive inverses are unique using only the definition of "additive inverse", we will know that this holds in these other scenarios as well. This is the true power of mathematics: concepts usually apply much more generally beyond the specific setting you're interested in.

One more point regarding the additive inverse $-\mathbf{x}$ of $\mathbf{x}$. Certainly, in coordinates this is just going to be the vector whose coordinates are the negatives of those of $\mathbf{x}$, or in other words, $-\mathbf{x}$ will be the result of multiplying $\mathbf{x}$ by the scalar -1. However, if we didn't have such expressions in coordinates to work with, and if we couldn't picture vectors geometrically, why should it be true that the result of the scalar multiplication $(-1)\mathbf{x}$ should give the additive inverse $-\mathbf{x}$ of $\mathbf{x}$? The point is that from this approach, $-\mathbf{x}$ is simply denoting the vector with the property that if you add it to $\mathbf{x}$ you get $\mathbf{0}$, and this definition makes no mention of any type of scalar multiplication. This is what another problem on the homework asks you to show, that $(-1)\mathbf{x}$ is indeed the vector satisfying the "additive inverse" property—once we can show this using only general properties of vectors and nothing about coordinates, the same argument will apply to other types of "spaces" as well.

Distributive properties. There are two distributive properties involving vector addition and scalar multiplication, one where we distribute scalars and another where we distribute vectors:

- for any $a \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$, and
- for any $a, b \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$, $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$.

Both are straightforward to verify simply by computing both sides of the stated equalities and verifying that they are equal, and both boil down the usual distributive property $a(b+c) = ab+ac$ of real numbers.

And of course, the same distributive properties hold even when we have more than two vectors, which hopefully by now we can guess that reason for this is "induction".

Proof that the first distributive properties generalizes. We leave the check of the base case

$$a(\mathbf{v}_1 + \mathbf{v}_2) = a\mathbf{v}_1 + a\mathbf{v}_2$$

to you. Suppose that this distributive property holds for any k vectors, and let $\mathbf{v}_1, \dots, v_{k+1}$ be $k+1$ vectors in $\mathbb{R}^m$. By the base case, we have:

$$a(\mathbf{v}_1 + \cdots + \mathbf{v}_{k+1}) = a[(\mathbf{v}_1 + \cdots + \mathbf{v}_k) + \mathbf{v}_{k+1}] = a(\mathbf{v}_1 + \cdots + \mathbf{v}_k) + a\mathbf{v}_{k+1}.$$

By the induction hypothesis, $a(\mathbf{v}_1 + \cdots + \mathbf{v}_k) = a\mathbf{v}_1 + \cdots + a\mathbf{v}_k$, so all together we get

$$a(\mathbf{v}_1 + \cdots + \mathbf{v}_{k+1}) = a(\mathbf{v}_1 + \cdots + \mathbf{v}_k) + a\mathbf{v}_{k+1} = a\mathbf{v}_1 + \cdots + a\mathbf{v}_k + a\mathbf{v}_{k+1},$$

showing that the distributive property holds for $k+1$ vectors. We conclude by induction that

$$a(\mathbf{v}_1 + \cdots + \mathbf{v}_n) = a\mathbf{v}_1 + \cdots + a\mathbf{v}_n$$

for any n vectors in $\mathbb{R}^m$. □

Exercise. Prove that $(a_1+\cdots+a_n)\mathbf{v} = a_1\mathbf{v}+\cdots+a_n\mathbf{v}$ for any $\mathbf{v} \in \mathbb{R}^m$ and n scalars $a_1, \ldots, a_n \in \mathbb{R}$.

Exercise. Show *geometrically* that the distributive property $a(\mathbf{x}+\mathbf{y}) = a\mathbf{x}+a\mathbf{y}$ holds; that is, draw $a(\mathbf{x}+\mathbf{y})$ and draw $a\mathbf{x}+a\mathbf{y}$, and justify based on your drawing that these are the same.

Back to lines. Now that we've built up some basic properties of vectors, we return to an observation we made earlier regarding lines, only now we look at it from a new point of view. The observation as the following: for a line $ax+by=c$ and points $\left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right]$ on it, the values of m and n such that

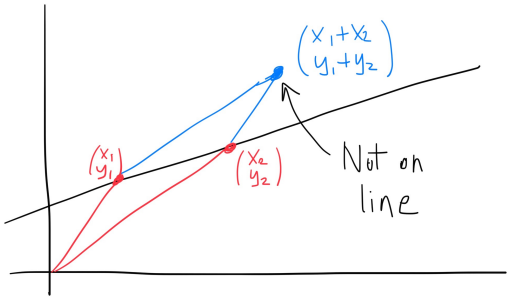
$$\begin{bmatrix} mx_1+nx_2 \\ my_1+ny_2 \end{bmatrix}$$

is still on the line are those for which $m+n=1$ in the $c \neq 0$ case, and in the $c=0$ case there is no restriction.

The point now is that we can view the above vector as the result of scaling the two original ones by m and n respectively and then adding the results:

$$\begin{bmatrix} mx_1+nx_2 \\ my_1+ny_2 \end{bmatrix} = m\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + n\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}.$$

So, why does it make sense geometrically that we get no restrictions on m and n when $c=0$ while we do get a restriction when $c \neq 0$? When $c \neq 0$, it is not true that adding two vectors on the line $ax+by=c$ must always give a vector on the same line, nor is it true that scaling a vector on the line must result in another vector on the line:



(Notice that when we say a vector is "on" a line, we mean that the endpoint of the vector—namely the point corresponding to the vector in coordinates—is on the line; we do *not* mean that the arrow representing the vector literally lies on the line.) Only certain types of scalings geometrically result in something on the same line. However, in the $c=0$ case, in fact adding *any* two points on the line gives something on the same line, and scaling anything on the line gives something still on it.

This is really the *key* difference between those lines for which $c \neq 0$ and those for which $c=0$, or in other words, between lines which don't pass through the origin and those which do. This type of distinction will play a big role in many of the topics we'll see going forward.

Lecture 5: Linear Combinations

Warm-Up 1. We show that for any scalar $a \in \mathbb{R}$, $a\mathbf{0} = \mathbf{0}$; that is, multiplying the zero vector by any scalar results in the zero vector. This is clear if we work with coordinates:

$$a\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a0 \\ \vdots \\ a0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

but the point is that will justify this in a "coordinate-free" way. By doing so we know that the same fact will be true in other settings where addition and scalar multiplication make sense but where we don't necessarily have "obvious" coordinates to work with. (What I mean by this will be clear once we talk about *vector spaces*.)

Note that $\mathbf{0} + \mathbf{0} = \mathbf{0}$. Thus

$$a\mathbf{0} = a(\mathbf{0} + \mathbf{0}) = a\mathbf{0} + a\mathbf{0}.$$

Whatever $a\mathbf{0}$ is, it has an additive inverse, and adding this additive inverse to both sides gives

$$\mathbf{0} = a\mathbf{0}$$

as claimed. (To be clear, after adding $-a\mathbf{0}$ to both sides we get $\mathbf{0} = a\mathbf{0} + \mathbf{0}$, and the right side is $a\mathbf{0}$ by the additive identity property of $\mathbf{0}$.)

Exercise. Show that for any $\mathbf{v} \in \mathbb{R}^n$, $0\mathbf{v} = \mathbf{0}$. Do this in a way which does not involve the coordinate representation of $\mathbf{v}$.

Warm-Up 2. The Warm-Up and Exercise above show that if $a = 0$ or $\mathbf{v} = \mathbf{0}$, then $a\mathbf{v} = \mathbf{0}$. Here we prove the *converse*: if $a\mathbf{v} = \mathbf{0}$, then $a = 0$ or $\mathbf{v} = \mathbf{0}$. This says that the *only* way in which multiplying a scalar and a vector can give the zero vector is if either the scalar or vector were zero to begin with.

How exactly do we prove something which claims one of two (or more) things can happen? The key is to consider the possible cases as to what a can be: either $a = 0$ or $a \neq 0$. If $a = 0$, then certainly "$a = 0$ or $\mathbf{v} = \mathbf{0}$" is true since at least one is zero, so there is nothing to show in this case. So, we really know have to think about the case when $a \neq 0$—then we want to show that $\mathbf{v}$ must be zero. This is a general strategy for proving statements of the form "if P, then Q or R" where P, Q, R are some statements: we instead show that "if P and Q is false, then R". The reason as to why this works coming from considering the only possible cases for Q: either Q is true, in which case there was really nothing to show after all, or Q is false, in which case we must show that R must be true. In this particular example, we could instead consider what happens with $\mathbf{v}$: either $\mathbf{v} = 0$, in which case there is nothing to show, or $\mathbf{v} \neq 0$, in which case we would have to show that $a = 0$. This would be an alternate approach to the one we use below.

So, to show that if $a\mathbf{v} = \mathbf{0}$, then $a = 0$ or $\mathbf{v} = \mathbf{0}$, we assume that $a\mathbf{v} = \mathbf{0}$ and $a \neq 0$, and show that $\mathbf{v} = \mathbf{0}$. Since $a \neq 0$, the scalar $\frac{1}{a}$ is defined, so multiplying $a\mathbf{v} = \mathbf{0}$ through by $\frac{1}{a}$ is possible—this gives

$$\frac{1}{a}(a\mathbf{v}) = \frac{1}{a}\mathbf{0}, \text{ so } \mathbf{v} = \mathbf{0}$$

since by the first Warm-Up we know that $\frac{1}{a}\mathbf{0} = \mathbf{0}$.

Back to intersecting lines. Previously we considered the system of linear equations:

$$\begin{aligned} x + 2y &= 0 \\ -3x - 2y &= 8 \end{aligned}$$

which has exactly one solution: $x = -4, y = 2$. Geometrically this means that the lines defined by the two given equations in $\mathbb{R}^2$ intersect in the point $(-4, 2)$ and nowhere else.

However, there is another way to interpret this system, which is perhaps more interesting. Consider both sides of the given system as the components of vectors, so we have the vector equality:

$$\begin{bmatrix} x + 2y \\ -3x - 2y \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}.$$

Now, the left-hand side can be expressed further using vector addition and scalar multiplication as:

$$\begin{bmatrix} x \\ -3x \end{bmatrix} + \begin{bmatrix} 2y \\ -2y \end{bmatrix} = x \begin{bmatrix} 1 \\ -3 \end{bmatrix} + y \begin{bmatrix} 2 \\ -2 \end{bmatrix},$$

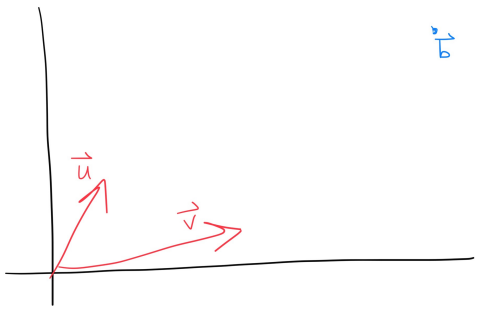
so the original linear system is the same as the vector equation

$$x \begin{bmatrix} 1 \\ -3 \end{bmatrix} + y \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \end{bmatrix}.$$

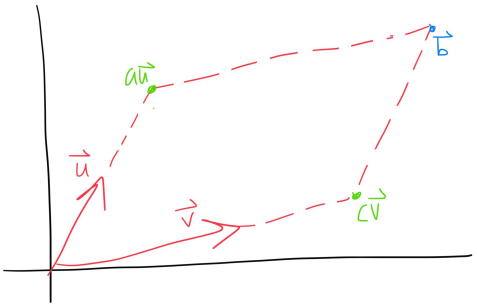
A solution of the original system gives scalars x, y for which this vector equation is true, and conversely scalars x, y for which this vector equation is true in turn gives a solution of the original system.

The left-hand side of this vector equation is called a *linear combination* of the vectors $\left[\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 \\ -2 \end{smallmatrix}\right]$, so the fact that this equation has a solution $(x, y) = (-4, 2)$ means that the vector $\left[\begin{smallmatrix} 0 \\ 8 \end{smallmatrix}\right]$ can be expressed as a linear combination of $\left[\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right]$ of $\left[\begin{smallmatrix} 2 \\ -2 \end{smallmatrix}\right]$.

Viewing linear combinations geometrically. Consider the vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ and the point (or vector) $\mathbf{b} \in \mathbb{R}^2$:



We claim that $\mathbf{b}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Indeed, the question is whether we can find a multiple of $\mathbf{u}$ and a multiple of $\mathbf{v}$ which add up to $\mathbf{b}$. Multiples of $\mathbf{u}$ lie along the line through the origin and $\mathbf{u}$ and multiples of $\mathbf{v}$ lie along the line through the origin and $\mathbf{v}$, and by "eyeballing" the picture we can draw a parallelogram with edges along these lines whose fourth corner is at $\mathbf{b}$:



We will soon remove the vagueness of "eyeballing" and see how to justify facts like this precisely.

In fact, given *any* point in $\mathbb{R}^2$, you can convince yourself that it is possible to draw a parallelogram with edges along the lines through $\mathbf{u}$ and $\mathbf{v}$ whose fourth corner will be at that point (recall that negative scalar multiples are drawn in the direction *opposite* the given vector), so anything in $\mathbb{R}^2$ is a linear combination of the $\mathbf{u}$ and $\mathbf{v}$ drawn above. We will see that this essentially reflects the fact that $\mathbf{u}$ and $\mathbf{v}$ are not parallel.

Definition. A *linear combination* of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ is an expression of the form

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$$

where $a_1, a_2, \dots, a_k \in \mathbb{R}$ are scalars.

Example. Let's see how to work with this definition. Suppose that $\mathbf{b} \in \mathbb{R}^n$ is a linear combination of $\mathbf{u}, \mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ and that $\mathbf{u}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$. We claim that $\mathbf{b}$ is then a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$ alone.

Since $\mathbf{b}$ is a linear combination of $\mathbf{u}, \mathbf{v}_1, \dots, \mathbf{v}_k$, there exist scalars $a, a_1, \dots, a_k \in \mathbb{R}$ such that

$$\mathbf{b} = a\mathbf{u} + a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k.$$

Since $\mathbf{u}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$, there are scalars $c_1, \dots, c_k$ such that

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k.$$

Substituting this expression in place of $\mathbf{u}$ gives:

$$\begin{aligned}\mathbf{b} &= a\mathbf{u} + a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k \\ &= a(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) + a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k \\ &= (ac_1 + a_1)\mathbf{v}_1 + \cdots + (ac_k + a_k)\mathbf{v}_k,\end{aligned}$$

so $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$ as claimed.

The point of this result is to say that if one vector $\mathbf{u}$ in a collection of vectors is a linear combination of the others, then $\mathbf{u}$ can be thrown away and anything which could be expressed as a linear combination of the original vectors can still be expressed as a linear combination of the remaining vectors, so that in a sense $\mathbf{u}$ was "redundant".

Exercise. Prove the following generalization of the above fact: if $\mathbf{u}_1, \dots, \mathbf{u}_\ell \in \mathbb{R}^n$ are each linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$, then anything which can be expressed as a linear combination of

$\mathbf{u}_1, \ldots, \mathbf{u}_\ell, \mathbf{v}_1, \ldots, \mathbf{v}_k$ can be expressed as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ alone. Induction (on the number of "$\mathbf{u}$" vectors) might be a nice way to do this.

Example we didn't do in class. Suppose that none of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ can be expressed as a linear combination of the others and suppose that $\mathbf{u} \in \mathbb{R}^n$ is a vector which is not a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Then we claim that none of the vectors in the longer list

$$\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}$$

can be expressed as a linear combination of the others. Note the subtlety: we know for sure that the last vector $\mathbf{u}$ in this list can't be expressed as a linear combination of the rest since this is part of what we are assuming, and we know that none of the $\mathbf{v}_i$'s can be expressed as a linear combinations of the $\mathbf{v}$'s alone, but why couldn't one of the $\mathbf{v}_i$'s be expressible as a linear combination of the $\mathbf{v}$'s *and* $\mathbf{u}$? This requires justification.

Suppose to the contrary that $\mathbf{v}_i$ for some $1 \leq i \leq k$ is a linear combination of

$$\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k, \mathbf{u}.$$

(Note that this list excludes $\mathbf{v}_i$.) So, for some scalars $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k, a \in \mathbb{R}$ we have

$$\mathbf{v}_i = a_1\mathbf{v}_1 + \cdots + a_{i-1}\mathbf{v}_{i-1} + a_{i+1}\mathbf{v}_{i+1} + \cdots + a_k\mathbf{v}_k + a\mathbf{u}.$$

If $a = 0$, the final term isn't there so this expresses $\mathbf{v}_i$ as a linear combination of the $\mathbf{v}$'s alone, which is not possible by assumption. Thus we must have $a \neq 0$. But then we can rearrange terms to get

$$a\mathbf{u} = \mathbf{v}_i - a_1\mathbf{v}_1 - \cdots - a_{i-1}\mathbf{v}_{i-1} - a_{i+1}\mathbf{v}_{i+1} - \cdots - a_k\mathbf{v}_k,$$

which in turn using the fact that $a \neq 0$ gives:

$$\mathbf{u} = -\frac{a_1}{a}\mathbf{v}_1 - \cdots - \frac{a_{i-1}}{a}\mathbf{v}_{i-1} + \frac{1}{a}\mathbf{v}_i - \frac{a_{i+1}}{a}\mathbf{v}_{i+1} - \cdots - \frac{a_k}{a}\mathbf{v}_k.$$

But this now expresses $\mathbf{u}$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$, which contradicts the assumption that $\mathbf{u}$ was not expressible as such a linear combination. Since the assumption that $\mathbf{v}_i$ for some $1 \leq i \leq k$ is a linear combination of

$$\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k, \mathbf{u}$$

leads to a contradiction, we conclude that none of the vectors

$$\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}$$

is a linear combination of the others.

Remark. In words, the above example shows that if we no vectors in some list are linear combinations of the others and we tack on a new vector which is not a linear combination of the vectors in the original list, then the vectors in the new list still have the property that none of the vectors is a linear combination of the rest. The importance of this example and the one before it will become clear once we talk about the notions of *span* and *linear independence*.

Affine combinations. We return to considering a line $ax + by = c$ in $\mathbb{R}^2$ which does not pass through the origin, so $c \neq 0$. If $\left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right]$ are on this line, we have seen that

$$\begin{bmatrix} mx_1 + nx_2 \\ my_1 + ny_2 \end{bmatrix} = m\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + n\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

is also on the line only when $m+n=1$. Thus, not all linear combinations of $\left[\begin{smallmatrix} x_1 \\ y_1 \end{smallmatrix}\right]$ of $\left[\begin{smallmatrix} x_2 \\ y_2 \end{smallmatrix}\right]$ remain on the line, but only those involving coefficients which add up to 1.

Such linear combinations will come up often enough for us that we give them a special name: an *affine combination* of $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ is a linear combination $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ such that $c_1 + \cdots + c_k = 1$. We'll see the types of situations in which affine combinations arise soon enough.

Lecture 6: Linear Systems

Warm-Up. Consider a *hyperplane* $a_1x_1 + \cdots + a_nx_n = 0$ passing through the origin in $\mathbb{R}^n$. To be clear, we speaking about a hyperplane we require that at least one coefficient $a_1, \dots, a_n$ is nonzero, and to say that it passes through the origin means that the point where all x_i coordinates equal zero satisfies the given equation; a hyperplane which does not pass through the origin would have equation $a_1x_1 + \cdots + a_nx_n = b$ where $b \neq 0$. In the case of $\mathbb{R}^2$, a hyperplane is just a line, and in the case of $\mathbb{R}^3$ a hyperplane is an ordinary plane, so a hyperplane in general is a generalization of a line or plane to higher dimensions.

We show that there are $n-1$ vectors we can find with the property that *any* point on this hyperplane is a linear combination of those $n-1$ vectors. Looking at a special case—that of a (hyper)plane in $\mathbb{R}^3$—can help suggest what to do in general. Say we had the plane

$$2x + y + 3z = 0.$$

The key realization is that such an equation allows us to express one variable in terms of the other two, say $y = -2x - 3z$. Thus any point on this plane must look like

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ -2x - 3z \\ z \end{bmatrix}.$$

But we can rewrite this second expression as

$$\begin{bmatrix} x \\ -2x - 3z \\ z \end{bmatrix} = \begin{bmatrix} x \\ -2x \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -3z \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix},$$

so any point on the plane must look like

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix}$$

for some $x, z \in \mathbb{R}$, which says that any point on this plane is a linear combination of the concrete vectors $\left[\begin{smallmatrix} 1 \\ -2 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 0 \\ -3 \\ 1 \end{smallmatrix}\right]$.

Now we return to the general setting. Since $a_1x_1 + \cdots + a_nx_n = 0$ is a hyperplane, at least one coefficient is nonzero—say $a_i \neq 0$. Then the components of any point on this hyperplane satisfy

$$a_ix_i = -a_1x_1 - \cdots - \widehat{a_ix_i} - \cdots - a_nx_n$$

where the notation $\widehat{a_ix_i}$ means that this term is omitted. (For instance, $a_1x_1 + \widehat{a_2x_2} + a_3x_3$ means $a_1x_1 + a_3x_3$. The point of this notation is that it is a little simpler to write and read than something like

$$a_ix_i = -a_1x_1 - \cdots - a_{i-1}x_{i-1} - a_{i+1}x_{i+1} - \cdots - a_nx_n,$$

which is how we originally wrote this in class.) Since $a_i \neq 0$, this gives

$$x_i = -\frac{a_1}{a_i}x_1 - \cdots - \widehat{x_i} - \cdots - \frac{a_n}{a_i}x_n.$$

Thus any point on the hyperplane must look like

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ -\frac{a_1}{a_i}x_1 - \cdots - \widehat{x_i} - \cdots - \frac{a_n}{a_i}x_n \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ -\frac{a_1}{a_i} \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ -\frac{a_2}{a_i} \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ -\frac{a_n}{a_i} \\ \vdots \\ 1 \end{bmatrix},$$

where the first vector in the final expression encodes all the coefficients of x_1, the second all the coefficients of x_2, and so on. There are $n-1$ terms in this final expression since there is one term for each of the variables $x_1, \ldots, \widehat{x_i}, \ldots, x_n$ excluding x_i, so this expressions shows that any point on the hyperplane in question is a linear combination of the $n-1$ vectors

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ -\frac{a_1}{a_i} \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ -\frac{a_2}{a_i} \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ -\frac{a_n}{a_i} \\ \vdots \\ 1 \end{bmatrix},$$

was was to be shown. We say that these vectors *span* the hyperplane, a term for which we'll give a precise definition soon.

Question to think about. Can we always find $n-1$ vectors different than those above which also span the hyperplane? Is it possible to have *fewer* than $n-1$ vectors which span the hyperplane?

Linear systems. As we've seen, questions which ask whether some vector is a linear combination of others boil down to questions about systems of linear equations, since adding up the terms on the left of

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{b}$$

and equation the resulting components with those on the right results in a system of m equations (where m is the number of components of each vector involved, so that each vector is in $\mathbb{R}^m$) and the n variables $x_1, \ldots, x_n$. So, we had better understand how to solve linear systems in general.

An $m \times n$ *linear system* (or a *system of m linear equations in n variables*) is a collection of equations of the form

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + \cdots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

where all the a_{ij} and the b_i are scalars. (Notice that in this notation, the first index on a coefficient a_{ij} tells us which equation it comes from while the second index tells us to which variable it is attached.) There is no restriction on these scalars, meaning that having some of them be zero is allowed. A *solution* of this linear system is a vector in $\mathbb{R}^m$ with entries $x_1, \ldots, x_n$ which satisfy all equations simultaneously.

Elementary operations. The basic approach to finding all solutions of a linear system comes from performing certain operations which 1) result in simpler equations by "eliminating" variables, and 2) do not change the possible solutions. We will refer to the following operations as *elementary operations*:

- swap one equation for another,
- multiply an equation by a nonzero scalar, and
- add a multiple of one equation to another.

Exercise. Show that the first type of elementary operation above can be obtained via a sequence of operations of only the second and third types.

Theorem. The key fact is that elementary operations do not alter solutions of linear systems. To be precise, suppose that

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + \cdots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

is a linear system and consider the system obtained via one elementary operation. Then any solution of the original system is also a solution of the new system, and conversely any solution of the new system is in turn a solution of the original system.

Proof. We prove that the second type of operation—multiplying a row by a nonzero scalar—does not alter solutions. The analogous fact for the third type of operation is left to the homework, and the fact that row swaps do not alter solutions can either be seen as "obvious" or can be derived from the result in the Exercise above that row swaps can be described using the other two types of elementary operations.

Suppose that $(x_1', x_2', \ldots, x_n')$ is a solution of the original system, so that

$$a_{11}x_1' + \cdots + a_{1n}x_n' = b_1$$
$$a_{21}x_1' + \cdots + a_{2n}x_n' = b_2$$
$$\vdots$$
$$a_{m1}x_1' + \cdots + a_{mn}x_n' = b_m.$$

Note that we are making a distinction between a variable x_i and the value x_i' we are substituting into it. The system obtained by multiplying the i-th row by a nonzero scalar $r \neq 0$ has as its i-th equation:

$$ra_{i1}x_1 + \cdots + ra_{in}x_n = b_i.$$

Since this is the only equation in the new system which is different from the original system, the values $(x_1', \ldots, x_n')$ still satisfy all other equations in the new system.

To show that these values also satisfy the new i-th equation, we simply plug in each x_i' in place of each x_i and verify that the required equation is satisfied:

$$ra_{i1}x_1' + \cdots + ra_{in}x_n' = r(a_{i1}x_1' + \cdots + ra_{in}) = rb_i$$

where we know that $a_{i1}x_1' + \cdots + ra_{in} = b_i$ since we are assuming $(x_1', \ldots, x_n')$ is a solution of the original system. Thus $(x_1', \ldots, x_n')$ does satisfy the new i-th equation $ra_{i1}x_1 + \cdots + ra_{in}x_n = b_i$, so $(x_1', \ldots, x_n')$ is a solution of the new system.

Now we show that if $(x_1', \ldots, x_n')$ is a solution of the new system, then it is also a solution of the original system. We can do this using a similar algebraic approach as before where we take that $(x_1', \ldots, x_n')$ satisfies the equation

$$ra_{i1}x_1' + \cdots + ra_{in}x_n' = rb_i$$

and show then that it must satisfy $a_{i1}x_1' + \cdots + a_{in}x_n' = b_i$ as well, but to save some algebra we can argue as follows. Note that the old i-th equation $a_{i1}x_1 + \cdots + a_{in}x_n = b_i$ can be obtained by multiplying the new i-th equation $ra_{i1}x_1 + \cdots + ra_{in}x_n = rb_i$ by the nonzero scalar $\frac{1}{r}$, so the old system can be obtained from the new system by this elementary operation. But we just showed in the first part that if $(x_1', \ldots, x_n')$ is a solution of the new system, then it is also a solution of any system obtained by multiplying an equation of the new system by a nonzero scalar, so this says right away that $(x_1', \ldots, x_n')$ is indeed a solution of the old system. We conclude that the old system and the new one have exactly the same solutions. □

Exercise. Finish the proof of the theorem above, by showing that adding a multiple of one row to another does not change the possible solutions of a linear system.

Remark. The direction in the theorem above which shows that a solution of the new system gives a solution of the old system depended on the fact that the old system can be obtained from the new one still via elementary operations. This is also true for the other types of elementary operations, meaning that any elementary operation has an *inverse* operation which does the "opposite" of what the original operation does. For instance, the operation which swaps two equations is its own inverse, and the operation which adds r times row i to row j is the operation which adds $-r$ times row i to row j.

Lecture 7: Matrices and Row Reduction

Warm-Up. We show that the linear systems

$$\begin{aligned} ax + by + cz &= d \\ fx + gy + hz &= j \\ mx + ny + pz &= q \end{aligned}$$

and

$$\begin{aligned}(a-2f)x+\quad &(b-2g)y+\quad &(c-2h)z=\quad &d-2j\\ (2a+3f)x+\quad &(2b+3g)y+\quad &(2c+3h)z=\quad &2d+3j\\ (a+2f+m)x+&(b+2g+n)y+&(c+2h+p)z=&d+2j+q\end{aligned}$$

have the same solutions. To do so, all we need to show is that the second system is obtainable from the first via elementary operations.

Certainly, the new first equation is obtained via the single elementary operation which takes -2 times the second equation and adds it to the first. Now, to get the new second equation we need to take 2 times the first and add it to 3 times the second, but this is not an allowed *single* elementary operation since we are multiplying not only the equation which we are adding by a scalar but also the equation which is being added on to. However, the point is that we can view this as the result of two elementary operations: first multiply the second equation by 3, and then add 2 times the first to it.

Finally, note that adding the first equation plus 2 times the second equation to the third in the original system gives the new third equation—again this process does not consist of a single elementary operation, but is indeed something we can describe using a sequence of two elementary operations: add 2 times the second equation to the third to give as a new third equation:

$$(2f+m)x+(2g+n)y+(2h+p)z=2j+q,$$

and then add the original first equation to this in order to obtain the required third equation

$$(a+2f+m)x+(b+2g+n)y+(c+2h+p)z=d+2j+q.$$

Since the second system can be obtained from the first via sequences of (possibly more than one) elementary operation, it and the original system have the same solutions.

Remark. We will often combine two or more elementary operations in one step when solving a system of equations, but note that at times it will be important to distinguish between performing a single elementary operation vs a sequence of such operations.

Example. Here is an example we used to come up with some linear system we were interested in solving, with the point being to illustrate yet further how linear systems might arise in practice. Suppose we want to find all polynomials $p(x)$ of degree at most 3 satisfying the properties:

$$p(1)=2,\ p(2)=0,\ \text{and}\ \int_0^2 p(x)\,dx=4.$$

If we let $p(x)=ax^3+bx^2+cx+d$ for some unknown scalars $a,b,c,d\in\mathbb{R}$, these properties place the following restrictions on a,b,c,d:

$$a+b+c+d=2,\ 8a+4b+2c+d=0,\ \text{and}\ 4a+\frac{8}{3}b+2c+2d=4.$$

Thus finding the required polynomials boils down to finding the values of a,b,c,d satisfying the linear system

$$\begin{aligned}a+\ \ b+\ \ c+\ \ d&=2\\ 8a+4b+2c+\ \ d&=0\\ 4a+\tfrac{8}{3}b+2c+2d&=4.\end{aligned}$$

Actually, since we know that our elementary operations do not alter possible solutions, let's go ahead and clear fraction (since fractions are harder to do algebra with) in the third equation by multiplying the third equation by 3, and solve

$$\begin{aligned} a+\ \ b+\ \ c+\ \ d &= 2 \\ 8a+4b+2c+\ \ d &= 0 \\ 12a+8b+6c+6d &= 12 \end{aligned}$$

instead. We do this using the process of *Gauss-Jordan elimination*, performed on the *augmented matrix* of our system:

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 2 \\ 8 & 4 & 2 & 1 & 0 \\ 12 & 8 & 6 & 6 & 12 \end{array}\right],$$

which is the (in this case 3×5) matrix which encodes all coefficients of our system as well as the values to the right of the equals sign. The vertical lines before the final columns are just to separate visually the portion of the augmented matrix which involves coefficients of variables and the portion which involves values to the right of the equals sign.

Gauss-Jordan elimination. The goal of Gauss-Jordan elimination is to transform (or *row-reduce*) the given matrix into one in so-called *reduced row-echelon form* via *elementary row operations*, which are the elementary operations we saw earlier for equations only now performed on the rows of a matrix. The process of performing elementary row operations to a matrix is also called *row reduction*. As the "elimination" in the name of this process suggests, at each step we get a matrix which describes a linear system in which more and more variables have been eliminated.

We won't go into a description of Gauss-Jordan elimination here, but will instead defer to the book *and* to my previous lecture notes for Math 290, which you can find on canvas. I strongly urge to look at these, not only to clarify things in the book but also to see more examples worked out in detail. You should also check the book or my notes for a precise definition of what it means for a matrix to be in "reduced row-echelon form".

Question to think about. Are reduced row-echelon forms unique? That is, if a matrix is transformed into a matrix A in reduced row-echelon form by some elementary row operations, and if by going through a different set of row operations it is transformed into a matrix B which also happens to be in reduced row-echelon form, must A and B be the same?

Back to example. The Gauss-Jordan elimination process applied to the augmented matrix in our polynomial example yields the following matrices at various steps, where an arrow indicates the result of some row operations:

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 2 \\ 8 & 4 & 2 & 1 & 0 \\ 12 & 8 & 6 & 6 & 12 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 2 \\ 0 & -4 & -6 & -7 & -16 \\ 0 & -4 & -6 & -6 & -12 \end{array}\right] \to \left[\begin{array}{cccc|c} 4 & 0 & -2 & -3 & -8 \\ 0 & -4 & -6 & -7 & -16 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right]$$

$$\to \left[\begin{array}{cccc|c} 4 & 0 & -2 & 0 & 4 \\ 0 & -4 & -6 & 0 & 12 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & -1/2 & 0 & 1 \\ 0 & 1 & 3/2 & 0 & -3 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right].$$

The first nonzero entry of each row is called a *pivot* (note that the book also calls these "leading variables"), and the point of each step is to make all entires above and below a pivot equal to zero. The final matrix above is in reduced row-echelon form.

What's the point? We know that the linear system corresponding to the final matrix has the same solutions as the one corresponding to the original matrix since row operations do not change solution sets. The final reduced row-echelon form is the augmented matrix of:

$$\begin{aligned} a \ - \tfrac{1}{2}c \ &= 1 \\ b + \tfrac{3}{2}c \ &= -3 \\ d &= 4 \end{aligned},$$

which tells us after solving for a in the first equation and b in the second that any solution must look like

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 1 + \frac{1}{2}c \\ -3 - \frac{3}{2}c \\ c \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \\ 0 \\ 4 \end{bmatrix} + c \begin{bmatrix} 1/2 \\ -3/2 \\ 1 \\ 0 \end{bmatrix}.$$

Thus this, and hence the original, system has infinitely many solutions, since for each arbitrary value of c the above expression will give values of a, b, d which satisfy the given system of equations. The variable c is called a *free variable* since it does not correspond to a pivot, so when writing the general form of a solution we are expressing all *pivot* variables in terms of the non-pivot (or free) variables. The fact that there is only one free variable tells us that the set of solutions actually forms a line in $\mathbb{R}^4$, but we'll come back to this later.

To answer the original question, the polynomials $p(x)$ of degree at most 3 which satisfy

$$p(1) = 2, \ p(2) = 0, \text{ and } \int_0^2 p(x)\,dx = 4$$

are those of the form

$$p(x) = \left(1 + \frac{1}{2}c\right)x^3 + \left(-3 - \frac{3}{2}c\right)x^2 + cx + 4$$

where $c \in \mathbb{R}$ is any real number.

Other things that can happen. Consider the linear system with augmented matrix

$$\left[\begin{array}{ccccc|c} 1 & 2 & 1 & 4 & -1 & 5 \\ -2 & -4 & -2 & -3 & -3 & -5 \\ 3 & 6 & 5 & 10 & -4 & 14 \\ -1 & -2 & 1 & -2 & -4 & 2 \end{array}\right].$$

When going through the row-reduction process at some point you might end up with

$$\left[\begin{array}{ccccc|c} -2 & -4 & 0 & -10 & 1 & -11 \\ 0 & 0 & 2 & -2 & -1 & -1 \\ 0 & 0 & 0 & -4 & 4 & -8 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

This is not in reduced row-echelon form yet for various reasons (pivots aren't equal to 1, not all entries above a pivot are zero), but we can already tell that the corresponding linear system has no solution: the final row corresponds to the equation $0 = 1$, which can't possibly be true.

If there is a solution but there are no free variables, then we are in the scenario where there is exactly one solution. So now we can see the algebraic reason as to why any linear system must have zero, exactly one, or infinitely many solutions: either we have a row in the reduced-row echelon

form corresponding to an impossible equation, so there are no solutions, or we don't, in which case we have exactly one solution if there are no free variables and we have infinitely many if there are free variables. In particular, if we have at least two solutions then there *must* be a free variable, so that in fact we must infinitely many solutions.

All of this works just as well for linear systems involving *complex* coefficients, complex variables, and complex solutions. We'll see an example of this next time.

Lecture 8: More on Matrices and Linear Systems

Warm-Up 1. Note that I made a slight change to the version we did in class: here, the lower right entry is $-k-4$ instead of $-k+2$. This will give a scenario in which there are infinitely many solutions for some k.

We determine the values of k for which the system with augmented matrix

$$\left[\begin{array}{cccc} 2 & -1 & 2 & 1 \\ -4 & 2 & -4 & -2 \\ 4 & -k-2 & 7 & 4 \\ -6 & k^2+3 & k^2-4k-6 & -k-4 \end{array}\right]$$

has zero, one, or infinitely many solutions. Row reducing gives:

$$\left[\begin{array}{cccc} 2 & -1 & 2 & 1 \\ -4 & 2 & -4 & -2 \\ 4 & -k-2 & 7 & 4 \\ -6 & k^2+3 & k^2-4k-6 & -k-4 \end{array}\right] \to \left[\begin{array}{cccc} 2 & -1 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -k & 3 & 2 \\ 0 & k^2 & k^2-4k & -k-1 \end{array}\right]$$

$$\to \left[\begin{array}{cccc} 2 & -1 & 2 & 1 \\ 0 & -k & 3 & 2 \\ 0 & 0 & k^2-k & k-1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

To be clear, in the final step we multiplied the third row by k and added it to the fourth row, and then swapped some rows. This is not in reduced row-echelon form, but it is in a form which will allow us answer the question. The key point is that we can already determine where the pivots in the reduced row-echelon form would be.

In order to have no solutions, the third row must be of the form

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & \text{nonzero} \end{array}\right]$$

since this would correspond to an impossible equation. Such a row occurs when $k^2-k=0$ and $k-1 \neq 0$, so this happens for $k=0$. Thus for $k=0$ the system has no solutions. For $k=1$, the third row consists of all zeroes and the reduced matrix is thus:

$$\left[\begin{array}{cccc} 2 & -1 & 2 & 1 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

In this case the third column of the reduced row-echelon form will not contain a pivot, so it corresponds to a free variable. Since there is at least one solution—because there is no impossible equation—there are infinitely many solutions when $k=1$.

Finally we consider what happens for $k \neq 0, 1$. In this case there will be a pivot in each column corresponding to a variable, meaning that there are no free variables and hence there cannot be infinitely many solutions. Since there will be no impossible equation, there will be at least solution, so we conclude that there is exactly one. Another way to see this is as follows. The third row gives

$$(k^2 - k)z = k - 1,$$

which since $k^2 - k \neq 0$ can be solved to give a specific value of z. The second row gives

$$-ky + 3z = 2,$$

and since $k \neq 0$, plugging in the previously-obtained value for z will give a way to solve explicitly for a unique value of y. Then plugging this y value and the previous z value into the equation corresponding to the first row will give an equation which can be solved to obtain a unique value of x. Thus we get unique values for x, y, z, so there is only one solution.

Warm-Up 2. We solve the *complex* linear system:

$$\begin{aligned} 2z + 3iw &= 0 \\ (1+i)z - w &= -i. \end{aligned}$$

Reducing the augmented matrix gives:

$$\left[\begin{array}{cc|c} 2 & 3i & 0 \\ 1+i & -1 & -i \end{array}\right] \to \left[\begin{array}{cc|c} 2 & 3i & 0 \\ 0 & 2-3i & -2i \end{array}\right] \to \left[\begin{array}{cc|c} -4+6i & 0 & 6 \\ 0 & 2-3i & -2i \end{array}\right],$$

so the reduced row-echelon form is

$$\left[\begin{array}{cc|c} 1 & 0 & 6/(-4+6i) \\ 0 & 1 & -2i/(2-2i) \end{array}\right].$$

Thus the system has a unique solution:

$$z = \frac{6}{-4+6i}, \; w = -\frac{2i}{2-2i}.$$

The key takeaway is that this is no different than solving a system with real coefficients, except for the use of complex numbers. Rewriting the given system in vector form as

$$z\begin{bmatrix} 2 \\ 1+i \end{bmatrix} + w\begin{bmatrix} 3i \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -i \end{bmatrix},$$

we can interpret the solution we found as saying that those are the coefficients needed to express $\left[\begin{smallmatrix} 0 \\ -i \end{smallmatrix}\right]$ as a linear combination of $\left[\begin{smallmatrix} 2 \\ 1+i \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 3i \\ -1 \end{smallmatrix}\right]$.

Matrix equations. Next time we will prove that reduced row-echelon forms are *unique* in a certain sense. Before that, we take note of some basic facts.

A key observation is that any linear system can be written as a *matrix equation* of the form

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

where A is the *coefficient matrix* of the system encoding the coefficients of all variables, $\mathbf{x}$ is the unknown vector to-be-solved for encoding the variables, and $\mathbf{b}$ encodes the numbers to the right of

the equals signs. For an $m \times n$ system, A will be an $m \times n$ matrix, $\mathbf{x}$ will be a vector in $\mathbb{R}^n$, and $\mathbf{b}$ a vector in $\mathbb{R}^m$. The *product* $A\mathbf{x}$ of a matrix and a vector is defined precisely so that $A\mathbf{x} = \mathbf{b}$ encodes the linear system. Check the book or my Math 290 notes for details.

The upshot is that we know three ways of viewing the same type information: a system of linear equations, and linear combination vector equation, and a matrix equation. Each point of view will be useful going forward.

Row equivalence. We say that two matrices A and B (of the same size) are *row equivalent* if one can be obtained from the other using elementary row operations. Note that if B can be obtained from A using some row operations, then A can be obtained from B by performing the *inverse* of each of those row operations in the opposite order.

We'll talk more about row equivalence next time, but for now here is a key point: if A and B are row equivalent, then the equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solutions. Indeed, the augmented matrix of the first system is of the form

$$\big[A \mid \mathbf{0}\big]$$

where $\mathbf{0}$ denotes a column of zeroes, and performing the row operations which transform A into B on this augmented matrix will transform it into

$$\big[B \mid \mathbf{0}\big]$$

since row the column of zeroes at the beginning will remain zeroes throughout. (Note that if $\mathbf{b} \neq 0$, performing these row operations to the augmented matrix

$$\big[A \mid \mathbf{b}\big]$$

does *not* necessarily give $\big[B \mid \mathbf{b}\big]$ since in this case the final column $\mathbf{b}$ could very well end up being something different.) Since $\big[A \mid \mathbf{0}\big]$ can be transformed into $\big[B \mid \mathbf{0}\big]$ using row operations, the corresponding systems $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solutions as claimed. This fact will be important next time when proving the uniqueness of reduced row-echelon forms.

Question to think about. If $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solutions, must A and B be row equivalent?

Lecture 9: Reduced Echelon Form

Warm-Up. I made a slight change as compared to the version we did in class: instead of asking whether the system in question has a unique solution, we ask how many solutions it has.

We define the *rank* of a matrix to be the number of pivots in its reduced row-echelon form. Note that for this definition to make sense, we have to know that this number of pivots does not depend on the reduced echelon form we use, but this will follow from the fact we'll prove in a bit that reduced echelon forms are unique.

Suppose that A is a 3×4 matrix of rank 3. How many solutions does the equation $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ have? We show it has infinitely many. Consider the linear system with augmented matrix $\big[A \mid \mathbf{b}\big]$. The first observation is that since A has 4 columns and rank 3, the reduced echelon form of A has a column without a pivot, and so this corresponds to a free variable. Thus it is not possible for $A\mathbf{x} = \mathbf{b}$ to have exactly one solution. Now, since A only has 3 rows, each row of the reduced echelon form must contain a pivot, so an impossible equation of the form

$$0 = \text{ something nonzero}$$

cannot arise. (This would require the reduced echelon form of A to have a row of all zeroes, and this row would then not contain a pivot.) Thus $A\mathbf{x} = \mathbf{b}$ cannot have zero solutions, and hence it must have infinitely many solutions.

As a side note, observe that the reduced echelon form of a 3×4 matrix of rank 3 must be of the form:

$$\begin{bmatrix} 1 & 0 & 0 & ? \\ 0 & 1 & 0 & ? \\ 0 & 0 & 1 & ? \end{bmatrix}, \begin{bmatrix} 1 & 0 & ? & 0 \\ 0 & 1 & ? & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & ? & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{ or } \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where a question mark denotes a location which can have any possible value.

Theorem. (Uniqueness of Reduced Row-Echelon Form) Suppose that a matrix A is row equivalent to matrices B and C, both of which are in reduced row-echelon form. Then B and C must be the same. This says that when performing row operations on A to obtain a matrix in reduced row-echelon form, there is only one such reduced form with which we can end up. Thus, it makes sense to talk about *the* reduced row-echelon form of A, which is usually denoted by $\text{rref}(A)$.

Remark. Before giving the proof, we give a comment or two on the proof itself. This is the most difficult proof we have come across yet, and is the first proof which involves more than simply playing around with equations and definitions in that it requires a firm understanding of various concepts. This is not a proof you would be expected to memorize or be able to come up with on your own in a first course such as this—what *is* expected is that you can read through the proof and understand why it works. The proof illustrates and brings together many topics we've seen.

Proof of Theorem. (This proof is based on a proof due to John Alongi, which in turn is based on a proof due to Thomas Yuster. This proof basically just clarifies certain steps in their proofs.)

Since A is row equivalent to B, the equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have the same solutions, and since A is row equivalent to C, $A\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ have the same solutions. Thus $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ have the same solutions.

We proceed using induction on the numbers of columns of the matrices involved. For the base case, suppose that A (and hence B and C) are $m \times 1$ matrices. There are only two possible $m \times 1$ matrices in reduced row-echelon form: the one with a 1 as the first entry and zeroes elsewhere, and the one with all entries equal to 0:

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

so B and C must each be one of these. But the first matrix above has only one solution to

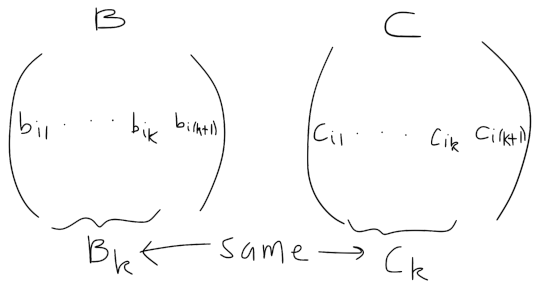
$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} x = \mathbf{0},$$

namely $x = 0$, while the second has any $x \in \mathbb{R}$ as a solution. Thus since $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ have the same solutions, it cannot be the case that B is one of the matrices above and C is the other, so they must each be the same one and hence are equal.

Now for the inductive hypothesis assume that the reduced row-echelon form of any $m \times k$ matrix, for some $k \geq 1$, is unique, meaning that there is only one matrix in reduced row-echelon form to

which an $m \times k$ matrix can be row equivalent. Suppose that A is an $m \times (k+1)$ matrix, so B and C are as well. Denote by A_k, B_k, C_k the $m \times k$ matrices obtained by removing the $(k+1)$-st columns of A, B, C respectively. The row operations which transform A into B will transform A_k into B_k, and those which transform A into C will transform A_k into C_k, so we get that A_k is row equivalent to both B_k and C_k. Also, since B and C are in reduced row-echelon form, B_k and C_k are as well since removing a column does not mess with any of the properties defining what it means for a matrix to be in reduced row-echelon form. Hence by in the inductive hypothesis we conclude that $B_k = C_k$.

Thus the first k columns of B and C are the same, so all that remains to be able to conclude that $B = C$ is to show that the final columns of B and C are the same. Suppose instead that $B \neq C$, so that the final columns of B and C are different. In particular, if we denote the entries of B by b_{ij}, where b_{ij} is the entry in the i-th row and j-th column, and we denote the entries of C similarly by c_{ij}, suppose that the i-th entries $b_{i(k+1)}$ and $c_{i(k+1)}$ in the final columns of B and C respectively are different. For a visual aid, here is what our matrices thus look like:



Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{k+1} \end{bmatrix} \in \mathbb{R}^{k+1}$$

be any solution to $B\mathbf{x} = \mathbf{0}$, and hence to $C\mathbf{x} = \mathbf{0}$ as well. The i-th entry of the equation $B\mathbf{x} = \mathbf{0}$ looks like

$$b_{i1}x_1 + \cdots + b_{ik}x_k + b_{i(k+1)}x_{k+1} = 0$$

and the i-th entry of $C\mathbf{x} = \mathbf{0}$ says

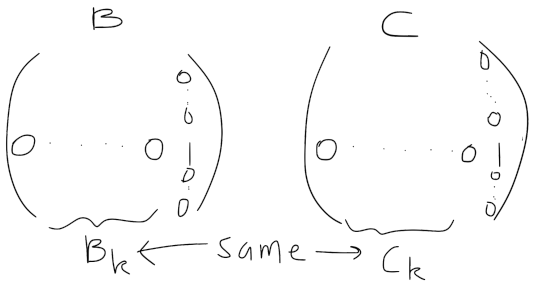
$$c_{i1}x_1 + \cdots + c_{ik}x_k + c_{i(k+1)}x_{k+1} = 0.$$

Since $B_k = C_k$ we have $b_{ij} = c_{ij}$ for all $j = 1, \ldots, k$, so the equations above imply that

$$b_{i(k+1)}x_{k+1} = c_{i(k+1)}x_{k+1}.$$

Since $b_{i(k+1)} \neq c_{i(k+1)}$, the only way this can be true is when $x_{k+1} = 0$. Thus any solution of $B\mathbf{x} = 0$ or $C\mathbf{x} = \mathbf{0}$ must have $x_{k+1} = 0$, which says that x_{k+1} cannot be a free variable since it can only have one specific value. Hence the columns to which x_{k+1} correspond must have pivots, so the final columns of B and C contain a pivot.

Since B and C are in reduced row-echelon form, every other entry in these finals columns must be zero, so the final columns of B and C have a 1 in one location and zeroes elsewhere. The rows in which the pivots 1 occur must be rows where B_k and C_k contain all zeroes since otherwise the 1 in the final column would not be the first nonzero entry in that row:



But then the pivots 1 must occur in the same row for B as in C since otherwise one of these matrices would have a row of all zeroes above a nonzero row, which is not possible if B and C are in reduced row-echelon form. We conclude that the final columns of B and C are in fact the same, so $B = C$, contradicting the assumption that $B \neq C$. Thus B and C must have been the same all along, so $B = C$ and the proof is finished. □

Corollary of the proof. Note that really the only key property we used in the proof is that is that $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ are supposed to have the same solutions. Thus, what we have actually shown is the following: if B and C are both in reduced row-echelon form and $B\mathbf{x} = \mathbf{0}$ and $C\mathbf{x} = \mathbf{0}$ have the same solutions, then $B = C$. This is really the main technical point to remember.

Lecture 10: Span and Linear Independence

Warm-Up 1. We show that if matrices A and B are row equivalent, then they have the same reduced row-echelon form. This is one direction of Problem 2 on Homework 3.

If A and B are row equivalent, there is a sequence of row operations $r_1, \dots, r_k$ which transform A into B:

$$A \to \text{ intermediate matrix } \to \cdots \to B.$$

There is also a sequence of row operations $s_1, \dots, s_\ell$ which transform B into its reduced row-echelon form $\operatorname{rref}(B)$:

$$B \to \cdots \to \operatorname{rref}(B).$$

Performing the operations $r_1, \dots, r_k$ and then the operations $s_1, \dots, s_\ell$ hence transform A into $\operatorname{rref}(B)$:

$$A \to \cdots \to B \to \cdots \to \operatorname{rref}(B).$$

Thus A is row equivalent to the matrix $\operatorname{rref}(B)$ in reduced row-echelon form, so by the uniqueness of reduced row-echelon forms we must have $\operatorname{rref}(A) = \operatorname{rref}(B)$ as claimed.

Warm-Up 2. We show that a system of linear equations $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\operatorname{rank}(A) = \operatorname{rank}(A \mid \mathbf{b})$, where $(A \mid \mathbf{b})$ denotes the augmented matrix corresponding to the system $A\mathbf{x} = \mathbf{b}$. (Recall that the rank of a matrix is the number of pivots in its reduced row-echelon form.)

The key point is that when row-reducing the augmented matrix $(A \mid \mathbf{b})$ to its reduced row-echelon form, the "A" portion reduces to the reduced row-echelon form of A:

$$(A \mid \mathbf{b}) \to (\operatorname{rref} A \mid ?)$$

where "?" denotes the vector obtained by performing to $\mathbf{b}$ the same operations which transform A into $\operatorname{rref} A$. This system will thus have a solution as long as there is no row corresponding to an equation of the form

$$0 = \text{nonzero}.$$

In order for this to occur, the "?" column would have to contain a pivot since its final entry would be the first nonzero entry in the bottom row of the reduced matrix (rref A |?). If this final column has a pivot, then there is one more pivot in this reduced augmented matrix than there is in rref A, so $\operatorname{rank}(A \mid \mathbf{b}) > \operatorname{rank}(A)$ in this case.

Thus, $A\mathbf{x} = \mathbf{b}$ has a solution if and only if the reduced row-echelon form of $(A \mid \mathbf{b})$ does not have a row consisting of all zeroes and then a 1, which happens if and only if the final column is does not have a pivot, which thus occurs if and only if the number of pivots in $\operatorname{rref}(A \mid \mathbf{b})$ is the same as the number of pivots in rref A, which is the same as saying that $\operatorname{rank}(A) = \operatorname{rank}(A \mid \mathbf{b})$.

Why is a plane a plane? Consider the equation $2x - y + 3z = 0$ in $\mathbb{R}^3$. We have repeatedly stated that such equations describes planes, but now we are the point at which we can ask: why does this describe a plane? In other words, what is a good definition of "plane"? In this case we can just have a computer sketch the set of all points satisfying $2x - y + 3z = 0$ and see visually that this appears to be a plane, but in math we strive for more precision that this.

The key observation is one which we have already seen: it is possible to find two vectors on this plane such that any other point on it can be written as a linear combination of those two vectors. Indeed, after solving for x, we see that any point on this plane must be of the form:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{1}{2}y - \frac{3}{2}z \\ y \\ z \end{bmatrix} = y \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -3/2 \\ 0 \\ 1 \end{bmatrix},$$

so any point on this plane is a linear combination of

$$\begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -3/2 \\ 0 \\ 1 \end{bmatrix}.$$

We say that these vectors *span* the plane in question.

Span. The *span* of some given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is the set of all linear combinations of those vectors. We denote this span by $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, so using set notation:

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k \mid a_1, \ldots, a_k \in \mathbb{R}\}.$$

Back to planes. So, letting V denote the plane $2x - y + 3z$, we could say

$$V = \operatorname{span}\left\{ \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3/2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Perhaps it is the fact that V is spanned by two vectors which guarantees it is a plane? Almost—there is one additional property these vectors have which guarantees their span will be a plane: *linear independence*.

To see why this is needed, consider a line in $\mathbb{R}^2$ such as $y + x = 0$. In this case, any point on the line can be written as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -x \end{bmatrix} = x \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

so the line is spanned by a single vector. However, it is also true that any point on this line can be written as

$$\begin{bmatrix} x \\ -x \end{bmatrix} = -x \begin{bmatrix} 1 \\ -1 \end{bmatrix} + x \begin{bmatrix} 2 \\ -2 \end{bmatrix},$$

so this line is *also* spanned by $\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}2\\-2\end{smallmatrix}\right]$. If we simply defined a plane as a space spanned by two vectors, then this line would qualify as a plan as well, which is absurd.

However, the point is that since the second vector $\left[\begin{smallmatrix}2\\-2\end{smallmatrix}\right]$ here is already a linear combination (i.e. multiple) of the first vector $\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$, any linear combination of these two can be expressed as a linear combination of the first alone:

$$a\begin{bmatrix}1\\-1\end{bmatrix} + b\begin{bmatrix}2\\-2\end{bmatrix} = (a+2b)\begin{bmatrix}1\\-1\end{bmatrix}.$$

Thus, what makes V a plane is that it is spanned by two vectors, neither of which is a multiple of the other. (Although, we'll see next time that this *still* doesn't give the most general notion of "plane".) The notion of linear independence makes this precise.

Linear dependence and independence. A collection of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ is *linearly dependent* if one can be written as a linear combination of the others. This collection is *linearly independent* if none is a linear combination of the others.

Remark about the book. Note that the book does not talk about linear dependence and independence until Chapter 3, whereas we're still essentially looking at material from Chapter 1. You can check the end of Section 3.2, specifically the "Bases and Linear Independence" subsection, to see the linear independence material, just keep in mind that we haven't yet spoken about all the concepts which show up there.

Removing dependent vectors does not change span. Here is a key property: if $\mathbf{u} \in \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, then

$$\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}\} = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}.$$

We actually saw this property previously in Lecture 5 on "Linear Combinations", only we didn't phrase it in terms of spans. Check there to see the proof.

The point is that when we have linearly dependent vectors, it will always be possible to remove one—namely one which is a linear combination of the others—without changing the overall span. This is what is happening in the line example above: $\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}2\\-2\end{smallmatrix}\right]$ are linearly dependent, so throwing either away does not change the span.

Example. The general solution of the system $A\mathbf{x} = \mathbf{0}$ where

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

is given by

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix} = \begin{bmatrix}x_1\\-2x_3+x_5\\x_3\\-4x_5\\x_5\end{bmatrix} = x_1\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix} + x_3\begin{bmatrix}0\\-2\\1\\0\\0\end{bmatrix} + x_5\begin{bmatrix}0\\1\\0\\-4\\1\end{bmatrix}.$$

This the set of solutions is spanned by

$$\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-2\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\-4\\1\end{bmatrix}.$$

The first vector here is not a linear combination of the other two since there is no way to obtain the 1 in the first entry using the zeroes in the first entries of the other two vectors, the second is not a linear combination of the other two because of the 1 as its third entry, and the third is not a linear combination of the others because of the 1 in its fifth entry. Thus these vectors are linearly independent.

Question to think about. The set of solutions of $A\mathbf{x} = \mathbf{0}$ above is spanned by three linearly independent vectors—is it possible that there are two *other* linearly independent vectors which still span the same set of solutions? We'll come back to this later; the answer is no, but this is a deep fact which will take some work to prove, and is related to problem of giving a precise meaning of the word *dimension*.

Theorem. Given some vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$, checking to see if they are linearly dependent or independent by checking one at-a-time whether each is a linear combination of the others is tedious and time-consuming. However, we don't need to do this, as we have the following way to check for linear dependence/independence in one swing:

> $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly dependent if and only if there are scalars $a_1, \dots, a_k \in \mathbb{R}$, at least one of which is zero, such that $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$. Equivalently, $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent if the only way in which $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$ can be true is to have all scalars $a_1, \dots, a_k$ equal zero.

This was actually on Homework 2, only we didn't phrase using the terms linear "dependence" or "independence". Check the solutions to Problem 3 for a proof.

The upshot is that to check if vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly dependent or independent, we simply solve $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$ for the scalars $a_1, \dots, a_k$ and see whether or not they all have to be zero.

Proposition. Our final fact for today says that we can extend a linearly independent list to a longer linearly independent list by throwing in a vector not in the span of the original vectors: if $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly independent and $\mathbf{u} \in \mathbb{R}^n$ is not in $\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, then $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}$ are linearly independent. This is essentially the "Example we didn't do in class" from Lecture 5, but let's give a proof here for completeness.

Proof. Consider the equation

$$a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k + a\mathbf{u} = \mathbf{0}.$$

We can rewrite this as

$$-a\mathbf{u} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k.$$

If $a \neq 0$, then dividing by a gives

$$\mathbf{u} = -\frac{a_1}{a}\mathbf{v}_1 - \cdots - \frac{a_k}{a}\mathbf{v}_k,$$

which would mean that $\mathbf{u}$ *is* a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$ and so is in their span; since we are assuming u is not in this span, we must have $a = 0$. But if $a = 0$ the equation we had previously for $a\mathbf{u}$ becomes

$$\mathbf{0} = a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k.$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent, the only way this equation can hold is to have $a_1 = \cdots = a_k = 0$. Thus the only way $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k + a\mathbf{u} = \mathbf{0}$ can be true is to have all coefficients $a_1 = \cdots = a_k = a = 0$, so $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}$ are linearly independent as claimed. □

Lecture 11: Solutions of Linear Systems

Warm-Up 1. We find the value(s) of k for which the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -k \\ 2 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 3 \\ -2k-2 \\ k+4 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} 4 \\ 2-4k \\ k+10 \end{bmatrix}$$

are linearly independent. Thus, we must determine when

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}$$

has only the trivial solution $c_1 = c_2 = c_3 = 0$. This vector equation is the same as the matrix equation

$$\begin{bmatrix} 1 & 3 & 4 \\ -k & -2k-2 & 2-4k \\ 2 & k+4 & k+10 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

and reducing the augmented matrix gives

$$\left[\begin{array}{ccc|c} 1 & 3 & 4 & 0 \\ -k & -2k-2 & 2-4k & 0 \\ 2 & k+4 & k+10 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 3 & 4 & 0 \\ 0 & k-2 & 2 & 0 \\ 0 & 0 & k & 0 \end{array}\right].$$

For $k = 0$ and $k = 2$, there are free variables, so there are infinitely many scalars satisfying

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}$$

in this case, and so $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly dependent for $k = 0, 2$. For $k \neq 0, 2$, there are pivots in each column so the only solution is $c_1 = c_2 = c_3 = 0$ and the given vectors are linearly independent in this case.

Warm-Up 2. We show that the complex vectors

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \begin{bmatrix} i \\ 1 \end{bmatrix}, \ \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

are linearly dependent over $\mathbb{C}$, by which we mean that one can be written as a complex linear combination of the others. Row reducing the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & i & 1 & 0 \\ 1 & 1 & -i & 0 \end{array}\right]$$

yields infinitely many solutions, which we know will be true without reducing given the fact that there are more columns than rows, so there will be at least one free variable. In particular, one possible solution gives

$$i\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} i \\ 1 \end{bmatrix} + (1-i)\begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and solving for one vector in terms of the rest expresses it as a complex linear combination of the others.

However, we claim now that the given vectors are in fact linearly independent *over* $\mathbb{R}$, which means that it is not possible to express any vector as a *real* linear combination of the others, where

"real linear combination" means a linear combination involving real coefficients. Equivalently, the only *real* scalars $c_1, c_2, c_3 \in \mathbb{R}$ satisfying

$$c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} i \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

are $c_1 = c_2 = c_3 = 0$. Indeed, comparing first entries on both sides gives the requirement that

$$c_1 + ic_2 + c_3 = 0,$$

and since c_1, c_3 are real this says that ic_2 must be real as well, which it is not unless $c_2 = 0$ since c_2 is real. Now that we know $c_2 = 0$, comparing second entries gives

$$c_1 - ic_3 = 0,$$

which gives $c_1 = ic_3$, so $c_2 = c_3 = 0$ since otherwise we would have something real equaling something not real. Since we must have $c_1 = c_2 = c_3 = 0$, we conclude that the given vectors are linearly independent over $\mathbb{R}$.

Remark. We'll see more later on about the distinction between being linearly independent over $\mathbb{C}$ vs over $\mathbb{R}$ when we talk about abstract *vector spaces*.

Planes redux. Based on last time, can try to define a "plane" to be a space which is spanned by two linearly independent vectors. However, consider the plane $x + y + z = 1$ in $\mathbb{R}^3$, or really any plane which doesn't pass through the origin. We ask: are there two vectors (linearly independent or not) which span this plane?

The answer is no. For one reason, the span of any vectors will always contain the zero vector, so if we did have that the plane $x+y+z = 1$ was equal to $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ then this plane would have to pass through the origin, which it doesn't. For another reason, note that any span has the property that adding two vectors in it results in another vector in the span. In particular, if $\mathbf{u} = a\mathbf{v}_1 + b\mathbf{v}_2$ and $\mathbf{w} = c\mathbf{v}_1 + d\mathbf{v}_2$ are both in $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, we have:

$$\mathbf{u} + \mathbf{w} = (a\mathbf{v}_1 + b\mathbf{v}_2) + (c\mathbf{v}_1 + d\mathbf{v}_2) = (a + c)\mathbf{v}_1 + (b + d)\mathbf{c}_2,$$

which is also a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ and so is in $\operatorname{span}\{\mathbf{v}_1\mathbf{v}_2\}$. However, adding two points on the plane $x + y + z = 1$ does not give another point on the plane: $(1, 0, 0)$ and $(0, 1, 0)$ are both on this plane but their sum $(1, 1, 0)$ is not.

Thus, a "space spanned by two linearly independent vectors" is not an appropriate characterization of a plane since it doesn't account for planes which don't pass through the origin. To fix this we must better understand the relation between equations of the form $A\mathbf{x} = \mathbf{0}$ and those of the form $A\mathbf{x} = \mathbf{b}$.

Homogeneous equations. A linear system of the form $A\mathbf{x} = \mathbf{0}$ is called a *homogeneous* system. There are two key properties of such systems: their set of solutions are "closed under linear combinations", and their set of solutions can be described as spans.

To say that the set of solutions is *closed under linear combinations* means that any linear combination of solutions is still a solution. Indeed, suppose that $\mathbf{x}_1, \dots, \mathbf{x}_k$ are solutions of $A\mathbf{x} = \mathbf{0}$ and that $c_1, \dots, c_k$ are any scalars. Then

$$A(c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k) = A(c_1\mathbf{x}_1) + \cdots + A(c_k\mathbf{x}_k) = c_1A\mathbf{x}_1 + \cdots + c_kA\mathbf{x}_k = c_1\mathbf{0} + \cdots + c_k\mathbf{0} = \mathbf{0},$$

so $c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k$ is also a solution of $A\mathbf{x} = \mathbf{0}$. As for the fact that the set of solutions can be described as a span, this follows from the usual way we solve linear systems using Gauss-Jordan elimination: after expressing the pivot variables in terms of the free variables and then "factoring out" the free variables, we're left with a linear combination expression which describes every possible solution.

Remark. Above we used the fact that the following distributive property holds:

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$$

where A is a matrix and $\mathbf{x}, \mathbf{y}$ are vectors. Of course such a property should hold if we expect matrix multiplication to behave like other types of multiplication we're used to, but it is important to realize why this works.

We could compute both sides of the above equation to verify that we get the same answer, but actually I claim we've already essentially done this. Denoting the columns of A by

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix},$$

recall that a product $A\mathbf{x}$ can be written as a linear combination of the columns of A using the entries of $\mathbf{x}$ as coefficients. Thus, denoting the entries of $\mathbf{x}$ by x_i and those of $\mathbf{y}$ by y_i, we have:

$$A(\mathbf{x} + \mathbf{y}) = (x_1 + y_1)\mathbf{v}_1 + \cdots + (x_n + y_n)\mathbf{v}_n$$

and

$$A\mathbf{x} + A\mathbf{y} = (x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n) + (y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n).$$

Thus the fact that these two expressions are equal comes down to the distributive property

$$(x_i + y_i)\mathbf{v}_i = x_i\mathbf{v}_i + y_i\mathbf{v}_i$$

of scalar multiplication.

Similarly, the fact we also used above that for a scalar c we have

$$A(c\mathbf{x}) = cA\mathbf{x}$$

reflects the property $(cx_i)\mathbf{v}_i = c(x_i\mathbf{v}_i)$ of scalar multiplication.

Inhomogeneous equations. A linear system of the form $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b} \neq \mathbf{0}$ is called an *inhomogeneous* system. In this case, the set of solutions it not closed under arbitrary linear combinations, it is only closed under affine combinations.

Indeed, suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are solutions of $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b} \neq \mathbf{0}$. First, if $c_1, \ldots, c_k$ are scalars such that $c_1 + \cdots + c_k = 1$, then

$$A(c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k) = c_1A\mathbf{x}_1 + \cdots + c_kA\mathbf{x}_k = c_1\mathbf{b} + \cdots + c_k\mathbf{b} = (c_1 + \cdots + c_k)\mathbf{b} = \mathbf{b},$$

so the affine combination $c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k$ is also a solution of $A\mathbf{x} = \mathbf{b}$. Conversely, if $c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k$ is also a solution, from above we get the requirement that

$$(c_1 + \cdots + c_k)\mathbf{b} = \mathbf{b},$$

which since $\mathbf{b} \neq 0$ requires that $c_1 + \cdots + c_k = 1$. This says that the only types of linear combination which still yields a solution are the affine ones.

Theorem. Finally we state the relation between solutions of homogeneous and inhomogeneous equations. Suppose that $\mathbf{x}_0$ is a solution of the inhomogeneous equation $A\mathbf{x} = \mathbf{b}$. Then *any* solution of this inhomogeneous equation is of the form

$$\mathbf{x}_0 + \mathbf{y}$$

where $\mathbf{y}$ is a solution of the corresponding homogeneous equation $A\mathbf{x} = \mathbf{0}$. In other words, translating the solutions of $A\mathbf{x} = \mathbf{0}$ by a particular solution of $A\mathbf{x} = \mathbf{b}$ yields all solutions of $A\mathbf{x} = \mathbf{b}$. We'll prove this result next time as our Warm-Up.

Lecture 12: Linear Transformations

Warm-Up. We prove the result we stated at the end of last time: if $\mathbf{x}_0$ is a solution of the inhomogeneous equation $A\mathbf{x} = \mathbf{b}$, then any solution of $A\mathbf{x} = \mathbf{b}$ is of the form

$$\mathbf{x}_0 + \mathbf{y}$$

where $\mathbf{y}$ is a solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. Note that any such expression *is* a solution of $A\mathbf{x} = \mathbf{b}$ since

$$A(\mathbf{x}_0 + \mathbf{y}) = A\mathbf{x}_0 + A\mathbf{y} = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

but the claim here is that *all* solutions of $A\mathbf{x} = \mathbf{b}$ arise in this way. (This is why this fact is almost but not quite what was proved in Problem 7 of Homework 3.)

Let $\mathbf{x}_1$ be any solution of $A\mathbf{x} = \mathbf{b}$. If we want to express this as

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{y},$$

note that there is only one possible thing which $\mathbf{y}$ could be: namely $\mathbf{y} = \mathbf{x}_1 - \mathbf{x}_0$. The only thing to check is that this $\mathbf{y}$ is in fact a solution of $A\mathbf{x} = \mathbf{0}$, which it is since

$$A(\mathbf{x}_1 - \mathbf{x}_0) = A\mathbf{x}_1 - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

(This *was* proved in Problem 7 of Homework 3.) Thus, to summarize,

$$\mathbf{x}_1 = \mathbf{x}_0 + (\mathbf{x}_1 - \mathbf{x}_0)$$

indeed expresses the arbitrary solution $\mathbf{x}_1$ of $A\mathbf{x} = \mathbf{b}$ as $\mathbf{x}_0$ plus a solution of $A\mathbf{x} = \mathbf{0}$.

The structure of solutions of linear systems. We've actually already seen the fact above in action, although we didn't explicitly point it out until now. For instance, consider again the plane $x + y + z = 1$. Points on this plane can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 - y - z \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \underbrace{y \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}}_{\text{point on } x+y+z=0},$$

where the first term is a specific point on the given plane, while the linear combination portion gives all possible points on the corresponding homogeneous equation $x + y + z = 0$. Thus we can now give a true definition of a "plane": a *plane* in $\mathbb{R}^n$ is a set of vectors which can be expressed as

$$\mathbf{x}_0 + c_1\mathbf{v}_1 + c_2\mathbf{v}_2$$

where $\mathbf{x}_0$ is some specific point and $\mathbf{v}_1, \mathbf{v}_2$ are linearly independent. In other words, we take the plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$ and then translate it by adding $\mathbf{x}_0$ to all points. We'll come back to this definition later when we talk about the more general notion of an *affine subspace* of $\mathbb{R}^n$.

Similarly, consider the linear system $A\mathbf{x} = \mathbf{b}$ with augmented matrix:

$$\left[\begin{array}{ccccc|c} 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

The solutions are given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_1 \\ -2x_3 - 3x_5 + 1 \\ x_3 \\ -4x_5 - 2 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -2 \\ 0 \end{bmatrix} + \underbrace{x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}}_{\text{solution of } A\mathbf{x}=\mathbf{0}},$$

where again we have a particular solution of $A\mathbf{x} = \mathbf{b}$ plus an arbitrary solution of $A\mathbf{x} = \mathbf{0}$.

So, to sum up pretty much everything we've done these first few weeks, here are the key takeaways:

1. Solutions of an inhomogeneous system $A\mathbf{x} = \mathbf{b}$ are given by adding to a particular solution arbitrary solutions of the corresponding homogeneous system $A\mathbf{x} = \mathbf{0}$,
2. Given a homogeneous system $A\mathbf{x} = \mathbf{0}$, linearly independent vectors can be found which span the set of all solutions.

If you look back through examples and homework problems from the first few weeks, you'll no doubt notice that many of them were actually just special cases of these key facts.

Linear transformations. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if it has the following properties:

- $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and
- $T(c\mathbf{x}) = cT(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

We refer to the first property as "preservation of addition" and the second as "preservation of scalar multiplication". The main examples are so-called *matrix transformations*, which are those functions given by a formula of the form

$$T(\mathbf{x}) = A\mathbf{x}$$

where A is an $m \times n$ matrix. In fact, the most basic fact about linear transformations is that they are *always* induced by a matrix in this way, so that every linear transformation is in fact a matrix transformation. We'll prove this next time.

Remark. At least for now, much of the material we'll see regarding linear transformations is the same as that in the book or as what is covered in Math 290. For this reason, I'll won't include everything we talk about in class in these notes, but will instead suggest you check my old Math 290 notes (and the book) for completeness. In particular, the "rotation by 90°" example we looked

at in class can be found in my Math 290 notes, as well as some other examples we looked at of transformations which are/aren't linear.

In addition, and as I also mention in my Math 290 notes, note that the book's definition of "linear transformation" is different from ours: the book *defines* a linear transformation to be one which is induced by a matrix, and then later shows that this is equivalent to the requirements that $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ and $T(c\mathbf{x}) = cT(\mathbf{x})$. The approach we're taking is much more common, and is the only definition which also makes sense in the *infinite-dimensional* setting, which we'll talk about soon enough.

Affine transformations. Here is one thing not covered in the book nor in my Math 290 notes. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = x + 1.$$

This function is *not* linear according to our definition since

$$f(x+y) = (x+y)+1 \text{ and } f(x)+f(y) = (x+1)+(y+1) = x+y+2$$

are not the same. This might seem strange: the graph of f is the line $y = x + 1$, which we do normally think of as being a "linear" thing, and yet here we're saying that f is not linear.

If f is not a linear transformation, then what is it? The answer is that such a map is an *affine transformation*, which (using SAT analogies) are to linear transformations what affine combinations are to linear combinations. I'll save the definition of "affine transformation" for the homework, where you'll be asked to show that they are all of form "something linear plus a constant".

Lecture 13: More on Linear Transformations

Induction is everywhere. Let's point out a minor point which is important to realize nonetheless. Last time we mentioned that linear transformations "preserve linear combinations", in the sense that

$$T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + \cdots + c_kT(\mathbf{v}_k),$$

which comes from first breaking up the sum as

$$T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = T(c_1\mathbf{v}_1) + \cdots + T(c_k\mathbf{v}_k)$$

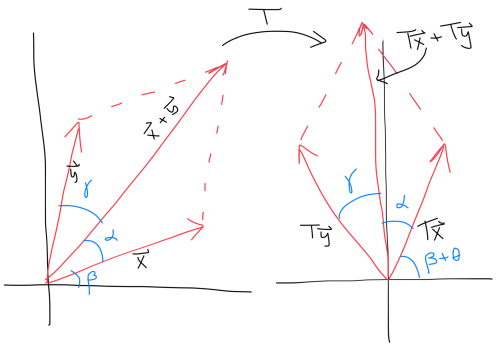
and then using the fact that T preserves scalar multiplication to pull each scalar out.

However, technically this above "breaking up the sum" property is not quite what we gave as the "preservation of addition" property of linear transformations, since in the definition we only required that this could be done for a sum of *two* things and not k things in general. But the point is that requiring this property for only a sum of two things *does* imply it holds for any number of things as a result of induction—namely the type of induction where we use the base case in the induction step. This is all hidden in the "$\cdots$" showing up in the above expressions, and is not something we'll go through the trouble of proving every single time. The upshot is that whenever we do something using "$\cdots$", keep in mind that there might be some induction hidden in there.

Warm-Up 1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation which rotates a vector by an angle θ around the origin. (We are using the word "transformation" here just to mean "function", so that "transformation" by itself does not necessarily mean "linear transformation".) We show that T is

linear. (Saying that T is "linear" is just a shorthand way of saying that it is a linear transformation.) We do this purely using some geometry, although there are algebraic ways of seeing this as well.

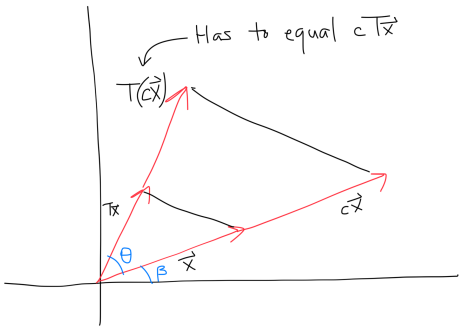
Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}$, we must first show that $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$. Let's be clear about what this is actually saying: on the left we first add $\mathbf{x}$ and $\mathbf{y}$ and then rotate the resulting vector by θ around the origin, whereas on the left we first rotate each of $\mathbf{x}$ and $\mathbf{y}$ separately and then add the results—the claim is that we end up with the same result either way. We use the following picture as guides, with various angles drawn in:



The claim is that rotating the vector $\mathbf{x}+\mathbf{y}$ drawn on the left by θ results in the vector $T\mathbf{x}+T\mathbf{y}$ drawn on the right. To justify this, we only need to justify that the parallelogram on the left produces the parallelogram on the right after rotation. All we need are the following properties of rotations: they "preserve" the angle between two vectors and they "preserve" length, meaning that the angle between two vectors is the same after rotation as it is before, and that the length of a vector is the same after rotation as it is before. (Later we'll use these ideas to give a precise definition of "rotation".)

Consider the triangles with vertices $\mathbf{0}$, (the endpoint of) $\mathbf{x}$, and (the endpoint of)$\mathbf{y}$ in the first picture and $\mathbf{0}$, $T\mathbf{x}$, $T\mathbf{y}$ in the second. These triangles are congruent since two of their sides have the same length ($T\mathbf{x}$ has the same length as $\mathbf{x}$ and $T\mathbf{y}$ has the same length as $\mathbf{y}$) and the included angles have the same measure $\gamma + \alpha$. But in the first picture, the triangle with vertices $\mathbf{0}, \mathbf{x}, \mathbf{y}$ is congruent to the one with verticies $\mathbf{x}, \mathbf{y}, \mathbf{x}+\mathbf{y}$ since these two triangles split the given parallelogram in half. Similarly, in the second picture the triangle with vertices $\mathbf{0}, T\mathbf{x}, T\mathbf{y}$ is congruent to the one with vertices $T\mathbf{x}, T\mathbf{y}, T\mathbf{x}+T\mathbf{y}$, so we conclude that all four triangles thus mentioned are congruent. This implies that the two parallelograms are congruent, so the one on the right is indeed obtained by rotating the one on the left. Hence the diagonal $T\mathbf{x}+T\mathbf{y}$ of the second parallelogram is $T(\mathbf{x}+\mathbf{y})$, so T preserves addition as claimed.

Next we must show that $T(c\mathbf{x}) = cT(\mathbf{x})$ for any scalar $c \in \mathbb{R}$. Again, to be clear, the left side is the result of first scaling $\mathbf{x}$ by c and then rotating the result, while the right side is the result of first rotating $\mathbf{x}$ and then scaling the result by c. Consider the picture:



Take the triangle with vertices $\mathbf{0}, \mathbf{x}, T\mathbf{x}$ and the one with vertices $\mathbf{0}, c\mathbf{x}, T(c\mathbf{x})$. These triangles share a common angle θ and are each isosceles since $\mathbf{x}, T\mathbf{x}$ have the same length and $c\mathbf{x}, T(c\mathbf{x})$ have the same length. Hence they are similar, so since one of edge of the larger triangle is obtained by scaling $\mathbf{x}$ by c, the other edge must also be obtained by scaling $T\mathbf{x}$ by c, so $cT\mathbf{x} = T(c\mathbf{x})$. Hence T preserves scalar multiplication, so it is linear.

Warm-Up 2. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. We show that $T(\mathbf{0}) = \mathbf{0}$, so that any linear transformation must send the zero vector to the zero vector. If we make use of the fact we'll prove next that linear transformations are always induced by matrices, this is just the statement that $A\mathbf{0} = \mathbf{0}$ for any matrix A. However, we want an argument works without making use of this fact, since, as mentioned last time, at some point we'll run into settings where it is not true that linear transformations are always matrix transformations.

Here are two approaches. First, since $0 \cdot \mathbf{0} = \mathbf{0}$ (scalar zero times zero vector equals zero vector) and T preserves scalar multiplication, we have

$$T(\mathbf{0}) = T(0 \cdot \mathbf{0}) = 0 \cdot T(\mathbf{0}) = \mathbf{0}$$

since $0\mathbf{v} = \mathbf{0}$ for any vector $\mathbf{v}$. Second, since $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and T preserves addition, we have

$$T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0}).$$

Whatever $T(\mathbf{0})$ is, it has an additive inverse, and adding this inverse (i.e. subtracting $T(\mathbf{0})$) to both sides gives

$$\mathbf{0} = T(\mathbf{0})$$

as claimed. Since we only used the linearity properties of T and not the fact that it is given by a matrix, these same proofs will work in the more general settings we'll come too.

Theorem. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then the fact is that there exists an $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, so that every linear transformation is a matrix transformation. The proof makes use of some key ideas we've seen previously: writing a vector as a linear combination of other vectors, and recognizing that linear combination expresses can be written in terms of matrix products.

Proof. Denote the entries of $\mathbf{x} \in \mathbb{R}^n$ by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Since T is linear, we have:

$$\begin{aligned} T(\mathbf{x}) &= T\left(x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right) \\ &= x_1 T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 T \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n T \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & T \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} & \cdots & T \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & T \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} & \cdots & T \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \end{bmatrix} \mathbf{x} \end{aligned}$$

where in the third line we have used the fact that a product $A\mathbf{x}$ where A is a matrix can be written as a linear combination of the columns of A with coefficients the entries of $\mathbf{x}$. Thus for the $m \times n$ matrix

$$A = \begin{bmatrix} T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & T \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} & \cdots & T \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \end{bmatrix}$$

we indeed have $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, so T is a matrix transformation. □

Remark. The matrix derived above is called the *standard matrix* of T. The vectors consisting of 1's and 0's used above are important enough that we'll give them special names: $\mathbf{e}_i \in \mathbb{R}^n$ denotes the vector with a 1 in the i-th position and zeroes elsewhere, so the standard matrix of T has columns $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$. Thus, to find the standard matrix of any linear transformation (i.e. the matrix so that multiplying by that matrix has the same affect as applying the transformation), we simply apply T to the special vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and use the results as columns.

Rotations. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation which rotates $\mathbb{R}^2$ by an angle θ about the origin. Since this is linear (as we saw in the Warm-Up), it must be induced by some matrix A. Rotating $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ by θ gives $\left[\begin{smallmatrix} \cos\theta \\ \sin\theta \end{smallmatrix}\right]$ while rotating $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ gives $\left[\begin{smallmatrix} -\sin\theta \\ \cos\theta \end{smallmatrix}\right]$, so

$$T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad \text{and} \quad T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

(Check the book or my Math 290 lecture notes to see why these are correct.) Thus the standard matrix of T is

$$A = \begin{bmatrix} T\begin{bmatrix}1\\0\end{bmatrix} & T\begin{bmatrix}0\\1\end{bmatrix} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

meaning that multiplying a vector by this matrix has the same effect as rotating it by θ around the origin.

Other geometric things. There are many other geometric transformations which can be described in terms of matrices. In particular, our book goes through a description of orthogonal projections, reflections, scalings, and shears. We'll make use of these from time to time, and I'll ask that you read Section 2.2 in the book and my old 290 notes to understand what these transformations are.

Example. Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation such that

$$T\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2\\-1\end{bmatrix} \quad \text{and} \quad T\begin{bmatrix}0\\3\end{bmatrix} = \begin{bmatrix}-1\\3\end{bmatrix}.$$

From this information alone we determine the standard matrix of T.

For instance, by the linearity of T we have

$$\begin{bmatrix}-1\\3\end{bmatrix} = T\begin{bmatrix}0\\3\end{bmatrix} = 3T\begin{bmatrix}0\\1\end{bmatrix},$$

so

$$T\begin{bmatrix}0\\1\end{bmatrix} = \frac{1}{3}\begin{bmatrix}-1\\3\end{bmatrix} = \begin{bmatrix}-1/3\\1\end{bmatrix}.$$

Then since

$$\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}0\\1\end{bmatrix},$$

the linearity of T gives

$$T\begin{bmatrix}1\\0\end{bmatrix} = T\begin{bmatrix}1\\1\end{bmatrix} - T\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}2\\-1\end{bmatrix} - \begin{bmatrix}-1/3\\1\end{bmatrix} = \begin{bmatrix}7/3\\-2\end{bmatrix}.$$

The values of $T\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $T\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ give the columns of the standard matrix of T, so the standard matrix is

$$\begin{bmatrix}7/3 & -1/3\\-2 & 1\end{bmatrix}.$$

Thus T is defined explicitly by

$$T(\mathbf{x}) = \begin{bmatrix}7/3 & -1/3\\-2 & 1\end{bmatrix}\mathbf{x}.$$

You can check that this indeed satisfies $T\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] = \left[\begin{smallmatrix}2\\-1\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\3\end{smallmatrix}\right] = \left[\begin{smallmatrix}-1\\3\end{smallmatrix}\right]$.

Remark. The point of the example above is the following: knowing only two pieces of information determined T completely. In general, for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, knowing n pieces of information—namely the values of $T(\mathbf{e}_1), \dots, T(\mathbf{e}_n)$—is enough to fully determine the behavior of T. This is what is special about linear transformations, and we'll see how to put this fact to great use.

Linear over $\mathbb{C}$ vs over $\mathbb{R}$. The definition of a linear transformation works just as well when considering spaces consisting of *complex* vectors: a function $T : \mathbb{C}^n \to \mathbb{C}^m$ is a (complex) linear transformation if

$$T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) \quad \text{and} \quad T(a\mathbf{x}) = aT(\mathbf{x}) \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n \text{ and } a \in \mathbb{C}.$$

Here, $\mathbb{C}^n$ denotes the space of complex vectors with n complex entries.

For instance, the function $T : \mathbb{C}^2 \to \mathbb{C}^2$ defined by

$$T\begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} -iz + w \\ 2z + 3iw \end{bmatrix} = \begin{bmatrix} -i & 1 \\ 2 & -3i \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix}$$

is linear. (Indeed, this is a complex matrix transformation, and is linear for the same reason real matrix transformations are.) In this setting it is also true that any complex linear transformation $\mathbb{C}^n \to \mathbb{C}^m$ is induced by an $m \times n$ complex matrix.

However, consider the function $f : \mathbb{C} \to \mathbb{C}$ defined by

$$f(a+ib) = b \text{ where } a, b \in \mathbb{R},$$

so f is the function which outputs the imaginary part of a complex number. This satisfies $T(z+w) = T(z) + T(w)$, so it preserves addition. But note that

$$f(i(a+ib)) = f(-b+ia) = a \text{ whereas } if(a+ib) = ib,$$

so $f(iz) \neq if(z)$. Thus f does not preserve complex scalar multiplication, so it is not complex linear. However, if we restrict the types of scalars we consider to only real numbers, we do indeed have

$$f(c(a+ib)) = f(ca+icb) = cb = cf(a+ib) \text{ for } c \in \mathbb{R}.$$

Hence the "preservation of scalar multiplication" property *does* hold for real scalars, even though it does not hold for arbitrary complex scalars. Because of this we, we would say that f is *real* linear but not complex linear, or that it is linear over $\mathbb{R}$ but not linear over $\mathbb{C}$.

We'll see more on the distinction between linear over $\mathbb{R}$ vs over $\mathbb{C}$ later when we discuss linear transformations in a more general setting.

Lecture 14: Yet More on Linear Transformations

Today we spent 15 minutes going over Quiz 3, then 20 or so minutes on the Warm-Up, and finally a little less than 10 minutes on compositions of linear transformations. Here I'll only include the Warm-Up in all its glory and save the material on compositions for Friday's lecture notes, where it fits in better.

The point of the Warm-Up, and hence of this lecture, is to further delve into the idea that linear transformations are completely determined by a finite amount of data, only now we'll amend this to say a finite amount of "linearly independent" data. This Warm-Up is WAY too long for an exam, and probably too difficult as well, but I strongly suggest you go through and understand it well since it really illustrates some important concepts.

Warm-Up. For $k = 0, 1, 2$, we determine how many linear transformations $T : \mathbb{R}^3 \to \mathbb{R}^2$ there are which satisfy the following properties:

$$T\begin{bmatrix} k \\ k^2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \; T\begin{bmatrix} 2k \\ 4k \\ k \end{bmatrix} = \begin{bmatrix} k \\ -k \end{bmatrix}, \; T\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

First, consider $k = 2$. Then the requirements are:

$$T\begin{bmatrix}2\\4\\1\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix},\ T\begin{bmatrix}4\\8\\2\end{bmatrix} = \begin{bmatrix}2\\-2\end{bmatrix},\ T\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}.$$

However, since T is supposed to be linear, we should have:

$$T\begin{bmatrix}4\\8\\2\end{bmatrix} = T\left(2\begin{bmatrix}2\\4\\1\end{bmatrix}\right) = 2T\begin{bmatrix}2\\4\\1\end{bmatrix} = 2\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}2\\4\end{bmatrix},$$

which does not agree with the second requirement. Thus we conclude that there are no such linear transformations when $k = 2$.

$\mathbf{k = 1}$ case. Now consider $k = 1$. The requirements become:

$$T\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix},\ T\begin{bmatrix}2\\4\\1\end{bmatrix} = \begin{bmatrix}1\\-1\end{bmatrix},\ T\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}.$$

So far there is no immediate way to see why such a T could not exist, as we had in the $k = 2$ case. To find such a T, we find its standard matrix by computing

$$T\begin{bmatrix}1\\0\\0\end{bmatrix},\ T\begin{bmatrix}0\\1\\0\end{bmatrix},\ \text{and } T\begin{bmatrix}0\\0\\1\end{bmatrix}.$$

To compute each of these, we try to write each $\mathbf{e}_i$ as a linear combination of the three vectors on which we know how T behaves:

$$\mathbf{e}_i = c_1\begin{bmatrix}1\\1\\1\end{bmatrix} + c_2\begin{bmatrix}2\\4\\1\end{bmatrix} + c_3\begin{bmatrix}1\\2\\3\end{bmatrix}.$$

The point is that, if we can do this, the linearity of T will give

$$T(\mathbf{e}_i) = c_1T\begin{bmatrix}1\\1\\1\end{bmatrix} + c_2T\begin{bmatrix}2\\4\\1\end{bmatrix} + c_3T\begin{bmatrix}1\\2\\3\end{bmatrix},$$

which we will be able to compute since we are given the three outputs needed here.

First we solve

$$\begin{bmatrix}1\\0\\0\end{bmatrix} = c_1\begin{bmatrix}1\\1\\1\end{bmatrix} + c_2\begin{bmatrix}2\\4\\1\end{bmatrix} + c_3\begin{bmatrix}1\\2\\3\end{bmatrix}$$

for c_1, c_2, c_3. Reducing the corresponding augmented matrix gives

$$\left[\begin{array}{ccc|c}1&2&1&1\\1&4&2&0\\1&1&3&0\end{array}\right] \to \left[\begin{array}{ccc|c}1&0&0&2\\0&1&0&-1/5\\0&0&1&-3/5\end{array}\right],$$

so $c_1 = 2, c_2 = -\frac{1}{5}, c_3 = -\frac{3}{5}$. Thus

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} - \frac{3}{5} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

so

$$\begin{aligned} T \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} &= 2T \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} T \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} - \frac{3}{5} T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\ &= 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \frac{3}{5} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 6/5 \\ 18/5 \end{bmatrix}, \end{aligned}$$

which gives the first column of the standard matrix of T.

Note that in the course of solving the required linear system above we saw that

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \\ 1 & 1 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which implies that the vectors

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

are linearly independent. Since these are three linearly independent vectors in $\mathbb{R}^3$, they must span $\mathbb{R}^3$, which is how we know that not only $\mathbf{e}_1$ but $\mathbf{e}_2$ and $\mathbf{e}_3$ as well are expressible as linear combinations of them. We'll use this in a bit to explicitly compute $T(\mathbf{e}_2)$ and $T(\mathbf{e}_3)$ in a manner similar to that above, but let us point the following out now: we now know that there will be a linear transformation T satisfying the required properties, and moreover that there is only one! Indeed, we already that $T(\mathbf{e}_1)$ could only have one possible value, and similarly $T(\mathbf{e}_2)$ and $T(\mathbf{e}_3)$ have only one possible value each, so the standard matrix of T will be completely determined.

The main point. The overarching point here is that since the given vectors are linearly independent, there can be no trouble arising as in the $k = 2$ case since the only possible such trouble comes from the fact that linear transformations should satisfy

$$T(c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k) = c_1T(\mathbf{x}_1) + \cdots + c_kT(\mathbf{x}_k),$$

which gives restrictions whenever some input vector is a linear combination of others. Since that is not the case in the $k = 1$ case, there are no restrictions beyond the three stipulated in the problem itself.

This is what we mean by saying that T is completely determined by a finite amount of "linearly independent" data—in this case three pieces of linearly independent data. What will happen in the $k = 0$ case, which we'll look at soon, is that the given requirements will actually only give *two* pieces of linearly independent data, meaning that there will be such a linear transformation but it won't be uniquely determined, so there will be infinitely many possible such transformations.

Back to the k = 1 case. To compute $T(\mathbf{e}_2)$, we reduce:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 1 & 4 & 2 & 1 \\ 1 & 1 & 3 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 2/5 \\ 0 & 0 & 1 & 1/5 \end{array}\right],$$

which gives

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{5} \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} + \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Thus

$$\begin{aligned} T \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} &= -T \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{5} T \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} + \frac{1}{5} T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\ &= - \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{2}{5} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \frac{1}{5} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -2/5 \\ -11/5 \end{bmatrix}, \end{aligned}$$

which gives the second column of the standard matrix of T. Finally, to compute $T(\mathbf{e}_3)$ we reduce:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 1 & 4 & 2 & 0 \\ 1 & 1 & 3 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1/5 \\ 0 & 0 & 1 & 2/5 \end{array}\right],$$

so

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} + \frac{2}{5} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix},$$

and thus

$$\begin{aligned} T \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} &= 0T \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} T \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} + \frac{2}{5} T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\ &= -\frac{1}{5} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \frac{2}{5} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1/5 \\ 3/5 \end{bmatrix} \end{aligned}$$

is the third column of the standard matrix of T.

We conclude that when $k = 1$, there is only one linear transformation satisfying the given requirements that it is explicitly given by

$$T(\mathbf{x}) = \begin{bmatrix} 6/5 & -2/5 & 1/5 \\ 18/5 & -11/5 & 3/5 \end{bmatrix} \mathbf{x} \text{ for } \mathbf{x} \in \mathbb{R}^3.$$

As a sanity check, go ahead and verify that this matrix indeed satisfies

$$\begin{bmatrix} 6/5 & -2/5 & 1/5 \\ 18/5 & -11/5 & 3/5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 6/5 & -2/5 & 1/5 \\ 18/5 & -11/5 & 3/5 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

and

$$\begin{bmatrix} 6/5 & -2/5 & 1/5 \\ 18/5 & -11/5 & 3/5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

as required in the $k = 1$ case.

$\mathbf{k = 0}$ case. Finally we consider $k = 0$. The requirements become

$$T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \; T \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \; T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

There are no impossibilities arising from these conditions; in particular, the second condition is satisfied by any linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$. However, the three given inputs are linearly dependent since the second input can be written as a linear combination of the other two. Indeed, the fact that

$$T \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

apart from being satisfied by any linear T, can also be derived from either of the other two conditions by writing

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

and using the linearity of T. Thus we only have two "linearly independent" pieces of data:

$$T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We claim that in this case there are infinitely many linear transformations satisfying these given requirements. Indeed, the given vectors cannot span all of $\mathbb{R}^3$, so there are vectors in $\mathbb{R}^3$ which are in the span of these two; denote such a vector by $\mathbf{b} \in \mathbb{R}^3$. Since the two given inputs are linearly independent, we will then have that

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \; \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \; \mathbf{b}$$

are linearly independent. Since these are three linearly independent vectors in $\mathbb{R}^3$, they must actually span $\mathbb{R}^3$ so every vector, in particular each $\mathbf{e}_i$, can be written as a linear combination of each. Writing each $\mathbf{e}_i$ as

$$\mathbf{e}_i = c_1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + c_3 \mathbf{b}$$

will then allow us to compute $T(\mathbf{e}_i)$ using

$$T(\mathbf{e}_i) = c_1 T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + c_2 T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + c_3 T(\mathbf{b}).$$

(In particular note that $\mathbf{e}_3$ is already the first of the given input vectors, so only really need to compute $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$.)

However, we have no information given as to what $T(\mathbf{b})$ could be, and since $\mathbf{b}$ is linearly independent from the given input vectors, there will be no restrictions on $T(\mathbf{b})$ whatsoever! Each possible value of $T(\mathbf{b})$ will give a different value for $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$, which is why we will have infinitely many linear transformations satisfying the given conditions in this case.

To take a concrete example, take $\mathbf{b}$ to be the vector

$$\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

You can check that

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

are indeed linearly independent. Solving

$$\mathbf{e}_2 = c_1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

for c_1, c_2, c_3 will give:

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = -\frac{3}{2} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

Thus, recalling that there are no restrictions on what $T(\mathbf{b})$ can be, if

$$T \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix},$$

then

$$\begin{aligned} T \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} &= -\frac{3}{2} T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \frac{1}{2} T \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{1}{2} T \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ &= -\frac{3}{2} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} a \\ b \end{bmatrix} \\ &= \begin{bmatrix} (-2-a)/2 \\ (-5-b)/2 \end{bmatrix}. \end{aligned}$$

Thus any matrix of the form

$$\begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & T(\mathbf{e}_3) \end{bmatrix} = \begin{bmatrix} a & (-2-a)/2 & 1 \\ b & (-5-b)/2 & 2 \end{bmatrix}$$

will induce a transformation satisfying the given requirements, so there are infinitely many such linear transformations in the $k = 0$ case.

Lecture 15: Products

Warm-Up 1. Given linear transformations $S, T : \mathbb{R}^n \to \mathbb{R}^m$, we define their *sum* to be the function $S + T : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$(S + T)\mathbf{x} = S\mathbf{x} + T\mathbf{x} \text{ for } \mathbf{x} \in \mathbb{R}^n.$$

In other words, $S + T$ is the function which adds together the outputs of S and T. We show that $S + T$ is linear.

First, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have:

$$\begin{aligned}(S + T)(\mathbf{x} + \mathbf{y}) &= S(\mathbf{x} + \mathbf{y}) + T(\mathbf{x} + \mathbf{y}) \\ &= (S\mathbf{x} + S\mathbf{y}) + (T\mathbf{x} + T\mathbf{y}) \\ &= (S\mathbf{x} + T\mathbf{x}) + (S\mathbf{y} + T\mathbf{y}) \\ &= (S + T)(\mathbf{x}) + (S + T)(\mathbf{y})\end{aligned}$$

where the second line follows from the fact that S and T preserve addition. Hence $S + T$ preserves addition.

Second, for any $\mathbf{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$, we have:

$$\begin{aligned}(S + T)(a\mathbf{x}) &= S(a\mathbf{x}) + T(a\mathbf{x}) \\ &= aS\mathbf{x} + aT\mathbf{x} \\ &= a(S\mathbf{x} + T\mathbf{x}) \\ &= a(S + T)(\mathbf{x}),\end{aligned}$$

so $S+T$ preserves scalar multiplication. (The second line uses the fact that S and T preserve scalar multiplication.) Since $S + T$ preserves addition and scalar multiplication, it is linear as claimed.

Sums of matrices. Using the same notation as above, let A be the matrix of S and B the matrix of T. Since $S + T$ is linear, it too must be represented by a matrix; the question is: how does the matrix of $S + T$ relate to the matrix of S and the matrix of T. The unsurprising answer is that the matrix of $S + T$ is given by the *sum* $A + B$, which is the matrix obtained by simply adding an entry of A to the corresponding entry of B.

Thus the sum is defined so that

$$(S + T)\mathbf{x} = (A + B)\mathbf{x}.$$

Note that this is the same as

$$(A + B)\mathbf{x} = A\mathbf{x} + B\mathbf{x},$$

which you should view as a new distributive property. (Previously, we only had such a property when taking a matrix times the sum of vectors.) The point is that $A + B$ is defined the way it is precisely so that this property holds. Using induction, the same is true when we take the sum of more than two matrices.

Warm-Up 2. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^k$ are linear transformations. The *composition* of S and T is the function $ST : \mathbb{R}^n \to \mathbb{R}^k$ defined by first applying T and then S:

$$(ST)\mathbf{x} = S(T\mathbf{x}).$$

We show that ST is linear as well.

First, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, using the fact that S and T preserve addition we have:

$$\begin{aligned}(ST)(\mathbf{x}+\mathbf{y}) &= S(T(\mathbf{x}+\mathbf{y})) \\ &= S(T\mathbf{x}+T\mathbf{y}) \\ &= S(T\mathbf{x})+S(T\mathbf{y}) \\ &= (ST)\mathbf{x}+(ST)\mathbf{y}\end{aligned}$$

so ST preserves addition. Second, for any $x \in \mathbb{R}^n$ and $a \in \mathbb{R}$, we have

$$(ST)(a\mathbf{x}) = S(T(a\mathbf{x})) = S(aT\mathbf{x}) = aS(T\mathbf{x}) = a(ST)\mathbf{x},$$

so ST preserves scalar multiplication, where we have used the fact that S and T do. Thus ST is linear as claimed.

Products of matrices. As when discussing sums, we can ask here what the relation between the matrix of ST is in relation to the matrix of S and the matrix of T. Denoting the $k \times m$ matrix of S by A and the $m \times n$ matrix of T by B, the *product* of A and B is the matrix AB giving the standard matrix of ST. This matrix AB is of size $k \times n$ and the definition guarantees that

$$(AB)\mathbf{x} = A(B\mathbf{x})$$

is true. Indeed, this equality *is* the definition of the matrix AB. The important point to remember is that matrix multiplication is defined the way it is precisely so that it corresponds to compositions of linear transformations.

The question now is: how do we actually compute the product AB directly? You can see how this is done in the book or my Math 290 notes. To derive the necessary formulas, you can figure out what $(ST)(\mathbf{e}_1), \dots, (ST)(\mathbf{e}_n)$ are and then use the fact that these should give the columns of the matrix of ST. Alternatively, note that if we denote the columns of B by $\mathbf{b}_1, \dots, \mathbf{b}_n$, then:

$$\begin{aligned}(AB)\mathbf{x} &= A\left[\begin{bmatrix}\mathbf{b}_1 & \dots & \mathbf{b}_n\end{bmatrix}\begin{bmatrix}x_1 \\ \vdots \\ x_n\end{bmatrix}\right] \\ &= A(x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n) \\ &= x_1A\mathbf{b}_1 + \cdots + x_nA\mathbf{b}_1 \\ &= \begin{bmatrix}A\mathbf{b}_1 & \cdots & A\mathbf{b}_n\end{bmatrix}\begin{bmatrix}x_1 \\ \vdots \\ x_n\end{bmatrix} \\ &= \begin{bmatrix}A\mathbf{b}_1 & \cdots & A\mathbf{b}_n\end{bmatrix}\mathbf{x},\end{aligned}$$

which says that AB should be the matrix whose columns are $A\mathbf{b}_1, \dots, A\mathbf{b}_n$, which is true. Again, check the book or my Math 290 notes to see examples of explicit computations.

Example 1. This and the final example are meant to emphasize the link between matrix multiplication and compositions of linear transformations. Let A be the following 2×2 matrix:

$$A = \begin{bmatrix}1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2\end{bmatrix}.$$

We compute A^{100}, which means the product of A with itself 100 times. Of course, the wrong way to do this is to sit down and start multiplying A by itself a bunch of times—the better way is to use the fact that A^{100} would be the matrix describing the composition of the corresponding transformation with itself 100 times.

In fact, A describes the rotation transformation $\mathbb{R}^2 \to \mathbb{R}^2$ by the angle $\pi/3$. The key point is that rotation by $\pi/3$ six times results in a total rotation by 2π, which is the same as doing nothing. Thus, every time we rotate a multiple of 6 times, we have done nothing. This implies that rotating 96 times is the same as doing nothing, so

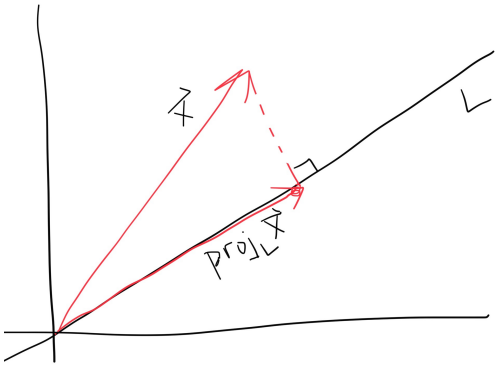
$$\text{rotating 100 times} \;=\; \text{rotating 4 times} \;=\; \text{rotating by } 4\pi/3 \text{ once.}$$

Thus A^{100} should be the matrix of a rotation by $4\pi/3$, which is:

$$A^{100} = \begin{bmatrix} \cos\frac{4\pi}{3} & -\sin\frac{4\pi}{3} \\ \sin\frac{4\pi}{3} & \cos\frac{4\pi}{3} \end{bmatrix} = \begin{bmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{bmatrix}.$$

Actually, for this specific example, it's not too hard to multiply A by itself a few times and recognize a pattern, but the approach we used here will be much simpler when dealing with more complex examples.

Example 2. Let $P : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which orthogonally projects a vector onto the a line L:



We compute P^4. In fact, we claim that $P^2 = P$, so $P^k = P$ for any $k \geq 1$.

For a given $\mathbf{x} \in \mathbb{R}^2$, $P\mathbf{x}$ is the orthogonal projection $\text{proj}_L \mathbf{x}$ of $\mathbf{x}$ onto L, as drawn above. The key point is that if we apply this projection to a vector which is on L to begin with, nothing happens:

$$\text{proj}_L \mathbf{y} = \mathbf{y} \text{ if } \mathbf{y} \text{ is on } L.$$

Thus, since $P\mathbf{x}$ will always be on L, we have:

$$P(P\mathbf{x}) = P\mathbf{x} \text{ for any } \mathbf{x} \in \mathbb{R}^2,$$

so $P^2\mathbf{x} = P\mathbf{x}$ for any $\mathbf{x}$. But this says that the matrices P^2 and P must be the same since they correspond to the same transformation, meaning that they have the same effect when applied to any vector. Thus $P^2 = P$ as claimed, so higher powers of P will be equal to P as well, which is something you can justify precisely using induction.

Lecture 16: Inverses

Warm-Up 1. We find all 2×2 matrices A such that

$$A \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} A.$$

(We say that A and $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ *commute* when this is true. The point is that matrix multiplication is not commutative in general, so not all matrices A will satisfy this property.) Writing down an arbitrary expression for A:

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

the point is that the condition that A and $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ commute give a linear system which the entries of A must satisfy.

We have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} a+3b & 2a+4b \\ c+3d & 2c+4d \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+2c & b+2d \\ 3a+4c & 3b+4d \end{bmatrix}.$$

Thus for A and $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ to commute we must have

$$\begin{bmatrix} a+3b & 2a+4b \\ c+3d & 2c+4d \end{bmatrix} = \begin{bmatrix} a+2c & b+2d \\ 3a+4c & 3b+4d \end{bmatrix},$$

so the entries of A must satisfy the following linear system:

$$\begin{aligned} 3b - 2c &= 0 \\ 2a + 3b - 2d &= 0 \\ 3a + 3c - 3d &= 0 \end{aligned}.$$

Solving this gives the general solution as being of the form

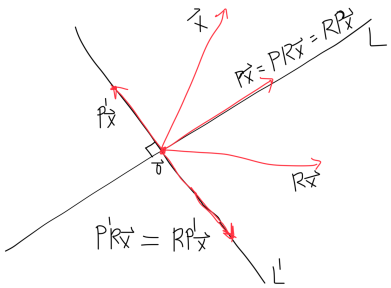
$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} -c+d \\ \frac{2}{3}c \\ c \\ d \end{bmatrix},$$

so we conclude that the matrices which commute with $\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$ are those which are of the form

$$A = \begin{bmatrix} -c+d & \frac{2}{3}c \\ c & d \end{bmatrix}$$

where $c, d \in \mathbb{R}$.

Warm-Up 2. Consider the perpendicular lines L and L' drawn below, and let P be the matrix of the orthogonal projection onto L, P' the matrix of the orthogonal projection onto L', and R the reflection across the line L:



We compute all the possible products of these three: $PP'R, PRP', P'RP, P'PR, RPP', RP'P$. In fact, we claim that all of these products equal the zero matrix.

The point is that instead of figuring out what each matrix P, P', R actually looks like and then computing each product by hand, we use the fact that matrix products correspond to compositions of linear transformations. Take any $\mathbf{x} \in \mathbb{R}^2$. Then $P\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto L, and so in particular is a vector on L itself. Orthogonally projecting any vector on L onto L' results in zero since L and L' are perpendicular, so $P'P\mathbf{x} = \mathbf{0}$. Finally, reflecting the zero vector still gives the zero vector, so we conclude that

$$RP'P\mathbf{x} = \mathbf{0} \text{ for all } \mathbf{x} \in \mathbb{R}^2.$$

Thus $RP'P$ describes the transformation which sends everything to $\mathbf{0}$, so $RP'P$ is the zero matrix. Similarly, $PR\mathbf{x}$ is obtained by first reflecting $\mathbf{x}$ across L and then projecting the result onto L. (Note that the result of this is the same as projecting $\mathbf{x}$ itself, so $RP\mathbf{x} = P\mathbf{x}$.) Thus $PR\mathbf{x}$ is again a vector on L, so $P'PR\mathbf{x} = \mathbf{0}$ since projecting something on L onto L' gives the zero vector. Hence $P'PR$ is the zero matrix.

And so on arguing in a similar manner will show that the result of applying any of the desired products to any vector will always give the zero vector in the end. To do one more: $PRP'\mathbf{x}$ means we first project $\mathbf{x}$ onto L', then reflect the result across L, and finally project the result onto L. In this case, $P'\mathbf{x}$ is on L', $RP'\mathbf{x} = -P'\mathbf{x}$ since reflecting a vector on L' across the perpendicular L just flips its direction, so $PRP'\mathbf{x} = -PP'\mathbf{x} = \mathbf{0}$ since $-P'\mathbf{x}$ is on L' and projecting it onto L gives $\mathbf{0}$ since L and L' are perpendicular. Thus PRP' sends any vector to the zero vector, so PRP' is the zero matrix.

Exercise. Finish checking that the remaining products above are also zero.

Invertible matrices. We say that an $n \times n$ matrix A is *invertible* if there exists an $n \times n$ matrix B such that

$$AB = I_n \quad \text{and} \quad BA = I_n$$

where I_n denotes the $n \times n$ identity matrix. (Note that at first glance we need to consider both equalities $AB = I$ and $BA = I$ separately since matrix multiplication is not commutative, so one being satisfied doesn't immediately guarantee the other is as well. We'll actually see next time that for *square* matrices, if one of the above equalities is satisfied the other must be as well, but this is highly nontrivial fact. Indeed, this is NOT true when we consider linear transformations between *infinite-dimensional* spaces, which we'll talk about soon enough.) If such a B exists, we call it the *inverse* of A and denote it by A^{-1}.

Exercise. Show that if a matrix has an inverse, it has only one; in other words, show that inverses are unique if they exist.

Invertible transformations. The point is that invertible matrices correspond to *invertible* linear transformations, where we say that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible if there exists a linear transformation $S : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$(ST)\mathbf{x} = \mathbf{x} \text{ for all } \mathbf{x} \in \mathbb{R}^n \text{ and } (TS)\mathbf{y} = \mathbf{y} \text{ for all } \mathbf{y} \in \mathbb{R}^m.$$

This says that S and T "undo" each other, so that if T sends $\mathbf{x}$ to $\mathbf{y}$, $S = T^{-1}$ sends $\mathbf{y}$ back to $\mathbf{x}$. The equalities $AB = I$ and $BA = I$ in the definition of an invertible matrix are the matrix versions of the above definition of invertible transformations.

To get a feel for the notion of an inverse, note that the inverse of rotation by θ is rotation by $-\theta$ and the inverse of a reflection is the same reflection itself since performing a reflection twice in a row gives back the original vector.

Exercise. Show that orthogonal projections are not invertible.

Why only square matrices? In the definition we gave of "invertible matrix", note that we assumed A had to be a square matrix, with the same number of rows and columns. This is no accident: it is only these types of matrices which have the chance of being invertible! Said another way, only linear transformations $\mathbb{R}^n \to \mathbb{R}^n$ have the hope of being invertible, so that a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ with $n \neq m$ is never invertible.

To be precise, suppose that A is an $n \times m$ matrix with $n < m$. It is possible to find an $m \times n$ matrix B such that $AB = I_n$, as you will show on Homework 5. However, for such a B it is never going to be true that $BA = I_m$, which you will also show on Homework 5. In other words, for non-square matrices we can never have both $AB = I$ and $BA = I$ at the same time, so we only consider invertibility for square matrices.

Computing inverses. We'll talk about the importance of invertible matrices next time, so for right now we will simply outline the method used to compute them. In the 2×2 case there is a simply formula which you should know by heart:

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

which exists if and only if $ad - bc \neq 0$. We'll talk more about the mysterious quantity "$ad - bc$" next quarter when we discuss *determinants*.

In the more general $n \times n$ setting such a formula is not simple to memorize, which is fine because the method we now describe gives more insight into what's special about invertible matrices than a memorized formula would. To compute A^{-1}, we first setup a large $n \times 2n$ augmented matrix $\begin{bmatrix} A \mid I \end{bmatrix}$ having A on the left and the $n \times n$ identity matrix on the right. Then we row reduce this matrix until the portion on the left becomes I—the fact is that the resulting matrix on the right is precisely A^{-1}:

$$\begin{bmatrix} A \mid I \end{bmatrix} \to \begin{bmatrix} I \mid A^{-1} \end{bmatrix}.$$

Note that this assumes that A can be row-reduced to the identity, so that $\text{rref}(A) = I$. We'll see next time that this indeed *must* be true for any invertible matrix.

Why does this process work? Consider the augmented matrix $\begin{bmatrix} A \mid I \end{bmatrix}$ as representing a linear system

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

where the first n columns correspond to the entries of $\mathbf{x}$ and the final n columns correspond to the entries of $\mathbf{y}$. After reducing to something of the form $\left[\, I \mid B \,\right]$ our system becomes

$$I\mathbf{x} = B\mathbf{y}.$$

Thus the matrix B obtained has the property that

$$A\mathbf{x} = \mathbf{y} \text{ if and only if } \mathbf{x} = B\mathbf{y},$$

which says precisely that if A is the matrix of a linear transformation T, B describes the inverse transformation T^{-1}. Indeed, we have:

$$(BA)\mathbf{x} = B(A\mathbf{x}) = B\mathbf{y} = \mathbf{x} \text{ and } (AB)\mathbf{y} = A(B\mathbf{y}) = A\mathbf{x} = \mathbf{y},$$

so BA and AB are both identity matrices. Thus, $B = A^{-1}$ as claimed.

What's to come? Check the book or my Math 290 notes for examples of computing explicit inverses using the method described above. So far we know that if $\text{rref}(A) = I$, then A is invertible since the this method produces a matrix which satisfies the properties required of A^{-1}. Next time we will see that the converse is true: if A is invertible, then $\text{rref}(A) = I$. Thus saying that a matrix is invertible if the *same* as saying that $\text{rref}(A) = I$. This is the key behind the many different ways we have of characterizing invertibility, some of which we'll talk about next time, and others which are yet to come. In some sense, a good bulk of linear algebra deals with understanding the types of properties which are equivalent to invertibility.

Lecture 17: More on Invertibility

Even though we phrase the first two examples as Warm-Ups, they each actually contain important observations about invertible matrices, which we'll include as part of the main theorem we discuss afterwards.

Warm-Up 1. We show that the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

is not invertible. In the course of doing so we will derive a key fact: if $\text{rref}(A) \neq I$, then A is not invertible.

Say we didn't know beforehand that A was not invertible and tried to find its inverse using the method described last time. Reducing $\left[\, A \mid I \,\right]$ gives:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right].$$

At this point we can see that it will not be possible to continue reducing the left portion to the 3×3 identity matrix, since $\text{rref}(A)$ will only have two pivots. Thus, this method for finding the inverse of A does not work.

However, this is not yet enough alone to say that A is not invertible since it could be that the inverse of A could be found by some other method apart from the one we described. Why is it

exactly that if this method doesn't succeed we know for sure A is not invertible? The key comes in recognizing that if A is invertible, then any equation $A\mathbf{x} = \mathbf{b}$ always has exactly one solution, since multiplying both sides on the left by A^{-1} gives

$$\mathbf{x} = A^{-1}\mathbf{b}$$

has the only $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{b}$. In particular, $A\mathbf{x} = \mathbf{0}$ can only have the $\mathbf{x} = \mathbf{0}$ solution if A is invertible, which we can see by taking $\mathbf{b} = \mathbf{0}$ above.

Based on the row-operations we have done so far, we can see that for the given matrix A, $A\mathbf{x} = \mathbf{0}$ will have infinitely many solutions since the reduced form will have a free variable. Since $A\mathbf{x} = \mathbf{0}$ has more than only the $\mathbf{x} = \mathbf{0}$ solution, A cannot be invertible as claimed. What we have shown is that if $\text{rref}(A) \neq I$, then A is not invertible since $A\mathbf{x} = \mathbf{0}$ will have more than one solution.

Warm-Up 2. Suppose that A and B are both $n \times n$ matrices such that $AB = I$. We show that $BA = I$. In other words, the equation $AB = I$ alone guarantees that A and B are both invertible and that they are each other's inverse. This is amazing: matrix multiplication is not commutative, and yet in this case $AB = I$ does imply $BA = I$. The fact that A and B are *square* matrices is crucial, as is the fact that they are matrices of a finite size. (As I alluded to last time, the analogous result won't be true in infinite-dimensional settings.)

Note that we cannot proceed as follows: multiply both sides of $AB = I$ by A^{-1} on the left to get $A^{-1}AB = A^{-1}I$, so $B = A^{-1}$, and then $BA = A^{-1}A = I$ as claimed. This assumes that we already know A^{-1} exists, which is *not* a given. Indeed, the point of the this Warm-Up is to show that $AB = I$ does imply A and B are both invertible. The moral is: never introduce inverses until you know for sure that they exist.

Instead, we use the alternate characterization of "invertible" derived in the first Warm-Up. Suppose that $\mathbf{x} \in \mathbb{R}^n$ satisfies

$$B\mathbf{x} = \mathbf{0}.$$

Multiplying both sides on the left by A gives

$$A(B\mathbf{x}) = A\mathbf{0}, \text{ so } (AB)\mathbf{x} = \mathbf{0}.$$

But $AB = I$, so this final equation becomes $\mathbf{x} = \mathbf{0}$. Thus the only solution of $B\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. This means that $\text{rref}(B) = I$ since otherwise $B\mathbf{x} = \mathbf{0}$ would have infinitely many solutions. Thus the method we described last time for finding the inverse of B will work, so B is invertible.

Now that we know B^{-1} exists, we can multiply both sides of $AB = I$ on the right by B^{-1} go get $A = B^{-1}$, so that $BA = BB^{-1} = I$ as claimed.

Amazingly Awesome Theorem. This theorem goes by many names: the Invertible Matrix Theorem, the Inverse Function Theorem, and others; there is no set standard. I am using the name I've used for over 10 years now, which is meant to emphasize how special—and amazingly awesome—this theorem really is. (As I said in class, some of the Math 290 instructors have started using then name "amazingly awesome" as well, so don't be surprised if years from now you see a Wikipedia entry about this.)

Here is the statement. Suppose that A is an $n \times n$ matrix. Then the following conditions are equivalent, meaning that they all imply and are implied by one another:

(a) A is invertible

(b) $\text{rref}(A) = I_n$

(c) $\text{rank}(A) = n$

(d) the only solution of $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$

(e) the columns of A are linearly independnet

(f) for any $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ has a unique solution

(g) the columns of A span $\mathbb{R}^n$

(h) the linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is surjective

(i) the linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is injective

In the second-to-last condition, to say that T is *surjective* means that for any $b \in \mathbb{R}^n$ there exists $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{b}$, and in the final condition to say that T is *injective* means that if $T(\mathbf{x}) = T(\mathbf{y})$, then $\mathbf{x} = \mathbf{y}$.

Thus, all conditions above mean exactly the same thing, and are all different ways of characterizing what it means for a matrix to be invertible. The key one is really the second, from which all the others follow. As I mentioned, understanding why this theorem is true and the ways (to be seen) in which it can be extended is one of the key goals of linear algebra.

Proof. We went through some of these in class, but here we'll give a (mostly) complete proof. I would expect that proving that any of these imply any of the others is something you should all be able to understand how to do when it comes to exam-type material.

We have already seen that (a) and (b) are equivalent based on the method we described for computing inverses and the Warm-Ups. That (b) and (c) are equivalent is also clear: $\text{rank}(A) = n$ if and only if $\text{rref}(A)$ has n pivots, which seen A is $n \times n$ is true if and only if there are no rows of all zeroes, which is true if and only if $\text{rref}(A) = I$. That (b) and (d) are equivalent is also something we've essentially worked out already: the only solution of $A\mathbf{x} = \mathbf{0}$ being $\mathbf{x} = \mathbf{0}$ is true if and only if $\text{rref}(A)$ has no rows of all zeroes so that the augmented matrix $\begin{bmatrix}\text{rref}(A) \mid \mathbf{0}\end{bmatrix}$ has no free variables, which is true if and only if $\text{rref}(A) = I$. Thus each of the first four conditions are equivalent to one another.

Condition (e) is a rephrasing of (d): the columns of $A = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ are linearly independent if and only if the only scalars $x_1, \ldots, x_n$ satisfying

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0}$$

are $x_1 = \cdots = x_n = 0$, which is true if and only if the only $\mathbf{x} \in \mathbb{R}^n$ satisfying $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$ since $A\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ where $x_1, \ldots, x_n$ are the entries of $\mathbf{x}$.

(a) $\implies$ (f): Suppose A is invertible. For any $\mathbf{b} \in \mathbb{R}^n$ set $\mathbf{x} = A^{-1}\mathbf{b}$. Then $A\mathbf{x} = A(A^{-1}\mathbf{b}) = \mathbf{b}$, so $A\mathbf{x} = \mathbf{b}$ has a solution, namely $\mathbf{x} = A^{-1}\mathbf{b}$. If $\mathbf{x} \in \mathbb{R}^n$ is any vector satisfying $A\mathbf{x} = \mathbf{b}$, then multiplying on the left by A^{-1} gives $A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}$, so $\mathbf{x} = A^{-1}\mathbf{b}$, showing that $A\mathbf{x} = \mathbf{b}$ has only one solution.

(f) $\implies$ (g): Suppose that for any $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ has a solution. Since $A\mathbf{x}$ is the same as $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ where $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are the columns of A and $x_1, \ldots, x_n$ are the entries of $\mathbf{x}$, this says that for any $\mathbf{b} \in \mathbb{R}^n$ there are scalars $x_1, \ldots, x_n$ such that $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{b}$. Hence any $\mathbf{b} in \mathbb{R}^n$ is a linear combination of the columns of A, so the columns of A span $\mathbb{R}^n$.

(g) $\implies$ (a): Suppose that the columns of A span $\mathbb{R}^n$. If $\text{rref}(A)$ wasn't I, then $\text{rref}(A)$ would have a row of all zeroes, which would mean that there are certain vectors $\mathbf{b} \in \mathbb{R}^n$ for which $A\mathbf{x} = \mathbf{b}$ wouldn't have a solution. This in turn would mean that $\mathbf{b}$ is not a linear combination of

the columns of A, so since the columns of A span $\mathbb{R}^n$ this cannot happen. Hence $\text{rref}(A) = I$, so A is invertible since (b) $\implies$ (a).

(d) $\iff$ (i): Suppose that the only solution of $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$ and let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ be two vectors such that $A\mathbf{x}_1 = A\mathbf{x}_2$. Then $A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{0}$, so $A(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$. Since $\mathbf{x}_1 - \mathbf{x}_2$ is thus a solution of $A\mathbf{x} = \mathbf{0}$, we have $\mathbf{x}_1 - \mathbf{x}_2 = 0$, so $\mathbf{x}_1 = \mathbf{x}_2$ and $T\mathbf{x} = A\mathbf{x}$ is injective. Conversely, suppose that $T(\mathbf{x}) = A\mathbf{x}$ is injective. If $\mathbf{x} \in \mathbb{R}^n$ satisfies $A\mathbf{x} = \mathbf{0}$, then $A\mathbf{x} = A\mathbf{0}$. Since T is injective, this implies that $\mathbf{x} = \mathbf{0}$, so the only solution of $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.

Finally, the fact that property (h) is equivalent to invertibility is left to Homework 6, but is simply a restatement of some other properties listed here. □

Computing inverses, redux. We finish with another justification that the method we descried for computing inverses actually works. Suppose that A satisfies $\text{rref}(A) = I$. There are then elementary row operations which transform A into I, and as discussed on Problem 9 of Homework 5 there are corresponding elementary matrices such that performing one of these elementary row operations is the same as multiplying by the corresponding elementary matrix. Thus there are elementary matrices $E_1, \dots, E_m$ such that

$$E_m \cdots E_1 A = I.$$

Since all matrices are square, the second Warm-Up from today shows that A is invertible and

$$E_m \cdots E_1 = A^{-1}.$$

But thinking of this product on the left as

$$E_m \cdots E_1 = E_m \cdots E_1 I,$$

we see that the product $A^{-1} = E_m \cdots E_1$ is obtained by taking the row operations which transform A into I and applying them to I, which is precisely what happens to the right side of the augmented matrix $\left[A \mid I\right]$ when reducing so that the left side becomes I. Thus the resulting portion on the right side of $\left[I \mid A^{-1}\right]$ is indeed the inverse of A as claimed.

Lecture 18: Vector Spaces

Warm-Up. Suppose that A is an invertible $n \times n$ matrix. We show that the *transpose* of A, which is the matrix A^T obtained by turning the rows of A into columns, is also invertible. We will take it for granted that the following is true: $(AB)^T = B^T A^T$, so that the transpose of a product is the product of transposes in the *reverse* order. We will talk more about transposes when we talk about the dot product, and will see why $(AB)^T = B^T A^T$ is true then.

If A is invertible, there exists an $n \times n$ matrix B (which is just the inverse of A) such that $AB = I$. Taking the transpose of both sides gives

$$(AB)^T = I^T \implies B^T A^T = I$$

since $I^T = I$. But since A^T and B^T are square matrices, the second Warm-Up from last time shows that A^T must be invertible as claimed, with B^T being the inverse of A^T.

Here is another way to see this. We can show directly that $(A^{-1})^T$ satisfies the defining properties of the inverse of A^T. We compute:

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I \text{ and } A^T (A^{-1})^T = (A^{-1}A)^T = I^T = I,$$

so $(A^{-1})^T$ is the inverse of A^T, and hence A^T is invertible. (This shows that $(A^T)^{-1} = (A^{-1})^T$.)

Vector spaces. A vector space is meant to be a type of "space" which shares many of the same properties $\mathbb{R}^n$ has. The key points are that a vector space as an "addition" operation and a "scalar multiplication" operation which satisfies the same basic properties (associativity, commutativity, distributive, etc.) addition and scalar multiplication of vectors in $\mathbb{R}^n$ satisfy.

To be precise, suppose that V is a set equipped with two operations: one we call "addition" which takes two elements of V and outputs another element of V, and the other we call "scalar multiplication" which takes a scalar in $\mathbb{K}$ ($\mathbb{K}$ will denote either or $\mathbb{R}$ or $\mathbb{C}$ here) and an element of V and outputs another element of V. We say that V is a *vector space over* $\mathbb{K}$ if the following properties hold:

(a) $(u+v)+w = u+(v+w)$ for all $u, v, w \in V$

(b) $u+v = v+u$ for all $u, v \in V$

(c) there exists an element $0 \in V$ such that $u+0 = u = 0+u$ for any $u \in V$

(d) for any $u \in V$, there exists an element $-u \in V$ such that $u+(-u) = 0 = (-u)+u$

(e) $a(bu) = (ab)u$ for any $a, b \in K$ and $u \in V$

(f) $a(u+v) = au+av$ for any $a \in \mathbb{K}$ and $u, v \in V$

(g) $(a+b)u = au+bu$ for any $a, b \in \mathbb{K}$ and $u \in V$

(h) $1u = u$ for any $u \in V$.

Thus, as stated, the "addition" and "scalar multiplication" operations on V should satisfy the same types of properties as do addition and scalar multiplication on $\mathbb{R}^n$.

Some important things to note. First, we will refer to the elements of V as *vectors* in V, although they are not necessarily "vectors" in the sense of $\mathbb{R}^n$ as we've previously used the term; that is, a "vector" in a vector space V is simply an element of V. Second, in most concrete examples we'll be interested in, "addition" and "scalar multiplication" look very similar to what we ordinarily refer to as addition and scalar multiplication, but the point is that in general these operations could be anything whatsoever, as long as together they satisfy the properties listed above—for instance, there is a problem on the homework which takes the "addition" of x and y to be xy instead of $x+y$. We'll see another example next time. Third, we refer to the element $0 \in V$ in property (c) as the *zero vector* of V, even thought it might not be "zero" as we're used to—the point is that the zero vector is completely characterized by the property given in (c), which says that it should be the additive identity. Finally, (d) says that we use the notation $-u$ to denote the additive inverse of u, which might not literally be "negative u".

Other properties. We have seen a few other properties of addition and scalar multiplication in $\mathbb{R}^n$, such as:

$$0u = \mathbf{0},\ a\mathbf{0} = \mathbf{0},\ (-1)\mathbf{v} = -\mathbf{v}.$$

The first two were Warm-Ups during the second week of class, and the third was on Homework 1. The point is that since the proofs we gave of these properties only depended on distributive and other such properties, the same arguments show that these equalities hold in general vector spaces as well. This is why it was important to give proofs of these which did not depend on writing an element in $\mathbb{R}^n$ as a column with n entries, but rather proofs using only the properties of addition

and scalar multiplication given above. Similarly, in Homework 1 we showed that property (b)—the commutativity of addition—it follows from the other properties, so this is true in a general vector space as well.

Main examples. The main examples of vector spaces are the ones we've been seeing all along: $\mathbb{R}^n$ and $\mathbb{C}^n$ with their usual addition and scalar multiplications. To be precise, $\mathbb{R}^n$ is a vector space over $\mathbb{R}$ (also called a *real* vector space), while $\mathbb{C}^n$ is usually thought of as a vector space over $\mathbb{C}$ (also called a *complex* vector space), even though $\mathbb{C}^n$ can also be considered to be a vector space over $\mathbb{R}$, as we'll see in some examples next time. The difference just comes in the types of scalars we allow ourselves to use.

Spaces of matrices. We let $M_{m,n}(\mathbb{R})$ denote the set of $m \times n$ matrices with entires in $\mathbb{R}$. With the usual definitions of matrix addition and multiplication of a matrix by a scalar, $M_{m,n}(\mathbb{R})$ is a vector space over $\mathbb{R}$. The "zero vector" in this space is the $m \times n$ zero matrix, and the "negative" of a matrix A is the matrix obtained by changing the signs of all the entries of A. (So, a "vector" in $M_{m,n}(\mathbb{R})$ would mean an $m \times n$ matrix in this context.) For square matrices, we will usually use $M_n(\mathbb{R})$ instead of $M_{n,n}(\mathbb{R})$ to denote the space of $n \times n$ matrices over $\mathbb{R}$.

Similarly, $M_{m,n}(\mathbb{C})$ denotes the space of $m \times n$ matrices with complex entries, and is a vector space over $\mathbb{C}$ with the usual addition and scalar multiplication of matrices.

Spaces of polynomials. We let $P_n(\mathbb{R})$ denote the space of polynomials with real coefficients of degree at most n:

$$P_n(\mathbb{R}) := \{a_0 + a_1 x + \cdots + a_n x^n \mid a_0, \ldots, a_n \in \mathbb{R}\}.$$

With the usual addition of polynomials and the usual scalar multiplication, this is a vector space over $\mathbb{R}$. The space $P(\mathbb{R})$ (without a subscript) is the set of *all* polynomials with real coefficients, with no restriction on the degree. In any of these, the "zero vector" is the zero polynomial 0.

Similarly, $P_n(\mathbb{C})$ denotes the space of polynomials with complex coefficients of degree at most n. This is a complex vector space with the usual addition and scalar multiplication.

Spaces of functions. We let $F(\mathbb{R})$ denote the space of all functions from $\mathbb{R}$ to $\mathbb{R}$:

$$F(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is a function}\}.$$

With the usual addition and scalar multiplication of functions, this is a vector space over $\mathbb{R}$.

Lecture 19: More on Vector Spaces

Warm-Up. Define an "addition" operation $\oplus$ and a "scalar multplication" $\odot$ on $\mathbb{R}$ via

$$x \oplus y = \sqrt[3]{x^3 + y^3 + 1} \quad \text{and} \quad a \odot x = \sqrt[3]{a(x^3 + 1) - 1}$$

for all "vectors" $x, y \in \mathbb{R}$ and "scalars" $a \in \mathbb{R}$. This gives a vector space structure on $\mathbb{R}$ over $\mathbb{R}$. We will verify part of this.

First we determine the "zero vector" of this vector space. This should be an element $y \in \mathbb{R}$ such that

$$x \oplus y = x \text{ for all } x \in \mathbb{R}.$$

Using our definition of $\oplus$, this gives the requirement that

$$\sqrt[3]{x^3 + y^3 + 1} = x \text{ for all } x \in \mathbb{R},$$

which after solving implies that $y = -1$. Thus -1 is the "zero vector" of this vector space:

$$x \oplus (-1) = \sqrt[3]{x^3 + (-1)^3 + 1} = \sqrt[3]{x^3} = x.$$

Next we determine the "additive inverse" of each element. The "additive inverse" of $x \in \mathbb{R}$ should be the element $y \in \mathbb{R}$ such that

$$x \oplus y = \text{zero vector}.$$

(Note that this condition was written in our vector space axioms as $x + y = 0$, but the key point is that "0" here denotes the zero vector of the space and not necessarily the *number* 0. This is why I wrote out "zero vector" explicitly above, to avoid confusion.) In our case, we determined the zero vector to be -1, so we want y to satisfy

$$x \oplus y = \sqrt[3]{x^3 + y^3 + 1} = -1.$$

Solving for y gives

$$y = \sqrt[3]{-x^3 - 2}.$$

We verify:

$$x \oplus \sqrt[3]{-x^3 - 2} = \sqrt[3]{x^3 + \left(\sqrt[3]{-x^3 - 2}\right)^3 + 1} = \sqrt[3]{x^3 - x^3 - 2 - 1} = -1$$

as required, so $\sqrt[3]{-x^3 - 2}$ is the "additive inverse" of x.

Let us also verify one of the distributive properties: for any $a \in \mathbb{R}$ and $x, y \in \mathbb{R}$,

$$a \odot (x \oplus y) = (a \odot x) \oplus (a \odot y),$$

which is the analog of $a(x + y) = ax + ay$ using our redefined operations. The left side is:

$$\begin{aligned} a \odot (x \oplus y) &= a \odot \sqrt[3]{x^3 + y^3 + 1} \\ &= \sqrt[3]{a\left(\sqrt[3]{x^3 + y^3 + 1}^3 + 1\right) - 1} \\ &= \sqrt[3]{a(x^3 + y^3 + 1 + 1) - 1} \\ &= \sqrt[3]{a(x^2 + y^3 + 2) - 1}. \end{aligned}$$

On the other hand:

$$\begin{aligned} (a \odot x) \oplus (a \odot y) &= \sqrt[3]{a(x^3 + 1) - 1} \oplus \sqrt[3]{a(y^3 + 1) - 1} \\ &= \sqrt[3]{\sqrt[3]{a(x^3 + 1) - 1}^3 + \sqrt[3]{a(y^3 + 1) - 1}^3 + 1} \\ &= \sqrt[3]{a(x^3 + 1) - 1 + a(y^3 + 1) - 1 + 1} \\ &= \sqrt[3]{a(x^3 + y^3 + 2) - 1}, \end{aligned}$$

so $a \odot (x \oplus y) = (a \odot x) \oplus (a \odot y)$ as claimed.

Now, thinking back to Homework 1, we showed there that

$$(-1)\mathbf{v} = -\mathbf{v},$$

and the point is that the proof we gave using only the distributive property, additive inverse propety, and so on works more generally in *any* vector space. That is, in any vector space it should be true that multiplying the scalar -1 by a vector should give the additive inverse of that vector, no matter what our "addition" and "scalar multiplication" operations actually are. Indeed, note that

$$(-1) \odot x = \sqrt[3]{-1(x^3+1)-1} = \sqrt[3]{-x^3-2},$$

which we worked out earlier is indeed the additive inverse of x.

Also, back in the second week of class we showed that $0\mathbf{v} = \mathbf{0}$, meaning that multiplying any vector by the scalar 0 should give the zero vector of that space. Again, the proof we gave using only the "vector space" axioms now works more generally, so it should be true in this specific example as well. Indeed:

$$0 \odot x = \sqrt[3]{0(x^3+1)-1} = \sqrt[3]{-1} = -1,$$

which is the zero vector of this space.

Remark. The Warm-Up above illustrates the point of the time we spent verifying such "obvious" properties as

$$0\mathbf{v} = \mathbf{0} \text{ and } (-1)\mathbf{v} = -\mathbf{v},$$

which is: although such properties might be straightforward to see in $\mathbb{R}^n$ using components, the proofs we gave avoiding coordinates show that these properties are true in any vector space whatsoever, which is far from obvious since "addition" and "scalar multiplication" might be crazy-looking operations in general.

The example in the Warm-Up didn't just come out of nowhere—we'll see later when we talk about *isomorphisms* how I came up with it and how you can similarly come up with tons of other "crazy" examples.

Linear combinations and independence. The definition of *linear combination* and *linearly independent* we had previously for $\mathbb{R}^n$ works just as well in any vector space. To be clear, for a vector space V over $\mathbb{K}$ a linear combination of $v_1, \dots, v_k \in V$ is an expression of the form

$$c_1 v_1 + \cdots + c_k v_k \text{ where } c_1, \dots, c_k \in \mathbb{K}.$$

The set of all such linear combinations of $v_1, \dots, v_k$ is called their *span*, just as it was in the $\mathbb{R}^n$ case. Similarly, we can talk about affine combinations in an arbitrary vector space.

A collection of vectors $v_1, \dots, v_k \in V$ are linearly independent over $\mathbb{K}$ if none is a linear combination of the others over $\mathbb{K}$, or equivalently if the only scalars $a_1, \dots, a_k \in \mathbb{K}$ satisfying

$$a_1 v_1 + \cdots + a_k v_k = 0$$

are $a_1 = \cdots = a_k = 0$. Similarly, we can talk about vectors be affinely independent in an arbitrary vector space.

Complex vector spaces are real vector spaces. If V is a vector space over $\mathbb{C}$, we can always consider it to be a vector space over $\mathbb{R}$ as well since "scalar multplication" still makes sense if we restrict to only multiplying by real scalars, which are included among the possible complex scalars. (Note, however, that a vector space over $\mathbb{R}$ cannot generally also be considered to be a vector over $\mathbb{C}$, since knowing how to multiply by real scalars does not guarantee that multiplying by complex scalars is also be possible.)

Consider $\mathbb{C}^n$. As a vector space over $\mathbb{C}$ it is spanned by $\mathbf{e}_1, \dots, \mathbf{e}_n$ since

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = z_1\mathbf{e}_1 + \cdots + z_n\mathbf{e}_n.$$

However, these vectors alone do not span $\mathbb{C}^n$ viewed as a vector space over $\mathbb{R}$—in particular, the above expression is not a valid linear combination over $\mathbb{R}$ since the coefficients $z_1, \dots, z_n$ involved are not necessarily real. Writing each z_k as $z_k = a_k + ib_k$ with $a_k, b_k \in \mathbb{R}$, we have:

$$\begin{aligned} \begin{bmatrix} a_1 + ib_1 \\ \vdots \\ a_n + ib_n \end{bmatrix} &= (a_1 + ib_1)\mathbf{e}_1 + \cdots + (a_n + ib_n)\mathbf{e}_n \\ &= a_1\mathbf{e}_1 + b_1(i\mathbf{e}_1) + \cdots + a_n\mathbf{e}_n + b_n(i\mathbf{e}_n), \end{aligned}$$

which *is* now a valid linear combination over $\mathbb{R}$ since the coefficients used $a_1, b_1, \dots, a_n, b_n$ are all real. (The point is that we absorb the i into the vector being used.) This says that any element of $\mathbb{C}^n$ can be written as a linear combination of

$$\mathbf{e}_1,\ i\mathbf{e}_1,\ \dots,\ \mathbf{e}_n,\ i\mathbf{e}_n$$

over $\mathbb{R}$, so that these $2n$ vectors (each of the $\mathbf{e}_i$'s along with what you get when you multiply each by i) span $\mathbb{C}^n$ as a vector space over $\mathbb{R}$.

Another example. The space $P_2(\mathbb{C})$ of polynomials of degree at most 2 with complex coefficients is a complex vector space, but also a real vector space. As a complex vector space, it is spanned by $1, x, x^2$ since any $p(x) = P_2(\mathbb{C})$ is of the form

$$p(x) = a_0 + a_1x + a_2x^2$$

for some $a_0, a_1, a_2 \in \mathbb{C}$. (Think of the constant term a_0 as $a_0 \cdot 1$.) However, as a real vector space it is spanned by $1, i, x, ix, x^2, ix^2$ since each $a_k = b_k + ic_k$ with $b_k, c_k \in \mathbb{R}$ gives

$$p(x) = (b_0 + ic_0) + (b_1 + ic_1)x + (b_2 + ic_2)x^2 = b_0 + c_0(i) + b_1x + c_1(ix) + b_2x^2 + c_2(ix^2),$$

which is a linear combination over $\mathbb{R}$. Note that as a vector space over $\mathbb{C}$ we had 3 spanning vectors, but as a vector space over $\mathbb{R}$ we had 6, and in the $\mathbb{C}^n$ case, there were n spanning vectors over $\mathbb{C}$ but $2n$ over $\mathbb{R}$; this "times 2" behavior is not a coincidence, as you'll show on a future homework.

What about $\mathbb{K}^\infty$? By $\mathbb{K}^\infty$ we mean the set of all "infinitely long" vectors, or "vectors" with an infinite number of components:

$$\mathbb{K}^\infty := \{(x_1, x_2, x_3, \dots) \mid \text{each } x_i \text{ is in } K\}.$$

(Contrary to previous notation, sometimes we'll write such vectors as rows when it saves space.) Another way of saying this is that $\mathbb{K}^n$ is the space of infinite sequences of elements of $\mathbb{K}$. We ask: are there finitely many vectors which span $\mathbb{K}^\infty$?

Note that, at first glance we might think we can write an expression like:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n + \cdots.$$

However, this doesn't really make sense: what does it mean to add together infinitely many vectors? In calculus, we can try to make sense of adding together infinitely many numbers using the idea of an *infinite series*, but what would the analogous thing be for vectors? Moreover, not all such series are "convergent", meaning they don't necessarily always give *finite* values. The upshot is that we would do NOT call the above expression a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \ldots$: the term "linear combination" always refers to a combination of *finitely* many vectors. By the *span* of an infinite collection of vectors $\mathbf{e}_i$ we thus mean the set of all linear combinations of finitely many of them.

Going back to the question we asked, it is certainly not true that there are finitely many vectors among the $\mathbf{e}_i$ which span all of $\mathbb{K}^\infty$. Indeed, if we take finitely many of the $\mathbf{e}_i$, say:

$$\mathbf{e}_{n_1}, \ldots, \mathbf{e}_{n_k},$$

then any linear combination of these will have zeroes in the locations after the $\max\{n_1, \ldots, n_k\}$ location, since each of the $\mathbf{e}_{n_i}$ have zeroes after these locations. Thus such combinations cannot possibly give everything in $\mathbb{K}^\infty$, so finitely many of the $\mathbf{e}_i$ vectors can never span $\mathbb{K}^\infty$.

However, how do we know that we couldn't find some finitely many *other* vectors, none of which are among the $\mathbf{e}_i$, which *would* span $\mathbb{K}^n$? It turns out that this is not possible, so that $\mathbb{K}^n$ is an example of what we will call an *infinite-dimensional* vector space. We'll come back to seeing how we can justify something like this after we more about linear independence.

Lecture 20: Subspaces

Warm-Up. Let $F(\mathbb{R})$ denote the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. Consider the subset U of $F(\mathbb{R})$ consisting of those functions which are twice-differentiable and whose second derivatives satisfy the equation $f'' + f = 0$:

$$U := \{f \in F(\mathbb{R}) \mid f'' \text{ exists and } f'' + f = 0\}.$$

(The equation $f'' + f = 0$ is what is known as a *differential equation*, so U is the se of solutions to this differentiable equation.) We claim that U is a vector space over $\mathbb{R}$ when equipped with the ordinary addition and scalar multiplication of functions.

For instance, the distributive property $a(f + g) = af + ag$ for $a \in \mathbb{R}$ and $f, g \in U$ says that the function $a(f + g)$ should be the same as the function $af + ag$. Two functions are the same when they give the same value on a common input, so we verify:

$$[a(f + g)](x) = a(f + g)(x) = a(f(x) + g(x)) = af(x) + g(x)$$

and

$$(af + ag)(x) = (af)(x) + (ag)(x) = af(x) + ag(x),$$

which are the same. Thus $a(f + g)$ and $af + ag$ give the same output, so $a(f + g) = af + ag$.

However, rather than check every single tedious vector space property, we will assume we know already that $F(\mathbb{R})$ is a vector space over $\mathbb{R}$, which it is. (Checking all the properties will go along the lines of checking the distributive property above.) The point is that since U is a subset of $F(\mathbb{R})$, any properties which already hold in $F(\mathbb{R})$ will also hold automatically in U. Thus the associativity, commutativity, and distributive properties hold automatically for U. In fact, to verify that U is a vector space over $\mathbb{R}$ we really only need to check three things:

- $0 \in U$, where 0 denotes the zero vector of $F(\mathbb{R})$,
- if $f, g \in U$, then $f + g \in U$, and

- if $f \in U$ and $a \in \mathbb{R}$, then $af \in U$.

The first property is needed to guarantee that U does have a zero vector, and the second and third properties are needed to guarantee that addition and scalar multiplication on U make sense: we would not have a valid "addition" if adding two things in a space produced something not in that space, and similarly for scalar multiplication. After these properties are verified, the rest of the vector space properties hold automatically because they hold in the larger space $F(\mathbb{R})$. (There is one subtle point about the existence of additive inverses, which we'll point out afterwards.)

The zero vector of $F(\mathbb{R})$ is the *zero function*, which is the function 0 which sends everything to 0: $0(x) = 0$ for all $x \in \mathbb{R}$. This zero function is twice-differentiable (as all constant functions are), and its second derivative satisfies

$$0'' + 0 = 0 + 0 = 0,$$

so 0 satisfies the requirements needed to be in U. Thus U contains the zero vector of $F(\mathbb{R})$.

If $f, g \in U$, then f, g are each twice-differentiable and satisfy

$$f'' + f = 0 \text{ and } g'' + g = 0.$$

Thus $f + g$ is twice differentiable since $(f + g)'' = f'' + g''$, and

$$(f + g)'' + (f + g) = f'' + g'' + f + g = (f'' + f) + (g'' + g) = 0 + 0 = 0,$$

so $f + g$ satisfies the defining equation required of elements of U. Hence $f + g \in U$.

If $f \in U$ and $a \in \mathbb{R}$, then f is twice-differentiable and $f'' + f = 0$. Thus af is twice-differentiable since $(af)'' = af''$, and

$$(af)'' + (af) = af'' + af = a(f'' + f) = a0 = 0,$$

so $af \in U$ since it satisfies the required properties of elements of U. We conclude that U is indeed a vector space over $\mathbb{R}$.

What about negatives? For a vector space V, it is required that if $v \in V$, then its additive inverse $-v$ is also in V, and note that we didn't check this explicitly in the Warm-Up example. If we want to call U in the Warm-Up a vector space, we do in fact have to know that if $f \in U$, then $-f \in U$. However, the point is that this follows from the properties we showed above U does posses: we have shown previously that in any vector space $(-1)v = -v$ for any element v, so the fact that "if $f \in U$, then $-f \in U$" is encoded in the third property we listed above, namely that "if $f \in U$, then $af \in U$" for any $a \in \mathbb{R}$. Thus, this is not something we have to check separately.

What about zero? After the discussion above you might think: well then, why do we require that $0 \in U$ as the first condition in the Warm-Up since third condition implies that this is true by taking the scalar to be 0: if $f \in U$, then $0f \in U$ and $0f$ is the always the zero vector. The answer is: this is absolutely correct, and really the only reason we require $0 \in U$ separately is to guarantee that U is not *empty*, meaning that it actually has something in it.

Subspaces. Let V be a vector space over $\mathbb{K}$. A subset U of V is a *subspace* of V if it has the following properties:

- $0 \in U$,
- if $u, v \in U$, then $u + v \in U$ (we say that U is *closed under addition*), and

- if $u \in U$ and $a \in \mathbb{K}$, then $au \in U$ (we say that U is *closed under scalar multiplication*.

The point is that these properties alone guarantee U is itself a vector space over $\mathbb{K}$, since the other vector space axioms are automatically satisfied because they are already true in the "larger" space V.

The moral is: subspaces are just vector spaces which sit inside of other vector spaces.

Back to $f'' + f = 0$. So, the point of the Warm-Up was really to show that the space U of functions satisfying $f'' + f = 0$ is a subspace of $F(\mathbb{R})$. For example, $\sin x$ is one such function, and $\cos x$ is another. It turns out that *any* function f satisfying $f'' + f = 0$ is a linear combination of $\sin x$ and $\cos x$:

$$\text{there exist } a, b \in \mathbb{R} \text{ such that } f(x) = a \sin x + b \cos x.$$

Thus, we would say that $\sin x$ and $\cos x$ span U, so $U = \text{span}\{\sin x, \cos x\}$. You will verify this rigorously on Homework 7. More generally, such claims are also true for other types of differential equations, but you would need a full course on differential equations to prove this in general. The point for us is that this provides a connection between the study of differential equations and linear algebra.

Subspaces of $\mathbb{R}^n$. Wha are some subspaces of $\mathbb{R}^2$? One is $\{\mathbf{0}\}$, so the set which contains only the origin. Another is all of $\mathbb{R}^2$. Also, any line passing through the origin is a subspace of $\mathbb{R}^2$, since adding two points on such a line gives something on the same line, and scaling anything on such a line gives something on the same line. It turns out that this is it: the only subspaces of $\mathbb{R}^2$ are $\{\mathbf{0}\}$, lines through the origin, and $\mathbb{R}^2$ itself.

Note that a line which *doesn't* pass through the origin is not a subspace, since in particular it does not contain the zero vector. (Such lines are still special though, we'll see why next time.) The points on the parabola $y = x^2$ do not form a subspace either, since a parabola is not closed under addition nor scalar multplication.

More generally, the only subspaces of $\mathbb{R}^n$ are: $\{\mathbf{0}\}$, lines through the origin, planes through the origin, higher-dimensional analogues through the origin, hyperplanes through the origin, and $\mathbb{R}^n$ itself. We'll be able to prove this rigorously after we take about *bases*.

Another example. Let U be the subset of $P_n(\mathbb{C})$ consisting of all *even* polynomials, where a polynomial $p(x)$ is even if $p(-x) = p(x)$. We show that this is a subspace of $P_n(\mathbb{C})$.

First, the zero polynomial 0 is even since there are no x's to replace with $-x$ at all, so $p(-x) = p(x)$ becomes $0 = 0$, which is true. If $p(x), q(x) \in U$, then

$$p(-x) + q(-x) = p(x) + q(x)$$

since $p(x)$ and $q(x)$ are each even, so $p(x) + q(x)$ is even and is hence in U. Thus U is closed under addition. Finally, if $p(x) \in U$ is even and $a \in \mathbb{C}$, then

$$ap(-x) = ap(x),$$

so $ap(x)$ is even and hence in U, so U is closed under scalar multiplication. Thus U is a subspace of $P_n(\mathbb{C})$ as claimed.

What are some elements of U which span U? Le us think about the $n = 4$ case only Writing an arbitrary polynomial as

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4,$$

$p(x)$ is even if and only if

$$p(-x) = a_0 - a_1x + a_2x^2 - a_3x^3 + a_4x^4 = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4,$$

which by comparing coefficients requiers that $a_1 = -a_1$ and $a_3 = -a_3$. Thus a_1 and a_3 must be zero, so an even polynomial is of the form

$$p(x) = a_0 + a_2x^2 + a_4x^4,$$

or in other words, an even polynomial has only even degree terms. This shows that $1, x^2, x^4$ span U over $\mathbb{C}$, whereas $1, i, x^2, ix^2, x^4, ix^4$ would span U over $\mathbb{R}$.

Lecture 21: More on Subspaces

Subspaces are closed under linear combinations. Although the definition of a subspace only says that

$$x + y \in U \text{ if } x, y \in U,$$

it follows by induction that *any* sum of finitely many elements of U is still in U. This, together the being closed under scalar multiplication property, implies that U is closed under arbitrary linear combinations. So, we can also characterize subspaces of a vector space as those subsets which are closed under linear combinations:

$$\text{if } v_1, \ldots, v_k \in U \text{ and } c_1, \ldots, c_k \in \mathbb{K}, \text{ then } c_1v_1 + \cdots + c_kv_k \in U.$$

The point is that rather than having to check this more general property, the subspace definition gives us simpler conditions which are enough to check.

Warm-Up 1. Let U be the subset of $M_n(\mathbb{R})$ consisting of those matrices A which equal their own transpose:

$$U := \{A \in M_n(\mathbb{R}) \mid A^T = A\}.$$

Such a matrix is said to be *symmetric*, so U is the set of $n \times n$ symmetric matrices. We show that U is a subspace of $M_n(\mathbb{R})$.

We use the fact that the transpose of a sum is the sum of individual transposes:

$$(A + B)^T = A^T + B^T$$

and that the transpose of a scalar multiple of a matrix is the scalar multiple of the transpose:

$$(cA)^T = cA^T.$$

Both of these are not hard to see using the definition of the transpose as "switch rows and columns", but later we will see the *true* characterization of transposes and will see the "real" reason why transposes have these properties, in addition to the property that $(AB)^T = B^TA^T$.

Turning the rows of the zero matrix into columns still gives the zero matrix, so $0^T = 0$ and 0 is symmetric, so $0 \in U$. If $A, B \in U$, then $A^T = A$ and $B^T = B$, so

$$(A + B)^T = A^T + B^T = A + B,$$

and hence $A + B \in U$. Finally, if $A \in U$ and $c \in \mathbb{R}$, then

$$(cA)^T = cA^T = cA,$$

so cA equals its own transpose and thus $cA \in U$. Hence U is closed under scalar multiplication, so we conclude that the space of $n \times n$ symmetric matrices is a subspace of $M_n(\mathbb{R})$.

Concretely, in say the $n = 3$ case, a symmetric matrix must satisfy

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^T = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

which means that a symmetric matrix must look like

$$\begin{bmatrix} a & b & c \\ b & e & f \\ c & f & i \end{bmatrix}.$$

The name "symmetric" comes from the fact that the entries are symmetric across the diagonal.

Warm-Up 2. Suppose that V is a vector space over $\mathbb{K}$ and that U and W are subspaces of V. We show that the *intersection* $U \cap W$ of U and W is also a subspace of V, where the intersection of two sets is the set of things they have in common:

$$U \cap W := \{x \in V \mid x \in U \text{ and } x \in W\}.$$

First, since U is a subspace of V we know that $0 \in U$, and since W is a subspace of V we know that $0 \in W$. Thus $0 \in U \cap W$, so $U \cap W$ contains the zero vector of V. Next, suppose that $x, y \in U \cap W$. This means that $x, y \in U$ and $x, y \in W$. Since U and W are each closed under addition, we have

$$x + y \in U \text{and } x + y \in W.$$

Thus $x + y \in U \cap W$, so $U \cap W$ is closed under addition. Finally, suppose that $x \in U \cap W$ and $a \in \mathbb{K}$. Since U and W are each closed under scalar multiplication, we have

$$ax \in U \text{ and } ax \in W$$

so $ax \in U \cap W$. We conclude that $U \cap W$ is a subspace of V.

For instance, a plane through the origin in $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$, and the intersection of two such planes, which might be say a line through the origin, is also a subspace of $\mathbb{R}^3$. More generally, if we have subspaces defined as the set of solutions to some homogeneous equations, their intersection is the set of solutions to the system obtained by putting all equations together.

Exercise. The *union* $U \cup W$ of two subspaces U and W of a vector space V is the set of things you get when you throw everything in U together with everything in W:

$$U \cup W := \{x \in V \mid x \in U \text{ or } x \in W\}.$$

The union of two subspaces is *not* necessarily itself a subspace. Indeed, consider the x-axis and y-axis in $\mathbb{R}^2$—each of these are subspaces but their union is not. But, it might be possible for $U \cup W$ to indeed be a subspace; for instance, this happens if $U = W$.

Show that $U \cup W$ is a subspace of V if and only if $U \subseteq W$ or $W \subseteq U$. (The notation $A \subseteq B$ means that A is a *subset* of B, which means that everything in A is also in B. So, the claim is that $U \cup W$ is a subspace of V if and only if one of U or W is fully contained in the other.)

Spans are subspaces. Let V be a vector space over $\mathbb{K}$ and let $v_1, \dots, v_k \in V$. Then $\operatorname{span}\{v_1, \dots, v_k\}$ is a subspace of V. Moreover, $\operatorname{span}\{v_1, \dots, v_k\}$ is the *smallest* subspace of V containing $v_1, \dots, v_k$, whereby "smallest" we mean that if U is *any* subspace of V containing $v_1, \dots, v_k$, then

$$\operatorname{span}\{v_1, \dots, v_k\} \subseteq U,$$

so any subspace containing $v_1, \dots, v_k$ must contain their entire span. (This is what is special about spans.)

Proof. Since

$$0_V = 0v_1 + \cdots + 0v_k,$$

the zero vector 0_V of V is in $\operatorname{span}\{v_1, \dots, v_k\}$. (Note that here we use a subscript to denote the space of which 0 is the zero vector. This is to not confuse the two uses of "0" in this expression: on the left the zero vector, whereas on the right 0 is denoting the scalar $0 \in \mathbb{K}$. The right side equals the zero vector since $0v = 0_V$ for any v in any vector space, so the right side is just a sum of a bunch of zero vectors.)

If $x, y \in \operatorname{span}\{v_1, \dots, v_k\}$, then each of x and y is a linear combination of $v_1, \dots, v_k$:

$$x = c_1v_1 + \cdots + c_kv_k \text{ and } y = d_1v_1 + \cdots + d_kv_k$$

for some $c_1, \dots, c_k, d_1, \dots, d_k \in \mathbb{K}$. Then

$$x + y = (c_1 + d_1)v_1 + \cdots + (c_k + d_k)v_k,$$

so $x + y \in \operatorname{span}\{v_1, \dots, v_k\}$, and hence spans are closed under addition.

Finally, using the same x as above, if $a \in \mathbb{K}$ then

$$ax = a(c_1v_1 + \cdots + c_kv_k) = (ac_1)v_1 + \cdots + (ac_k)v_k \in \operatorname{span}\{v_1, \dots, v_k\},$$

so spans are closed under scalar multiplication. We conclude that $\operatorname{span}\{v_1, \dots, v_k\}$ is a subspace of V as claimed.

If U is any subspace of V containing $v_1, \dots, v_k$, the fact that U is closed under linear combinations says that

$$c_1v_1 + \cdots + c_kv_k \in U$$

for any $c_1, \dots, c_k \in \mathbb{K}$. But such expressions make up all elements of $\operatorname{span}\{v_1, \dots, v_k\}$, so U contains everything in this span as required in order to say that $\operatorname{span}\{v_1, \dots, v_k\}$ is the smallest subspace of V containing $v_1, \dots, v_k$. □

Kernels are subspaces. The *kernel* of an $m \times n$ matrix is the set $\ker A$ of solutions to $A\mathbf{x} = \mathbf{0}$:

$$\ker A := \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}.$$

Weeks ago when talking about homogeneous equations of this form we in fact showed that the set of solutions of such an equation is always closed under linear combinations, which in our new language means that $\ker A$ is a subspace of $\mathbb{R}^n$.

For instance, consider the following subset of $\mathbb{R}^3$:

$$U := \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; 2x = y + z \text{ and } 3y = 2z \right\}.$$

It can be checked that this is a subspace of $\mathbb{R}^3$ by verifying the definition of subspace directly, or we can simply note that this space is the same as the kernel of

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

Indeed, a vector is in U if and only it satisfies the system

$$\begin{aligned} 2x - \ y - \ z &= 0 \\ 3y - 2z &= 0 \end{aligned},$$

which is true if and only if it satisfies the matrix equation

$$\begin{bmatrix} 2 & -1 & -1 \\ 0 & 3 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since U is the kernel of a matrix, it is automatically a subspace.

Affine subspaces. Consider now the following subset

$$W := \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \;\middle|\; 2x = y + z + 1 \text{ and } 3y = 2z + 2 \right\}$$

of $\mathbb{R}^3$. This is the same as the set of solutions of the inhomogeneous equation

$$\begin{bmatrix} 2 & -1 & -1 \\ 0 & 3 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

When discussing solutions of inhomogeneous equations previously we saw that the set of such solutions is NOT closed under arbitrary linear combinations, it is only closed under affine combinations.

We say that a nonempty subset W of a vector space V over $\mathbb{K}$ is an *affine subspace* of V if it is closed under affine combinations: if $v_1, \ldots, v_k \in W$ and $c_1, \ldots, c_k \in \mathbb{K}$ satisfy $c_1 + \cdots + c_k = 1$, then $c_1 v_1 + \cdots + c_k v_k \in W$. Thus, the set W above is an affine subspace of $\mathbb{R}^3$. More generally, the set of solutions of an inhomogeneous equation $A\mathbf{x} = \mathbf{b}$, where A is $m \times n$, is an affine subspace of $\mathbb{R}^n$. For instance, any line in $\mathbb{R}^2$ or $\mathbb{R}^3$ which doesn't pass through the origin is an affine subspace, as is any plane in $\mathbb{R}^3$ which doesn't pass through the origin.

Sometimes to distinguish between an "affine subspace" and our previous notion of subspace, we will use the term *linear subspace* to mean what we previously referred to as simply "subspace". By default, if we use the term subspace without any adjective in front, we will take it to always mean linear subspace—whenever we want to refer to an affine subspace we will always use the adjective "affine" in front.

Lecture 22: Bases

Warm-Up. Suppose that V is a vector space over $\mathbb{K}$ and that U is a (linear) subspace of V. Let $b \in V$ and define $b + U$ to be the set of all things in V obtained by adding b to elements of U:

$$b + U := \{b + u \mid u \in U\}.$$

(We say that $b+U$ is the *translation* of U by b.) We show that $b+U$ is an affine subspace of V.

First, since U is a subspace of V, $0 \in U$ so $b = b+0 \in b+U$. Hence $b+U$ is nonempty. Now, suppose that $v_1, \dots, v_k \in b+U$ and that $c_1, \dots, c_k \in \mathbb{K}$ are scalars such that $c_1 + \cdots + c_k = 1$. Since each v_i is in $b+U$, we can write each v_i as

$$v_i = b + u_i \text{ for some } u_i \in U$$

by the form elements of $b+U$ take. Thus:

$$\begin{aligned} c_1 v_1 + \cdots + c_k v_k &= c_1(b+u_1) + \cdots + c_k(b+u_k) \\ &= (c_1 + \cdots + c_k)b + (c_1 u_1 + \cdots + c_k u_k) \\ &= b + (c_1 u_1 + \cdots + c_k u_k). \end{aligned}$$

Since U is a linear subspace of V, it is closed under arbitrary linear combinations so

$$c_1 u_1 + \cdots + c_k u_k \in U.$$

Thus the expression above for $c_1 v_1 + \cdots + c_k v_k$ is of the form "b plus something in U", so

$$c_1 v_1 + \cdots + c_k v_k \in b+U.$$

Hence $b+U$ is closed under affine combinations, so it is an affine subspace of V.

Exercise. Show that *any* affine subspace of V is of the form $b+U$ for some $b \in V$ and some (linear) subspace U of V. (This is on Homework 7.)

Affine = linear plus translation. As a consequence of the Warm-Up, any line, plane, hyperplane, other higher-dimensional analogou in $\mathbb{R}^n$ which do not pass through the origin are all affine subspaces of $\mathbb{R}^n$ since each of these can be described by translating a line, plane, hyperplane, other higher-dimensional analogue which *does* pass through the origin. More generally, the set of solutions to an inhomogeneous equation $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix, is an affine subspace of $\mathbb{R}^n$ since it is obtained by translating the set of solutions of the homogeneous equations $A\mathbf{x} = \mathbf{0}$ by a particular solution of $A\mathbf{x} = \mathbf{b}$.

Every usage of the word "affine" we have seen so far in this class fits into the general idea that affine things are linear things plus a translation. This is true for affine subspaces as we see from the Warm-Up and related Exercise, it is true for the affine transformations defined on Homework 4 (where the result was that affine transformations are of the form $T(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, and it is true when considering the relation between affinely independent versus linearly independent vectors as in Problem 4 of Homework 4; Problem 4 of Homework 7 also relates to this. The point is that affine things are things we normally would have called "linear" in other contexts, only they aren't really linear given the restrictive definition of "linear" we now use; affine things are things which are "essentially" linear up to translation.

Finite-dimensionality. We say a vector space V over $\mathbb{K}$ is *finite-dimensional* if there are finitely many vectors in V which span V. For instance, each of $\mathbb{K}^n, P_n(\mathbb{C}), M_{m,n}(\mathbb{K})$ are finite-dimensional, as are any subspaces of these. We say that V is *infinite-dimensional* over $\mathbb{K}$ if it is not finite-dimensional over $\mathbb{K}$.

For an example of something which is not finite-dimensional, consider the space $P(\mathbb{K})$ of *all* polynomials with coefficients in $\mathbb{K}$, where we place no restriction on the degree as opposed to $P_n(\mathbb{K})$. (Note that the term "polynomial" still requires that each polynomial have finitely many terms.)

We claim that $P(\mathbb{K})$ is not finite-dimensional, meaning that no finite collection of polynomials can span all of $P(\mathbb{K})$.

Indeed, let $p_1(x), \ldots, p_k(x)$ be any finite number of polynomials in $P(\mathbb{K})$. Any linear combination of these:

$$c_1 p_1(x) + \cdots + c_k p_k(x)$$

has degree at most equal to the maximum of the degrees of the individual $p_i(x)$:

$$\deg(c_1 p_1(x) + \cdots + c_k p_k(x)) \leq \max\{\deg p_1(x), \ldots, \deg p_k(x)\},$$

since no higher powers of x can occur simply as a result of adding scalar multiples of the $p_i(x)$. Thus no polynomial with a degree larger than this maximum can be in the span of $p_1(x), \ldots, p_k(x)$, so these polynomials do not span $P(\mathbb{K})$. This was an arbitrary finite number of polynomials, we conclude that no finite number of polynomials can span $P(\mathbb{K})$, so $P(\mathbb{K})$ is infinite-dimensional.

Bases. Consider the plane $x + y + z = 0$, which defines a subspace of $\mathbb{R}^3$. This is spanned by, say:

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \\ 3 \end{bmatrix},$$

but of course this is not the smallest possible spanning set since the first two vectors alone also span the plane. The issue is that the four vectors we gave are linearly dependent, and we have seen that whenever we have linearly dependent vectors we can throw away any which are linear combinations of the others without changing the overall span. In this case, we can throw away the third and fourth vectors—which are each linear combinations of the first two—and still be left with vectors which span the plane. The first two vectors are linearly independent, and thus it is not possible to throw one away and still have the remaining vector span the plane.

We say that a collection S of vectors in a (finite or infinite-dimensional) vector space V is a *basis* for V if:

- the vectors in S span V, and
- the vectors in S are linearly independent.

A basis for a vector space can be thought of as a "minimal" spanning set for the space, and also as a "maximal" linearly independent set. (We'll make these notions precise later on.) Thus, in the example above, the vectors

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

form a basis for the plane $x + y + z = 0$ in $\mathbb{R}^3$. There are many other possible bases for this same plane, and indeed for vector spaces in general.

Example. Consider the space V of 3×3 symmetric matrices over $\mathbb{R}$:

$$V := \{A \in M_3(\mathbb{R}) \mid A^T = A\}.$$

We find a basis for this space. First, note that any 3×3 symmetric matrix has the form

$$\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \text{ where } a, b, c, d, e, f \in \mathbb{R}.$$

Think of the scalars a, b, c, d, e, f as "free variables" in the sense that there is no restriction placed on them. Factoring out these "free variables" gives

$$\begin{aligned}\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} &= a \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &\quad + c \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &\quad + e \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} + f \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},\end{aligned}$$

which shows that any 3×3 symmetric matrix can be written as a linear combination of the six matrices on the right, so these six matrices span V. Call these matrices $A_1, \ldots, A_6$ in the order in which they appear.

To check that these six matrices are linearly independent, we suppose $c_1, \ldots, c_6 \in \mathbb{R}$ are scalars such that

$$c_1 A_1 + \cdots + c_6 A_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and show that all c_i have to be zero. First, A_1 is the only matrix which could give a nonzero entry in the upper-left corner of the sum on the left, so since this entry must be 0 in the zero matrix we must have $c_1 = 0$. Since A_2 is the only matrix which could possibly give a nonzero entry in either the first row, second column position or the second row, first columns position and each of these must be zero in the zero matrix, we must have $c_2 = 0$. Continuing in a similar way we conclude that all c_i are zero, where the point is that each of these matrices has a 1 in a location where all the others have a 0. Thus $A_1, \ldots, A_6$ are linearly independent, so they form a basis of V.

Another example. Let V be the space of polynomials $p(x)$ in $P_3(\mathbb{K})$ satisfying $p(2) = 0$:

$$V := \{p(x) \in P_2 3(\mathbb{K}) \mid p(2) = 0\}.$$

We claim that $(x-2), (x-2)^2, (x-2)^3$ form a basis of V. First, to show that these are linearly independent, suppose that $c_1, c_2, c_3 \in \mathbb{K}$ satisfy

$$c_1(x-2) + c_2(x-2)^2 + c_3(x-2)^3 = 0.$$

Note that the third term is the only which could give degree 3 term, which is $c_3 x^3$. Since the right side 0 has no degree 3 term, we must have that the left side shouldn't either, so $c_3 = 0$. We're left with

$$c_1(x-2) + c_2(x-2)^2 = 0.$$

The second term is the only which could give a degree 2 term, which $c_2 x^2$. Since the right side has no such term, we must have $c_2 = 0$. Thus we get

$$c_1(x-2) = 0.$$

Since the right side has no x term at all, we must have $c_1 = 0$, so all coefficients in our original equation are 0 and hence $(x-2), (x-2)^2, (x-3)^3$ are linearly independent.

Now we show that these three polynomials span V, simply by brute force. (We'll see "better" ways of showing something like this soon which can avoid a lot of messy computations.) Suppose that

$$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \in V,$$

so $p(x)$ satisfies $p(2) = 0$. We must show that $p(x)$ can be written as a linear combination of $(x-2), (x-2)^2, (x-3)^3$, or in other words that there exist scalars $c_1, c_2, c_3 \in \mathbb{K}$ such that

$$a_0 + a_1x + a_2x^2 + a_3x^3 = c_1(x-2) + c_2(x-2)^2 + c_3(x-2)^3.$$

Multiplying out the right side gives

$$(-2c_1 + 4c_2 - 8c_3) + (c_1 - 4c_2 + 12c_3)x + (c_2 - 6c_3)x^2 + c_3x^3,$$

so comparing coefficients with those of $p(x)$ shows that c_1, c_2, c_3 must satisfy

$$-2c_1 + 4c_2 - 8c_3 = a_0,\ c_1 - 4c_2 + 12c_3 = a_1,\ c_2 - 6c_3 = a_2,\ c_3 = a_3.$$

Also, since $p(2) = 0$, we must have

$$a_0 + 2a_1 + 4a_2 + 8a_3 = 0,\ \text{so } a_0 = -2a_1 - 4a_2 - 8a_3.$$

Thus c_1, c_2, c_3 must satisfy the linear system:

$$\begin{aligned} -2c_1 + 4c_2 &= -2a_1 - 4a_2 - 8a_3 \\ c_1 - 4c_2 + 12c_3 &= a_1 \\ c_2 - 6c_3 &= a_2 \\ c_3 &= a_3. \end{aligned}$$

Solving for this c_1, c_2, c_3 shows that there is a solution, meaning that there are scalars c_1, c_2, c_3 such that

$$p(x) = c_1(x-2) + c_2(x-2)^2 + c_3(x-2)^3.$$

Thus $(x-2), (x-2)^2, (x-2)^3$ span V, so these three polynomials form a basis of V.

For those of you who had Calculus BC or know something about *Taylor series*, note that the type of expression we want in this case:

$$p(x) = c_1(x-2) + c_2(x-2)^2 + c_3(x-2)^3$$

takes the form of a Taylor series centered at 2. From this point of view, the general theory of Taylor series will say write away that such coefficients c_1, c_2, c_3 exist, and specifically it turns out that

$$c_k = \frac{p^{(k)}(2)}{k!}(x-2)^k$$

where $p^{(k)}$ denotes the k-th derivative of p. This is NOT the way I had in mind above when I said we'll be able to show that $(x-2), (x-2)^2, (x-2)^3$ span V in a simpler way soon, so no worries if you're unfamiliar with Taylor series.

Lecture 23: Dimension

Warm-Up. Let V be a finite dimensional vector space over $\mathbb{K}$. We show that a collection $v_1, \ldots, v_n \in V$ forms a basis for V if and only if for every $v \in V$ there exist *unique* scalars $c_1, \ldots, c_n \in \mathbb{K}$ such that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

The key word here is "unique", with the point being that bases provide a unique way of expressing each element of V, as opposed to spanning sets in general where they may be more than one way of writing an element of V as such a linear combination.

Suppose first that $v_1, \ldots, v_n$ is a basis of V and let $v \in V$. Since $v_1, \ldots, v_n$ span V, there exist scalars $c_1, \ldots, c_n \in \mathbb{K}$ such that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

Suppose now that $d_1, \ldots, d_n \in \mathbb{K}$ are possibly *different* scalars such that

$$v = d_1 v_1 + \cdots + d_n v_n.$$

We must show that each d_i is the same as the corresponding c_i. Since

$$c_1 v_1 + \cdots + c_n v_n = d_1 v_1 + \cdots + d_n v_n,$$

we have

$$(c_1 - d_1) v_1 + \cdots + (c_n - d_n) v_n = 0.$$

Since $v_1, \ldots, v_n$ are linearly independent, each of the coefficients above must be zero, so

$$c_i - d_i = 0, \text{ and hence } c_i = d_i \text{ for each } i.$$

Thus the scalars needed to express v as a linear combination of $v_1, \ldots, v_n$ are unique as claimed.

Conversely suppose that $v_1, \ldots, v_n$ have the property that each $v \in V$ can be written as a linear combination of $v_1, \ldots, v_n$ in a unique way. Saying that each $v \in V$ can be written as a linear combination of $v_1, \ldots, v_n$ (forgetting about uniqueness for now) means that $v_1, \ldots, v_n$ span V. To how that $v_1, \ldots, v_n$ are linearly independent, suppose that $c_1, \ldots, c_n \in \mathbb{K}$ are scalars such that

$$c_1 v_1 + \cdots + c_n v_n = 0.$$

We must show that each c_i is zero. Since we can also express the zero vector as

$$0 = 0v_1 + \cdots + 0v_n,$$

we have

$$c_1 v_1 + \cdots + c_n v_n = 0v_1 + \cdots + 0v_n.$$

The uniqueness of the scalars needed to express 0 as a linear combination of $v_1, \ldots, v_n$ now guarantees that $c_i = 0$ for all i, so $v_1, \ldots, v_n$ are linearly independent and hence form a bass for V as claimed.

Linearly dependent means non-uniqueness. To be clear, here is an example where a linearly dependent spanning set gives rise to non-unique linear combinations. Consider the subspace of $\mathbb{R}^3$ consisting of all vectors of the form

$$\begin{bmatrix} a \\ 2a \\ b \end{bmatrix}.$$

This subspace is spanned by

$$\begin{bmatrix}1\\2\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\2\\1\end{bmatrix}.$$

When expressing a vector in this subspace as a linear combination of these spanning vectors, there are infinitely many ways of doing so, here are a few:

$$\begin{aligned}\begin{bmatrix}a\\2a\\b\end{bmatrix} &= a\begin{bmatrix}1\\2\\0\end{bmatrix} + b\begin{bmatrix}0\\0\\1\end{bmatrix} + 0\begin{bmatrix}1\\2\\1\end{bmatrix}\\ &= (a-b)\begin{bmatrix}1\\2\\0\end{bmatrix} + (b-a)\begin{bmatrix}0\\0\\1\end{bmatrix} + b\begin{bmatrix}1\\2\\1\end{bmatrix}\\ &= \left(a-\frac{b}{2}\right)\begin{bmatrix}1\\2\\0\end{bmatrix} + \frac{b}{2}\begin{bmatrix}0\\0\\1\end{bmatrix} + \frac{b}{2}\begin{bmatrix}1\\2\\1\end{bmatrix}.\end{aligned}$$

This point is that these three vectors do not form a basis for this space since they are linearly dependent. Throwing away the third vector does leave us with a basis, and the uniqueness property given in the Warm-Up will hold.

Size of a basis. We are working up towards proving that any bases of a finite dimensional vector space must consist of the same number of vectors. Bases themselves aren't unique: a given vector space has infinitely many possible bases. For instance,

$$\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} \quad \left\{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}3\\1\end{bmatrix}\right\} \quad \left\{\begin{bmatrix}1\\-20\end{bmatrix}, \begin{bmatrix}\pi\\e\end{bmatrix}\right\}$$

are three different bases for $\mathbb{R}^2$. Note that each of these contain two vectors, which will be true of any basis of $\mathbb{R}^2$. This will allow us to finally give a precise meaning to the word "dimension".

Linearly independent size vs spanning size. First we show that the size of any linearly independent set of vectors in V must be less than or equal to the size of any spanning set of V. To be clear, the claim is that if $v_1, \ldots, v_k \in V$ are linearly independent and $w_1, \ldots, w_m \in V$ span V, then $k \leq m$. This specific proof is due to John Alongi.

Suppose to the contrary that $k > m$. Since $w_1, \ldots, w_m$ span V, we can write each v_i as a linear combination of $w_1, \ldots, w_m$:

$$v_i = a_{1i}w_1 + \cdots + a_{mi}w_m \text{ for some } a_{1i}, \ldots, a_{mi} \in \mathbb{K}.$$

Consider the $m \times k$ matrix A whose i-th column consists of the m scalars $a_{1i}, \ldots, a_{mi}$:

$$A = \begin{bmatrix}a_{11} & \cdots & a_{1k}\\ \vdots & \ddots & \vdots\\ a_{m1} & \cdots & a_{mk}\end{bmatrix}.$$

Since $k > m$, this matrix has more columns than rows and hence there exist $x_1, \ldots, x_k \in \mathbb{K}$, at least one of which is nonzero, such that

$$\begin{bmatrix}a_{11} & \cdots & a_{1k}\\ \vdots & \ddots & \vdots\\ a_{m1} & \cdots & a_{mk}\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_k\end{bmatrix} = \begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}.$$

(In other words, the system $A\mathbf{x} = \mathbf{0}$ has a nonzero solution since $\operatorname{rref}(A)$ has at least one column without a pivot.)

For the scalars $x_1, \dots, x_k$ above, we have (using the previous linear combination expressions we had for the v_i):

$$\begin{aligned} x_1v_1 + \cdots + x_kv_k &= x_1(a_{11}w_1 + \cdots + a_{m1}w_m) + \cdots + x_k(a_{1k}w_1 + \cdots + a_{mk}w_m) \\ &= (x_1a_{11} + \cdots + x_ka_{1k})w_1 + \cdots + (x_1a_{m1} + \cdots + x_ka_{mk})w_m. \end{aligned}$$

But the coefficients in this final expression are precisely the entries of the product

$$\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mk} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix},$$

which are all zero by the choice of $x_1, \dots, x_k$. Thus

$$x_1v_1 + \cdots + x_kv_k = 0w_1 + \cdots + 0w_m = 0,$$

showing that $v_1, \dots, v_k$ are linearly dependent since at least one of the x_i is nonzero. This contradicts our assumption that $v_1, \dots, v_k$ are linearly independent, so we must have $k \leq m$ as claimed.

Remark. Note that what we have actually shown in the proof above is that whenever we have more vectors than the number of vectors in a spanning set, those vectors must be linearly dependent. In particular, any list containing more vectors than that in a basis for a vector space must be linearly dependent.

All bases have the same size. We can now show that any two bases of a finite dimensional vector space must have the same number of vectors. Suppose that V is finite dimensional and that $v_1, \dots, v_n$ and $w_1, \dots, w_m$ are both bases of V. Noting first that $v_1, \dots, v_n$ are linearly independent and that $w_1, \dots, w_m$ span V, the result above shows that $n \leq m$. Flipping things around noting that $w_1, \dots, w_m$ are linearly independent and that $v_1, \dots, v_n$ span V, we also get $m \geq n$. Since $n \leq m$ and $m \leq n$, we conclude that $m = n$ as claimed.

Dimension. We define the *dimension* $\dim V$ of a vector space to be the number of vectors in a basis for V. The fact that all bases have the same number of vectors shows that this definition makes sense. Sometimes to emphasize the types of scalars we are considering, we use $\dim_{\mathbb{K}} V$ to denote the dimension of V considered as a vector space over $\mathbb{K}$. Note that $\mathbb{K}$ matters: a problem on Homework 7 shows that if $\dim_{\mathbb{C}} V = n$, then $\dim_{\mathbb{R}} V = 2n$.

So far this definition of "dimension" only makes sense for finite dimensional spaces. We'll say some things about bases for infinite dimensional spaces next time. Note also that the dimension of V is essentially the number of "independent parameters" needed to specify an element of V; for instance, the space of 3×3 symmetric matrices is 6-dimensional since such a matrix must of the form

$$\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix},$$

and so it requires 6 independent values a, b, c, d, e, f to specify a 3×3 symmetric matrix.

The zero vector space. We define the dimension of $\{0\}$ to be zero. After all, what could a basis for $\{0\}$ actually be? A first guess might be to say that 0 itself forms a basis, but it seems strange

that we should consider a single point to be "1-dimensional". Indeed, any set of vectors containing 0 is actually linearly dependent *by convention*, even when 0 is the only vector in that set.

So, we end up with the strange-looking fact that a basis for $\{0\}$ is given by the empty set! This basis has zero elements, so $\dim\{0\} = 0$. Really, we say that the empty set gives a basis of $\{0\}$ just so that saying $\dim\{0\} = 0$ makes sense, but you should take this to be simply a quirk of the definition of *dimension*. Apart from considering a single point to be a 0-dimensional object geometrically, we'll see other reasons why we should consider $\dim\{0\}$ to be 0, namely to make certain formulas work out nicely.

Lecture 24: More on Bases and Dimension

Old problems are now easy. Before the Warm-Up, let's point out that many problems we considered before, say on homeworks, are now relatively straightforward. For instance, consider a hyperplane

$$a_1x_1 + \cdots + a_nx_n = 0$$

passing through the origin in $\mathbb{R}^n$. Problem 10b on Homework 3 actually shows that this hyperplane is $(n-1)$-dimensional, and Problem 10c, which was a difficult problem back then, is not just a consequence of the fact that fewer vectors than the dimension of a vector space cannot possibly span that space since any spanning set must have at least as many vectors as that in a basis. I encourage you to go back through old problems and examples and see how many of them can be rephrased using newer material.

Warm-Up. We show that a vector space V over $\mathbb{K}$ is infinite dimensional if and only if there exists an infinite collection $v_1, v_2, v_3, \ldots$ of linearly independent vectors in V. Recall that our definition of "infinite dimensional" is "not finite dimensional", so "can't be spanned by finitely many vectors".

Suppose that V is infinite dimensional. Then $V \neq \{0\}$, so there exists a nonzero vector $v_1 \in V$. Since V is infinite dimensional, v_1 alone does not span V, so there exists $v_2 \in V$ such that v_2 is not in $\operatorname{span}\{v_1\}$. Then v_1, v_2 are linearly independent. Again, these two cannot span V since V is not finite dimensional, so there exists $v_3 \in V$ not in $\operatorname{span}\{v_1, v_2\}$. Then v_1, v_2, v_3 are linearly independent, and continuing in this manner we end up with infinitely many linearly independent vectors. (To be a little more precise, suppose by induction that we have constructed so far k linearly independent vectors $v_1, \ldots, v_k$. These do not span V, so picking v_{k+1} not in their span gives $k+1$ linearly independent vectors. Thus we can find n linearly independent vectors for any n, which all together gives infinitely many such vectors.)

Conversely suppose that there exists a list of linearly independent vectors $v_1, v_2, v_3, \ldots$ in V. If V was finite dimensional, there would exist finitely many vectors $w_1, \ldots, w_k$ in V which span V, but then $v_1, \ldots, v_{k+1}$ would be a list of linearly independent vectors containing more elements than the spanning set $w_1, \ldots, w_k$. This is not possible, so V must be infinite dimensional.

$\mathbb{K}^\infty$ is infinite dimensional. We can now show that $\mathbb{K}^\infty$ is infinite dimensional. This is an immediate consequence of the Warm-Up once we note that

$$e_1,\ e_2,\ e_3,\ e_4, \ldots$$

is a list of infinitely many linearly independent vectors of $\mathbb{K}^\infty$, where e_i denotes the vector with a 1 in the i-th location and zeroes elsewhere.

Now, as discussed a while ago, the vectors $e_1, e_2, e_3, \ldots$ do NOT span $\mathbb{K}^\infty$ since it is not true that everything in $\mathbb{K}^\infty$ can be written as a linear combination of *finitely* many of these vectors,

which is a requirement when we talk about the span of an infinite set of vectors. Thus, $e_1, e_2, \dots$ do not give a basis for $\mathbb{K}^\infty$. Contrast this with the space of polynomials $P(\mathbb{K})$, for which

$$1,\ x,\ x^2,\ x^3, \dots$$

does give an honest basis. (The point is that any polynomial only has finitely many terms, whereas an element of $\mathbb{K}^\infty$ might contain infinitely many nonzero entries.)

So, what would a basis for $\mathbb{K}^\infty$ actually be? Does it even have a basis? It turns out one can show, using some advanced results in *set theory*, that *every* infinite dimensional vector space indeed *does* have a basis! (We'll show in a bit without much trouble that finite dimensional spaces always have bases.) However, the amazing thing is that even though we know $\mathbb{K}^\infty$ has a basis, it is NOT possible to actually write one down explicitly, which is kind of mind-blowing. We won't discuss this further in this course, but this is related to the notion of *cardinality* and the idea one infinite set can in a sense be "larger" than another infinite set. To be precise, the claim is that $\mathbb{K}^\infty$ has *uncountable* dimension, whereas something like $P(\mathbb{K})$ has *countable* dimension. Look up "countable" and "uncountable" to learn more, or we can talk about it in office hours sometime. The proof that infinite dimensional spaces have bases uses what's called the *axiom of choice*, which we can also talk about in office hours if you're interested. Good stuff.

Spanning sets can be reduced to bases. Suppose that $v_1, \dots, v_k$ is a basis for a vector space V. If this vectors are linearly independent, they give a basis for V. If they are linearly dependent, one is a linear combination of the others, so throwing it away still results in vectors which span V. If this smaller list is linearly independent, it gives a basis for V, and if not we can throw out another vector and still be left with spanning vectors. Continuing in this way we eventually end up with a basis for V, so the result is that any set of vectors which spans V can be *reduced* to a basis.

Finite dimensional spaces have bases. In particular, if V is finite dimensional, there are finitely many vectors $v_1, \dots, v_k$ which span V. By the result above, this list can be reduced to a basis, so V has a basis. This basis will contain finitely many vectors, which justifies our use of the term "finite dimensional" to describe spaces which can be spanned by finitely many vectors.

Linearly independent sets can be extended to bases. Suppose that $v_1, \dots, v_k$ is a list of linearly independent vectors in a finite dimensional space V. If these vectors span V, they form a basis. If not, there is a vector $v_{k+1} \in V$ which is not in their span, and then $v_1, \dots, v_k, v_{k+1}$ is still linearly independent. If these span V they form a basis, and if not we can find another vector to append to end up with $v_1, \dots, v_k, v_{k+1}, v_{k+2}$ being linearly independent. Continuing in this way we eventually end up with linearly independent vectors $v_1, \dots, v_n$ which span V, so they give a basis for V. (This process must stop since V is not infinite dimensional, so it is not possible to find a list of infinitely many linearly independent vectors in V.) This shows that any linearly independent list of vectors in V can be "extended" to a basis of V.

Subspaces of finite dimensional spaces. Finally, we justify some seemingly "obvious" facts, but which aren't actually obvious at all. Note the way that we use what we've built up to actually prove these facts.

Suppose that V is finite dimensional and that U is a subspace of V. First we claim that U itself must be finite dimensional. If U instead were infinite dimensional, by the Warm-Up we would be able to find infinitely any linearly independent vectors in U, but seeing as these would also be linearly independent vectors in V, this contradicts the fact that V is finite dimensional. Thus U must be finite dimensional.

Second, we claim that $\dim U \leq \dim V$. Let $u_1, \ldots, u_k$ be a basis for U. Since these vectors are linearly independent in V as well, they can be extended to a basis

$$u_1, \ldots, u_k, v_1, \ldots, v_\ell$$

for V. But then $\dim U = k \leq k + \ell = \dim V$, as claimed. In the fact, the only subspace of V which has the same dimension as V is V itself: if U is a subspace of a finite dimensional vector space V and $\dim U = \dim V$, then $U = V$.

Lecture 25: Linear Transformations Redux

Warm-Up. Suppose that V is a finite dimensional vector space, with $n = \dim V$. Let $v_1, \ldots, v_n \in V$. We justify the fact that if $v_1, \ldots, v_n$ are linearly independent, then they automatically span V and so form a basis, and if instead $v_1, \ldots, v_n$ span V, then they are automatically linearly independent and so form a basis. We've seen previously that these facts are true of n vectors in $\mathbb{R}^n$, but now we are saying that they hold in any finite dimensional space.

Suppose $v_1, \ldots, v_n$ are linearly independent. Then this list can be extended to a basis of V. But since any basis of V must contain $n = \dim V$ vectors, the resulting basis obtained by extending this list must be $v_1, \ldots, v_n$ itself, so these vectors already form a basis of V and hence span V. Said another way: if $v_1, \ldots, v_n$ didn't span V, extending them to a basis would give a basis with more than n vectors, which is not possible.

Suppose $v_1, \ldots, v_n$ span V. Then this list can be reduced to a basis of V. But since any basis of V must contain n vectors, the basis obtained after reducing must be $v_1, \ldots, v_n$ itself, so these vectors already form a basis and hence are linearly independent. Said another way: if $v_1, \ldots, v_n$ weren't linearly independent, reducing this list to a basis would yield a basis with fewer than n vectors, which is not possible.

Affine dimension. To finish up our material on dimension, we give a definition of the affine dimension of an affine subspace of V. If W is an affine subspace of V, we know from Homework 7 that we can write W as the translate of a linear subspace U: $W = b + U$ for some $b \in W$. We define the *affine dimension* of W to be the (linear) dimension of U. With this definition, we can now give a completely general definition of the terms "line", "plane", and "hyperplane". A *line* in a vector space V is a 1-dimensional affine/linear subspace of V, a *plane* in V is a 2-dimensional affine/linear subspace of V, and a *hyperplane* in V is an $(n-1)$-dimensional affine/subspace of V where $n = \dim V$. These definitions agree with the usual concept of "line" and "plane" in $\mathbb{R}^3$.

Note that the affine dimension of W is NOT the number of vectors in an affine basis for W as defined on Homework 7—it is one *less* than the number of vectors in an affine basis. Check the solutions to Homework 7 to see a further discussion of this, and note that Problem 1b on Homework 1 was really about showing that a line which doesn't pass through the origin in $\mathbb{R}^2$ has an affine basis consisting of any two distinct points on that line, or in other words this problem shows that such a line is a 1-dimensional affine subspace of $\mathbb{R}^2$.

Linear transformations. A function $T : V \to W$ from a vector space V over $\mathbb{K}$ to a vector space W over $\mathbb{K}$ is a *linear transformation* if it preserves addition and scalar multiplication in the sense that:

- $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$, and
- $T(cv) = cT(v)$ for all $v \in V$ and $c \in \mathbb{K}$.

This is the same definition we had for linear transformations $\mathbb{R}^n \to \mathbb{R}^m$, only now we're phrasing it in the more general setting of abstract vector spaces.

There is one thing to note now which wasn't an issue when we dealt only with $\mathbb{R}^n$ previously: the addition and scalar multiplication used on either side of these conditions are the ones taking place in the appropriate vector space. Meaning, the "+" on the left side of the first condition indicates the addition operation on V whereas the "+" on the right side indicates the addition on W, and cv on the left of the second condition uses the scalar multiplication on V whereas $cT(v)$ on the right uses the scalar multiplication of W. The point is that linear transformations relate the two additions to one another and the two scalar multiplication to one another, so in a sense a linear transformation "preserves" the vector space structure.

Example 1. Let $\mathbb{R}^+$ denote the set of positive real numbers equipped with the "addition" and "scalar multiplication" operations defined by:

$$x \oplus y = xy \quad \text{and} \quad r \odot x = x^r.$$

We claim that the function $T : \mathbb{R} \to \mathbb{R}^+$ defined by $T(x) = e^x$ is a linear transformation, where the domain $\mathbb{R}$ has the usual addition and scalar multiplication. The "preservation of addition" property in this case looks like:

$$T(x + y) = T(x) \oplus T(y),$$

since as stated before the $+$ on the right side in the definition of "linear transformation" denotes the addition of the second space. In our case, we have:

$$T(x + y) = e^{x+y} = e^x e^y = T(x)T(y) = T(x) \oplus T(y)$$

as required. For the scalar multiplication property, we have:

$$T(cx) = e^{cx} = (e^x)^c = T(x)^c = c \odot T(x)$$

as required as well. Hence T is a linear transformation.

In fact, this property is what motivates the definition of the vector space structure on $\mathbb{R}^+$ in the first place. The function $T : \mathbb{R} \to \mathbb{R}^+$ is invertible (with inverse $g(x) = \log x$), so the question is: what would "addition" and "scalar multiplication" on $\mathbb{R}^+$ have to be in order to make f linear? The addition operation $\oplus$ on $\mathbb{R}^+$ would have to satisfy the property that

$$T(x + y) = e^{x+y} \text{ is the same as } T(x) \oplus T(y) = e^x \oplus e^y,$$

which says that $\oplus$ must be ordinary multiplication, and the scalar multiplication operation $\odot$ must satisfy the requirement that

$$T(cx) = e^{cx} \text{ is the same as } c \odot T(x) = c \odot e^x,$$

which says that $c \odot e^x$ must be $(e^x)^c$.

Looking back to the Warm-Up from Lecture 19, the addition and scalar multiplication operations

$$x \oplus y = \sqrt[3]{x^3 + y^3 + 1} \quad \text{and} \quad a \odot x = \sqrt[3]{a(x^3 + 1) - 1}$$

defined there came from wanting the invertible function $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = x^3 + 1$ to be linear: i.e. these are the only operations on (the second copy of) $\mathbb{R}$ for which $x \mapsto x^3 + 1$ is a linear transformation.

Example 2. Let $C^\infty(\mathbb{R})$ denote the set of infinitely-differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, which is a vector space over $\mathbb{R}$. (It is a subspace of the space of all functions from $\mathbb{R}$ to $\mathbb{R}$.) We show that the function $L : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ defined by

$$L(f) = f'' - 2f' + f$$

is a linear transformation. The essential point is that the operation of taking a derivative is a linear one. We have

$$\begin{aligned} L(f+g) &= (f+g)'' - 2(f+g)' + (f+g) \\ &= f'' + g'' - 2f' - 2g' + f + g \\ &= (f'' - 2f' + f) + (g'' - 2g' + g) \\ &= L(f) + L(g), \end{aligned}$$

so L preserves addition. Also, if $c \in \mathbb{R}$ is a scalar:

$$L(cf) = (cf)'' - 2(cf)' + (cf) = cf'' - 2cf' + cf = c(f'' - 2f' + f) = cL(f),$$

so L preserves scalar multiplication and is hence linear.

Such linear transformations are called *linear differential operators*, and are the main object of study in advanced differential equations courses since so-called *linear* differential equations can be rephrased as looking at solutions of an equation of the form $L(f) = g$ where L is some linear differential operator. The upshot is that by recasting the study of differential equations in terms of linear algebra, a new range of tools are available to study them. We won't go into this in any more detail, but mention it just to point out one of the main uses of linear algebra in other areas.

Example 3. Define the function $T : M_2(\mathbb{R}) \to \mathbb{R}^2$ by

$$T\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+d \\ ad-bc \end{bmatrix}.$$

This is not linear, since for instance it does not preserve scalar multiplication:

$$T\left(k\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} ka+kd \\ k^2ad - k^2bc \end{bmatrix} \text{ is not equal to } kT\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ka+kd \\ kad - kbc \end{bmatrix}$$

unless $k = 1$ or 0. Also, T does not preserve addition, as you should check on your own.

Properties of linear transformations. We list some properties that all linear transformations have, which should remind you of properties we previously saw for matrices. If $T : V \to W$ is a linear transformation, then:

- $T(0_V) = 0_W$, where 0_V and 0_W denote the zero vectors of V and W respectively,
- T preserves arbitrary linear combinations,
- T is completely determined by what it does to a basis of V,
- The set $\{v \in V \mid T(v) = 0\}$ of solutions to $T(v) = 0$ is a subspace of V, and
- For fixed $w \in W$, the set $\{v \in V \mid T(v) = w\}$ of solutions to $T(v) = w$ is an affine subspace of V if there is at least one solution.

The first property is the analog of $A\mathbf{0} = \mathbf{0}$ for matrix transformations, the second property follows by induction, the fourth property is the analog of the claim that the set of solutions of a homogeneous equation $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$ when A is $m \times n$, and the fifth is the analog of the claim that the set of solutions of an inhomogeneous equation $A\mathbf{x} = \mathbf{b}$ is an affine subspace of $\mathbb{R}^n$ when A is $m \times n$ and there is at least one solution. Proving these will be left as an exercise.

The third property is something we saw for matrices only we didn't necessarily make it as explicit as we are here. The claim is that if $v_1, \dots, v_n$ is a basis of V, then knowing $T(v_1), \dots, T(v_n)$ is enough to determine what T does to *any* input. For matrices, we previously referred to this as the fact that matrices are determined by a "finite amount of linearly independent data". Indeed, if $v \in V$ is any vector, we can write it as a linear combination of the basis vectors in a unique way:

$$v = c_1 v_1 + \cdots + c_n v_n,$$

and since T is linear we get

$$T(v) = c_1 T(v_1) + \cdots + c_n T(v_n),$$

so knowing $T(v_1), \dots, T(v_n)$ allows us to compute $T(v)$ for any $v \in V$ as claimed.

Exercise. Prove the remaining properties listed above.

Definitions. Let $T : V \to W$ be a linear transformation. We say that T is

- *injective* if $T(x) = T(y)$ implies $x = y$,
- *surjective* if for any $w \in W$ there exists $v \in V$ such that $T(v) = w$, and
- *invertible* if it is both injective and surjective, or equivalently if there is an inverse linear transformation $T^{-1} : W \to V$ such that

$$T^{-1}(T(v)) = v \text{ for all } v \in V \text{ and } T(T^{-1}(w)) = w \text{ for all } w \in W.$$

Again, all these definitions mimic something we already saw for matrices. Invertible linear transformations are also called *isomorphisms*, and we'll talk about them more in the coming days.

Lecture 26: Images and Kernels

Definitions. We start with two definitions. For a linear transformation $T : V \to W$, the *kernel* of T is the subspace $\ker T$ of V consisting of all vectors which T sends to 0:

$$\ker T := \{v \in V \mid T(v) = 0\}.$$

The *image* of T is the subspace $\operatorname{im} T$ of W consisting of all the elements of W which are attained as actual outputs of T:

$$\operatorname{im} T := \{w \in W \mid \text{there exists } v \in V \text{ such that } T(v) = w\}.$$

The fact that $\ker T$ is a subspace of V was mentioned last time as the property that the solutions of $T(v) = 0$ form a subspace of V, and the fact that $\operatorname{im} T$ is a subspace of W is something we'll now prove.

Proof that images are subspaces. First, since $T(0_V) = 0_W$ is true for any linear transformation, $0_W \in \operatorname{im} T$ since there is something in V, namely 0_V, which is sent to 0_W. Second, if $w_1, w_\in \operatorname{im} T$, then there exist $v_1, v_2 \in V$ such that

$$T(v_1) = w_1 \text{ and } T(v_2) = w_2,$$

so

$$T(v_1 + v_2) = T(v_1) + T(v_2) = w_1 + w_2,$$

showing that $w_1 + w_2 \in \operatorname{im} T$ and hence $\operatorname{im} T$ is closed under addition. Finally, if $w \in \operatorname{im} T$ and $a \in \mathbb{K}$ is a scalar, there exists $v \in V$ such that $T(v) = W$, so that

$$T(av) = aT(v) = aw,$$

meaning that $aw \in \operatorname{im} T$ and hence $\operatorname{im} T$ is closed under scalar multiplication. Thus $\operatorname{im} T$ is a subspace of W as claimed. □

Warm-Up 1. Suppose that $T : V \to W$ is a linear transformation. We show that for $x, y \in V$, $T(x) = T(y)$ if and only if $x \in y + \ker T$, where $y + \ker T$ is the affine subspace of V obtained by translating the linear subspace $\ker T$ by y. For the backwards direction, if $x \in y + \ker T$, we can write x as

$$x = y + v \text{ for some } v \in \ker T,$$

so

$$T(x) = T(y + v) = T(y) + T(v) = T(y)$$

since $v \in \ker T$. Thus $x \in y + \ker T$ implies $T(x) = T(y)$.

For the forward direction, suppose $T(x) = T(y)$. Then $T(x - y) = T(x) - T(y) = 0$, so $x - y \in \ker T$. Thus

$$x = y + (x - y)$$

expresses x as y plus something in the kernel of T, so $x \in y + \ker T$ as claimed.

Note that for T to be injective, we need $T(x) = T(y)$ to imply $x = y$. Since we have $x \in y + \ker T$, in order for this to force x and y to be the same, we need $\ker T$ to only consist of the zero vector: i.e. if $x = y + v$ for some $v \in \ker T$ and $x = y$, we must have $v = 0$. Thus we get as a consequence of the Warm-Up that

$$T \text{ is injective if and only if } \ker T = \{0\}.$$

The point is that the kernel of T measures the extent to which T fails to be injective: if $\ker T$ is 0-dimensional, T is injective, while if $\ker T$ is positive dimensional then T is not injective. Moreover, if say $\ker T$ is k-dimensional, then for any $y \in V$ there is an entire k-dimensional affine subspace $y + \ker T$ of V consisting of elements which map to the same thing as y does, so in a sense as $\dim(\ker T)$ gets larger and larger, T becomes "more and more" non-injective.

Warm-Up 2. Consider the function $T : M_n(\mathbb{K}) \to M_n(\mathbb{K})$ defined by

$$T(A) = 2(A - A^T),$$

so that T sends a matrix A to the matrix $2(A - A^T)$. We show that T is linear, and then compute its kernel, the dimension of its kernel, its image, and the dimension of its image.

The fact that T is linear comes from the fact that the operation of taking transposes is a linear one in the sense that $(A + B)^T = A^T + B^T$ and $(cA)^T = cA^T$. We have:

$$T(A + B) = 2[(A + B) - (A + B)^T]$$

$$\begin{aligned}
&= 2(A + B - A^T - B^T) \\
&= 2(A - A^T) + 2(B - B^T) \\
&= T(A) + T(B),
\end{aligned}$$

so T preserves addition. If $c \in \mathbb{K}$ is a scalar,

$$T(cA) = 2[(cA) - (cA)^T] = 2(cA - cA^T) = c2(A - A^T) = cT(A),$$

so T preserves scalar multiplication and is thus linear as claimed.

Now, $\ker T$ consists of those matrices which satisfy $T(A) = 0$. Thus we are looking at matrices A for which

$$2(A - A^T) = 0, \text{ or } 2A = 2A^T, \text{ or } A = A^T.$$

Thus $A \in \ker T$ if and only if $A = A^T$, so we find that the kernel of T consists of the $n \times n$ symmetric matrices:

$$\ker T = \{A \in M_n(\mathbb{K}) \mid A^T = A\}.$$

Such a matrix is of the form

$$\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{12} & \ddots & & \vdots \\
\vdots & & \ddots & \vdots \\
\vdots & & & a_{(n-1)n} \\
a_{1n} & \cdots & a_{(n-1)n} & a_{nn}
\end{bmatrix}$$

where the entries below the diagonal correspond to ones above the diagonal. This has

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

independent parameters (obtained by adding up the number of independent parameters in each column), which are the values of a_{ij} for $i \leq j$, so we conclude that

$$\dim(\ker T) = \frac{n(n+1)}{2}.$$

The image of T consists of all matrices $B \in M_n(\mathbb{K})$ which can be expressed as

$$B = 2(A - A^T) \text{ for some } A \in M_n(\mathbb{K}).$$

Note that all such matrices must in fact satisfy $B^T = -B$:

$$B^T = 2(A - A^T)^T = 2(A^T - A) = -2(A - A^T) = -B$$

where we use the fact that $(A^T)^T = A$. This says that anything in the image of T must be *skew-symmetric* (the name given to matrices satisfying $B^T = -B$), so the image is contained in the space of all skew-symmetric $n \times n$ matrices. In fact, we claim that $\operatorname{im} T$ is in fact *equal* to the entire space of $n \times n$ skew-symmetric matrices:

$$\operatorname{im} T = \{B \in M_n(\mathbb{K}) \mid B^T = -B\}.$$

One way to see this is to note that if B is skew-symmetric, then $A = \frac{1}{4}B$ satisfies

$$T(A) = 2(A - A^T) = 2\left(\frac{1}{4}B - \frac{1}{4}B^T\right) = \frac{1}{2}(B - B^T) = \frac{1}{2}(B - (-B)) = B,$$

so $B \in \operatorname{im} T$, showing that any skew-symmetric matrix is in the image of T.

Another way to show that $\operatorname{im} T$ is the space of skew-symmetric matrices proceeds as follows. As noted above, $\operatorname{im} T$ is a subspace of the space of skew-symmetric matrices, so if we can show that this space has the same dimension as $\operatorname{im} T$, then we know they must be the same. Any skew-symmetric $n \times n$ matrix is of the form

$$\begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ -a_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \vdots & & & a_{(n-1)n} \\ -a_{1n} & \cdots & -a_{(n-1)n} & 0 \end{bmatrix},$$

where all diagonal entries are 0 and the entries below the diagonal are the *negatives* of the corresponding entry above the diagonal. In this case, there are

$$1 + 2 + \cdots + (n-1) = \frac{(n-1)n}{2}$$

independent parameters, namely the values of a_{ij} for $i < j$, so the space of skew-symmetric $n \times n$ matrices has dimension $\frac{(n-1)n}{2}$. We will see later (using the so-called "Rank-Nullity Theorem") how we can easily show that $\operatorname{im} T$ also has dimension $\frac{(n-1)n}{2}$, which will finish off this alternate justification that the image of T is the space of skew-symmetric matrices. For now, as a clue of what's to come, note that

$$\dim M_n(\mathbb{K}^2) - \dim(\ker T) = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2},$$

which is the claimed dimension of $\operatorname{im} T$—this is no accident, hint hint.

Image and kernel of a matrix. If A is an $m \times n$ matrix, then A defines a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ via $T(\mathbf{x}) = A\mathbf{x}$, and we refer to the image and kernel of this transformation as the image and kernel of A itself. So,

$$\ker A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\},$$

which is the space of solutions of $A\mathbf{x} = \mathbf{0}$. If we go through the process of solving this equation using row operations, we end up with a general solution of the form

$$\mathbf{x} = c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k$$

where $c_1, \ldots, c_k$ are the free variables in the row reduced-echelon form of the augmented matrix $(A \mid \mathbf{0})$, and $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are the vectors obtained after "factoring" out the free variables. This vectors thus span $\ker A$, and since each has a 1 in a location where the others have a zero, they are linearly independent hence form a basis for $\ker A$. We conclude that

$$\dim(\ker A) = \text{ the number of free variables in } A\mathbf{x} = \mathbf{0}.$$

The image of A consists of all $\mathbf{b} \in \mathbb{R}^m$ for which there exists $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{b}$, or in other words all $\mathbf{b} \in \mathbb{R}^m$ for which $A\mathbf{x} = \mathbf{b}$ has a solution. But we have seen that saying $A\mathbf{x} = \mathbf{b}$ has a solution is the same as saying that $\mathbf{b}$ is a linear combination of the columns of A, so we find that

$$\operatorname{im} A = \operatorname{span}\{\text{columns of A}\}$$

is an alternate description of $\operatorname{im} A$. A problem on Homework 8 asks you to show that a basis for $\operatorname{im} A$ is obtained by taking the columns of A which correspond to the pivot columns of $\operatorname{rref}(A)$, so the number of such columns is given by $\operatorname{rank} A$. Thus we conclude that

$$\dim(\operatorname{im} A) = \operatorname{rank} A,$$

which finally gives the "real" meaning behind the notion of rank.

Note that in this case:

$$\dim(\operatorname{im} A) + \dim(\ker A) = n$$

if A is $m \times n$, which is a special case of the so-called "rank-nullity theorem", which was also mentioned in the second Warm-Up above. We'll come back to this in a few days.

Example. Say that P is the 2×2 matrix of the orthogonal projection of $\mathbb{R}^2$ onto a line L in $\mathbb{R}^2$. The image of P is the line L itself, since it is only points on this line which are obtained as outputs of P, and the kernel of P is the line perpendicular to L, since points on this perpendicular line are the ones which project to $\mathbf{0}$. Thus $\operatorname{rank} P = 1$, which you could also see but working out the standard matrix of P and row-reducing; the point is that seeing that $\operatorname{rank} P = 1$ is much simpler when determining $\operatorname{im} P$ geometrically than when row reducing P.

Note the fact that P is not invertible is reflected in the fact that $\operatorname{rank} P < 2$. Indeed, non-invertible transformations in general are ones which "collapse" dimensions, sending higher-dimensional spaces onto lower-dimensional ones.

Amazingly Awesome continued. An $n \times n$ matrix A is invertible if and only if $\ker A = \{0\}$, or equivalently if and only if $\dim(\ker A) = 0$. Also, A is invertible if and only if $\operatorname{im} A = \mathbb{R}^n$, or equivalently $\dim(\operatorname{im} A) = n$. The first condition is just rephrasing the statement that $A\mathbf{x} = \mathbf{0}$ has only the $\mathbf{x} = \mathbf{0}$ solution, while the second rephrasing the statement that the columns of A span $\mathbb{R}^n$.

Lecture 27: Isomorphisms

Warm-Up 1. The actual Warm-Up we did in class was not correct as stated without the assumption we give below on the size of $\dim W$. I encourage you to see if you can figure out why on your own, and we can talk about it in office hours. For now, I'll note that there is something wrong with our claim that the kernel of the transformation we defined consisted of U alone. The solution below is correct.

Suppose that V is a finite-dimensional vector space over $\mathbb{K}$ and that U is a subspace of V. Let W be a vector space of dimension at least $\dim V - \dim U$. We construct a linear transformation $T : V \to W$ whose kernel is precisely U. This shows that any subspace of V arises as the kernel of some linear transformation.

Let $u_1, \dots, u_k$ be a basis for U, and extend it to a basis $v_1, \dots, v_\ell$ of V. Pick any ℓ linearly independent vectors $w_1, \dots, w_\ell$ in W (such vectors exist by our assumption that $\dim W \geq \dim V - \dim U$) and define T by saying

$$T(u_i) = 0 \text{ for each } i, \; T(v_j) = w_j \text{ for each } j,$$

and then extending T linearly to all of V. To say that we *extend* T *linearly* means that we define T on the rest of V in the only possible way we can if we want to ensure that T is linear: to be precise, since $u_1, \dots, u_k, v_1, \dots, v_\ell$ is a basis of V, any $x \in V$ can be written in unique way as

$$x = c_1 u_1 + \cdots + c_k u_k + d_1 v_1 + \cdots + d_\ell v_\ell,$$

and then in order for T to be linear we *must* define $T(x)$ as

$$T(x) = c_1T(u_1) + \cdots + c_kT(u_k) + d_1T(v_1) + \cdots + d_\ell T(v_\ell).$$

This is a reflection of the fact that linear transformations are completely determined by what they do to a basis, so if we specify what T does to a basis there is only one possible way to define T on other inputs if T is to be linear.

In this case, by how we defined $T(u_i)$ and $T(v_j)$ we get

$$T(x) = d_1w_1 + \cdots + d_\ell w_\ell.$$

Thus $T(x) = 0$ if and only if $d_1w_1 + \cdots + d_\ell w_\ell = 0$, which is true if and only if $d_1 = \cdots = d_\ell = 0$ since $w_1, \ldots, w_\ell$ are linearly independent. Thus $x \in \ker T$ if and only if x can be written as a linear combination of $u_1, \ldots, u_k$ alone, which means that $x \in U$ since these vectors span U. Hence $\ker T = U$ as desired.

Warm-Up 2. Suppose that U is a subspace of $\mathbb{R}^n$. As a counterpart to the previous Warm-Up we now show that there is an $n \times n$ matrix whose image is U. The key point is to recall that the image of a matrix is the same as the span of its columns.

Let $\mathbf{u}_1, \ldots, \mathbf{u}_k$ be a basis of U. Then the $n \times k$ matrix having these vectors as its columns has image equal to U, since this image is spanned by $\mathbf{u}_1, \ldots, \mathbf{u}_k$, which span U. However, this matrix is not $n \times n$ unless U was actually $\mathbb{R}^n$ itself. To fix this, we can simply throw in extra columns which don't alter the span of the columns; for instance, let A be the matrix having $\mathbf{u}_1, \ldots, \mathbf{u}_k$ as the first k columns and then enough columns of $\mathbf{0}$ to get an $n \times n$ matrix overall:

$$A = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

The image of A is the span of $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{0}, \ldots, \mathbf{0}$, which is the same as the span of $\mathbf{u}_1, \ldots, \mathbf{u}_k$, which is U. Any matrix of this type where the remaining $n - k$ columns are linearly combinations of $\mathbf{u}_1, \ldots, \mathbf{u}_k$ will work.

Isomorphisms. Recall that a linear transformation $T : V \to W$ is an *isomorphism* if it is injective and surjective, or equivalently if it is invertible. When such an isomorphism exists, we say that V and W are *isomorphic* vector spaces. The point is that isomorphisms "preserve" various properties of vector spaces, so that isomorphic vector spaces are in some sense the "same", or at least behave in the same way, as we'll see.

Example. Define a function $T : P_2(\mathbb{R}) \to \mathbb{R}^3$ by

$$a_0 + a_1x + a_2x^2 \mapsto \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix},$$

where this notation is just another way of saying

$$T(a_0 + a_1x + a_2x^2) = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}.$$

So, T simply associates to a polynomial in $P_2(\mathbb{R})$ the vector in $\mathbb{R}^3$ which encodes its coefficients. You can check that T is linear, which amounts to noticing that adding polynomials corresponds to

adding coefficients and multiplying a polynomial by a scalar amounts to multiplying the coefficients by that scalar. Since the only polynomial which is sent to $\mathbf{0}$ is the zero polynomial (the one all of whose coefficients are zero), $\ker T = \{0\}$ so T is injective. Moreover, given any vector in $\mathbb{R}^3$, we can find a polynomial which is sent to that vector by taking the polynomial whose coefficients are precisely the entries of the given vector, T is surjective. Thus T is an isomorphism, so $P_2(\mathbb{R})$ and $\mathbb{R}^3$ are isomorphic.

The idea is that even though $P_2(\mathbb{R})$ and $\mathbb{R}^3$ appear to be different since they consist of different types of objects, they should really be thought of as being the "same" since the way in which addition and scalar multiplication (the operations which determine the vector space structure) behave in one mimics the way in which they behave in the other. In other words, polynomials in $P_2(\mathbb{R})$ are fully characterized by their coefficient vectors, so there's not much point in distinguishing between them.

Here is a concrete problem we can now more easily answer using this approach. A while back we considered the subspace U of $P_2(\mathbb{R})$ consisting of those polynomials $p(x)$ which satisfy $p(2) = 0$:

$$U = \{p(x) \in P_2(\mathbb{R}) \mid p(2) = 0\}.$$

We showed that $(x-2), (x-2)^2$ formed a basis for U, which in particular means that U is 2-dimensional. The justification that $(x-2), (x-2)^2$ are linearly independent was straightforward, but it took more work to show that these spanned all of U. We didn't do this part in class, and in the lecture notes it boiled down to doing a bunch of algebra. Here we now give a much simpler justification of the fact that $(x-2), (x-2)^2$ form a basis of U.

We start with the fact that $(x-2), (x-2)^2$ are linearly independent, which again is not hard to show. Now, if we could show that $\dim U = 2$ in a way which didn't depend on finding an explicit basis at first, we would be able to say immediately that $(x-2), (x-2)^2$ formed a basis of U and hence spanned U. The point is that instead of working with U directly, we work with the subspace of $\mathbb{R}^3$ to which it corresponds under the isomorphism T. In particular, so that a polynomial $p(x) = a_0 + a_1 x + a_2 x^2$ is in U requires that

$$p(2) = a_0 + 2a_1 + 4a_2 = 0.$$

Thus, after applying T, the vectors in $\mathbb{R}^3$ which arise as the outputs corresponding to the inputs in U are those of the form

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \text{ such that } a_0 + 2a_1 + 4a_2 = 0.$$

The space of such vectors forms a basis of $\mathbb{R}^3$, which we call the *image of U under T*:

$$U \mapsto \left\{ \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \in \mathbb{R}^2 \;\middle|\; a_0 + 2a_1 + 4a_2 = 0 \right\}.$$

(The image of a subspace is the set of all outputs obtained by applying the transformation to inputs which come from that subspace.) The point is that isomorphisms preserve dimensions, so that $\dim U$ is equal to the dimension of its image. The image consists of all solutions of the linear system

$$a_0 + 2a_1 + 4a_2 = 0,$$

and the space of solutions to this 2-dimensional since there are two free variables in the reduced row-echelon form. Thus $\dim U = 2$, so $(x-2), (x-2)^2$ form a basis for U since they are two linearly independent vectors in this 2-dimensional space.

The overarching idea behind isomorphisms is that questions which might not be so straightforward to answer in one vector space might become simpler to address if we instead rephrase it in terms of an isomorphic vector space.

Things preserved under isomorphisms. An isomorphism $T : V \to W$ preserves the following notions:

subspaces, linear independence, bases, dimension, and other "linear algebraic" concepts

in the sense that linearly independent vectors in V are sent to linearly independent vectors in W, subspaces of V are sent to subspaces of W, bases of subspaces of V are sent to bases of the corresponding subspaces of W, the dimension of a subspace of V equals the dimension of its "image" in W, and so on.

Another example. Let V be the space of 3×3 symmetric matrices over $\mathbb{R}$:

$$V = \{A \in M_3(\mathbb{R}) \mid A^T = A\}.$$

This space is isomorphic to $\mathbb{R}^6$, which reflects the fact that it takes 6 independent parameters to specify an element of V. Concretely, a specific isomorphism $V \to \mathbb{R}^6$ is given by

$$\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \mapsto \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix}.$$

Again, the point is that instead of carrying around the extra "baggage" which comes with working with matrices, we can interpret such a symmetric matrix corresponding to an element of $\mathbb{R}^6$, and in this way questions about 3×3 symmetric matrices can be rephrased in terms of questions about $\mathbb{R}^6$, where they might be simpler to address.

Lecture 28: Rank-Nullity Theorem

Warm-Up. Suppose that $S : V \to W$ is an isomorphism and that $T : V \to V$ is any linear transformation. The composition $B := STS^{-1}$ then gives a linear transformation $W \to W$, and we will encode all of these maps in the diagram:

$$\begin{array}{ccc} V & \xrightarrow{\quad S \quad} & W \\ {\scriptstyle T}\downarrow & & \downarrow{\scriptstyle B = STS^{-1}} \\ V & \xrightarrow{\quad S \quad} & W \end{array}$$

We claim that under S, $\ker T$ maps onto $\ker B$ and $\operatorname{im} T$ maps onto $\operatorname{im} B$, which is what it means to say that isomorphisms "preserve" kernels and images: anything in the kernel of T on the V side of this diagram corresponds to something in the kernel of B on the W side, and anything in the image on the V side corresponds to something in the image on the W side. (Think of the transformation

$B = STS^{-1}$ as following the top arrow from right to left, then the left arrow, and then the bottom arrow. I'm just using B to denote this composition to keep from having to always write STS^{-1}.)

To say that $\ker T$ maps onto $\ker B$ means two things: applying S to something in $\ker T$ gives something in $\ker B$, and second anything in $\ker B$ comes from applying S to something in $\ker T$. First, suppose that $v \in \ker T$. Then

$$B(Sv) = (STS^{-1})(Sv) = S(Tv) = S(0) = 0,$$

so $Sv \in \ker B$. Thus anything in $\ker T$ is sent to something in $\ker B$. Now suppose that $w \in \ker B$. Since S, being an isomorphism, is surjective, there exists $v \in V$ such that $Sv = w$. Since $Bw = 0$, we have:

$$0 = Bw = (STS^{-1})(Sv) = S(Tv).$$

Since S is injective, the only thing which maps to 0 is 0, so the above equality implies that $Tv = 0$. Thus the thing which maps to $w \in \ker B$ is itself in $\ker T$, so anything in $\ker B$ arises by applying T to something in $\ker T$. We conclude that S maps $\ker T$ onto $\ker B$ as claimed.

To say that $\operatorname{im} T$ maps onto $\operatorname{im} B$ means two things: applying S to something in $\operatorname{im} T$ gives something in $\operatorname{im} B$, and second anything in $\operatorname{im} B$ comes from applying S to something in $\operatorname{im} T$. Let $v' \in \operatorname{im} T$, meaning that there exists $v \in V$ such that $T(v) = v'$. Then

$$B(Sv) = (STS^{-1})(Sv) = S(Tv) = Sv',$$

showing that Sv' is in the image of B since $Sv \in W$ is something which B sends to Sv'. Thus applying S to something in $\operatorname{im} T$ gives something $\operatorname{im} B$. Now suppose that $w' \in \operatorname{im} B$, so there exists $w \in W$ such that $Bw = w'$. Then

$$w' = Bw = S(T(S^{-1}w)).$$

Here, $S^{-1}w \in V$ and thus $T(S^{-1}w) \in \operatorname{im} T$, so $T(S^{-1}w)$ is an element of $\operatorname{im} T$ which S sends to w', so anything in $\operatorname{im} B$ arises by applying S to something in $\operatorname{im} T$. Thus S maps $\operatorname{im} T$ onto $\operatorname{im} B$ as claimed.

Remark. It might take a good while to make sense of the above proofs, especially if you're not used to arguments which move from one vector space to another and back. That's okay: we'll see more of this type of idea when we talk about *coordinates*.

Example. Let us actually use the result of the Warm-Up to justify something we did in a previous Warm-Up. In the second Warm-Up of Lecture 26 we considered the linear transformation $T : M_n(\mathbb{K}) \to M_n(\mathbb{K})$ defined by

$$T(A) = 2(A - A^T).$$

We showed there that the kernel of T was the space of all $n \times n$ symmetric matrices:

$$\ker T = \{A \in M_n(\mathbb{K}) \mid A^T = A\},$$

and we also justified there that the dimension of this kernel is $\frac{n(n+1)}{2}$. We also showed using a more brute-force computation that the image of T was the space of $n \times n$ skew-symmetric matrices, and now we see how we can justify this using a dimension count instead.

As we said back in Lecture 26, anything in the image of T definitely is skew-symmetric, so $\operatorname{im} T$ is a subspace of the space of skew-symmetric matrices:

$$\operatorname{im} T \subseteq \{C \in M_n(\mathbb{K}) \mid C^T = -C\}.$$

Thus to show that these two sets are actually equal, all we need to do is show they have the same dimension. The space of skew-symmetric $n \times n$ matrices has dimension $\frac{n(n-1)}{2}$ as shown in Lecture 26, so we only to justify that $\dim(\operatorname{im} T)$ is $\frac{n(n-1)}{2}$ as well.

The point is that we can figure out $\dim(\operatorname{im} T)$ by passing from $M_n(\mathbb{K})$ to an isomorphic vector space instead, and then applying the result of the Warm-Up. In particular, consider the isomorphism

$$S : M_n(\mathbb{K}) \to \mathbb{K}^{n^2}$$

which associates to an $n \times n$ matrix the vector in $\mathbb{K}^{n^2}$ which encodes all of its entires—to be precise, the isomorphism we have in mind is:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \mapsto \begin{bmatrix} a_{11} \\ \vdots \\ a_{1n} \\ \vdots \\ a_{n1} \\ \vdots \\ a_{nn} \end{bmatrix}.$$

Consider the diagram:

$$\begin{array}{ccc} M_n(\mathbb{K}) & \xrightarrow{S} & \mathbb{K}^{n^2} \\ \Big\downarrow{\scriptstyle T} & & \Big\downarrow{\scriptstyle B = STS^{-1}} \\ M_n(\mathbb{K}) & \xrightarrow{S} & \mathbb{K}^{n^2} \end{array}$$

The Warm-Up implies that $\dim(\operatorname{im} T)$ should be the same as $\dim(\operatorname{im} B)$, since $\operatorname{im} T$ is sent to $\operatorname{im} B$ under the isomorphism S and isomorphisms preserve dimensions.

So, we have recast the question of finding the dimension of the image of T into one which involves finding the dimension of the image of a transformation $\mathbb{K}^{n^2} \to \mathbb{K}^{n^2}$ instead. But I claim that this is now more straightforward: B is given by some $n^2 \times n^2$ matrix, and for matrices we know that

$$\dim(\operatorname{im} B) + \dim(\ker B) = \text{the \# of columns of } B$$

since $\dim(\operatorname{im} B)$ is given by the number of pivots columns in $\operatorname{rref}(B)$ and $\dim(\ker B)$ the number of free variables. In our case, we know that

$$\dim(\ker B) = \frac{n(n+1)}{2}$$

since this was the value of $\dim(\ker T)$, so

$$\dim(\operatorname{im} B) = n^2 - \frac{n(n-1)}{2} = \frac{n(n+1)}{2}.$$

Thus we conclude that $\dim(\operatorname{im} T) = \frac{n(n+1)}{2}$, so $\operatorname{im} T$ is the entire space of skew-symmetric $n \times n$ matrices as claimed.

Rank-Nullity Theorem. The example above used the idea that we can move from working with $M_n(\mathbb{K})$ to working with $\mathbb{K}^{n^2}$ instead, and that for linear transformations $B : \mathbb{K}^{n^2} \to \mathbb{K}^{n^2}$ we have

$$\dim(\operatorname{im} B) + \dim(\ker B) = \dim \mathbb{K}^{n^2}.$$

In fact, such an equation is true for maps between *any* finite-dimensional spaces, which is the content of the *Rank-Nullity Theorem*. The benefit is that we don't always have to use an isomorphism to recast questions about a vector space V into ones about $\mathbb{K}^n$ if we want to make use of this relation between the image and kernel of a linear transformations, but can continue to work with V alone.

Here is the statement. Suppose that V is finite-dimensional and that $T : V \to W$ is a linear transformation. Then

$$\dim(\operatorname{im} T) + \dim(\ker T) = \dim V.$$

(As a consequence of the proof, $\operatorname{im} T$ is finite dimensional even though W might not be, so $\dim(\operatorname{im} T)$ makes sense.) The *rank* of T is defined to be $\dim(\operatorname{im} T)$ (which agrees with the number of pivots in the reduced echelon form in the case of matrices) and the *nullity* of T is defined to be $\dim(\ker T)$, so the equality above can be written as

$$\operatorname{rank} T + \operatorname{nullity} T = \dim V,$$

which is where the name "rank-nullity theorem" comes from.

We'll prove this result next time, and finish today with some sample applications.

Previous example. We can avoid using isomorphisms in the previous example and directly show that the rank of the linear map $A \mapsto 2(A - A^T)$ from $M_n(\mathbb{K})$ to $M_n(\mathbb{K})$ as as we previously claimed. Since $M_n(\mathbb{K})$ has dimension n^2 over $\mathbb{K}$, we have by the rank-nullity theorem:

$$\dim(\operatorname{im} T) = \dim M_n(\mathbb{K}) - \dim(\ker T) = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

where we make use of the fact that $\ker T$ is the space of symmetric $n \times n$ matrices, as previously shown.

Old example revisited. We previously looked at the following example. Let U denote the subspace of $P_2(\mathbb{R})$ consisting of polynomials $p(x)$ which satisfy $p(2) = 0$. We showed that $(x - 2), (x-2)^2$ was a basis of this space, where the key point was showing that $\dim U = 2$ either directly or by using an isomorphism. Now we can do this a bit more simply using the rank-nullity theorem.

Let $T : P_2(\mathbb{R}) \to \mathbb{R}$ be the linear map defined by

$$p(x) \mapsto p(2).$$

(You should convince yourself that this is linear.) The kernel of T is precisely the subspace U we are interested in. The image of T is a subspace of $\mathbb{R}$, so it is either $\{0\}$ or $\mathbb{R}$ itself, and since there is something nonzero in the image (the result of applying T to any nonzero constant polynomial gives a nonzero value in the image), we must have that $\operatorname{im} T = \mathbb{R}$. Thus $\operatorname{rank} T = 1$ so

$$\dim(\ker T) = \dim P_2(\mathbb{R}) - \operatorname{rank} T = 3 - 1 = 2$$

and hence U is 2-dimensional.

Final example. Consider the linear map $T : P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ defined by

$$p(x) \mapsto p'(x).$$

The kernel consists of all polynomials which have derivative zero, so is the space of constant polynomials and hence has dimension 1. Thus by rank-nullity we get

$$\dim(\operatorname{im} T) = \dim P_n(\mathbb{R}) - \dim(\ker T) = (n+1) - 1 = n,$$

which implies that $\operatorname{im} T = P_{n-1}(\mathbb{R})$ since $\dim P_{n-1}(\mathbb{R}) = n$. Hence T is surjective, which is an elaborate way of showing that any polynomial has an antiderivative: for any polynomial $q(x)$, there exists a polynomial $p(x)$ such that $p'(x) = q(x)$. Of course, we don't need linear algebra to show this since we can write down such an antiderivative explicitly, but the rank-nullity theorem gives a new way of interpreting this fact.

Lecture 29: More on Rank-Nullity

Warm-Up. Let U be the subspace of $P_3(\mathbb{R})$ defined by

$$U := \{p(x) \in P_3(\mathbb{R}) \mid p(1) = 0 \text{ and } p(2) = 0\}.$$

Thus, U consists of those polynomials which have *both* 1 and 2 as roots. We find a basis of U.

Thinking about the relation between roots of a polynomial and ways of factoring that polynomial, any polynomial in U must have both $(x-1)$ and $(x-2)$ as factors. Some possible polynomials in U are:

$$(x-1)(x-2),\ (x-1)^2(x-2),\ (x-1)(x-2)^2.$$

The idea is that if we can figure out what the dimension of U must be some other way, we can reduce our problem of finding a basis to that of finding enough linearly independent polynomials in U. The rank-nullity theorem will allow us to determine the dimension of U without a lot of work.

Define the transformation $T : P_3(\mathbb{R}) \to \mathbb{R}^2$ by

$$p(x) \mapsto \begin{bmatrix} p(1) \\ p(2) \end{bmatrix}.$$

This is linear, as you can check, and has U as its kernel: the kernel consists of those polynomials $p9x)$ which are sent to the zero vector in $\mathbb{R}^2$, which requires that both components $p(1)$ and $p(2)$ be zero. Thus, by recasting U as the kernel of some linear transformation, we can make use of the rank-nullity theorem to find its dimension.

The image of T is a subspace of $\mathbb{R}^2$, and we claim that this image is in fact all of $\mathbb{R}^2$. Indeed, given any $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right] \in \mathbb{R}^2$, we can find scalars $a_0, a_1, a_2, a_3 \in \mathbb{R}$ such that

$$a_0 + a_1 + a_2 + a_3 = c \quad \text{and} \quad a_0 + 2a_1 + 4a_2 + 8a_3 = d$$

since the reduced echelon form of the augmented matrix

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & c \\ 1 & 2 & 4 & 8 & d \end{array}\right]$$

has a pivot in both rows. This says that given any $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$ we can find a polynomial

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

which T sends to $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$, so anything in $\mathbb{R}^2$ is in the image of T and hence $\operatorname{im} T = \mathbb{R}^2$ as claimed.

Thus

$$\dim(\ker T) = \dim P_3(\mathbb{R}) - \dim(\operatorname{im} T) = 4 - 2 = 2,$$

so U is 2-dimensional. Since $(x-1)(x-2)$ and $(x-1)^2(x-2)$ are two linearly independent (because they have different degrees) elements of U, they form a basis of U. (The polynomials $(x-1)(x-2)$ and $(x-1)(x-2)^2$ would also give a basis, as would $(x-1)^2(x-2)$ and $(x-1)(x-2)^2$.)

Proof of Rank-Nullity. One way to prove the rank-nullity theorem might be to pick isomorphisms between V and some $\mathbb{R}^n$ and between W and some $\mathbb{R}^m$, and then use these to rephrase $\dim(\ker T)$ and $\dim(\operatorname{im} T)$ in terms of some matrix instead. Of course, this assumes that W is finite-dimensional, which is actually not needed in the rank-nullity theorem, but even so it has the extra baggage of having to work with isomorphisms. Here we give a direct proofs which avoids isomorphisms.

To recall, the statement is that if V is finite-dimensional and $T : V \to W$ is a linear transformation, then $\dim V = \dim(\operatorname{im} T) + \dim(\ker T)$.

Proof. Let $u_1, \dots, u_k \in \ker T$ be a basis for $\ker T$, and extend this to a basis

$$u_1, \dots, u_k, v_1, \dots, v_\ell$$

of V, which can be done since $u_1, \dots, u_k$ are linearly independent and V is finite-dimensional we claim that $T(v_1), \dots, T(v_\ell)$ then form a basis of $\operatorname{im} T$.

First, let $w = T(v) \in \operatorname{im} T$. Since the u_i's and v_j's give a basis of V, they span V so we can write

$$v = c_1 u_1 + \cdots + c_k u_k + d_1 v_1 + \cdots + d_\ell v_\ell$$

for some $c_1, \dots, c_k, d_1, \dots, d_\ell \in \mathbb{K}$. Then

$$w = Tv = c_1 T(u_1) + \cdots + c_k T(u_k) + d_1 T(v_1) + \cdots + d_\ell T(v_\ell) = d_1 T(v_1) + \cdots + d_\ell T(v_\ell)$$

since each of $T(u_1), \dots, T(u_k)$ are zero given that the u_i's came from a basis of $\ker T$. This shows that anything in $\operatorname{im} T$ can be written as a linear combination of $T(v_1), \dots, T(v_\ell)$, so these vectors span $\operatorname{im} T$.

Now, to show that these vectors are linearly independent, suppose that

$$a_1 T(v_1) + \cdots + a_\ell T(v_\ell) = 0$$

for some $a_1, \dots, a_\ell \in \mathbb{K}$. We must show that each a_j is zero. By linearity of T we have

$$T(a_1 v_1 + \cdots + a_\ell v_\ell) = 0,$$

so $a_1 v_1 + \cdots + a_\ell v_\ell \in \ker T$. Hence we can express this in terms of the basis we have for the kernel of T, so

$$a_1 v_1 + \cdots + a_\ell v_\ell = b_1 u_1 + \cdots + b_k u_k$$

for some $b_1, \dots, b_k \in \mathbb{K}$. Rearranging gives

$$a_1 v_1 + \cdots + a_\ell v_\ell - b_1 u_1 - \cdots - b_k u_k = 0.$$

Since the u_i's and v_j's form a basis of V, they are linearly independent so we conclude that all coefficients in this final expression must be zero, and in particular $a_1 = \cdots = a_\ell = 0$. Thus $T(v_1), \dots, T(v_\ell)$ are linearly independent as claimed, so we conclude that they form a basis of $\operatorname{im} T$.

Using the bases for $\ker T, V$, and $\operatorname{im} T$ we have constructed above, we get

$$\dim(\operatorname{im} T) + \dim(\ker T) = \ell + k = \dim V$$

as claimed, since there are ℓ $T(v_j)$'s, k u_i's, and $k + \ell$ u_i's and v_j's together. □

Consequences. We can now give some immediate consequences of the Rank-Nullity Theorem, which mimic things we previously saw for matrices and which now hold for linear transformations between finite-dimensional vector spaces in general. Suppose that V and W are finite dimensional and that $T : V \to W$ is a linear transformation. Then:

- if $\dim V = \dim W$, then T is injective if and only if it is surjective,
- if T is injective, then $\dim V \leq \dim W$, and
- if T is surjective, then $\dim V \geq \dim W$.

In terms of matrices, the first is the analog of the claim that a square matrix is injective if and only if it is surjective, the second is the analog of the claim that an injective matrix must have more rows than columns, and the third the claim that a surjective matrix must have more columns than rows. To restate the second and third claims: if $\dim V > \dim W$, there can be no injective map $V \to W$, and if $\dim V < \dim W$, there can be no surjective map $V \to W$.

Indeed, by Rank-Nullity we have

$$\dim V = \dim(\ker T) + \dim(\operatorname{im} T).$$

If T is injective, $\dim(\ker T) = \{0\}$ so $\dim V = \dim(\operatorname{im} T)$ in this case. Since $\operatorname{im} T$ is a subspace of W, $\dim(\operatorname{im} T) \leq \dim W$ so $\dim V \leq \dim W$, which gives the second claim. If $\dim V = \dim W$, we get $\dim(\operatorname{im} T) = \dim W$ so $\operatorname{im} T = W$ and hence T is surjective, which gives the forward direction of the first claim.

Instead if T is surjective, then $\operatorname{im} T = W$ so $\dim(\operatorname{im} T) = \dim W$ and rank-nullity gives

$$\dim V = \dim(\ker T) + \dim W$$

in this case. Thus $\dim V \geq \dim W$, which gives the third claim. If $\dim V = \dim W$, then the equality above gives $\dim(\ker T) = 0$, so T is injective, which gives the backwards direction of the first claim.

Be careful in infinite dimensions. We finish we pointing out that the first claim above is not true in the infinite-dimensional setting, which helps to illustrate why it finite-dimensional spaces are so special. Define the transformations $L, R : \mathbb{R}^\infty \to \mathbb{R}^\infty$ by

$$L(x_1, x_2, x_3, \ldots) = (x_2, x_3, x_3, \ldots)$$

and

$$R(x_1, x_2, x_3, \ldots) = (0, x_1, x_2, \ldots).$$

("L" stands for "left shift" since this map takes an infinite sequence and shifts everything to the left one space after dropping the first term, and "R" stands for "right shift" since this map shifts everything to the right one space after inserting a zero in the first term.) These are both linear, as you can check. Moreover, L is surjective but not injective, while R is injective but not surjective, which shows that for a linear maps from an infinite-dimensional space to itself injective and surjective to not imply each other. (The point being that these two *do* imply each other for linear maps from a finite-dimensional space to itself as a consequence of the rank-nullity theorem.)

Incidentally, these maps also illustrate another property of compositions which doesn't hold in infinite dimensions. We previously saw that for *square* matrices A and B, then $AB = I$ then

$BA = I$ as well. The same is actually true of linear maps between any finite-dimensional spaces of the same dimension. However, in the infinite-dimensional example above, we have

$$LR = \text{identity} \quad \text{but} \quad RL \neq \text{identity},$$

showing that the analog of "$AB = I$ if and only if $BA = I$" is not true for linear transformations from an infinite-dimensional space to itself. The fact that there is not easy analog of the rank-nullity theorem in infinite dimensions is part of the reason why infinite-dimensional spaces are so much harder to work with in general.

Lecture 30: Coordinates

Warm-Up. Suppose that A is an $m \times n$ matrix. We claim the linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ which A defines is injective if and only rank $A = n$ and hence can only be possible if $m \geq n$, and surjective if and only if rank $A = m$ and hence can only be possible if $m \leq n$. These are actually both things we could have justified a long time ago using reduced row-echelon forms and pivots, but here we approach them using the rank-nullity theorem, which gives a much simpler justification.

Rank-Nullity gives

$$n = \dim(\operatorname{im} A) + \dim(\ker A).$$

A is injective if and only if $\ker A = \{\mathbf{0}\}$, which is true if and only if $\dim(\ker A) = 0$, which is thus true if and only if $\dim(\operatorname{im} A) = n$, or equivalently rank $A = n$. In particular, this means that A must have as many rows as columns since if $m < n$ then rank $A \leq m < n$. The point is that an injective transformation cannot map an n-dimensional space into a lower dimensional one.

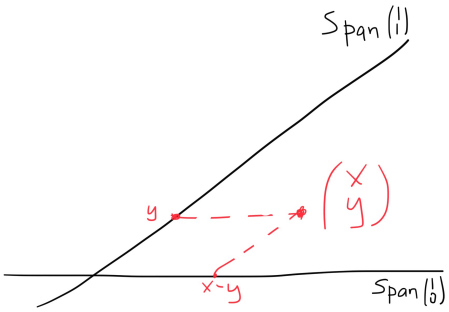
A is surjective if and only if $\operatorname{im} A = \mathbb{R}^m$, which is true if and only if $\dim(\operatorname{im} A) = m$, or equivalently rank $A = m$. (This part does not require rank-nullity.) In particular, since $n = \dim(\operatorname{im} A) + \dim(\ker A)$, this implies that $n - m = \dim(\ker A)$ is nonnegative, so $n \geq m$. The point is that a surjective transformation cannot map an n-dimensional space onto a higher dimensional one.

Motivation for coordinates. The usual way we draw $\mathbb{R}^2$ as the xy-plane uses a specific basis, namely the basis consisting of $\mathbf{e}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\mathbf{e}_2 = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. When we describe a point in $\mathbb{R}^2$ using coordinates $\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]$, what we really mean is that x and y are the scalars required to express this vector as a linear combination of this specific bases. Geometrically, this basis determines two sets of axes: span $\{\mathbf{e}_1\}$ and span $\{\mathbf{e}_2\}$, and the coordinates x, y tells us how far we have to move along these two axes to reach our given point.

However, there is nothing inherently special about this specific basis, and we could do something similar using a different basis instead! Indeed, consider the basis for $\mathbb{R}^2$ consisting of $\mathbf{v}_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\mathbf{v}_2 = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$. Visually, this gives us a new set of axes: span $\{\mathbf{v}_1\}$ and span $\{\mathbf{v}_2\}$, which are the usual x-axis and the line $y = x$ respectively. The point is that we can also use "coordinates" with respect to these set of axes to describe any vector in $\mathbb{R}^2$. If $\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right] \in \mathbb{R}^2$ (here we are taking x, y to be the usual x and y coordinates), we have

$$\begin{bmatrix}x\\y\end{bmatrix} = (x - y)\begin{bmatrix}1\\0\end{bmatrix} + y\begin{bmatrix}1\\1\end{bmatrix},$$

so we would say that the *coordinates* of $\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]$ relative to this new basis are $x - y$ and y. Geometrically, this means we have to move a distance of $x - y$ "along" the $\mathbf{v}_1$-axis and y "along" the $\mathbf{v}_2$-axis to reach our given point:



This idea works more generally in any finite-dimensional vector space, since all we need to talk about coordinates is a specific choice of basis. This will give us a new interpretation of the notion of an *isomorphism*, and we'll lead to further instances in which picking the "right" isomorphism can help to simplify a problem.

Definition of coordinates. Let V be a finite-dimensional vector space over $\mathbb{K}$ and let $v_1, \dots, v_n$ be a basis of V. The *coordinates* of a vector $v \in V$ relative to this basis are the scalars $a_1, \dots, a_n \in \mathbb{K}$ which satisfy

$$v = a_1 v_1 + \cdots + a_n v_n.$$

The fact that such scalars exist and are unique comes from the fact that $v_1, \dots, v_n$ forms a basis of V. The vector in $\mathbb{K}^n$ which encodes these scalars is known as the *coordinate vector* of v relative to the basis $\mathcal{B} = \{v_1, \dots, v_n\}$:

$$[v]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

(The subscript $\mathcal{B}$ is just indicating which basis we are taking coordinates with respect to.) The point is that by working with coordinates, we can in some sense interpret elements of V as if they were vectors in $\mathbb{K}^n$, which is an idea we'll elaborate on as we move forward.

Linearity of coordinates. One key property of coordinate vectors is that they are *linear*, in the sense that they satisfy the following properties:

$$[x + y]_{\mathcal{B}} = [x]_{\mathcal{B}} + [y]_{\mathcal{B}} \text{ and } [cx]_{\mathcal{B}} = c[x]_{\mathcal{B}} \text{ for } x, y \in V \text{ and } c \in \mathbb{K}.$$

Indeed, say that

$$[x]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \text{ and } [y]_{\mathcal{B}} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

are the coordinate vectors of x and y respectively. This means that

$$x = a_1 v_1 + \cdots + a_n v_n \text{ and } y = b_1 v_1 + \cdots + b_n v_n.$$

Then

$$x + y = (a_1 + b_1) v_1 + \cdots + (a_n + b_n) v_n,$$

so the coordinates of $x+y$ relative to $\mathcal{B}$ are $a_1+b_1,\dots,a_n+b_n$ and hence

$$[x+y]_{\mathcal{B}} = \begin{bmatrix} a_1+b_1 \\ \vdots \\ a_n+b_n \end{bmatrix} = [x]_{\mathcal{B}} + [y]_{\mathcal{B}}$$

as claimed. Moreover, if $c \in \mathbb{K}$, then $cx = ca_1v_1 + \cdots + ca_nv_n$, so

$$[cx]_{\mathcal{B}} = \begin{bmatrix} ca_1 \\ \vdots \\ ca_n \end{bmatrix} = c[x]_{\mathcal{B}},$$

also as claimed.

Isomorphisms with $\mathbb{K}^n$. To put the above facts into the right context, consider the function $V \to \mathbb{K}^n$ which assigns to an element of V its coordinate vector relative to the basis $\mathcal{B}$:

$$v \mapsto [v]_{\mathcal{B}}.$$

The fact that $[x+y]_{\mathcal{B}} = [x]_{\mathcal{B}} + [y]_{\mathcal{B}}$ says that this function preserves addition, while the fact that $[cx]_{\mathcal{B}} = c[x]_{\mathcal{B}}$ says that this function preserves scalar multiplication. Hence this function is a linear transformation, which we call the *coordinate transformation* associated to $\mathcal{B}$. In other words, the operation of "taking coordinates" is a linear one.

In fact, we claim that this coordinate transformation is actually an isomorphism. Indeed, if $[v]_{\mathcal{B}} = \mathbf{0}$, then

$$v = 0v_1 + \cdots + 0v_n = \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0},$$

so the only vector with all coordinates equal to 0 is the zero vector, and hence the coordinate transformation is injective. Second, given any

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{K}^n,$$

we can always find an element of V which has this specific vector as its coordinate vector—namely whatever

$$v = a_1v_1 + \cdots + a_nv_n$$

is. This says that coordinate transformations are surjective, so they are isomorphisms as claimed. The point is that working with coordinates (after picking a basis) indeed gives a way to think of V as if it were $\mathbb{K}^n$, so questions about V can be recast as questions about $\mathbb{K}^n$ instead.

Moreover, we claim that two *different* choices of basis $\mathcal{B} = \{v_1,\dots,v_n\}$ and $\mathcal{B}' = \{w_1,\dots,w_n\}$ give *different* coordinate transformations, or equivalently different isomorphisms between V and $\mathbb{K}^n$. Or said another way, if $\mathcal{B}$ and $\mathcal{B}'$ describe the same coordinate transformation, then $\mathcal{B}$ and $\mathcal{B}'$ are actually the same. Indeed, if $\mathcal{B}$ and $\mathcal{B}'$ give the same coordinate transformation, then

$$[x]_{\mathcal{B}} = [x]_{\mathcal{B}'} \text{ for any } x \in V.$$

In particular, $[v_i]_{\mathcal{B}} = [v_i]_{\mathcal{B}'}$ for any basis vector in $\mathcal{B}$. The coordinate vector of v_i relative to $\mathcal{B}$ is

$$[v_i]_{\mathcal{B}} = \mathbf{e}_i$$

since to express v_i in terms of the basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ we need coefficients as follows:

$$v_i = 0v_1 + \cdots + 1v_i + \cdots + 0v_n.$$

Thus $\mathbf{e}_i$ would also be the coordinate vector of $\mathbf{v}_i$ relative to the basis $\mathcal{B}'$, meaning that the expression for v_i as a linear combination of $w_1, \ldots, w_n$ is

$$v_i = 0w_1 + \cdots + 1w_i + \cdots + 0w_n = w_i,$$

and hence $v_i = w_i$ for each i. This shows that $\mathcal{B}$ and $\mathcal{B}'$ are in fact the same basis. The point here is that a basis is completely determined by the coordinate transformation it induces: given a coordinate transformation $V \to \mathbb{K}^n$, there can only be *one* basis which gives rise to it.

Examples. For the basis $\mathcal{B} = \{1, x, \ldots, x^n\}$ of $P_n(\mathbb{K})$, the coordinates of a polynomial $p(x)$ relative to $\mathcal{B}$ are simply the coefficients of the various powers of x in that polynomial, so the coordinate transformation $P_n(\mathbb{K}) \to \mathbb{K}^{n+1}$ this induces is

$$a_0 + a_1 x + \cdots + a_n x^n \mapsto \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

For the basis $\mathcal{B}' = \{E_{11}, E_{12}, \ldots, E_{mn}\}$ of $M_{mn}(\mathbb{K})$ (where E_{ij} is the $m \times n$ matrix have a 1 in the ij-th location and zeroes everywhere else), the coordinates of a matrix relative to $\mathcal{B}'$ are simply the entries of that matrix, so the resulting coordinate transformation $M_{mn}(\mathbb{K}) \to \mathbb{K}^{mn}$ is

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \mapsto \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Linear transformations via coordinates. As our first real indication that coordinates are useful things to consider, we can now answer the following question: linear transformations $\mathbb{R}^n \to \mathbb{R}^m$ are always described by matrices, so is there an analogous claim that can be made about linear transformations $V \to W$ between finite-dimensional vector spaces in general? In other words, in what sense are linear transformations between arbitrary vector spaces "described" by matrices?

Suppose that V is an n-dimensional vector space over $\mathbb{K}$ with basis $\mathcal{B}$ and that $T : V \to V$ is a linear transformation. (We'll consider more general linear transformations later.) Let $S : V \to \mathbb{K}^n$ denote the isomorphism determined by the coordinate transformation obtained by $\mathcal{B}$:

$$S(v) = [v]_{\mathcal{B}} \text{ for all } v \in V.$$

We have the following diagram:

$$\begin{array}{ccc} V & \xrightarrow{\quad S \quad} & \mathbb{K}^n \\ {\scriptstyle T}\downarrow & & \downarrow{\scriptstyle STS^{-1}} \\ V & \xrightarrow{\quad S \quad} & \mathbb{K}^n \end{array}$$

The linear transformation on the right $STS^{-1} : \mathbb{K}^n \to \mathbb{K}^n$ is the composition obtained by first using S^{-1} to move from the upper right to upper left, then T to move from the upper left to lower left, and then S to move from the lower left to the lower right—the standard matrix of this transformation is called the *matrix of T relative to $\mathcal{B}$* and is denoted by $[T]_{\mathcal{B}}$, so

$$[T]_{\mathcal{B}} = STS^{-1}.$$

The point is that this is the matrix which in a sense "characterizes" T itself. To be precise, this matrix satisfies the following property:

$$[Tv]_{\mathcal{B}} = [T]_{\mathcal{B}}[v]_{\mathcal{B}} \text{ for any } v \in V,$$

which we can see by following following the different arrows in the diagram above: starting with $v \in V$ in the upper left corner, applying S and then $[T]_{\mathcal{B}} = STS^{-1}$ (so following the top arrow and then the right arrow) gives $[T]_{\mathcal{B}}[v]_{\mathcal{B}}$, whereas applying T and then S (so following the left arrow and then the bottom arrow gives $[Tv]_{\mathcal{B}}$, so since these results should be the same we get the equality stated above. We'll see how to describe $[T]_{\mathcal{B}}$ more explicitly next time.

The idea is that to know what Tv actually is, all we need to know are its coordinates relative to $\mathcal{B}$, and this says that these coordinates can be obtained by multiplying the coordinate vector of v itself by the matrix $[T]_{\mathcal{B}}$. You should view this equation as analogous to $T(\mathbf{x}) = A\mathbf{x}$ for linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$; in our new setting it doesn't make sense to literally multiply an element of V by a matrix if V is not $\mathbb{K}^n$, but we are saying that nonetheless we can interpret T as it *were* the same as matrix multiplication by using coordinates to interpret elements of V in terms of vectors in $\mathbb{K}^n$. This is an incredibly fruitful idea, as we'll see next quarter.

Example. We finish with one example which starts to allude to the power of considering different types of coordinates. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which reflects vectors across the line $y = x$, and let $\mathbf{v}_1 = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ and $\mathbf{v}_2 = \left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$. We consider the non-standard basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ for $\mathbb{R}^2$. Then

$$[\mathbf{v}_1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad [\mathbf{v}_2]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Also, $T\mathbf{v}_1 = \mathbf{v}_1$ and $T\mathbf{v}_2 = -\mathbf{v}_2$, so

$$[T\mathbf{v}_1]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad [T\mathbf{v}_2]_{\mathcal{B}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Thus the matrix $[T]_{\mathcal{B}}$ of T relative to $\mathcal{B}$ must satisfy

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = [T]_{\mathcal{B}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ -1 \end{bmatrix} = [T]_{\mathcal{B}} \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

so the matrix of T is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Note that the form this matrix takes makes sense geometrically: if we consider a new set of axes for $\mathbb{R}^2$ given the line $y = x$ and the line $y = -x$, this matrix says that T does nothing to points on the first axis (because of the $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ column) and that T "flips" points on the second axis (because of the $\left[\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}\right]$ column), which is precisely what reflection across $y = x$ should do. If we view $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$ as giving a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ with respect to the usual x, y coordinates, it would describe reflection across the x-axis, so the point is that reflection across $y = x$ instead can *also* be described the *same* matrix, only we have to adjust the axes we are using to do so. We say that these two reflections are *similar* transformations, which is a term we'll give a precise meaning to next time.

Lecture 31: More on Coordinates

Warm-Up. Suppose that V is an n-dimensional vector space over $\mathbb{K}$ and that $S : V \to \mathbb{K}^n$ is an isomorphism. We show that there is a basis of V such that S *is* the coordinate transformation relative to this basis. This shows that *any* isomorphism from V to $\mathbb{K}^n$ can be interpreted as if it were a coordinate transformation—last time we showed that coordinate transformations always give isomorphisms, and here we are saying that all isomorphisms $V \to \mathbb{K}^n$ arise in this way. The point is that picking a basis of V to work with respect to (via coordinates) is the *same* as picking an isomorphism $V \to \mathbb{K}^n$.

The idea is that if we did have our basis $\mathcal{B} = \{v_1, \dots, v_n\}$ already in mind, we are requiring that applying S is the same as taking coordinates relative to this basis, so

$$S(v) = [v]_{\mathcal{B}} \text{ for all } v.$$

But from this we can reconstruct what the basis $\mathcal{B}$ must actually be: v_i must be the vector whose coordinate vector is $\mathbf{e}_i$, meaning that v_i must satisfy

$$S(v_i) = \mathbf{e}_i, \text{ or equivalently } v_i = S^{-1}(\mathbf{e}_i).$$

Thus the basis we are looking for must be

$$\mathcal{B} = \{S^{-1}(\mathbf{e}_1), \dots, S^{-1}(\mathbf{e}_n)\}.$$

Note that this is indeed a basis of V: since S is an isomorphism, $S^{-1} : \mathbb{K}^n \to V$ is an isomorphism, and we saw on Problem 3 of Homework 8 that isomorphisms sends bases to bases, so $S^{-1}(\mathbf{e}_1), \dots, S^{-1}(\mathbf{e}_n)$ forms a basis of V since $\mathbf{e}_1, \dots, \mathbf{e}_n$ forms a basis of $\mathbb{K}^n$.

We now verify that S is indeed the coordinate transformation determined by this basis. Take $v \in V$ and write it in terms of our basis as

$$v = a_1 S^{-1}(\mathbf{e}_1) + \cdots + a_n S^{-1}(\mathbf{e}_n).$$

Then

$$\begin{aligned} S(v) &= S(a_1 S^{-1}(\mathbf{e}_1) + \cdots + a_n S^{-1}(\mathbf{e}_n)) \\ &= a_1 S S^{-1}(\mathbf{e}_1) + \cdots + a_n S S^{-1}(\mathbf{e}_n) \\ &= a_1 \mathbf{e}_1 + \cdots + a_n \mathbf{e}_n \\ &= \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \\ &= [v]_{\mathcal{B}}, \end{aligned}$$

which shows that S is the desired coordinate transformation.

Matrix of linear map.

Derivative example.

Reflections.

Orthogonal projections.

Lecture 32: Change of Bases

Warm-Up. compositions

Example. diagonalizable

Similarity.

Diagonalizability.

Powers.

Quadratic forms.

Other vector spaces.