Math 4305 Notes
Linear Algebra

John McCuan

July 30, 2012

Introduction

This is a *problem based course*. You are expected to

1. Solve problems (primarily from the text by Curtis),
2. Compose detailed explanations of your solutions, and
3. Present your explanations on the board in class.

Points will be assigned to each problem and you will receive points for yor presentation based on correctness, clarity and efficiency (or elegance).

One comment on detail. Traditional lectures are not a prominent part of this course. It will be necessary for you to *read the text* on your own in order to acquire the needed information to solve the problems. It is a nicely written text; this activity should become routine with some practice. You are also required, however, to explain clearly what information from the text was needed to solve a problem.

Tuesday May 15, 2012

For the moment, I'm just putting something here from a past semester for you to read. You can also have a look at my notes on algebraic abstractions posted on the webpage.

1.2.1 A system of linear equations looks like

$$\begin{cases} x + y = 4 \\ 2x - 2y = 4 \end{cases}$$

This is a "two-by-two" system; two equations in two unknowns. The "row picture" consists of plotting the two lines represented by each equation individually. For the column picture, one lines up the two equations and

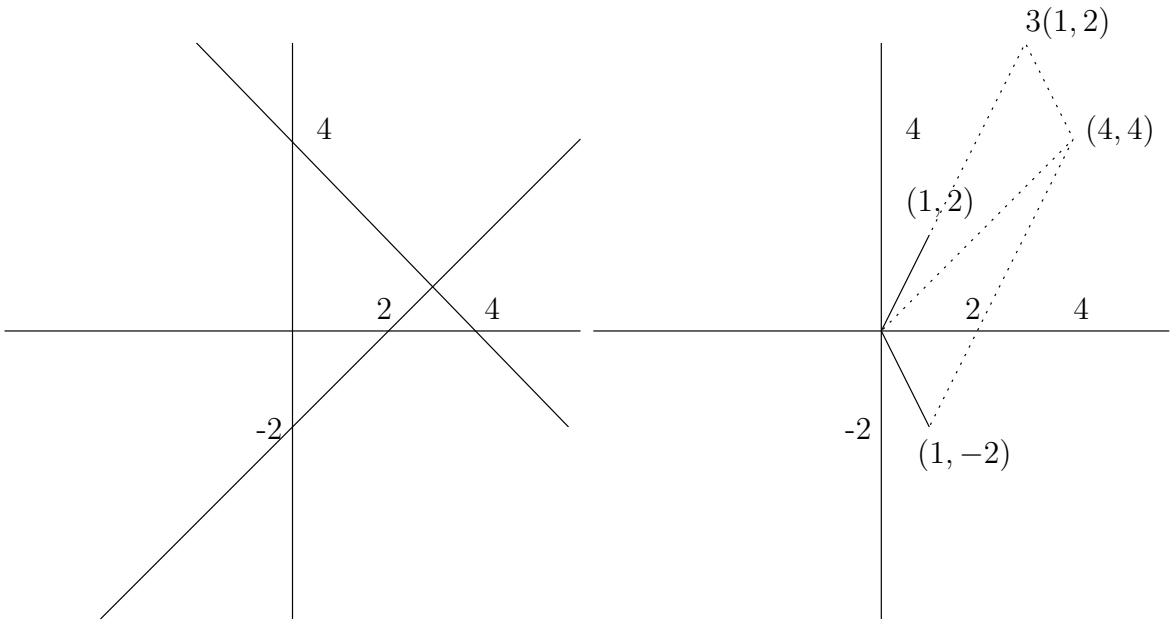

Figure 1: The Row Picture and the Column Picture.

considers the vectors formed by the coefficients:

$$x \begin{pmatrix} 1 \\ 2 \end{pmatrix} + y \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}.$$

It is easily checked that putting $x = 3$ and $y = 1$ in this vector equation works.

Remark Strang builds his text around solving systems of linear equations; for him, this is the main point of linear algebra.

1.2.1* If you change the first row, i.e., change the first plane, what is the change in the column vectors? If you vary the columns, what is the change in the row picture? Can you see a geometric or mechanical relation between the two?

Motivation

As mentioned just above, many teachers of linear algebra consider systems of linear algebraic equations as the primary motivation for linear algebra. (See for example the first sentence of Curtis' Preface.) I mentioned in class also that, as an undergraduate, I found that motivation rather uninspiring. I still do find it uninspiring to a certain extent.

I do find the subject pretty interesting, however. But the reason I find it interesting is a bit more complicated than just studying systems of linear algebraic equations.

As I have time, I will try to type in some notes on my motivation for studying linear algebra as I outlined it in class. What I would like to get across is built around the following:

1. I want to understand general "mappings" from $\mathbb{R}^n \to \mathbb{R}^m$ where n and m are integers. One nontrivial basic case to consider is when $n = m = 2$; these are maps of the plane into itself.
2. There is a way to "localize" or "linearize" the behavior of such a map. That is, there is a way to view the behavior of the map near a single point in some simplified way. (We know how to do this for real valued functions of one real variable from calculus; this is basically the definition of the derivative.)
3. For mappings of higher dimensional spaces, it is also possible to build simplified maps. They are linear maps built out of matrices containing the first (partial) derivatives of the original (nonlinear) map at a point.
4. The general class of linear maps, i.e., maps determined by matrix multiplication, is interesting and somewhat nontrivial to understand.

Maybe we should start with a review of something in my second point above. If you have a real valued function of one real variable, then we are accustomed to "graph the function." Actually, "graph" is more properly used in mathematics as a noun rather than a verb. The graph of a function is the set of all ordered pairs with first element in the domain of the function and second element given by the value of the function. In symbols, if f is a function with domain A, then the graph of f is

$$\{(x, f(x)) : x \in A\}.$$

This point set is important because the derivative of a function at a point x_0 in it's domain is interpreted as *the slope of the line which is tangent to the graph of the function at the point* $(x_0, f(x_0))$.

It's worth taking a few minutes of your time to see if you agree with that interpretation and why. (In order to do this properly, you'll need to remember, or look up, the definition of derivative, which I'm not giving you here.)

Closely related to that interpretation (picture) and the definition is the idea of *approximation.* That is to say, we can write down an expression for the function ℓ which gives the tangent line at $(x_0, f(x_0))$, and that can be interpreted as an approximation for the function f for domain values near x_0. Again, in symbols

$$f(x) \approx \ell(x) = f(x_0) + f'(x_0)(x - x_0).$$

Now, I want to bring in an idea which may be new to you:

> *This is a recipe for "building" a function ℓ which approximates f near a given point. That function ℓ is "built" out of two or three fixed things, namely*
>
> 1. *The given point x_0,*
> 2. *The value of f at the given point, and*
> 3. *The derivative of f at the given point.*
>
> *We want to emphasize the last one: The approximation is mainly built out of the derivative value $f'(x_0)$.*

Another point which is important in this discussion is that we want to emphasize the structure of the approximation. It is called a *linear* approximation. This can be a bit confusing because the function ℓ is not linear in general. Let me explain a bit more.

Always and forever, any function Λ which is linear will satisfy

1. $\Lambda(v+w) = \Lambda(v) + \Lambda(w)$, and
2. $\Lambda(cv) = c\Lambda(v)$.

When we write this, v and w are vectors in the domain of Λ and c is a scalar.

Today was mostly an administrative day. We did hit some nice points, however:

1. Linearity means $\Lambda(v+w) = \Lambda(v) + \Lambda(w)$ and $\Lambda(cw) = c\Lambda(w)$.
2. Linear maps $L : \mathbb{R}^n \to \mathbb{R}^m$ are given by matrix multiplication with the columns of the matrix given by the images of the standard basis vectors.
3. We talked a little about fields.

Next time I'm looking forward to having you show me what you can do.

Thursday May 17, 2012

(2.12)* on page 13 (Jevon R.) $ab = ac$ implies $b = c$ as long as $a \neq 0$.

Proof: Multiply both sides by a^{-1}.

1.2.1(a) (Gautam G.) The sum of the first n odds is n^2. We had a discussion of proof by induction. In summary, verify your assertion $A(n)$ for $n = 1$ and then show that

> $A(k)$ implies $A(k+1)$.

Then you've proved that $A(n)$ holds for $n = 1, 2, 3, \ldots$.

1.2.1(b) (David G.) Formula for the sum of the squares of the first n integers.

Well Ordering (Kowshik M.) The Well Ordering Principle (that every nonempty subset of nonnegative integers has a least element) implies the Principle of Mathematical Induction (which was proved in the following form: If a set S contains 1, and whenever you have $k \in S$, then you also have $k+1 \in S$, then $S = \{1, 2, 3, \ldots\}$.

The proof was nice. Let $T = \mathbb{N} \backslash S$, and assume $T \neq \phi$. By the Well Ordering Principle, there is a smallest element $a \in T$. Clearly, $a \neq 1$, since $1 \in S$. Therefore, $k = a - 1 \geq 1$ is an integer, and also $k = a - 1 \notin T$ since a is the smallest one. So, $k \in S$, but then $k + 1 = a \in S$ which is a contradiction.

1.2.3 (Aishwarye C.) This is the binomial expansion for an arbitrary field.

Let's see. When $n = 1$, you just have

$$(a+b)^1 = \left(\begin{array}{c} 1 \\ 0 \end{array} \right) a^1 + \left(\begin{array}{c} 1 \\ 1 \end{array} \right) b^1$$

which is true.

Next, assuming

$$(a+b)^k = \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k-j} b^j$$

we see that

$$\begin{aligned}
(a+b)^{k+1} &= (a+b)(a+b)^k \\
&= a \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k-j} b^j + b \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k-j} b^j \\
&= \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k-j+1} b^j + \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k-j} b^{j+1} \\
&= \left(\begin{array}{c} k \\ 0 \end{array} \right) a^{k+1} + \sum_{j=1}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k+1-j} b^j + \sum_{j=0}^{k-1} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k-j} b^{j+1} + \left(\begin{array}{c} k \\ k \end{array} \right) b^{k+1} \\
&= \left(\begin{array}{c} k+1 \\ 0 \end{array} \right) a^{k+1} + \sum_{j=1}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a^{k+1-j} b^j + \sum_{j=1}^{k} \left(\begin{array}{c} k \\ j-1 \end{array} \right) a^{k+1-j} b^j + \left(\begin{array}{c} k+1 \\ k+1 \end{array} \right) b^{k+1} \\
&= \left(\begin{array}{c} k+1 \\ 0 \end{array} \right) a^{k+1} + \sum_{j=1}^{k} \left[\left(\begin{array}{c} k \\ j \end{array} \right) + \left(\begin{array}{c} k \\ j-1 \end{array} \right) \right] a^{k+1-j} b^j + \left(\begin{array}{c} k+1 \\ k+1 \end{array} \right) b^{k+1} \\
&= \sum_{j=0}^{k+1} \left(\begin{array}{c} k+1 \\ j \end{array} \right) a^{k+1-j} b^j.
\end{aligned}$$

2.3.5 (Claudia R. with a bit of help from Aishwarye C.) $AB = -BA$ Proof: $AB = B - A = -(A - B)$ since $(B - A) + (A - B) = 0$. But then $-(A - B) = -BA$ (by definition).

2.3.6 (Dustin M.) $AB = CD$ if and only if there is some X with $C = A + X$ and $D = B + X$. (Translation)

Well, if $C = A+X$ and $D = B+X$, then $CD = D-C = B+X-(A+X) = B - A = AB$. So that direction is OK.

Conversely, if $AB = B - A = D - C = CD$, then we can add C and subtract B from both sides of the inner identity $B - A = D - C$ to obtain

$$C - A = D - B.$$

Let X be this common value. In particular, $C - A = X$, so $C = A + X$. Similarly, $D - B = X$, so $D = B + X$.

2.3.9 (Sarah M.) If $AB = CD$, then $AC = BD$.

Let's see. We are given $B - A = D - C$. Adding C to both sides and subtracting B from both sides, we get $C - A = D - B$. That's what we want.

Tuesday May 22, 2012

2.4.10 (Dustin M.) As pointed out in class, Curtis seems to have made a mistake on this one. What he probably meant to say was *Any finite collection of vectors which contains a set of linearly dependent vectors is linearly dependent.*

A reasonable proof of this fact, according to his definition might go like this:

Let the superset be called S and let the vectors in the linearly dependent subset be $v_1, \ldots, v_k$. By the definition of linear dependence, there are constants $a_1, \ldots, a_k$ not all zero with

$$\sum_{j=1}^{k} a_j v_j = 0.$$

We can denote the remaining vectors in S (if there are any) by $v_{k+1}, \ldots, v_n$. We need to show that this set is linearly dependent too. Just take $a_{k+1} = \cdots = a_n = 0$. Then it is clearly the case that

$$\sum_{j=1}^{n} a_j v_j = 0.$$

It is also clear that not all of the coefficients are zero, so we are done. □

There are a couple things to note:

1. It is really important to read the text carefully and critically, especially the definitions.
2. Taking the problem just as stated in the book, it is easy to give a counterexample as follows: If we are going to show that any such set is linearly dependent, then we must show it is finite. So all we need for a counterexample is a set which is infinite but contains a linearly dependent (finite) subset. Such an example is $\mathbb{R}$ which is a vector space over the field $\mathbb{R}$. A finite linearly dependent subset is $\{0\}$.
3. Really, it is useful to have a definition of linear dependence for an arbitrary set of vectors. The easiest way is to use Curtis' definition as a start and then say:

   Any set of vectors is linearly dependent if it contains some finite linearly dependent subset.

   This definition makes Curtis' problem trivial, but that is his problem. One should also check that the two definitions we have for *linearly dependent* are consistent. That is, the second definition shouldn't rule out or rule in any finite sets ruled out or ruled in respectively by the first definition.
4. There is still the question about an analogous statement for linearly independent vectors. I don't think we've nailed that one yet.
5. Also, it's a bit irritating to have a first definition for finite sets of vectors, and then a generalization of it. Can you give a definition that works for arbitrary sets from the start?

2.5.2 (David G.) Three vectors in $\mathbb{R}^2$ are linearly dependent.

We don't seem to have a solution here yet.

Let $x = (x_1, x_2)$, $y = (y_1, y_2)$, and $z = (z_1, z_2)$.

Obviously, if one of these vectors is the zero vector, then I can take a nonzero multiple of that one and zero multiples of the others to get the zero vector written as a nontrivial linear combination of the three, and we are done.

If any one of the first components is zero, we can assume it is the first one. Then our life is a little simpler. We're loooking at

$$x = (0, x_2),\ y = (y_1, y_2),\ \text{and } z = (z_1, z_2).$$

How about you show those three vectors are linearly dependent?

administrative I forgot to mention that we can have some homework problems graded. That's probably a good idea. I'll start assigning problems for you to work and turn in on Tuesdays starting in a week or so. Here's the first set: 1.2.4, 5; 2.3.1, 2, 10; 2.4.4, 4, 6, 7; 2.5.4, 5. (If you work them in class, then they don't need to be turned in.)

Thursday May 24, 2012

2.5.2 (Xian W., Noah C., Richard R.)

This problem still remains open. (See the discussion above.) Xian tried to use the fact that the column rank of a matrix is less than or equal to the row rank of a matrix and observed that the row rank of

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{pmatrix}$$

can be at most two. Therefore the row rank can be at most two as well, and the three vectors must be linearly dependent. Modulo the definitions of row and column rank (and matrices) which we'll get to in due time, this is basically correct reasoning. However, Xian was unable to provide a proof of the fact that column rank cannot exceed row rank.

Noah tried to use Theorem 5.3, and Sarah pointed out that Theorem 5.1 applies more directly. This approach is also correct in principle but was disallowed on the basis that we haven't gone through the proof of Theorem 5.1 and the problem explicitly says to give a proof "from the definitions," which means roughly that you're not allowed to quote Theorem 5.1.

Richard returned to David's approach, but wasn't able to convince us.

2.4.5 (Omar V.) The set

$$\Sigma = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 - x_2 + x_3 = 0\}$$

is a subspace and has a basis given by $\{(1,1,0),(1,0,-1)\}$.

Let's first show that every $(x_1, x_2, x_3) \in \Sigma$ is a linear combination of the vectors in our proposed basis. In fact,

$$x_2(1,1,0) - x_3(1,0,-1) = (x_2 - x_3, x_2, x_3) = (x_1, x_2, x_3)$$

with the last equality holding because $x_2 - x_3 = x_1$ (according to the definition of Σ). Thus, Σ is a generating set.

We also need to show linear independence. If $a(1,1,0) + b(1,0,-1) = (a+b, a, -b) = (0,0,0)$, then looking at the second and third components gives $a = 0$ and $b = 0$.

You should make sure you know how to write down the details of the assertion that Σ is a subspace.

2.6.5(a) and parts of 2.5.3 (Peter Y.)

Tuesday May 29, 2012

2.5.2 (David G.) I think we made a little progress today, but we're not quite there. What we have is something like this:

If one of the vectors (see notation above) is the zero vector, then we know the three vectors are linearly dependent because we can take the coefficient of the zero vector to be 1 and the other two coefficients to be zero. Thus, we can always assume that none of the vectors are the zero vector.

Next, if all the first components are zero (we called this something like CASE 0.), then none of the second components, x_2, y_2, or z_2, can be zero. Then, you can check that $0v_1 - z_2v_2 + y_2v_3 = 0$, and the last two coefficients are nonzero, so we're done.

Thus, if we're not in CASE 0, then by rearranging if necessary, we can always assume that $x_1 \neq 0$. With this understanding, we come to

CASE 1. $x_1y_2 - y_1x_2 \neq 0$.

First remember that we are already assuming here that $x_1 \neq 0$, so this is some kind of assumption on the interaction of the components of the first two vectors. In any case, it was observed that

$$v_2 - \frac{y_1}{x_1}v_1 = \frac{x_1y_2 - y_1x_2}{x_1}(0,1)$$

and

$$v_3 - \frac{z_1}{x_1}v_1 = \frac{x_1z_2 - z_1x_2}{x_1}(0,1).$$

Therefore, setting $\mu = -(x_1z_2 - z_1x_2)/(x_1y_2 - y_1x_2)$, we have

$$\mu\left(v_2 - \frac{y_1}{x_1}v_1\right) + v_3 - \frac{z_1}{x_1}v_1 = (0,0).$$

But this means

$$-\left(\frac{\mu y_1}{x_1} + \frac{z_1}{x_1}\right)v_1 + \mu v_2 + v_3 = (0,0).$$

Since that last coefficient (on the v_3 is $1 \neq 0$), this means $\{v_1, v_2, v_3\}$ is linearly dependent as desired.

But I guess there's one more case that we havne't handled. Next time?

2.6.5 (Peter Y.) Here we are looking at the polynomials x^2+2x+1, $2x+1$ and $2x^2 - 2x - 1$ in the vector space of polynomials of order no more than 2 over, for example, the reals.

As I pointed out in class last time, the easiest way to think about this vector space is in terms of "formal" polynomials. That is, you define arithmetic (adding polynomials and scalar multiplication) in the usual way, and think of x as always a symbolic variable, i.e., a bookkeeping placekeeper. In particular, equality for two polynomials means that you "collect like terms" in each and equate the coefficients.

(There is an alternative in which, for example, equality of two polynomials $p(x)$ and $q(x)$ is defined as having $p(x) = q(x)$ for every x in the field.

Then you prove that equality of polynomials is equivalent to equality of the coefficients. This is a fine definition and way to look at polynomials, but it is a bit more work—which you are welcomed to do.)

With our definition of "formal polynomials" in x, the space spanned by $\{1, x, x^2\}$ is essentially equivalent to $\mathbb{R}^3$ which of course is spanned by $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, and $\mathbf{e}_3 = (0, 0, 1)$.

Peter's solution was to form coefficient vectors from the polynomials and assemble them into a matrix:

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

As mentioned, the space is like $\mathbb{R}^3$ (or $\mathbb{F}^3$) and this really makes it look like that. In particular, Peter pointed out that the linear dependence/independence of the three polynomials is equivalent to the dependence/independence of the three row vectors in the matrix.

Peter then introduced "elementary row operations." There were three of them:

- Interchange two rows: $r_i \leftrightarrow r_j$,
- Replace a row with a nonzero multiple of itself: $r_i \to \mu r_i$,
- Replace a row with itself plus a nonzero multiple of a different row: $r_i \to r_i + \mu r_j$.

Two matrices are said to be *row equivalent* if one can be obtained from the other by a (finite) sequence of elementary row operations. Initially, this is defined asymetrically, i.e., as "one matrix is equivalent to a second if..." Then you have to show that the relation is symmetric. In fact, there are three properties that make it a bona fide "equivalence relation." These properties are the following:

- $A \sim A$ (reflexive),
- $A \sim B$ implies $B \sim A$ (symmetric),
- $A \sim B$ and $B \sim C$ implies $A \sim C$ (transitive).

We checked these things informally and talked a bit about the concept of "concatenation" in connection with the transitive property. It was pointed out that there is a reasonably tidy little theory of equivalence relations that is worth learning about. The main thing to know is about "equivalence classes." I think this is in your book, so we should go over it. If not, it may be found in a text like Halmos' "Naive Set Theory" and we should go over it. Since the full theory of equivalence relations is (apparently) not needed for the problem at hand, we'll let Peter off the hook for the moment.

Peter then quoted some theorem (Theorem 6.16?—I don't have the book in front of me). The theorem asserted, among other things, that if the matrix you have is row equivalent to one with a row of zeros, then the row vectors you started with were linearly dependent. We should probably make him prove at least that part of the theorem.

He proceeded to "row reduce" the matrix and got a row of zeros, so he was, more or less, done. I didn't record the row reduction, but we should all be able to do that sort of thing, so let's see if I can do it:

$$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 0 \\ -1 & -2 & 2 \end{pmatrix} \xrightarrow{r_2 \to r_2 - r_1, r_3 \to r_3 + r_1} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 3 \end{pmatrix}$$

$$\xrightarrow{r_3 \to r_3 + 3r_2} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

There's a row of zeros, so if I've done the reduction correctly, the original rows are linearly dependent (according to the theorem).

Actually, the theorem was a little more complicated than I have indicated, and maybe also required that the first two rows were in "echelon form," i.e., they were nonzero and each has first nonzero coefficient strictly to the right of the first nonzero coefficient of the previous row. You can see that I also happen to have gotten "echelon form" for the first two rows.

There was also some discussion about whether or not a matrix is just an ordered set of vectors (presumably of the same length). Apparently, there is some such definition in your text, and technically that might even be "the" way to define a matrix. But I don't think of a matrix that way. If I happen to "assemble" a matrix from a collection of row vectors, then I think of the matrix as a fundamentally new kind of object. But that's just me. My definition of a matrix would be that it is a "rectangular array" of elements. Then it has separate collections associated with it containing, for example, its rows or its columns.

There are a couple things it would be good to understand better.

One is the relation between elementary row operations and linear combinations of the rows. (It seems evident that there must be some relation.)

More broadly, it would be good to understand that theorem Peter quoted.

More specifically, it would be nice to be able to read off the coefficients of a linear combination which expresses the linear dependence/independence of the three vectors from the row reduction.

When I say "understand" a theorem, I almost always mean understand its proof, i.e., why the theorem is true. Of course, it takes some work just to understand the statement of a theorem, and it's good to understand what theorems say and how to use them correctly. But it's even better if you

understand the proof. And that is, to a large degree, what this course is about.

Finally, let me try to give a shorter proof:

Since none of the three polynomials is a multiple of any other one, we see that each pair of them spans a two dimensional subspace. Since

$$-2(x^2 + 2x + 1) + (2x^2 - 2x - 1) = -3(2x + 1)$$

We see that the middle polynomial is in the span of the first and third. In particular, the three are linearly dependent.

Since all the coefficients in the above relation are nonzero, it also says that any one of these polynomials can be expressed as a linear combination of the other two.

2.6.1 (Xian W.) This was, as I recall, just doing some row reductions. I assume everyone can do this kind of thing. If you don't feel like you can do it—or maybe even more importantly, if you feel like you can, but you can't—then it's is hereby your responsibility to learn how to do it.

This brings up a point about philosophy of teaching which might interest you. In my view, students have two primary responsibilities:

- Learn the material/master the skills in the course, and
- Communicate to the teacher that they understand the material and have mastered the skills.

Teachers also have two (main) responsibilities:

- Help the students learn the material, and
- Evaluate the students progress.

Of these four responsibilities, the most important one is the first one, and the burden (you'll note) falls on the student. It is very easy for "teaching" to be replaced with "busy work," i.e., stuff students can do to get an evaluation indicating they have learned/mastered the material when, in fact, they havent. I'm not into busy work, though sometimes I have been administratively required to assign it. For example, in some courses I've taught, the administrator has required that some specific percentage of the grade (say 15%) be based on attendance. So the student can show up to class and surf the internet and get credit for it. This really hurts the students in the long run, I think, but it seems to make a lot of people (students and administrators) happy.

Administrative: Let's include a final set of problem 2.6.1-4, and all the ones above and these will be due Tuesday June 5, 2012. I'll put them on the schedule page too. Let's make a goal of trying to understand everything up to section 2.6 by that same date. That's a good bit of material, but let's give it a try.

Thursday May 31, 2012

2.5.2 (David G.) CASE 2: $x_1y_2 = x_2y_1$.

If any one of the entries of v_1 or v_2 is zero, then notice that the common value of $x_1y_2 = x_2y_1$ is also zero. Since $x_1 \neq 0$, this means $y_2 = 0$ which in tern means $y_1 \neq 0$ and, hence, $x_2 = 0$. In this case, it is easy to check that

$$v_1/x_1 - v_2/y_2 + 0v_3 = (0, 0)$$

and since the first two coefficients are nonzero, we are done.

The final possibility is that none of the components of v_1 or v_2 vanishes. In this case, $x_1/x_2 = y_1/y_2$, and

$$v_1/x_2 - v_2/y_2 + 0v_3 = (x_1/x_2, 1) - (y_1/y_2, 1) = (0, 0),$$

so again we reach the same conclusion. □

Remark: This problem is closely related to the proof of Theorem 5.1. Can you further clarify the connection? In particular, how would David's solution be modified to parallel the proof given in the text of Theorem 5.1? What characteristics do they have in common and where does the reasoning differ?

2.7.3 (Richard R.) Every subspace of a finitely generated space is finitely generated and the subspace will have strictly smaller dimension unless the subspace equals its superspace.

We ran into trouble on this one partially because we do not yet know that *dimension*, i.e., the number of elements in a basis, is well defined. This is Theorem 5.3. Another problem was that we don't yet really understand Theorem 7.2: Every finitely generated vector space has a (finite) basis.

We did go over most of the definitions and technicalities to make sense of these questions. In particular, we should know what is meant by a *generating set* and *finitely generated*.[1] To be precise,

> *a subset G of a vector space is said to be a* generating set *if every vector in the vector space can be expressed as a linear combination of vectors in G.*

And...

> *A vector space is said to be* finitely generated *if it contains a generating set with finitely many elements.*

Once we have these down, we can see that

> *A* basis *B of vector space is a a linearly independent generating set.*

[1] In order to understand these concepts, it is important to understand what is meant by *linear dependenc/independence* and a *linear combination*.

Theorem 5.3 asserts that every finite basis has the same number of elements. Once that is known, that number of elements is the *dimension* of the vector space.

I think those are most of the relevant definitions. We seemed to get stuck pretty quickly on the proposition that a subspace of a finitely generated space is finitely generated due to the fact that the generators for the bigger space are not necessarily in the smaller space. Thus, it seems we are forced to somehow build a generating set in the smaller space, and it's not so obvious how to do that.

2.4.8 (Xian W.) This problem asks if it is always true that the union of vector subspaces is a subspace. An example suffices to answer the question: The span of $\mathbf{e}_1$ is the x-axis in $\mathbb{R}^2$ which is a subspace and the span of $\mathbf{e}_2$ is the y-axis which is another subspace. But the union is not closed under addition since $\mathbf{e}_1 + \mathbf{e}_2 = (1, 1)$ is not in the union.

Xian went further to generalize this observation: The union is a subspace if and only if one of the subspaces is a subspace of the other.

The key is to show that when neither is a subset of the other, then the union is not a subspace. Let the subspaces be X and Y and take $x \in X \backslash Y$ and $y \in Y \backslash X$. Then $x + y = z$ is not in the union because if we assumed that $z \in X$, then $y = z - x \in X$ which contradicts the fact that $y \notin X$. Similarly, assuming $z \in Y$ leads to the contradictory statement $x \in Y$. □

2.5.1 (Liangyi S.) If $\{b_1, \ldots, b_r\}$ is a basis for V, then none of the basis elements is the zero vector.

Assume one of the basis elements b_{i_0} is the zero vector, then taking zero coefficients for all the other basis elements and coefficient 1 for b_{i_0}, we get a linear combination of the basis elements

$$\sum_{i \neq i_0} 0 \cdot b_i + 1 \cdot b_{i_0} = 0$$

which expresses the zero vector but has a nonzero coefficient. This shows that $\{b_1, \ldots, b_r\}$ is a linearly dependent set of vectors and, thus, contradicts the fact that a set of basis vectors is linearly independent. □

Remarks: (1) This really shows that the zero vector cannot be in any linearly independent set.

(2) We were in good shape because we used the multiplicative identity as the coefficient of b_{i_0} and the zero in the field as the other coefficients. In this way, we could use the last axiom for vector spaces (Definition 3.1) and assertion (iv) from Theorem 3.5 to see that our linear combination simplified to be the zero vector.

You should, of course, check the proof of Theorem 3.5, and there is at least one technical property of vectors which doesn't seem to be listed there and you might try to prove it:

Can you show that *any* scalar times the zero vector is the zero vector?

Tuesday June 5, 2012

Row Reduction (Gautam G.) Gautam showed that the span of the rows of a matrix is unchanged under elementary row operations. Recall that the *span* of a collection of vectors is defined to be the collection of all (finite) linear combinations of those vectors. I don't think we have gone through the proof, but it can be shown (and you should check it) that the span of any collection of vectors is a vector space. In the particular case in which the vectors you are taking the span of are the rows of a matrix, then the resulting vector space is called the *row space*.

In any case, it is pretty clear that if you change the order of the rows, the row space of the resulting matrix is the same. Also, if you multiply a row by a nonzero constant, that doesn't change the span either.

Let's check that the span doesn't change if we replace the i-th row r_i with $r_i + cr_j$. It's clear that a linear combination of the new rows

$$\sum_{k \neq i} a_k r_k + a_i(r_i + cr_j)$$

is a linear combination of the old rows:

$$\sum_{k \neq j} a_k r_k + (a_j + ca_i) r_j.$$

Conversely, if we take a linear combination of the old rows

$$\sum a_k r_k,$$

then we can define coefficients b_k as follows:

$$\begin{cases} b_k = a_k \text{ for } k \neq j \\ b_j = a_j - ca_i. \end{cases}$$

Then the linear combination of the new rows

$$\sum_{k \neq i} b_k r_k + b_i(r_i + cr_j) = \sum_{k \neq i,j} a_k r_k + (a_j - ca_i) r_j + a_i r_i + ca_i r_j = \sum a_k r_k$$

is what we want, namely the linear combination of the old rows with which we started. □

Gautam also noted that the nonzero rows in a matrix in echelon form are linearly independent. Thus he concluded that one can find a basis for the row space of a matrix by row reducing the matrix to echelon form (which can always be done) and then taking the nonzero rows as the basis. (Technically, one might need to say that only the nonzero rows are in echelon form, but the terminology I've used is common.)

$c \cdot \vec{0} = \vec{0}$ (Liangyi S.) This is a property of vector spaces which should be in Theorem 3.5 but is not there.

$$\begin{aligned} c \cdot \vec{0} &= c \cdot \vec{0} + \vec{0} \quad (3.1-2) \\ &= c \cdot \vec{0} + (c \cdot \vec{0} - c \cdot \vec{0}) \quad (3.1-3) \\ &= (c \cdot \vec{0} + c \cdot \vec{0}) - c \cdot \vec{0} \quad \text{(associative)} \\ &= c(\vec{0} + \vec{0}) - c \cdot \vec{0} \quad (3.1-1) \\ &= c \cdot \vec{0} - c \cdot \vec{0} \quad (3.1-2) \\ &= \vec{0} \quad (3.1-3). \end{aligned}$$

2.6.4(c) (Claudia R.) We want to show that $-x^2 + x + 1$, $x^2 + 2x$, and $x^2 - 1$ are linearly independent as functions. This means that if we know

$$\lambda_1(-x^2 + x + 1) + \lambda_2(x^2 + 2x) + \lambda_3(x^2 - 1) = 0 \tag{1}$$

for every $x \in \mathbb{R}$, then $\lambda_1 = \lambda_2 = \lambda_3 = 0$. To see that this is the case, we can first take $x = 0$ to find that $\lambda_1 - \lambda_3 = 0$. Thus, replacing λ_3 with λ_1 in the original condition (1), we find

$$\lambda_1 x + \lambda_2(x^2 + 2x) = 0 \tag{2}$$

for every $x \in \mathbb{R}$. Next, we can take $x = -2$ to find that $\lambda_1 = 0$. But then taking $x = -1$, we have $-\lambda_2 = 0$. It follows that $\lambda_1 = \lambda_2 = \lambda_3 = 0$ and we are done.

2.6.4(a) (Xian W.) Here we are given three real valued functions f_1, f_2, and f_3 of a real variable, and three points x_1, x_2, and x_3. If you find out that the 3×3 matrix with i-th row given by $(f_i(x_1), f_i(x_2), f_i(x_3))$ has three linearly independent rows, then the three functions are linealry independent.

Assume not, then there are three constants λ_1, λ_2, and λ_3, not all zero for which

$$\sum \lambda_i f_i(x) = 0$$

for every $x \in \mathbb{R}$. In particular, this means that for each j

$$\sum_i \lambda_i f_i(x_j) = 0.$$

Therefore, letting r_i denote the i-th row of the matrix, we have

$$\sum_i \lambda_i r_i = \left(\sum_i \lambda_i f_i(x_1), \sum_i \lambda_i f_i(x_2), \sum_i \lambda_i f_i(x_3) \right) = (0,0,0).$$

This means $\{r_1, r_2, r_3\}$ is a linearly dependent set and contradicts the assumption that $\{r_1, r_2, r_3\}$ is a linearly independent set. $\square$

Thursday June 7, 2012

2.6.4(b) (Gautam G.) With notion as in the previous problem we assume here that the three functions are twice differentiable. If there is some x_* for which the matrix with i-th row

$$r_i = (f_i(x_*), f_i'(x_*), f_i''(x_*))$$

has linearly independent rows, then the three functions are linearly independent.

We again proceed by contradiction. Again, we assume there are three constants λ_1, λ_2, and λ_3, not all zero for which

$$\sum \lambda_i f_i(x) = 0$$

for every $x \in \mathbb{R}$. Differentiating with respect to x and evaluating at $x = x_*$, we find

$$\sum \lambda_i f_i'(x_*) = 0.$$

Similarly, differentiating twice with respect to x, we find

$$\sum \lambda_i f_i''(x) = 0$$

for every $x \in \mathbb{R}$. Evaluating at x_*, we get

$$\sum \lambda_i f_i''(x_*) = 0.$$

Thus,

$$\sum_i \lambda_i r_i = \left(\sum_i \lambda_i f_i(x_*), \sum_i \lambda_i f_i'(x_*), \sum_i \lambda_i f_i''(x_*) \right) = (0,0,0).$$

Again this means $\{r_1, r_2, r_3\}$ is a linearly dependent set and contradicts the assumption that $\{r_1, r_2, r_3\}$ is a linearly independent set. □

The Replacement Lemma (Theorem 7.4) and Theorems 5.1 and 5.3 (Kowshik M.) This is a pretty neat result. It says that if we have some vectors $v_1, v_2, \ldots, v_k$ and some linearly independent vectors $w_1, w_2, \ldots, w_\ell$ in the span of the v_j, then first of all $k \geq \ell$, and we can find a subcollection A' of $A = \{v_1, v_2, \ldots, v_k\}$ with ℓ elements such that

$$B = \{w_1, w_2, \ldots, w_\ell\} \cup (A \backslash A')$$

spans the same vector space as A.

The proof is by what you might call "finite induction" or "exhaustive induction." This works like induction, except there are really only finitely many cases to check.

The initial thing to prove is that we can replace one of the v's. In fact, since

$$w_1 = \sum a_j v_j$$

for some constants a_j, and we know w_1 is nonzero, it must be the case that some $a_{j_0} \neq 0$. It is pretty clear, of course, that the span of $\{w_1\} \cup \{v_j : j \neq j_0\}$ is in the span of the original v_j's. (Just look at the form of w_1 above.)

On the other hand, we have

$$v_{j_0} = \frac{1}{a_{j_0}} \left(w_1 - \sum_{j \neq j_0} a_j v_j \right) = \frac{1}{a_{j_0}} w_1 - \sum_{j \neq j_0} \frac{a_j}{a_{j_0}} v_j.$$

(Just look at the form of w_1 above.) Therefore, anything in the span of the original v_j's has the form

$$\sum c_j v_j = \sum_{j \neq j_0} c_j v_j + c_{j_0} \left(\frac{1}{a_{j_0}} w_1 - \sum_{j \neq j_0} \frac{a_j}{a_{j_0}} v_j \right).$$

Regrouping, we get

$$\sum c_j v_j = \sum_{j \neq j_0} \left(c_j - \frac{c_{j_0} a_j}{a_{j_0}} \right) v_j + \frac{c_{j_0}}{a_{j_0}} w_1.$$

That is to say, $\sum c_j v_j$ is in the span of the modified collection of vectors.

It will be noted that the main thing we used in the replacement procedure above was that one of the coefficients of one of the v_j's was nonzero. And the same reasoning will work to replace other vectors as long as this is the case. To be more precise, say we have already replaced a set $A' = \{v'_1, \ldots, v'_m\}$ of the vectors in A where $1 \leq m < \ell$, and we want to replace one more.

First of all, there must be more vectors in $A \backslash A'$ to replace. To see this, note that

$$w_{m+1} = \sum_{j=1}^{m} b_j w_j + \sum a_j v_j$$

where the second sum is taken over vectors $v_j \in A \backslash A'$. If all the coefficients a_j were zero, or there were no more v_j, then we would have

$$\sum_{j=1}^{m} b_j w_j - w_{m+1}$$

which contradicts the linear independence of the w_j.

From this point, we can take j_0 with $v_{j_0} \in A \backslash A'$ and a_{j_0}. Then we'll set $A'' = A' \cup \{v_{j_0}\}$, and show that

$$\operatorname{span}\left(\{w_1, \ldots, w_{m+1}\} \cup (A \backslash A'')\right) = \operatorname{span}\left(\{w_1, \ldots, w_m\} \cup (A \backslash A')\right) = \operatorname{span}(A).$$

The argument is very similar to the one above.

It's clear that you won't get anything new in this new span.

On the other hand, we know we have

$$v_{j_0} = \frac{1}{a_{j_0}} \left(w_{m+1} - \sum_{j=1}^{m} b_j w_j - \sum_{v_j \notin A',\ j \neq j_0} a_j v_j \right) = \frac{1}{a_{j_0}} w_{m+1} - \sum_{j=1}^{m} \frac{b_j}{a_{j_0}} w_j - \sum_{v_j \notin A',\ j \neq j_0} \frac{a_j}{a_{j_0}} v_j.$$

(Just look at the form of w_{m+1} above.) Therefore, anything in the span of $A \backslash A'$ has the form

$$\sum_{j=1}^{m} d_j w_j + \sum c_j v_j = \sum_{v_j \notin A',\ j \neq j_0} c_j v_j + c_{j_0} \left(\frac{1}{a_{j_0}} w_{m+1} - \sum_{j=1}^{m} \frac{b_j}{a_{j_0}} w_j - \sum_{v_j \notin A',\ j \neq j_0} \frac{a_j}{a_{j_0}} v_j \right).$$

Regrouping, we get

$$\sum_{j=1}^{m} d_j w_j + \sum c_j v_j = \sum_{j \neq j_0} \left(c_j - \frac{c_{j_0} a_j}{a_{j_0}} \right) v_j + \frac{c_{j_0}}{a_{j_0}} w_1.$$

That is to say, an arbitrary element from the old span is in the span of the modified collection of vectors.

□

Remark: The result above (the replacement lemma) is a generalization of Theorem 7.4.

Theorem 5.1: If you have $\{w_1, \ldots, w_\ell\} \subset \operatorname{span}\{v_1, \ldots, v_k\}$ and $\ell > k$, then $\{w_1, \ldots, w_\ell\}$ is linearly dependent.

Proof: Assume that $\{w_1, \ldots, w_\ell\}$ is linearly independent. Then use the replacement lemma to replace all of the vectors in $\{v_1, \ldots, v_k\}$ with $\{w_1, \ldots, w_k\}$. Then you end up with $w_{k+1} \in \operatorname{span}\{w_1, \ldots, w_k\}$, which contradicts the linear independence of $\{w_1, \ldots, w_\ell\}$. □

Theorem 5.3: If $\{w_1, \ldots, w_\ell\}$ and $\{v_1, \ldots, v_k\}$ are both linearly independent sets which span the same subspace, then $\ell = k$.

Proof: Assume $\ell < k$. Then Theorem 5.1 says that $\{v_1, \ldots, v_k\}$ is linearly dependent (a contradiction). Thus, $\ell \geq k$. But if you assume $\ell > k$, then you get a symmetric contradiction. □

Remark: This last result, Theorem 5.3, says that the notion of dimension is well defined for finitely generated vector spaces.

Here is another homework assignment for Tuesday June 12: 2.7.1, 4, 5, 6.

Tuesday June 12, 2012

2.7.3 (Xian W.) This effort to show that a subspace of a finitely generated space is finitely generated seemed to show some progress. The strategy (as I interpret it) was to show the following:

> *If a space S is not finitely generated, then given any integer n, there are n vectors $s_1, \ldots, s_n \in S$ which form a linearly independent set.*

If this claim is established, then one seeks a contradiction.

finite fields (Sarah M. and John R.) Sarah showed that no $\mathbb{Z}_n$ can be a field if $n = pq$ is composite as follows:

If p has a multiplicative inverse k, then $kp = mpq+1$. Thus, $(k-mq)p = 1$. But then you're multiplying two integers which are between 1 and $pq-1$ and getting 1, which is impossible.

Here is an alternative proof: Start as Sarah did. If p has a multiplicative inverse k, then $kp = 1$. (This is in the field (modular) arithmetic now, rather than in integer arithmetic.) But then $kpq = q$. (Multiply both sided by q.) On the other hand, $kpq = 0$, since kpq is a multiple of pq. Thus, $q = 0$ which is a contradiction.

Sarah also gave a partial proof that $\mathbb{Z}_p$ is a field when p is prime. In particular, she showed that each integer k between 1 and $p-1$ has a multiplicative inverse. The strategy was as follows: Look at the products mk for $m = 0, 1, \ldots, p-1$ and show they are all different. This means there must be p of them, and one of them must be 1.

In terms of real arithmetic, if two of the products km_1 and km_2 were the same, then you would have some $r \in \mathbb{Z}_p$ with $km_1 = q_1p + r$ and $km_2 = q_2p + r$. We can assume that $m_1 < m_2$, which I don't think Sarah mentioned, and then it follows that $q_1 < q_2$ as well. Then,

$$k(m_2 - m_1) = (q_2 - q_1)p.$$

We now have a product on the left of two integers which are both less than p equal to a product of two positive integers, and one of them is p. However, every positive integer greater than 1 is a unique product of primes. This means that since p is a factor on the right, it must also be a factor somewhere in $k(m_2 - m_1)$. But there are no factors of p on the left, since both k and $m_2 - m_1$ are smaller than p.

The last part of Sarah's argument is a proof that there are no *zero divisors* in $\mathbb{Z}_p$. This is a general property of fields:

> *If $ab = 0$ in a field, then either $a = 0$ and $b = 0$.*

Can you prove this?

It was also pointed out that there are other field properties that must be verified for $\mathbb{Z}_p$. John and I gave a shot at proving the distributive property. I think what we did was a bit too complicated:

In real arithmetic we can write

$$ab = pq_1 + r_1,$$

and

$$ac = pq_2 + r + 2.$$

It follows that in integer arithmetic

$$a(b + c) = ab + ac = p(q_1 + q_2) + r_1 + r_2.$$

Then, if $r_1 + r_2 = pq_0 + r_0$, we have

$$a(b + c) = ab + ac = (q_1 + q_2)p + r_1 + r_2 = (q_0 + q_1 + q_2)p + r_0.$$

That is to say, $a(b + c)$ and $ab + ac$ are the same (r_0) modulo p. In other words, they are the same in the $\mathbb{Z}_p$. Thus, in $\mathbb{Z}_p$,

$$a(b + c) = r_0 = ab + ac.$$

It is natural, in this situation to want to express $b + c$ by its modular equivalent, but this should be unnecessary if one has properly showed that the multiplication in $\mathbb{Z}_p$ is well defined. This, and several other properties, are worth writing down.

We also talked about the notion of a *field isomorphism*. That is a function ϕ which takes one field to another on a one-to-one and onto fashion and preserves the operations:

$$\phi(a + b) = \phi(a) + \phi(b) \quad \text{and} \quad \phi(ab) = \phi(a)\phi(b).$$

If there is such a function, then we say that the fields are "the same."

It can be shown that every finite field is $\mathbb{Z}_p$ up to isomorphism.

1.2.6 (Peter Y.) If we assume that 0 in a field has a multiplicative inverse, then we end up with the contradiction that every element in the field is 0.

Lemma: $a \cdot 0 = 0$ for every a in the field. (Note: This proof uses the assumption that 0 has a multiplicative inverse, but only for the case when $a = 0$. Can you give a proof in that case independent of the contradictory assumption of the problem?)

Proof: For every b,

$$\begin{aligned} b + a \cdot 0 &= b \cdot 1 + a \cdot 0 \\ &= b(a \cdot a^{-1}) + a \cdot 0 \\ &= a(ba^{-1} + 0) \\ &= aba^{-1} \\ &= b. \end{aligned}$$

This says that $a \cdot 0$ is behaving like an additive identity. Since the additive identity in a field is unique (proof?), we have established the lemma.

Next Peter used the lemma to solve the problem: $a = a \cdot 1 = a(0 \cdot 0^{-1}) = (a \cdot 0^{-1}) \cdot 0 = 0$. $\square$

2.7.4 (Omar V.) Here we want to see that two two-dimensional subspaces of $\mathbb{R}^3$ must intersect in at least a one-dimensional subspace. We didn't get to look at Omar's argument, so we'll do it next time.

2.5.2 (McCuan) Nobody seems to have gone back to the dreaded 2.5.2 and given a solution which parallels the proof of Theorem 5.1 (induction). I promised I would provide that, so here goes:

We first check that if $w_1 = c_1(1,0)$ and $w_2 = c_2(1,0)$, then $\{w_1, w_2\}$ is linearly dependent. This is easy since

$$c_2 w_1 - c_1 w_2 = (0,0),$$

and the only circumstances under which both coefficients is zero is if w_1 and w_2 are both the zero vector. Notice that the same kind of argument would work if we replaced $(1,0)$ with $(0,1)$ in the above assertion. OK, so we have a little lemma to use.

Now, imagine that $x = (x_1, x_2)$, $y = (y_1, y_2)$, and $z = (z_1, z_2)$. Using the previous reasoning, we can assume that $x_1 \neq 0$. (Otherwise, we have three vectors which are all multiples of $(0,1)$, and we know from the lemma that even a set with two of them would be linearly dependent.) Now, we set

$$w_1 = y - (y_1/x_1)x$$

and

$$w_2 = z - (z_1/x_1)x.$$

You will note that these two vectors are both multiples of $(0,1)$. It follows from the lemma that they form a linearly dependent pair. This means there are coefficients λ_1 and λ_2, not both zero, with

$$\sum \lambda_j w_j = (0,0).$$

Expanding, this means:

$$\lambda_1 y - (\lambda_1 y_1/x_1)x + \lambda_2 z - (\lambda_2 z_1/x_1)x = -(\lambda_1 y_1/x_1 + \lambda_2 z_1/x_1)x + \lambda_1 y + \lambda_2 z = (0,0).$$

Since this is a nontrivial linear combination of x, y, and z, we are done. $\square$

Thursday June 14, 2012

2.8.3 (Jevon R.) Given m equations

$$\sum_{j=1}^{n} \alpha_{ij} x_j = 0, \qquad i = 1, \ldots, m$$

for the n unknowns $x_1, \ldots, x_n$, we can form the coefficient matrix (α_{ij}) with columns

$$c_j = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{mj} \end{pmatrix}$$

and the unknown vector

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n, \end{pmatrix}$$

so that the system becomes equivalent to $(\alpha_{ij})x = 0 \in \mathbb{R}^m$.

Alternatively, we can express the same system as

$$\sum_{j=1}^{n} x_j c_j = 0 \in \mathbb{R}^m.$$

In this latter formulation, the existence of a nonzero solution x is equivalent to the assertion that $\{c_1, \ldots, c_n\}$ is a linearly dependent set in $\mathbb{R}^m$. If $n > m$, then this follows direclty from Theroem 5.1 since $\mathbb{R}^m$ is generated by $\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}$.

2.7.3 (Xian W.) *Any vector space which is not finitely generated, has linearly independent subsets of arbitrary finite size.*

Proof by induction on the number of elements in the set: For a single element set, simply take any nonzero element in the space. (If there is no such element, then the space is $\{0\}$, and we agree that it is finitely generated by convention.)

Now, say we have a linearly independent set $\{v_1, \ldots, v_k\}$ with k elements. Since the space is not finitely generated, there is some vector w which is not in the span of $\{v_1, \ldots, v_k\}$. Set $v_{k+1} = w$. We claim that $\{v_1, \ldots, v_{k+1}\}$ is linearly independent. To see this, consider a linear combination

$$\sum_{j=1}^{k+1} \lambda_j v_j = \sum_{j=1}^{k} \lambda_j v_j + \lambda_{k+1} w = 0.$$

We know that $\lambda_{k+1} = 0$, since otherwise, we have

$$w = \sum_{j=1}^{k} \left(-\frac{\lambda_j}{\lambda_{k+1}} \right) v_j$$

which is in the span of $\{v_1, \ldots, v_k\}$. Consequently, we have

$$\sum_{j=1}^{k} \lambda_j v_j = 0$$

which implies $\lambda_1 = \cdots = \lambda_k = 0$, since $\{v_1, \ldots, v_k\}$ is linearly independent. This establishes the lemma stated above.

Now, if S is a subspace of a finitely generated space T, then we know T is generated by some finite collection of vectors $\{v_1, \ldots, v_k\}$. Let $n > k$. If we assume S is not finitely generated, then we can find a collection of n linearly independent vectors in S. Since these vectors will also be in T, Theorem 5.1 says they must form a linearly dependent set. This is a contradiction.

Thus, S is finitely generated. Using Theorem 7.2, we know S and T both have bases. If the basis for S has more elements than that of T, then we again get a contradiction of Theorem 5.1, since the basis of S would be a linearly independent set in T with too many elements. Thus, the dimension of S cannot exceed that of T.

If equality holds between the dimensions of S and T, then there is a basis $\{v_1, \ldots, v_k\}$ of S with $k = \dim(T)$. If we assume there is some $w \in T \backslash S$, the the reasoning above implies that $\{v_1, \ldots, v_k, w\}$ is a linearly independent set. But this leads again to the same contradiction of Theorem 5.1 since $\{v_1, \ldots, v_k, w\}$ has too many elements for a linearly independent set in T. $\square$

zero divisors in a ring (Hadrien Glaude)

If F is a field, $a, b \in F$, and $ab = 0$, then $a = 0$ or $b = 0$.

Proof: Assume $ab = 0$ and $a \neq 0$. Then $b = a^{-1}ab = a^{-1} \cdot 0 = 0$. Similarly, if $ab = 0$ and $b \neq 0$, then $a = 0$. $\square$

Here are some (cool and important) definitions:

Group: A group is a set G with an operation $+$ which is associative and together they satisfy the following two properties: (1) There is an (additive) identity $0 \in G$, and (2) Every element has an (additive) inverse.

Note: Sometimes the operation in a group might be a kind of multiplication, as with the group of invertible matrices, or the group of nonzero elements of $\mathbb{R}$ under multiplication.

Note: A group is called *commutative* if $a + b = b + a$ for every pair of elements a and b in the group.

Ring: A ring A is a set with two associative operations, addition and multiplication such that the following are satisfied: (1) A is a commutative group with respect to addition, and (2) there is a multiplicative identity $1 \in A$, (3) multiplication distributes across addition: $a(b + c) = ab + ac$.

Note: We didn't say anything about multiplicative inverses in R. If you have multiplicative inverses for nonzero elements, then I think you get the property that $ab = 0$ implies $a = 0$ or $b = 0$. (Just use the proof above and show that $0 \cdot a = a \cdot 0 = 0$ in a ring like in a field.)

Note: A ring is called *commutative* if it is commutative with respect to the multiplication.

A nonzero element of a ring which can be multiplied by another nonzero element to get 0 is called a *zero divisor*. The nonexistence of zero divisors in a ring has a special name:

A ring is called an *integral domain* if $ab = 0$ implies $a = 0$ or $b = 0$.

With this new terminology, we can rephrase the notion of a field:

> *A* field *is a commutative ring A such that $A^* = A\backslash\{0\}$ is a group under multiplication.*

Here is a generalization of Sarah's argument that elements in $\mathbb{Z}_p$ have multiplicative inverses:

If A is a finite commutative ring and there are no zero divisors, then A is a (finite) field.

Proof: Given $a \in A^*$, define $f : A \to A$ by $f(x) = ax$.

First observe that f is one-to-one (injective): If $f(x) = f(y)$, then $ax = ay$. Thus, $ax - ay = a(x - y) = 0$. Since there are no zero divisors, and $a \neq 0$, we must have $x - y = 0$, i.e., $x = y$.

Next, we claim that f is onto (surjective). The reason is because the cardinality (number of elements) in $\{f(x) : x \in A\}$ is the same as the cardinality of A. In particular, there is some $x \in A$ with $f(x) = ax = 1$. □

8.1(h) (David G.) Given a system of m linear equations in n unknowns $x_1, \ldots, x_n$ as above, we can leave out the x's to form the *augmented matrix*:

$$\left(\begin{array}{ccc||c} \alpha_{11} & \cdots & \alpha_{1n} & b_1 \\ \vdots & & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} & b_m \end{array}\right).$$

Note that this system/equation is non-homogeneous: $(\alpha_{ij})x = b$.

One point that needs to be clear is that whenever you have a system like this, you can encode all the relevant information about the problem in this augmented array, and conversely, given an augmented $m \times (n + 1)$ array, you can construct a system of m equations for n unknowns.

Next, associated with such a system (or with an augmented array), there is a solution set $\Sigma = \{x \in \mathbb{R}^n : (\alpha_{ij})x = b\}$. David's first claim is that the solution set doesn't change under elementary row operations. In fact,

it's pretty clear that switching rows or multiplying a row by a nonzero constant doesn't change the solution set. Let's think carefully about what happens when we replace row i with row i plus λ times row j:

$$A' = \left(\begin{array}{ccc||c} \alpha_{11} & \cdots & \alpha_{1n} & b_1 \\ \vdots & & & \vdots \\ \alpha_{i1} + \lambda\alpha_{j1} & \cdots & \alpha_{in} + \lambda\alpha_{jn} & b_i + \lambda b_j \\ \vdots & & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} & b_m \end{array}\right).$$

Most of the equations are the same, so if x is a solution of the original system, then it will satisfy all of the equations except the i-th one for sure. The i-th one will be satisfied as well:

$$\sum_k (\alpha_{ij} + \lambda\alpha_{jk})x_k = \sum_k \alpha_{ij}x_k + \sum_k \lambda\alpha_{jk}x_k = b_i + \lambda b_j.$$

On the other hand, if x satisfies the new system associated with the modified matrix A', then it satisfies all the original equations except maybe the i-th one. And if we look at the i-th one, we see

$$\begin{aligned} \sum_k \alpha_{ij}x_k &= \sum_k \alpha_{ij}x_k + \lambda\sum \alpha_{jk}x_k - \lambda\sum \alpha_{jk}x_k \\ &= \sum_k (\alpha_{ij} + \lambda\alpha_{jk})x_k - \lambda\sum \alpha_{jk}x_k \\ &= b_i + \lambda b_j - \lambda b_j \end{aligned}$$

since the last sum is b_j by the j-th equation.

It was also pointed out that systems like this can be envisioned in terms of a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ given by $L(x) = (\alpha_{ij})x$. One is asking the question: What is the point set in $\mathbb{R}^n$ which maps onto a particular point $b \in \mathbb{R}^m$. If you want to understand how linear maps work, the relation to linear systems of equations is pretty clear. It has something to do with how close or how far the linear map is from being one-to-one. There are other basic relations which we will/should see soon.

Thus, we have established that the solutions set remains unchanged under elementary row operations. In particular, we can do some kind of row reduction in order to easily solve such systems. Exercise (h) gives an

example:

$$\left(\begin{array}{ccc||c} 2 & 1 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right) \xrightarrow{r_1 \to r_3} \left(\begin{array}{ccc||c} 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 2 & 1 & -1 & 0 \end{array}\right)$$

$$\xrightarrow{r_2 \to r_2 - r_1,\ r_3 \to r_3 - 2r_1} \left(\begin{array}{ccc||c} 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -1 \\ 0 & -1 & -3 & -2 \end{array}\right)$$

$$\xrightarrow{r_2 \to -r_2,\ r_3 \to -r_3} \left(\begin{array}{ccc||c} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 3 & 2 \end{array}\right)$$

$$\xrightarrow{r_3 \to r_3 - r_2} \left(\begin{array}{ccc||c} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

Now, starting at the last equation and working up, we see

$$x_3 = 1.$$

$$x_2 = 1 - 2x_3 = -1.$$

And

$$x_1 = 1 - x_2 - x_3 = 1.$$

The solution $(1, -1, 1)^T$ is easily seen to be a solution of the original system. In fact, we have shown it is the unique solution. □

uniqueness of 0 in a field (Dustin M.) This is a simplification of Dustin's original argument... and maybe a correction of it. If we assume there is some other element, say a, which behaves like an additive identity, then $x + a = x$ for every x. In particular, if we put the other additive identity 0 in for x we get $0 + a = 0$. But since 0 is acting like an additive identity too, $0 + a = a$ and we have $a = 0$. □

Tuesday June 19, 2012

2.9.3 (Claudia R.) I'm not exactly how Claudia's solution went, but here is one by row reduction:

$$\left(\begin{array}{cccc||c} -1 & 2 & 1 & 4 & 0 \\ 2 & 1 & -1 & 1 & 1 \end{array}\right) \xrightarrow{r_1 \to -r_1,\ r_2 \to r_2 - 2r_1} \left(\begin{array}{cccc||c} 1 & -2 & -1 & -4 & 0 \\ 0 & 5 & 1 & 9 & 1 \end{array}\right).$$

The last equation is now $5x_2 = 1 - x_3 - 9x_4$, and we can let x_3 and x_4 be any numbers. The first equation then gives $x_1 = 2x_2 + x_3 + 4x_4 = 2(1/5 - x_3/5 - 9x_4/5) + x_3 + 4x_4 = 2/5 + 3x_3/5 + 2x_4/5$. Thus,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2/5 + 3x_3/5 + 2x_4/5 \\ 1/5 - x_3/5 - 9x_4/5 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2/5 \\ 1/5 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 3/5 \\ -1/5 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 2/5 \\ -9/5 \\ 0 \\ 1 \end{pmatrix}.$$

It will be noted that this is a two-dimensional plane in $\mathbb{R}^4$ but not a subspace. This kind of set is sometimes called an em affine subspace due to the affine shift of the plane.

2.8.4 (Peter Y.) The objective here is to show that a homogeneous system of equations

$$\sum x_j c_j = 0$$

in n unknowns has a nontrivial solution if and only if the rank of the coefficient matrix is less than n.

If the system has a nontrivial solution, then this means exactly that the columns $\{c_1, \ldots, c_n\}$ are linearly dependent. If we denote by S the spaced spanned by these columns, then they are a generating set, and by Theorem 7.3, some subset of the columns provides a basis for S. On the other hand, the basis cannot consist of all elements of the set of columns, since that is a linearly dependent set. Thus, the number of elements in the basis, which is the rank of the coefficient matrix, is less than n the number of columns.

Conversely, if the rank of the coefficient matrix C is less than n, then we can use Theorem 5.1 to conclude that the columns are linearly dependent. To be precise, the space S spanned by the columns as above has a basis (and a generating set in particular) with fewer than n elements. Now Theorem 5.1 applies. □

If we haven't noted it above, let us note here that the *rank* of a matrix is given in Definition 8.6 as the dimension of the column space S defined above.

2.9.5 (Aishwarye C.) This one appears to still be open.

2.9.6 (Frank P.) If $(x, y) = (\alpha, \beta)$ and (γ, δ) are two points satisfying $Ax + By + C = 0$, then

$$A \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} + B \begin{pmatrix} \beta \\ \delta \end{pmatrix} + C \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since there are three unknowns and the rank can be no more than two, we know by problem 2.8.4 that there exists a nontrivial solution (A, B, C).

I don't really see an argument relating to other possible choices of (A, B, C). Perhaps the idea was to use Corollary 9.4. (Did anyone explain why this result is true?) If we can show that the rank is exactly two, then Corollary 9.4 says that the solution space is exactly one dimensional, which would mean that our original nontrivial soltuion $(A, B, C)^T$ would be a basis vector for the solution space, and we would get a one dimensional subspace of solutions as asserted in the problem.

Looking at the coefficient matrix

$$\begin{pmatrix} \alpha & \beta & 1 \\ \gamma & \delta & 1 \end{pmatrix},$$

we can see that the rank is exactly two as follows: First of all, the rank is at least one since $(1,1)^T$ is nonzero. Now, if the rank were one instead of two, then both other column vectors would be multiples of the last column vector. That means that the matrix would be

$$\begin{pmatrix} \alpha & \beta & 1 \\ \alpha & \beta & 1 \end{pmatrix}.$$

But then we would have $(\gamma, \delta) = (\alpha, \beta)$ which we do not have since we know the points are distinct. Thus, the rank is two and Corollary 9.4 gives that the solution space is one-dimensional.

The final assertion of this problem is that the line $\{(x,y) : Ax+By+C = 0\}$ passing through two distinct points is unique in $\mathbb{R}^2$. We have shown that any other line defined by an equation $A'x + B'y + C' = 0$ would be $\lambda(Ax + By + C) = 0$. Of course, if λ were zero, then $A'x + B'y + C' = 0$ would not define a line. Otherwise, $\lambda \neq 0$, and we get the same (unique) line. □

2.8.2 (David G.) This is another problem on systems of equations. The system in question is given by the augmented matrix

$$\left(\begin{array}{ccc||c} 3 & -1 & \alpha & 1 \\ 3 & -1 & 1 & 5 \end{array}\right) \xrightarrow{r_2 \to r_2 - r_1} \left(\begin{array}{ccc||c} 3 & -1 & \alpha & 0 \\ 0 & 0 & 1-\alpha & 4 \end{array}\right).$$

The last equation after reduction is $(1-\alpha)x_3 = 4$. First of all, if $\alpha = 1$, then there is no solution. That's clear.

If $\alpha \neq 1$, then we have $x_3 = 4/(1-\alpha)$ and $3x_1 - x_2 + x_3 = 3x_1 - x_2 + 4/(1-\alpha) = 0$. Making x_1 arbitrary, we get infinitely many solutions

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ 3x_1 + 4/(1-\alpha) \\ 4/(1-\alpha) \end{pmatrix} = \begin{pmatrix} 0 \\ 4/(1-\alpha) \\ 4/(1-\alpha) \end{pmatrix} + x_1 \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}.$$

subgroups of $\mathbb{Z}$ (Hadrian G.) A subgroup of a group is a subset which is a group under the same operation. As an exercise, you can check that a subset H of a group G is a subgroup if and only if (1) H contains the additive identity and (2) $a - b \in H$ whenever $a, b \in H$.

We are going to use the characterization of subgroups to show that *H is a subgroup of $\mathbb{Z}$ if and only if there is some natural number m such that $H = \{mn : n \in \mathbb{Z}\}$.*

Proof: If $H = m\mathbb{Z}$, then (1) $0 \in H$ and (2) If $a, b \in H$, then $a = pm$ and $b = qm$ for some p and q. Therefore, $a - b = (p-q)m \in m\mathbb{Z}$. It follows that H is a subgroup of $\mathbb{Z}$.

Conversely, if H is any subgroup...

Tuesday July 3, 2012

3.11.5 (Claudia R.) The objective here is to show the image under a linear function of a subspace is a subspace. The first thing to do is look at the form of the image:

$$L(V) = \{L(v) : v \in V\}.$$

Here I'm denoting the subspace by V to make the notation simple. It has been noted earlier that it is enough to show $L(V)$ is closed under addition and scalar multiplication. For addition, we take two arbitrary elements of $L(V)$ and try to express them as L of some element in V. Using the linearity of L this is not hard:

$$L(v) + L(w) = L(v + w).$$

Since V is closed under additon, we see that $v + w \in V$ and $L(v + w) \in L(V)$. Similarly,

$$\alpha L(v) = L(\alpha v),$$

and since V is closed under scalar multiples, $\alpha v \in V$ and $L(\alpha v) \in L(V)$.

3.11.8(a) (Gautam G.) Gautam's key observations are the following:

- A linear transformation is onto if every vector in the target space can be written as a linear combination of the columns of the matrix for the linear transformation.
- If any basis for the target can be expressed as a linear combination of the columns, then every other vector in the target can be as well.

For this problem, the target is $\mathbb{R}^2$ and the matrix is

$$\begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix}.$$

The first column minus the second column if $4\mathbf{e}_1$. And the first column plus three times the second column is $4\mathbf{e}_2$.

3.12.3(b) (Samantha A.) Here we want to solve

$$\begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The first equation implies that $x_1 = -x_2$, and the second equation then becomes $2x_2 + 2x_3 = 0$, or $x_2 = -x_3$. Setting $x_3 = a$, the solution set is

$$\Sigma = \left\{ \begin{pmatrix} a \\ -a \\ a \end{pmatrix} : a \in \mathbb{R} \right\}.$$

This is a straight line in $\mathbb{R}^3$ passing through the origin and the point $(1, -1, 1)^T$.

3.11.10 (John R.) An integration mapping on polynomials is defined by

$$I(f) = \sum_{j=0}^{k} \frac{a_j}{j+1} x^{j+1}$$

where $f = \sum a_j x^j$. Differention gives another linear mapping D with the usual formula. In particular, differentiating the expression above, we get

$$DI(f) = \sum_{j=0}^{k} a_j x^j,$$

so that D composed with I is the identity mapping on polynomials.

It should be noted that D is consequently onto since for any polynomial f, we have $I(f) = g$ is a polynomial, and $f = D(I(f)) = D(g)$. On the other hand, D is not one-to-one since the image of any constant polynomial under D is the zero polynomial.

The transformation I is one-to-one, because if we compare expressions like that above formally:

$$I(f) = \sum_{j=0}^{k} \frac{a_j}{j+1} x^{j+1} = \sum_{j=0}^{\ell} \frac{b_j}{j+1} x^{j+1} = I(g)$$

for some polynomial $g = \sum b_j x^j$, then we must have $k = \ell$ and $a_j/(j+1) = b_j/(j+1)$ for $j = 0, 1, \ldots, k$. That is, $a_j = b_j$, so that $f = g$. This antiderivative is not onto, however, because it is never the case, for example, that $I(f) = 1$.

Theorem 11.13 (David G.) This theorem says that a linear transformation is invertible if and only if it is one-to-one and onto. For clarity, let's generalize the result to $L(V, W)$ where V and W are possibly different vector spaces. For this, we need to generalize Definition 11.9:

> A linear transormation $T \in L(V, W)$ is *invertible* if there is some linear transformation $T^{-1} \in L(W, V)$ such that
>
> $$T \circ T^{-1} = \mathrm{id}_W \quad \text{and} \quad T^{-1} \circ T = \mathrm{id}_V.$$

If we prove the result for this definition, then the result stated in the text will follow as a special case.

For "only if" direction, we assume T is invertible. Then given any $w \in W$, we can set $v = T^{-1}(w) \in V$, and we find that $T(v) = w$. Thus, T is onto. On the other hand, if $T(v) = T(\tilde{v})$ for some v and $\tilde{v}$ in V, then we can apply T^{-1} to both sides, to see that $v = \tilde{v}$. Hence, T is one-to-one.

In the other direction, we must define the transformation T^{-1}, show it is well defined and linear. And then show the two composition conditions.

First for the definition, we set $T^{-1}(w) = v$ where v is some vector in V with $T(v) = w$.

We know there is such a vector v because T is onto. We claim that there is only one such v. In fact, if $\tilde{v}$ were another vector in V with $T(\tilde{v}) = w$, then we would have $T(v) = T(\tilde{v})$, and it would follow from the injectivity of T that $\tilde{v} = v$.

We have established that T^{-1} is a well defined function. (Remember the defintion: *T^{-1} is a rule or correspondece which assigns to each $w \in W$, a unique $v \in V$.*)

As David correctly points out, we must show that T^{-1} is linear. First we want to consider, say, $T^{-1}(w) + T^{-1}(\tilde{w})$. According to the definition, we can write $T^{-1}(w) = v$ and $T^{-1}(\tilde{w}) = \tilde{v}$, where

$$T(v) = w,$$

$$T(\tilde{v}) = \tilde{w},$$

and by the linearity of T,

$$w + \tilde{w} = T(v + \tilde{v}).$$

Using this last identity and the defintion of T^{-1}, we see that

$$T^{-1}(w + \tilde{w}) = v + \tilde{v}.$$

But $v + \tilde{v} = T^{-1}(w) + T^{-1}(\tilde{w})$, so we have that T^{-1} is additive on vectors in W.

Similarly, we see that

$$T(\alpha T^{-1}(w)) = \alpha T \circ T^{-1}(w),$$

and since $T^{-1}(w)$ is the vector whose image under T is w, this becomes

$$T(\alpha T^{-1}(w)) = \alpha w.$$

That is, $\alpha T^{-1}(w)$ is the vector whose image under T is *alphaw*. By the definition of T^{-1}, this means

$$T^{-1}(\alpha w) = \alpha T^{-1}(w).$$

We have now shown that T^{-1} is well defined and linear. It remains for Thursday to show that $T \circ T^{-1} = \mathrm{id}_W$ and the similar composition assertion with the reverse order.

Thursday July 5, 2012

Theorem 11.13 (David G.) Remember the definition: $T^{-1}(w)$ is the vector v such that $T(v) = w$. Thus, to compute

$$T \circ T^{-1}(w)$$

we look for $T(v)$ where v is a vector with $T(v) = w$. Well, I guess it's w, so we're done in this case. That is,

$$T \circ T^{-1} = \mathrm{id}_W.$$

On the other hand,

$$T^{-1} \circ T(v)$$

is the vector $\tilde{v}$ such that $T(\tilde{v}) = T(v)$. That vector is v.

Theorem 13.10 (Liangyi S.) Actually, this is only part of the proof of the theorem which says that a linear transformation of a finite dimensional vector space into itself is invertible if and only if it is surjective (onto).

Furthermore, Liangyi did a version of the theorem for matrices, which is apriori a special case, but turns out to be more or less the same thing.

Here's how it went.

By saying a matrix $A_{n\times n}$ is invertible, we mean there is a matrix $B = A^{-1}$ such that $AB = BA = I$ the identity matrix. If this is true, we want to show the rank of A is n (which is the same thing as being surjective).

It is enough to show the columns of A are linearly independent. Taking a linear combination of the columns which expresses the zero vector, we have $A\mathbf{x} = 0$ (where $\mathbf{x}$ is the column vector with the coefficients). But then we can apply A^{-1} to both sides to conclude $\mathbf{x} = 0$. Thus, there are no nontrivial linear combinations of the columns which give the zero vector, i.e., the columns are linearly independent.

This means the dimension of the image, i.e., the rank, is n.

Conversely, if we assume the rank of A is n, then there is some basis for the image $\mathbb{R}^n$ among the columns. Since the dimension of $\mathbb{R}^n$ is known to be n, however, it must be that the basis contains all the columns. In particular, the columns must be linearly independent. This, as above, means that the equation $A\mathbf{x} = 0$ has only the zero solution.

Now, we want to show injectivity of the mapping $L(\mathbf{x}) = A\mathbf{x}$. If we had $A\mathbf{x} = A\tilde{\mathbf{x}}$, then we would have $A(\mathbf{x} - \tilde{\mathbf{x}}) = 0$, but since we know there is only the zero solution, this gives $\mathbf{x} = \tilde{\mathbf{x}}$. □

compositions (Jevon R.) If $S : M \to N$ and $T : N \to P$ are (linear), one-to-one, and onto, then $T \circ S : M \to P$ is (linear), one-to-one, and onto. The linearity was not shown.

To show surjectivity, let $p \in P$. Since T is surjective, there is an $n \in N$ with $T(n) = p$. Next, since S is surjective, there is some $m \in M$ with $Sm = n$. Therefore, $T(S(m)) = T(n) = p$ and TS is onto.

To show injectivity, assume $TS(m_1) = TS(m_2)$. Since T is injective, this means $S(m_1) = S(m_2)$. But since S in injective, this means $m_1 = m_2$. □

Tuesday July 10, 2012

row and column rank (Xian W.) Let $A_{m\times n}$ be a matrix with column rank r. Denote the columns of A by $C_1, \ldots, C_n$ and a basis for the column space by $\{b_1, \ldots, b_r\}$. Each of the columns admits an expression

$$C_j = \sum_{i=1}^{r} \lambda_{ij} b_i.$$

The resulting coefficients make a matrix $\Lambda_{r\times n}$, and

$$A = B\Lambda$$

where $B_{m\times r}$ is the matrix with $b_1, \ldots, b_r$ in the columns.

Look at this product. Notice that the i-th row R_i of A is given by

$$R_i = \sum_{j=1}^{r} b_{ij}(\lambda_{j1}, \lambda_{j2}, \ldots, \lambda_{jn})$$

where the j-th entry of the column b_i is denoted by b_{ij}. This says that the rows of A are all linear combinations of the r rows of Λ. Since the rows of A are spanned by a set containing r row vectors, we see that the row rank of A cannot exceed the rank r.

This is a general result:

> *The row rank of a matrix is less than or equal to the (column) rank.*

Applying this result to A^T, we find that the rank of A, which is the row rank of A^T, is less than or equal to the rank of A^T. Since the rank of A^T is the row rank of A, we see that the rank of a matrix cannot exceed the row rank of a matrix as well. Thus, the row rank and the (column) rank must be equal.

This is a good proof, but I would still like for you to follow up on the proof using elementary row operations and the pivots of row echelon form. You need to understand the idea of isomorphism pretty well, and this is a good opportunity to do that.

3.13.1 (Sarah M.) Given a linear transformation $T : V \to W$ of a finite dimensional vector space V into a finite dimensional vector space W and given bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ for V and W respectively, the *definition of the matrix of T with respect to the bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$* is the following:

Express the image of each basis element of V in terms of the basis for W:

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

Then *the matrix A of T with respect to $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$* is the matrix with a_{ij} in the i-th row and j-th column.

Notice that if we write any vector $x \in V$ as $x = \sum x_j v_j$, then we can associate a column vector $\xi = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ with x, and we have

$$A\xi = \eta$$

where η is the column vector in $\mathbb{R}^m$ consisting of the coefficients of $T(x) = \sum y_i w_i$. To see this, simply note that

$$T(x) = \sum_j x_j T(v_j) = \sum_j x_j \sum_i a_{ij} w_i = \sum_i \sum_j a_{ij} x_j w_i = \sum_i (\sum_j a_{ij} x_j) w_i.$$

This exercise asks you (i.e., Sarah) to write down a linear transformation $T : V \to V$ with respect to two different bases. First of all, if $\{u_1, u_2\}$ is a basis and $T(u_1) = u_2$, $T(u_2) = u_1$, then the matrix with respect to $\{u_1, u_2\}$ is

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

Now, taking an alternative basis $w_1 = 3u_1 - u_2$, $w_2 = u_1 + u_2$, we can compute to find $u_1 = (w_1 + w_2)/4$ and $u_2 = (-w_1 + 3w_2)/4$. Using these relations, we find $T(w_1) = -w_1 + 2w_2$ and $T(w_2) = w_2$, so that the matrix of T with respect to $\{w_1, w_2\}$ is

$$B = \begin{pmatrix} -1 & 0 \\ 2 & 1 \end{pmatrix}.$$

If you think about this kind of example, you can "see" the linear transformation T as a transformation of $\mathbb{R}^2$ in different ways. One way to think about what you are seeing is that you are taking the particular basis chosen as the "standard basis" $\{\mathbf{e}_1, \mathbf{e}_2\}$ for $\mathbb{R}^2$. If you think about it like this, then the relation

$$\begin{cases} w_1 = 3u_1 - u_2 \\ w_2 = u_1 + u_2 \end{cases}$$

is saying that you are going to use a *change of basis matrix M* which sends the vector $3\mathbf{e}_1 - \mathbf{e}_2$ (which is the vector you think of as $3u_1 - u_2$

when you are using A) to $\mathbf{e}_1$ (which is the vector you think of as w_1 when you are using B). Similarly, M should send $\mathbf{e}_1 + \mathbf{e}_2$ to $\mathbf{e}_2$. The associated matrix M is not immediately obvious to write down since we are using the standard basis here. But if you think about how to write down the matrix associated with a transformation of $\mathbb{R}^2$ with respect to the standard basis, then it is really easy to write down

$$M^{-1} = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$$

since you know the images of $\mathbf{e}_1$ and $\mathbf{e}_2$ under the inverse of the change of basis. You can check that $B = MAM{-}1$. The matrix M^{-1} is referred to as X in the problem.

rank-nullity theorem (Claudia R.) This is a version of the rank-nullity theorem which says that when you have a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$, then

$$\dim L(\mathbb{R}^n) + \dim \ker(L) = n.$$

That is, the dimension of the kernel and the dimension of the image add up to the dimension of the domain.

Letting A denote the matrix of such a transformation there is a homogeneous system of equations associated with the kernel, namely, $Ax = 0$. Denoting the solution space of this system by Σ,

$$\Sigma = \{x : Ax = 0\}$$

and we want to show

$$\dim \Sigma = n - r$$

where r is the rank of A. (As we know, the rank of A is the dimension of the image of L.)

Claudia (and the book) wishes to begin by taking a basis for the column space from among the columns and then rearranging the columns so that the basis is given by the first r columns. This should really be carefully justified.

Let $C_1, \ldots, C_n$ denote the columns of A. We are going to start a little bit like Xian did and take a basis for the column space, however, since $C_1, \ldots, C_n$ is a generating set, we can use a theorem from the book (or the replacement lemma) to choose this basis from among the columns. Let's say the basis we get is

$$\{C_{j(1)}, C_{j(2)}, \ldots, c_{j(r)}\}.$$

Notice that this means there is a function $j : \{1, 2, \ldots, r\} \to \{1, \ldots, n\}$ which is one-to-one and assigns each integer in its domain to a particular column number from among the basis column numbers. What is desired is to rearrange the columns with the basis columns appearing first.

In order to do this, we can extend the domain of j in any manner to include $\{r+1,\ldots,n\}$ as long as the resulting function $j:\{1,\ldots,n\}\to\{1,\ldots,n\}$ is one-to-one and onto. Let's say we've done this. (To do it rigorously might require some kind of finite induction depending on your taste and how skeptical you happen to be at the time.)

Now, we have a new ordering of columns $\{b_1,\ldots,b_n\}$ with $b_k=C_{j(k)}$. In particular, if we put these columns in a matrix B, then the first r of them make a basis for the column space. We now work with B and attempt to prove that

$$\dim\Sigma_B=n-r$$

where $\Sigma_B=\{x:Bx=0\}$. This works as follows:

For $k=r+1,\ldots,n$, we can write

$$b_k=\sum_{\ell=1}^{r}\alpha_{k\ell}b_\ell.$$

This means that

$$u_k=\sum_{\ell=1}^{r}\alpha_{k\ell}\mathbf{e}_\ell-\mathbf{e}_k$$

is an element of Σ_B for $k=r+1,\ldots,n$. Notice that this is a collection of $n-r$ vectors $\{u_{r+1},\ldots,u_n\}$. We want to claim that this is a basis for Σ_B. This requires three things

1. Each vector u_k should be in Σ_B. (We've claimed this is true, but you should make sure you see why it's true.)
2. The collection should be linearly independent. (This is pretty easy to see since u_k has a -1 for it's k-th entry and u_m has a zero when $m\neq k$.)
3. Each vector in Σ_B can be written as a linear combination of the u_k. (This one is shown in the book, but we haven't done it yet.)

I'll do the second one carefully, and set you up to do the third one.

Let's say $\sum\lambda_k u_k=0$. Then choosing any particular m, the m-th component of the linear combination is $-\lambda_m$. This means that $\lambda_m=0$, so we have linear independence.

For the last one, let x be any element of Σ_B. This means that $Bx=0$. This means that a certain linear combination of the b_k's is zero. You need to show that it also means the vector x can be written as a linear combination of the u_ℓ's. This is a main point.

Markov Chain Matrix (Omar V.) Take a square matrix A for which the entries of each column sum to 1. We would like to show there is a vector x, the sum of whose entries sum to 1 and which satisfies

$$Ax=x.$$

Thursday July 12, 2012

Markov Chain Matrix (Omar V.) Given the square matrix above, we want to show there is a nonzero vector x with $(A-I)x = 0$.

Notice that the matrix $A-I$ has the property that each column has entries summing to zero. Thus, if we apply elementary row operations, we find

$$A - I \xrightarrow{r_n \to r_n - r_1} B_1 \xrightarrow{r_n \to r_n - r_2} \cdots \xrightarrow{r_n \to r_n - r_{n-1}} B_{n-1}$$

where B_{n-1} has the last row all zeros. Thus, the row rank of $A-I$ is less than n. In particular, by the rank-nullity theorem, the dimension of the kernel of $A-I$ is at least $n-(n-1) = 1$, and there is a nonzero vector x with $(A-I)x = 0$.

Such a vector is also called an *eigenvector* with eigenvalue 1, or a *fixed vector* for the matrix A, or a *fixed point* for the corresponding linear transformation.

We haven't yet shown that we can take the sum of the entries in x to be 1.

A^TA (Hadrian G.) Hadrian pointed out some interesting properties of the transpose matrix. First of all, given any two matrices that can be multiplied, one can check that

$$(AB)^T = B^TA^T.$$

As a consequence,

$$(A^TA)^T = A^TA,$$

and this means that A^TA is always symmetric. Furthermore, the nullspace associated with A^TA, namely $\Sigma = \{x : A^TAx = 0\}$, is the same as the null space associated with A.

In order to see this, first observe that $\{x : Ax = 0\} \subset \Sigma$. This is clear. On the other hand, if x satisfies $A^TAx = 0$, then

$$\|Ax\|^2 = (Ax)^T(Ax) = x^TA^TAx = x^T0 = 0.$$

Thus, $Ax = 0$. Thus, $\Sigma \subset \{x : Ax = 0\}$. □

3.13.7, 2, 8 (Dustin M.) If $T : V \to V$, then

$$T^2 = 0 \quad \Leftrightarrow \quad T(V) \subset n(T).$$

$\Leftarrow$: Assume $T(V) \subset n(T)$ and take any $x \in V$. Then $T(x) \in T(V)$, so $T(x) \in n(T)$. This means $T(T(x)) = T^2(x) = 0$. Since x was arbitrary, we have shown that $T^2 = 0$.

$\Rightarrow$: Assume $T^2 = 0$, and take $x \in T(V)$. We know that $x = T(v)$ for some $v \in V$. Thus, $T(x) = T^2(v) = 0$. This means that $x \in n(T)$. □

An earlier problem considers a transformation $S : u_1 \mapsto u_1 + u_2,\ u_2 \mapsto -u_1 - u_2$, where u_1 and u_2 provide a basis for the domain of S. If we compute $S(x) = S(x_1u_1 + x_2u_2)$ for an arbitrary element x in the domain, we find

$$S(x) = (x_1 - x_2)(u_1 + u_2).$$

This means, first of all, that the image is contained in $W = \text{span}\{u_1 + u_2\}$ which is a one-dimensional space since you can't sum basis vectors to get the zero vector.

On the other hand, taking an element of W is the same as taking $x_1 = x_2$ in our original computation. Thus, we see that $W \subset n(S)$. So this example falls into the category of problem 3.13.7 and gives the example sought in problem 3.13.8.

Tuesday July 17, 2012

4.14.4(c) (Noah C.) The basic objective here is to show that a composition of rotations of $\mathbb{R}^2$ is a rotation, the composition of two reflections of $\mathbb{R}^2$ is a rotation, and the composition of a reflection and a rotation is a reflection.

The definition of a rotation is a distance preserving linear transformation of $\mathbb{R}^2$ which has matrix (with respect to the standard basis) of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

The condition that the transformation is distance preserving implies that $a^2 + b^2 = 1$. (Why? Answer: Because the image of $\mathbf{e}_1$ is $(a, b)^T$.)

Conversely, if $a^2 + b^2 = 1$, then there is some angle θ such that $a = \cos\theta$ and $b = \sin\theta$. Then you can easily check that the matrix corresponds to a counterclockwise rotation of the plane through an angle θ.

We discussed similar considerations in regard to reflections of $\mathbb{R}^2$. The basic definition is that the transformation should be distance preserving and the corresponding matrix should have the form

$$\begin{pmatrix} a & b \\ b & -a \end{pmatrix}.$$

It is also true in this case that distance preserving is equivalent to the condition $a^2 + b^2 = 1$. We also discussed how any reflection could be decomposed into the composition of a reflection about one of the standard axes and a rotation, and that given a matrix like that above, there is some axis of reflection associated with the transformation.

Note that there is an angle θ associated with the a and the b appearing in the matrix in this case too. What is the axis of reflection in terms of this θ?

With these considerations out of the way, the rest of the problem is a computation. I will do the composition of a reflection and a rotation.

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} c & d \\ d & -c \end{pmatrix} = \begin{pmatrix} ac-bd & ad+bc \\ bc+ad & bd-ac \end{pmatrix}.$$

Since the upper left entry of the product $\alpha = ac - bd$ is the negative of the lower right entry, and the upper right entry $\beta = ad + bc$ is the same as the lower left entry, the matrix of the composition has the form

$$\begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix}.$$

We need to check that the transformation is distance preserving:

$$\alpha^2+\beta^2 = a^2c^2-2abcd+b^2d^2+a^2d^2+2abcd+b^2c^2 = (a^2+b^2)c^2+(b^2+a^2)d^2 = c^2+d^2 = 1.$$

powers of a linear transformation (Hadrian G.) Here Hadrian took a linear transformation L of a finite dimensional vector space and considered the sequences of inclusion

$$\ker L \subset \ker L^2 \subset \ker L^3 \subset \dots$$

and

$$\operatorname{Im} L \supset \operatorname{Im} L^2 \supset \operatorname{Im} L^3 \supset \dots.$$

To see the first inclusions, note that if $x \in \ker L^p$, then $L^p(x) = 0$. Thus, $L^{p+1}(x) = 0$. To see the second inclusions, note that if $x \in \operatorname{Im} L^p$, then $x = L^p(\xi)$ for some ξ in the domain of L. In particular, $x = L^{p-1}(L(\xi))$. Since $L(\xi)$ is also in the domain of L, we see that $x \in \operatorname{Im} L^{p-1}$. (The second argument only works for $p > 1$.)

Finally, Hadrian wants to show that these sequences stabilize. (If you don't know what it means for a sequence to stabilize, just follow the reasoning below, and that should become clear.)

Notice that the sequence of kernels implies a sequence of inequalities

$$\dim\ker L \leq \dim\ker L^2 \leq \dim\ker L^3 \leq \dots.$$

Furthermore, each of these inequalities is strict unless the consecutive kernels are actually equal. Thus, we have an increasing sequence which is bounded above by the dimension of the domain of L. It follows that there are some equalities in the sequence. Let p be the first power for which

$$\dim\ker L^p = \dim\ker L^{p+1}$$

and consequently

$$\ker L^p = \ker L^{p+1}.$$

Considering $x \in \ker L^{p+2}$, we note that

$$L^{p+1}(L(x)) = 0.$$

Thus, $L(x) \in \ker L^{p+1} = \ker L^p$. That is, $L^p(L(x)) = L^{p+1}(x) = 0$. That is, $x \in \ker L^{p+1}$. We have shown that $\ker L^{p+2} \subset \ker L^{p+1}$. Since the reverse inclusion is given in our original sequence, we have equality.

Extending this reasoning inductively, we see that $\ker L^r = \ker L^p$ for all $r \geq p$. This means that the first sequence stabilizes. We see immediately from this, and the rank-nullity theorem

$$\dim \operatorname{Im} L^r + \dim \ker L^r = \text{constant}$$

that the second sequence must also stabilize.

Thursday July 19, 2012

4.15.5 (Peter Y.) Given a finite orthonormal basis $\{u_1, \ldots, u_n\}$ for a vector space, any vector $u = \sum a_j u_j$ has coordinates

$$a_j = \langle u, u_j \rangle.$$

To see this, just take the inner product of both sides with u_k:

$$\langle u, u_k \rangle = \sum a_j \langle u_j, u_k \rangle = \sum a_j \delta_{jk} = a_k.$$

Note that we have used the notation

$$\delta_{jk} = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k \end{cases}$$

Also, if $v = \sum b_i u_i$, then

$$\langle u, v \rangle = \sum_{i,j} a_j b_i \langle u_j, u_i \rangle = \sum_j a_j \left(\sum_i b_i \delta_{ij} \right) = \sum_j a_j b_j.$$

distance preservation (Claudia R.) This was an interesting discussion of the assertion that a distance preserving function from $\mathbb{R}^2$ to $\mathbb{R}^2$ which fixes the origin must be linear. Here are some highlights:

- Curtis' proof seems to be wrong.
- (Hadrian G.) There is a polarization identity for the inner product:

$$\langle u, v \rangle = \frac{1}{2} \left(\|u + v\|^2 - \|u\|^2 - \|v\|^2 \right).$$

- We don't have a proof of this fact yet.

rotations and reflections (Gautam G.) It would be interesting to show that a linear transformation of $\mathbb{R}^2$ with matrix having the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

or

$$\begin{pmatrix} a & b \\ b & -a \end{pmatrix}$$

and satisfying $a^2 + b^2 = 1$ is distance preserving.

Tuesday July 24, 2012

4.15.7 (Xian W.) Orthogonal transformations are linear functions on $\mathbb{R}^n$ which preserve distance. This is equivalent to the following condition:

> *The matrix A of the transformation with respect to an orthonormal basis satisfies*
>
> $$A^T A = I.$$

Note: One must show the equivalence of this to the definition above, and one also needs to show that the secondary definition is not dependent on the choice of basis. (Can you do this?)

Since we already know that the linear transformations on $\mathbb{R}^n$ form a group, we only need to show closure and inverses.

To see closure, use the secondary definition. If the matrices of two orthogonal transformations are given by A and B, then the matrix of the composition (product) is given by $C = AB$. Now, we check that the composition transformation is orthogonal:

$$C^T C = (AB)^T AB = B^T A^T AB = B^T B = I.$$

Thus, the transormation generated by C is orthogonal and we have closure.

We next check inverses: If A is the matrix of an orthogonal transformation (with respect to an orthonormal basis), then—Hey wait a second, I think we need to show that A is invertible first before we show that its inverse is distance preserving. I don't remember that we did this. So maybe this one is still open. Sorry Xian.

It's true that we need to show invertibility. But let's go on assuming we have done that. Here is Xian's proof that the inverse (if it is well defined) is orthogonal: Let $x \in \mathbb{R}^n$. Then $T^{-1}(x) \in \mathbb{R}^n$. We need to show that the norm of $T^{-1}(x)$ is the same as the norm of x, then distance preserving follows. But since T is distance (and norm) preserving, we have

$$\|T^{-1}(x)\| = \|T(T^{-1}(x))\| = \|x\|.$$

This is what we wanted.

4.15.9 (Rui F.) Here we take a subspace W of a finite dimensional vector space V and we wish to show that

$$\dim W + \dim W^\perp = \dim V.$$

The strategy is to use the identity

$$\dim W + \dim W^\perp = \dim (W + W^\perp) + \dim (W \cap W^\perp)$$

from (7.5) on page 51. It would be good to prove (7.5).

In any case, we want to claim that

$$W + W^\perp = V \quad \text{and} \quad W \cap W^\perp = \{0\}.$$

To see the first assertion, it is enough to show reverse inclusion, since the forward inclusion is obvious from the fact that

$$W + W^\perp = \{w + \tilde{w} : w \in W,\ \tilde{w} \in W^\perp\}$$

and V is closed under addition. (In fact, it would be good to prove that this sum is a subspace.) Now, let $v \in V$. Taking an orthonormal basis $\{w_1, \ldots, w_m\}$ for W, we rewrite v as

$$v = a + (v - a)$$

where

$$a = \sum \langle v, w_j \rangle w_j.$$

An easy computation shows that $b \in W^\perp$. Since it is clear that $a \in W$, we have shown that $v = a + b \in W + W^\perp$.

It remains to prove that $W \cap W^\perp = \{0\}$. This is easy, since the definition of $W^\perp$ is

$$W^\perp = \{x \in V : \langle x, w \rangle = 0 \text{ for all } w \in W\}.$$

Thus, if w is in the intersection, it must satisfy

$$\|w\|^2 = \langle w, w \rangle = 0.$$

This means $w = 0$.

Note: It would also be a good idea to show that $W^\perp$ is a subspace.

rotations and reflections (Gautam G.) See the description of this problem above.

Let L be such a transformation and let the matrix (for example a rotation matrix) be denoted A, so that $L(\mathbf{x}) = A\mathbf{x}$. Then

$$\begin{aligned}
\|L(\mathbf{x})\|^2 &= \|(ax_1 - bx_2, bx_1 + ax_2)^T\|^2 \\
&= (ax_1 - bx_2)^2 + (bx_1 + ax_2)^2 \\
&= a^2x_1^2 + b^2x_2^2 + b^2x_1^2 + a^2x_2^2 \\
&= (a^2 + b^2)x_1^2 + (a^2 + b^2)x_2^2 \\
&= x_1^2 + x_2^2 \\
&= \|\mathbf{x}\|^2.
\end{aligned}$$

This means that L preserves norms. And for linear transformations this is enough to get distance preserving since

$$\|L(\mathbf{x}) - L(\mathbf{y})\| = \|L(\mathbf{x} - \mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|.$$

Gautam also showed that such transformations preserve inner product. This should provide the needed information to complete Claudia's (and Curtis') proof that distance preserving implies linear (affine).

determinants and parallelograms (Aishwarye C.) Aishwarye undertook to show that a linear transformation L of a parallelogram P has area given by

$$\operatorname{area} L(P) = \det A \operatorname{area}(P).$$

He ran out of time, and it will be interesting to see how his work proceeds.

Thursday July 26, 2012

Product Formula For Determinant (Liangyi S) This is an outline of the proof that for square matrices $\det AB = \det A \det B$:

1. First show that if E is an elementary row matrix, i.e., the matrix corresponding to an elementary row operation, then E has determinant as follows:
   (a) Type 1 (replace a row with itself plus a constant times another row)
   $$\det E = 1.$$
   (b) Type 2 (switch rows)
   $$\det E = -1.$$
   (c) Type 3 (multiply a row by $\lambda \neq 0$)
   $$\det E = \lambda.$$
2. Show that for an elementary row matrix
   $$\det EB = \det E \det B.$$
   This is the main step in some sense.
3. Show that A is invertible if and only if $\det A \neq 0$.
   To see this, observe that A can be reduced to echelon form A' using only type 1 and 2 elementary row operations. This corresponds to multiplying A on the left by a sequence of elementary row operations, and the determinants of such products can only differ from the determinant of A by a sign. Thus, the condition that $\det A \neq 0$ is equivalent to the condition that $\det A' \neq 0$. Note that A can also be

obtained by multiplication on the left by elementary row matrices; the inverses of the ones used to reduce A. For a matrix in echelon form, it is easy to see that the condition $\det A' \neq 0$ is equivalent to the condition that A' has full rank. But we already know that A' has full rank if and only if A has full rank. And we also know that A has full rank if and only if A is invertible. So we are done.

Next we come to the main proof:

4. If A is invertible, we ho ahead and use elementary row operations to reduce A to the identity. This means

$$A = E_1 \cdots E_k I.$$

Thus, $\det A = \det E_1 \cdots \det E_k$. Therefore,

$$\det AB = \det E_1 \cdots E_k B = \det A \det B$$

by the second step above.

5. If A is not invertible, then we know from the third step that $\det A = 0$. More precisely, when we reduce A to echelon form A', we see that A' ends with a row of zeros. This gives us:

$$\det AB = \det E_1 \cdots E_k A'B = \det A'B.$$

But looking at $A'B$, we see that the last row of this product must be all zeros. This means $\det A'B = 0$, and we are done. □

2.12.2 (Samantha A.) The question here is: Does matrix multiplication of $n \times n$ matrices satisfy the commutative property, $AB = BA$?

Samantha's answer is "Yes" when $n = 1$ since multiplication in a field commutes and that is what you're doing when $n = 1$.

For $n > 1$, the answer is "No." To see this, note that when $n = 2$

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

but

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

Thus, not all 2×2 matrices commute. It's also worth noting that if A and B represent 2×2 blocks and we use zeros to denote blocks of appropriate size to fill out an $n \times n$ matrix like this:

$$\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix},$$

then the block multiplication formula

$$\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} AB & 0 \\ 0 & 0 \end{pmatrix}$$

holds. Thus, not all $n \times n$ matrices will commute either.

Finally, it might be worth nothing that diagonal matrices commute with everything. Can you think of other matrices with this property?

Area and Determinants (Aishwaye C.) We also talked about various aspects of area and determinants. I think we convinced ourselves that in the 2×2 case, the determinant of a matrix with c_1 and c_2 in the columns gives the area of the parallelogram spanned by c_1 and c_2 up to a sign. The parallogram spanned by c_1 and c_2 is defined to be

$$P = \{ac_1 + bc_2 : 0 \leq a, b \leq 1\}.$$

We talked a bit about why this is a reasonable definition and what we might mean by "signed area" and orientation.

Incidentally, we also talked about rotations of $\mathbb{R}^2$ and verified directly that the product formula for determinants holds when the first matrix is a rotation. (*The more you know, the more you know.*)