LECTURE NOTES 10/13/05

JOSHUA P. BOWMAN

1. Abstract vector spaces

What is a vector? Loose definition we've been using so far: an object that can be added and scalar multiplied. Even this makes clear that the definition of a vector requires a *context,* i.e., a vector space. The examples we've seen:

- $\mathbb{R}^n$
- $\mathrm{Mat}_{k\times n}$

Here are other examples of sets where we can add and multiply by real numbers:

- $C[0,1]$, the set of continuous functions on the interval $[0,1]$. To add two functions together, we just add their values at each point, and likewise for scalar multiplication:
$$(f+g)(x) = f(x) + g(x),\ (cf)(x) = c(f(x)) \qquad \text{for all } x \in [0,1].$$
We know that adding two continuous functions gives another continuous function, and multiplying a function by a real number doesn't change its continuity.
- ℓ^0, the set of sequences (functions $\mathbb{N} \to \mathbb{R}$).
$$(a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots),$$
$$c(a_0, a_1, a_2, \dots) = (ca_0, ca_1, ca_2, \dots).$$
- $Poly_x^3$, the set of polynomials in x with degree at most 3. To add, we add coefficients, and likewise to multiply by a real. The sum of two such polynomials cannot have degree more than 3, so $Poly_x^3$ is closed under addition.

How can we talk about all these different things at once, and see what's the same about them? They are unified by the concept of a *vector space.*

Definition 1.1. A (*real*) *vector space* is a set V with addition and scalar multiplication defined that satisfy the following axioms:

$$\exists\ \mathbf{0} \in V \text{ such that } \forall\ \mathbf{v} \in V,\ \mathbf{v} + \mathbf{0} = \mathbf{v} \qquad \text{(identity)} \tag{1}$$

$$\forall\ \mathbf{v} \in V,\ \exists\ -\mathbf{v} \in V \text{ such that } \mathbf{v} + (-\mathbf{v}) = \mathbf{0} \qquad \text{(inverses)} \tag{2}$$

$$\forall\ \mathbf{v}, \mathbf{w} \in V,\ \mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v} \qquad \text{(commutativity)} \tag{3}$$

$$\forall\ \mathbf{u}, \mathbf{v}, \mathbf{w} \in V,\ c, d \in \mathbb{R},\ \begin{cases} (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) \\ c(d\mathbf{v}) = (cd)\mathbf{v} \end{cases} \qquad \text{(associativity laws)} \tag{4}$$

$$\forall\ \mathbf{v}, \mathbf{w} \in V,\ c, d \in \mathbb{R},\ \begin{cases} c(\mathbf{v} + \mathbf{w}) = c\mathbf{v} + c\mathbf{w} \\ (c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v} \end{cases} \qquad \text{(distributive laws)} \tag{5}$$

$$\forall\ \mathbf{v} \in V,\ 1\mathbf{v} = \mathbf{v} \qquad \text{(normalization)} \tag{6}$$

These axioms are chosen so that almost everything you know about $\mathbb{R}^n$ is true for any vector space. For example,

- A subspace of a vector space W is a subset W that is also itself a vector space; when adding and scalar multiplying things in W, you don't go outside of W. Looking at the examples above, we see that $Poly_x^3$ is a subspace of $C[0,1]$, because a polynomial is certainly continuous on $[0,1]$. You should convince yourself that the addition and scalar multiplication we've defined for $Poly_x^3$ is the same as the one for $C[0,1]$.
- Linear combination and span work just as before. A function in $Poly_x^3$ is a linear combination of the functions 1, x, x^2, and x^3; these four functions span $Poly_x^3$. The two sequences $(1,0,1,0,1,0,\dots)$ and $(0,1,0,1,0,1,\dots)$ in ℓ^0 span the set of sequences whose terms have alternating values.
- A finite set of vectors $\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ is linearly independent if
$$c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0} \implies c_1=\cdots=c_n=0.$$
A non-zero polynomial has only finitely many points on which it equals zero, so $a+bx+cx^2+dx^3 \in Poly_x^3$ can only equal 0 for all x if $a=b=c=d=0$. Thus $\{1,x,x^2,x^3\}$ is a linearly indepedent spanning set for $Poly_x^3$—hey, we have a basis!

As the last example above suggests, bases also work as before, and the "dimension" of a space is simply a measure of the size of a basis. Unfortunately, in "infinite-dimensional" spaces, some of what we know breaks down. This is mostly due to the fact that linear combinations must be *finite*. For example, in ℓ^0, let

$$\begin{aligned}\mathbf{e}_1&=(1,0,0,0,\dots)\\ \mathbf{e}_2&=(0,1,0,0,\dots)\\ \mathbf{e}_3&=(0,0,1,0,\dots)\\ &\vdots\end{aligned}$$

These vectors are linearly independent, because obviously

$$\sum_{j=1}^{k} a_j\mathbf{e}_{i_j}=(0,0,0,\dots) \iff a_1=\cdots=a_k=0.$$

But the vectors $\{\mathbf{e}_i\}$ *do not span* ℓ^0*,* because any linear combination of them will only have finitely many nonzero terms.

You may ask, why don't we simply allow infinite linear combinations? The reason is, once we write something like

$$\sum_{i=1}^{\infty} a_i\mathbf{e}_i,$$

we're no longer talking about a *sum;* we're talking about a *limit*, which means we need a way of measuring convergence. An abstract vector space doesn't come equipped with such a thing. There are infinite-dimensional spaces, called *Hilbert spaces,* in which one can make sense of the above expression, and these are a possible topic for the term project. We will not deal with questions of basis in infinite-dimensional spaces in this class.

2. Finite-dimensional spaces

Okay, let's step back from seeing what can go wrong. We should at least be able to understand completely our usual vector space notions when the dimension is finite. And such is the case, by the following proposition:

Proposition 2.1. *A vector space V is finite-dimensional, with dimension n, iff there exists a linear map $\Phi : \mathbb{R}^n \to V$ that is 1-to-1 and onto. Moreover, such a map sends a basis of $\mathbb{R}^n$ to a basis of V.*

(The book calls Φ a "concrete-to-abstract" function.) The point of the above proposition is that a vector space that can be spanned by only finitely many elements can be thought of as $\mathbb{R}^n$ in a *very specific way.* The map Φ tells us exactly which vectors in V to think of as our basis.

Let's look at our standby example, $Poly_x^3$. We've already seen that $\{1, x, x^2, x^3\}$ is a basis for this space, so the first map $\mathbb{R}^4 \to Poly_x^3$ that should come to mind is

$$\Phi : \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \mapsto a + bx + cx^2 + dx^3.$$

This almost looks stupid: it looks like the basis we've chosen for $Poly_x^3$ is just acting as "place-holders" for the $\mathbb{R}^4$ coordinates. That's what Φ does, in a sense: once we've chosen a set of linearly independent vectors in our abstract space, Φ associates them to the standard basis in $\mathbb{R}^n$. Once we've done that, we can treat V (or whatever subspace is spanned by the vectors we chose) *just like* $\mathbb{R}^n$.

Let's look at another example, which allows us to exploit the "bigness" of ℓ^0: As before, let $\mathbf{e}_i \in \ell^0$ be the sequence that is zero everywhere except in the $(i-1)$st term, which is 1. Fix n, and define $\Phi : \mathbb{R}^n \to \ell^0$ by

$$\Phi : \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \mapsto a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n = (a_1, a_2, \ldots, a_n, 0, 0, \ldots).$$

Thus, we can think of $\mathbb{R}^n$ as being the subspace of ℓ^0 composed of sequences that have 0 for the nth and higher terms. That is, ℓ^0 contains a copy of $\mathbb{R}^n$ for every n! This gives us a useful and accurate way to say $\mathbb{R}^m \subset \mathbb{R}^n$ when $m \leq n$: just tack on enough zeroes to vectors in $\mathbb{R}^m$ to make them have n entries.

3. Linear transformations

The axioms for a vector space also allow us to define linear transformations between vector spaces exactly as we always have. To emphasize:

Definition 3.1. Let V and W be vector spaces. A map $T : V \to W$ is *linear* iff

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) \qquad \text{for all } \mathbf{v}, \mathbf{w} \in V \tag{7}$$

$$\text{and} \qquad T(c\mathbf{v}) = cT(\mathbf{v}) \qquad \text{for all } \mathbf{v} \in V,\ c \in \mathbb{R}. \tag{8}$$

We've already seen examples in the case of the "concrete-to-abstract" functions. Here's one that gives a name to basic properties of the derivative that you've known for ages:

Example 3.2. Let $C^1(0,1)$ be the space of continuously differentiable functions on $(0,1)$. (This space, $C[0,1]$, and other spaces whose elements are functions are sometimes called *function spaces.*) Define $D : C^1(0,1) \to C(0,1)$ by

$$D(f) = f'.$$

Recall that $(f+g)' = f' + g'$, and $(cf)' = cf'$. Hence D, the *differentiation map,* is linear!

The kernel of a linear map is, as before, a subspace of the domain. In the previous example, the kernel is the set of constant functions, which is certainly a subspace: it's spanned by the function 1. But D is also *onto:* every continuous function has an antiderivative. Thus we don't have an analogue of the dimension formula (without more careful definitions, which we won't go into here). Once again, the problem is that $C^1(0,1)$ and $C(0,1)$ are infinite-dimensional (for example, they both contain all the polynomials). One of next week's exercises will probably ask you to think a bit more about this example, and to write a matrix for D just on the finite-dimensional space of polynomials of degree d or less.

4. Inner products

So what about length, orthogonality, and all that? The definition of a vector space doesn't give us any way to handle these things. And sometimes there's just no reasonable way to define them. We need some extra structure. Here's an example that generalizes $\mathbb{R}^n$.

Example 4.1 (An important subspace of ℓ^0)**.** Let's use the sequence structure of ℓ^0 to define the following:

$$\ell^2 = \left\{ (a_i)_{i=0}^{\infty} \in \ell^0 \;\middle|\; \sum_{i=0}^{\infty} |a_i|^2 < \infty \right\}.$$

This is called the space of *square summable sequences.* To show that it is a subspace of ℓ^0, we need to show that it's closed under addition and scalar multiplication. Scalar multiplication is easy: if $(a_i) \in \ell^2$, $c \in \mathbb{R}$, then

$$\sum_{i=0}^{\infty} |ca_i|^2 = \sum_{i=0}^{\infty} |c|^2 |a_i|^2 = |c|^2 \sum_{i=0}^{\infty} |a_i|^2 < \infty.$$

For addition, we first recall that $|ab| \le \frac{1}{2}(a^2 + b^2)$. Thus, if $(a_i), (b_i) \in \ell^2$,

$$\sum_{i=0}^{\infty} |a_i + b_i|^2 \le \sum_{i=0}^{\infty} \left(|a_i|^2 + 2|a_i b_i| + |b_i|^2\right) \le \sum_{i=0}^{\infty} \left(2|a_i|^2 + 2|b_i|^2\right) = 2\sum_{i=0}^{\infty} |a_i|^2 + 2\sum_{i=0}^{\infty} |b_i|^2 < \infty.$$

If (a_i) only had n non-zero terms, then $\sum_{i=0}^{\infty} |a_i|^2$ would just be the square of the length of (a_i) as a vector in $\mathbb{R}^n$. Since the length in $\mathbb{R}^n$ is defined by the dot product, we'd like to extend the dot product to ℓ^2. For $(a_i), (b_i) \in \ell^2$,

$$\langle (a_i), (b_i) \rangle = \sum_{i=0}^{\infty} a_i b_i$$

is called the *inner product* (or sometimes *dot product*) of (a_i) and (b_i). With it, the notions of length, orthogonality, and convergence once again become accessible. ℓ^2 is a Hilbert space.