MAT257 Notes

Jad Elkhaleq Ghalayini

September 7 2018

Today, I want to make some introductory remarks for motivation, and I want to show from the beginning why linear algebra will be important to us.

Differentiability

Let's recall for a moment what it means for a function to be differentiable at a point:

$$\exists f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

Of course, we can bring everything to the right and write

$$\lim_{h \to 0} \frac{f(a+h) - (f(a) + f'(a)h)}{h} = 0$$

The advantage of stating it this way is that we see that the linear function

$$f(a) + f'(a)h$$

is the *best linear approximation* to $f(a+h)$ at $h = 0$. In fact, it's the unique linear function $y = b + mh$ such that

$$\lim_{h \to 0} \frac{f(a+h) - (b + mh)}{h} = 0$$

Should this be called linear? I don't know. Maybe it would be better to call it affine.

From the point of view of linear algebra, we would say that a function which takes

$$h \mapsto f'(a) \times h$$

is a linear transformation. So the derivative can be thought of as a number, but it can also be thought of as a linear transformation from the real line to itself. And this is how we're going to define, and work with, derivatives in higher dimensions.

Functions of several variables

We're not going to be working on the real line, instead, we'll be working on $\mathbb{R}^n$: the space of n-tuples of real numbers. Instead of functions $f(x)$ of a single real variable x, we will hence be working with functions of n real variables, or if you like, functions of a single vector variable, where $x \in \mathbb{R}^n$, of the form

$$f(x) = f(x_1, ..., x_n)$$

Such a function is differentiable at a point $a = (a_1, ..., a_n)$ if "there's a thing that plays the role of the derivative." We could think about a deriative in $\mathbb{R}$ as a real number, but we could also think about it as a linear transformation, which is what we're going to do. That is, a derivative exists, if there exists a linear transformation

$$\lambda : \mathbb{R}^n \to \mathbb{R}$$

such that

$$\lim_{h \to 0} \frac{f(a+h) - (f(a) + \lambda(h))}{|h|} = 0$$

What's a linear transformation from $\mathbb{R}^n \to \mathbb{R}$. Well first of all, what's h? $h = (h_1, ..., h_n)$ is a vector. So λ can be thought of as a sort of row matrix such that

$$\lambda(h) = \begin{pmatrix} \lambda_1 & ... & \lambda_n \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

Sometimes, we'll write $\lambda(h)$ simply as λh. This right away is why linear algebra is important. As you can imagine, if we want to analyze the effect of a given derivative on this function, we're going to need to analyze this matrix, so linear algebra is going to be really crucial. In fact, a lot of the things that we do in the course can be done and were done classically with a really minimal amount of linear algebra. In fact this λ is a matrix, but it's entries are what we call the partial derivatives of f, so we can work only with a collection of partial derivatives. However, what we'll see is, from a conceptual perspective, there's a lot to be gained from the linear agebraic approach.

Analysis involves making approximations and estimations, like $\epsilon - \delta$ proofs in calculus. Of course, $\epsilon - \delta$ proofs involve measuring the size of a quantity or distance, and those are going to be really essential concepts in this course. That already shows up in the $|h|$ in the equation above, which is defined

$$|h| = \sqrt{h_1^2 + ... + h_n^2}$$

the <u>norm</u> of h. It's a measure of the size of h. If we write the norm

$$|x - y|$$

this is a measure of the size of the difference between x and y, or the <u>distance</u> between x and y. But especially in high dimension, even in one dimension but especially in high dimensions, there may be many different ways of measuring distance which are essentially equivalent.

Equivalent Norms

This is one norm, but in fact there are many different but equivalent norms. For example, some norms include

$$|x| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

$$||x|| = \max\{|x_1|, ..., |x_n|\}$$

$$|||x||| = \sum_{i=1}^{n} |x_i|$$

These are all equivalent, which means that if you pick any two of them, the first one is less than or equal to a constant times the second one. Why is that? What's the relationship between them?

It's a very good idea, in, anything, I guess, to understand things geometrically. And this actually can be seen just on the level of a picture of $\mathbb{R}^2$. After all, what if we take our first norm. What does it mean to say that the norm is less than or equal to r? It means that our point is inside a circle of radius r. On the other hand, what does it mean for the *second* norm to be less than or equal to r? It means that both norms have to be inside the square of side length r centered at the origin. And finally, what does it mean for the third norm to be less than or equal to r? That means we're inside the smaller square, or diamond, formed by the points

$$(0, r), (0, -r), (r, 0), (-r, 0)$$

Of course, there would be n-dimensional versions of this picture. But what does this picture say about the relationship between these norms? It tells us that

$$||x|| \leq |x| \leq |||x|||$$

How do we prove this analytically? We just take the square of everything!

Now how do we show that

$$|||x||| \leq a|x|$$

for some constant a? We can also think about this geometrically: take our circle of radius r. We need the first diamond which is outside of the circle. And what distance is the corner of the diamond from the origin? In the plane, it's $\sqrt{2}r$. In higher dimensions, it's $\sqrt{n}r$. So what this tells us is that

$$|||x||| \leq \sqrt{n}|x|$$

or, analytically,

$$|||x||| = \langle x, u(x) \rangle \leq |x||u(x)|$$

where

$$u_i(x) = \operatorname{sgn}(x_i) = \begin{cases} 1 & \text{if} \quad x_i \geq 0 \\ -1 & \text{if} \quad x_i < 0 \end{cases}$$

Now, how do we show that

$$|x| \leq a||x||$$

Well, let's again look at our square of side r. Clearly, the circle circumscribing this square has radius $\sqrt{n}r$, in n dimensions. We should finish this by doing the exercise of proving this analytically