Math Primer Write-Up

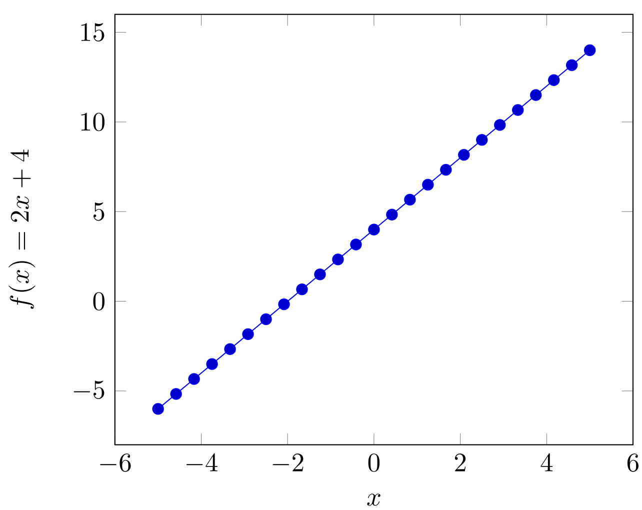
February 2018

1 Introduction

This primer should cover all the basic math you need to know for the lectures in SIGAI this semester. Please note that this won't be an in-depth covering of all the topics presented–our goal is to make sure you have a basic knowledge of how the math works so when we talk about advanced terms you roughly understand what's going on behind the scenes. Because of that, we won't be giving detailed explanations of how to perform these operations by hand, but rather will be talking about what each of these mathematical concepts mean.

2 Derivatives

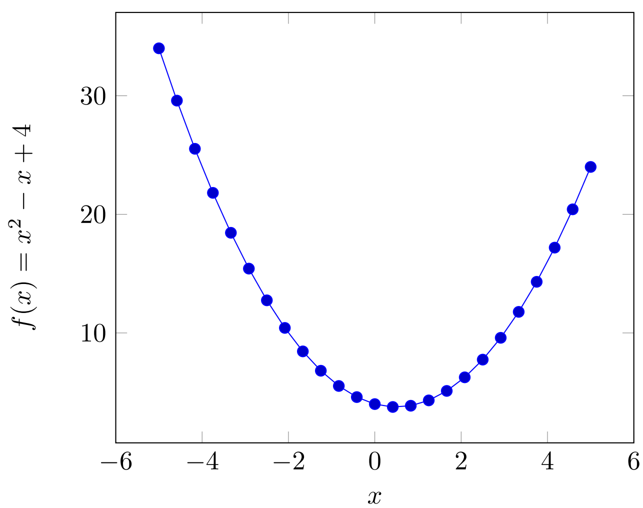
By this point in college, we should all be acquainted with the notion of slope. Take the graph below, for example:



It's pretty easy to find the slope of this graph. Recall that the definition of slope is simply the change in y over the change in x. In other words, for any two points (x_1, y_1) and (x_2, y_2) on a line, the slope of the line is:

$$\frac{y_2 - y_1}{x_2 - x_1}$$

However, what if the line we're looking at isn't straight? Consider the curve below:



Since our graph isn't a straight line, we can't use the ordinary slope definition. Instead of a linear function whose slope is the same at every point, the slope of this curve changes at different parts of the graph. The **derivative** helps us quantify the slope of a graph that isn't necessarily linear. There are lots of rules for derivatives, but the most basic two methods for finding derivatives are using the power rule and the chain rule.

The **power rule** is our method for finding the derivative of a polynomial. Let's look at the equation from the previous graph:

$$f(x) = x^2 - x + 4$$

To find the derivative of a polynomial like this, you subtract 1 from the exponent of each x, then multiply by the original exponent. It sounds complicated, but it's much simpler than it sounds. What this means is that, for an element like this:

$$x^k$$

we can find the derivative by multiplying by k and then subtracting 1 from the exponent of x. Thus the derivative of this is:

$$k(x^{k-1})$$

So we can find the derivative of the previous example as follows:

$$f(x) = x^2 - x + 4$$

To make it easier, we'll write everything as if it's being multiplied by a power of x.

$$f(x) = x^2 - x^1 + 4x^0$$

Then we can use the power rule on each term,

$$\begin{aligned} f(x) &= 2x^{2-1} - 1x^{1-1} + 0(4x^{0-1}) \\ &= 2x - 1 + 0 \\ &= 2x - 1 \end{aligned}$$

And that's the derivative for our equation.

It's more important to know what this means than how we found it. The derivative of our $f(x)$ (denoted by $\frac{df}{dx}$) is $2x - 1$. What this means is that, if we look at the point on $f(x)$ where $x = k$, the slope of our curve at that point is $2k - 1$. You can see on the graph earlier that our curve is flat right around $x = -1/2$. By plugging in $-1/2$ to the derivative, we see that the slope at $x = -1/2$ is actually 0, so our curve is flat at that point.

The key point here is that the derivative of a line is the slope of that line. Unlike the first definition of slope you learned, the derivative can measure the slope of *any* two-dimensional lines, not just simple ones.

The other important rule to know regarding derivatives is the chain rule. Here's the formal definition of the chain rule:
If we have an $f(x)$ such that $f(x) = g(h(x))$ for some functions $g(x)$ and $h(x)$, then the derivative $\frac{df}{dx}$ of $f(x)$ is equal to $\frac{dg}{dx} * \frac{dh}{dx}$.

This is another one of those definitions that seem hard but are actually easy. Take the following example:

$$f(x) = (3x^2 + x - 4)^3$$

Basically the chain rule says that if you have stuff raised to a power, you take the derivative of the inside ignoring the outside, and multiply it by the derivative of the outside ignoring the inside.
In this case, the derivative of the inside is

$$6x + 1$$

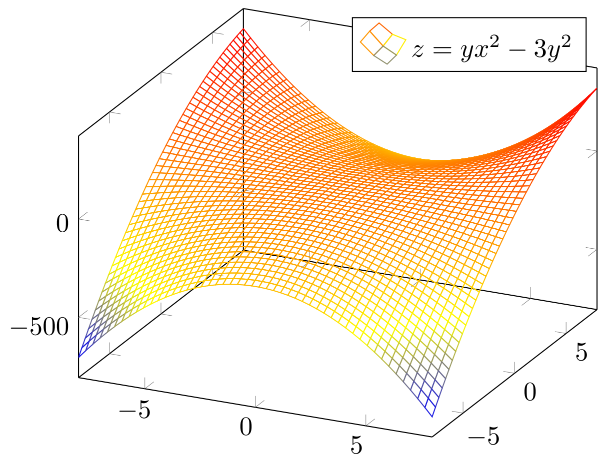
Note how we ignored the exponent on the outside of the parenthesis. The derivative of the outside leaving the inside alone is

$$3(3x^2 + x - 4)^2$$

Just the power rule where we ignore the terms on the inside of the parentheses. The power rule says that the derivative of the whole equation is these two derivatives times each other, so

$$(6x + 1) * 3(3x^2 + x - 4)^2$$

So now we have a general idea of what derivatives are. However, this definition of a derivative only works for two dimensional spaces. What if we wanted to find the slope of a surface in three dimensions?
This is where the notion of a **partial derivative** comes in. Let's look at the following surface:



This is a 3-D surface in 3 variables, so unfortunately we can't use the previously defined method for finding derivatives, since the graph will be changing at different rates depending on where we are. However, mathematicians came up with another fancy trick to find the slope of surfaces like this.
Note that if we take a cross section of this surface across one of the axes, we'll just get a 2-D graph. For example, lets look at the cross section of this graph where $x = 2$. The cross section of our graph is then the graph $f(y) = 4y - 3y^2$. If we take the derivative of this line, we'll find that the slope is $4 - 6y$. When we take a partial derivative, we're doing what we just did–you take one variable and treat every other variable as if it's a constant, then derive the equation. For example,

$$z = yx^2 - 3y^2$$

$$\frac{\partial z}{\partial y} = x^2 - 6y$$

$$\frac{\partial z}{\partial x} = 2yx$$

The notation "$\frac{\partial z}{\partial x}$" means that we deriving an equation while keeping all but x constant (the variable on the bottom is the one we are not keeping constant).

What this means is that, for any value of y, we have an equation detailing the slope of the cross section of that value of y. The chain rule as defined previously works exactly the same for partial derivatives.
A **gradient** is just a way to express partial derivatives together. The gradient of the equation defined before is just

$$\nabla(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, \frac{\partial z}{\partial z})$$

which basically means $\frac{\partial z}{\partial x}$ is the change in z for any change in x, $\frac{\partial z}{\partial y}$ is the change in z for any change in y, etc.

3 Vectors

A vector represents a force in a direction, and can be notated a couple ways. We can have coordinate notation:

$$\vec{v} = (x, y)$$

This represents a vector that goes from $(0, 0)$ to the point (x, y).
We can also have angle-magnitude notation:

$$\vec{v} = (r, \theta)$$

In this case, r is a value and θ is an angle. This means that our vector has a magnitude of r in the direction θ. If you think of a vector as a person pushing an object, magnitude refers to how hard they're pushing, and direction refers to what direction they're pushing.
Vectors in coordinate notation can be added and subtracted in a very simple way. Here's an example:

$$\vec{x} = (2, 4)$$

$$\vec{y} = (5, 3)$$

$$\vec{x} + \vec{y} = (2 + 5, 4 + 3) = (7, 7)$$

$$\vec{x} - \vec{y} = (2 - 5, 4 - 3) = (-3, 1)$$

The interesting thing about vectors is there's actually two different ways to multiply them together. The first of these is the **dot product**. The dot product essentially measures one vector contributes to the other vectors. In other words, the dot product of two vectors measures the interaction between similar dimensions of the two vectors. The #1 thing to remember about dot products is that a dot product is between two vectors and returns a scalar. You perform the dot product by multiplying like terms from coordinate notation, and then adding the products together.
For example:

$$\vec{x} = (2, 4)$$

$$\vec{y} = (5, 3)$$

$$\vec{x} \cdot \vec{y} = (2 \cdot 5) + (4 \cdot 3)$$
$$= 10 + 12 = 22$$

If we have two vectors that are perpendicular to each other, the dot product will be zero.
The other way to multiply vectors is with the **cross product**. Cross products measure the interaction between different dimensions of the two vectors. Cross products are rather complicated to explain succinctly, but here's the most important parts of the cross product:

- A cross product is denoted by the symbol $\times$ (ex. $\vec{x} \times \vec{y}$, read as "x cross y")
- A cross product between two vectors produces a **vector** as a result
- The cross product of two vectors produces a vector perpendicular to <u>both</u> vectors (so if we cross a vector on the x-axis with one on the y-axis, we'll get a vector on the z-axis)
- The cross product is not commutative, so $\vec{x} \times \vec{y} \neq \vec{y} \times \vec{x}$
- If two vectors are parallel, their cross product is 0

If you're interested in the specifics of calculating the cross product feel free to research it, there's a great explanation of dot and cross products at https://betterexplained.com/articles/cross-product/.
Here's essentially what you need to know:

- Dot products ($\vec{x} \cdot \vec{y}$) and cross products ($\vec{x} \times \vec{y}$) are operations between two vectors
- Dot products measure the interaction between similar dimensions of two vectors, and produce a scalar as output
- Cross products measure the interaction between different dimensions of the two vectors, and produce a vector as output
- Dot products are commutative, whereas cross products are not

4 Matrices

The last section we'll cover is a brief discussion on matrices. A matrix is the mathematical equivalent of an array from computer science. There's lots of operations and rules with them, but we'll just cover the basics of what you need to know.
Here's two matrices, just to see what they look like:

$$A = \begin{bmatrix} 1 & 3 \\ 5 & 4 \end{bmatrix}$$

$$B = \begin{bmatrix} 8 & 3 \\ 2 & 7 \end{bmatrix}$$

Essentially these are each equivalent to a 2×2 array in any programming language.
Addition and subtraction can be found by adding respective terms, like so:

$$\begin{aligned} & A + B \\ &= \begin{bmatrix} 1 & 3 \\ 5 & 4 \end{bmatrix} + \begin{bmatrix} 8 & 3 \\ 2 & 7 \end{bmatrix} \\ &= \begin{bmatrix} 1+8 & 3+3 \\ 5+2 & 4+7 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 6 \\ 7 & 11 \end{bmatrix} \end{aligned}$$

Subtraction works similarly. Note matrices can be of different sizes, but we cannot add or subtract different size matrices. Let's look at another matrix:

$$C = \begin{bmatrix} 9 & 3 & 2 \\ 1 & 4 & 6 \end{bmatrix}$$

The **transpose** of this matrix is a C^T such that the i'th row in C is the i'th column in C^T. What does this mean? Basically we take each row of a matrix and make it a column of a new matrix, and that's the transpose. We have to be careful to keep the rows in the right order (e.g. first row becomes first column, second becomes second column, etc.). Here's an example with our previous matrix, C:

$$C^T = \begin{bmatrix} 9 & 1 \\ 3 & 4 \\ 2 & 6 \end{bmatrix}$$

And that's all the math you should need to understand the SIGAI lectures. Feel free to contact myself or any of the SIGAI coordinators with any questions regarding this material. To reiterate what was said previously–it's more important to understand what these operations are doing than it is to know how to do them. The computer will do all these computations for you, we're just trying to make sure you roughly understand what the computer is doing and why it works.