Before beginning the official mathematical part of the study guide, I want to start with a summary of our goals and motivations to give you context for what follows. I highly recommend reading this first, because I think it will make it easier for you to understand the organization and themes of the pages that follow.

For us, mathematics is mainly interesting as a tool for studying real-world phenomena (motion, electricity, chemical reactions, and so on). Functions can be a useful way to create, describe, and analyze simplified models of particular parts of the real world.

Calculus is (for us) is the study of functions and their growth. At its core is the derivative, which associates a function to (the slope of) its best linear approximation at each point. This is a measure of the local *linear* growth/change of a function. For instance, velocity is the local instantaneous change in distance, and acceleration is the local linear change in velocity. In this course, we our goal is to develop familiarity with the derivative and how it is used both in theory and in practice.

This brings us to the two thematic "pillars" of this course: (1) exploring functions defined by their growth [rate equations] and (2) applying calculus to study functions with a view towards real-life applications [approximation]. The first is a little more abstract, and the second more concrete, but each informs the other and it is important to understand both!!

Summary of the study guide:

A. Basics of functions: common functions, especially polynomials and the notion of an inverse.

B. Limits and continuity. Limits are important quantifying "small errors" associated to approximations and the derivative.

C. Derivative as the slope of the "best linear approximation" [mentions alternative definitions].

D. Derivative rules from the viewpoint of lines. Derivative of polynomials.

E. Rate equations. Functions defined by rate equations (e^x and $\ln(x)$).

F. Approximation, both discrete and polynomial.

A. Functions.

A function is a rule that takes in an input and assigns some output. For us, the input and output are always going to be a number.

For instance, $f(x) = x^2 + 2x + 1$ is a function. The input is a number x, and the output is the value of $x^2 + 2x + 1$. For instance, $f(2) = 2^2 + 2 * 2 + 1 = 9$. In fact, this is an example of a quadratic function, which is a special kind of polynomial.

Functions often describe a relationship between two interesting numerical quantities. In physics, you might be interested in speed as a function of time. In chemistry, you might be interested in the concentration of some species as a function of time.

Some functions we will encounter in this course are: linear, quadratic, polynomials, square root, inverse functions, piecewise functions, and rational functions. Linear and quadratic functions are actually special kinds of polynomials, but they're also of independent interest. Later, we will define the exponential and logarithm functions with rate equations.

Here are descriptions of some important functions.

Linear: A linear function is one that can be written in the form $f(x) = mx + b$, for constants m and b. For instance, $f(x) = 2x + 1$ or $f(x) = -3x + 5$ are both linear functions. When graphed, these functions look like a line.

Quadratic: A quadratic function is one that can be written in the form $f(x) = ax^2 + bx + c$, for constants a, b, c, where we require that a is not zero (if $a = 0$, then we end up with a linear function!). For instance, $f(x) = x^2$ and $f(x) = x^2 - 2x - 3$ are both quadratic functions. When graphed, they make an upward- or downward-facing arc.

Polynomial: Both linear and quadratic functions are special kinds of polynomials. A polynomial is, roughly, any function that can be formed only by using addition and multiplication. The following are both polynomials:

$$f(x) = x^{55} - x^3 + 1$$

$$f(x) = (x^2 + 3x + 1)(x^4 - 3)$$

Every polynomial can be "expanded and recollected" into standard from: an expression only involving whole number powers of x (meaning things like x^3, x, x^8, but not x^{-1} or $x^{2.1}$ or $x^{1/2}$) multiplied by some numbers and added together. In this form, we say that the "degree" is the highest power of x that appears in the polynomial. For instance, quadratics all have degree 2.

Only the first example above is in this standard form. For two polynomials to be equal, they must have identical standard forms.

These functions are special because we can actually compute their values directly.

Square root: The square root of x, denoted $\sqrt{x}$, is defined to be the number that squares to x. Since the square of a number is always positive, this means that the square root can only handle positive inputs.

Inverse functions: Given two functions $f(x)$ and $g(x)$ such that $f(g(x)) = g(f(x)) = x$, we say that $f(x)$ and $g(x)$ are inverses to each other. The square root function is the inverse of the function $f(x) = x^2$ as long as we restrict only to positive inputs.

The exponential and logarithm functions are also inverses.

Absolute value: The absolute value of x is written $|x|$. It takes in the number x and makes it positive. For instance $|5| = 5$ doesn't change x because it is already positive, while $|-2| = 2$, changing from negative to positive two.

Piecewise functions: A piecewise function is any function that looks like a combination of unrelated functions, each defined on a portion of the x-axis. These are some of the only functions we've discussed which have discontinuities (points where one cannot take a limit or the limit doesn't agree with the function's value) because of the "jumps" that are allowed when moving from one region to another.

Rational functions: These are functions which can be written as a quotient of polynomials. For instance, $\frac{1}{x}$ and $\frac{x^2+1}{x+3}$ are both rational functions. These too can have discontinuities – division by zero is not allowed, so the function is not defined at any x value where the denominator is zero, and hence is discontinuous there.

Exponential functions: The function e^x is a special function defined by the rate equation

$$f'(x) = f(x)$$

$$f(0) = 1$$

It typically appears in problems about population growth or interest accumulation. It also has some special properties:

$$e^{a+b} = e^a e^b \qquad (e^x)^r = e^{rx}$$

More generally, for some constants A (initial population/amount) and r (growth/interest rate), the function Ae^{rx} is the solution to the rate equation

$$f'(x) = rf(x)$$

$$f(0) = A$$

Logarithm function: This function, written $\ln(x)$ can be defined in one of two ways. It can be defined as the inverse of the exponential function, so it is defined by the property $\ln(e^x) = e^{\ln(x)} = x$. Alternatively, it is the solution to the rate equation

$$f'(x) = \frac{1}{x}$$

$$f(1) = 0$$

Like the exponential function, it has some special properties:

$$\ln(ab) = \ln(a) + \ln(b) \qquad \ln(a^b) = b\ln(a)$$

All of these different kinds of functions can be combined into much more complicated functions with **addition**, **multiplication**, **division**, and **composition**. The latter is the trickiest. Given two functions, $f(x)$ and $g(x)$, their composition, written $f(g(x))$ is the function which takes x as input, transforms it to $g(x)$, and then finally outputs f evaluated at $g(x)$.

B. Limits and continuity.

Limits are a way of describing what happens to a function really really close to some point. Given a function $f(x)$ and an x-value a, we write

$$\lim_{x \to a} f(x)$$

for the value that f approaches as x is taken to be closer and closer to a. It might not exist: for instance if $f(x) = \frac{1}{x}$, then there is no limit as x approaches zero because the points on the graph aren't headed to a single spot as we move near zero. Other time it might not exist because $f(x)$ has a large jump (as can happen with piecewise functions).

If the limit above exists AND equals the value of the function at that point, then we say that $f(x)$ is continuous at a.

An important use of limits, and the reason we introduce them in this course, is that they allow us to quantify the notion that a function $f(x)$ is "much smaller" than $g(x)$ near some number. Suppose that both $f(a) = 0$ and $g(a) = 0$ and the two are continuous at a. Then we say that $f(x)$ is much smaller than $g(x)$ if

$$\lim_{x \to a} \frac{f(x)}{g(x)}$$

The reason is that because $g(x) \to 0$, the denominator gets very small, and hence tries to make the whole fraction very large. So $f(x)$ must be extremely small in order to not just balance that largeness, but decrease it to zero.

Examples are $f(x) = x^2$ and $g(x) = x$ at $a = 0$, more generally $(x - a)^n$ is smaller than any higher power of $(x - a)$ near a.

C. Definition of the derivative.

The derivative of $f(x)$ at a is denoted $f'(a)$, and it is the slope of the best linear approximation to $f(x)$ at $x = a$. To be precise, we want

$$f(x) = f(a) + f'(a)(x - a) + error(x)$$

where the error is "much smaller" than linear near a, in other words it is much smaller than $x - a$. Quantitatively, with limits, that means

$$\lim_{x \to a} \frac{error(x)}{x - a} = 0.$$

If such an approximation exists, then its slope is the derivative of $f(x)$ at a.

Note that we write the function as $f'(a)(x - a)$ to remind us that we are centering at a, and at zero there should be no contribution from the linear (or error) terms.

In practice, we tend to treat the error term as more or less zero near the point of interest. It behaves as such errors should: multiplied to anything, you get more error, and it changes nothing to which it is added. This is not a proof based class, so we will not be too careful with the error and its limit definition, but **it is still important that you are aware of, and point out somewhere, when there is an error term, even if later you will ignore it**.

(OPTIONAL) Another popular definition in calculus comes from rearranging the above. Subtracting $f(a)$ and dividing by $x - a$, we get

$$\frac{f(x) - f(a)}{x - a} = f'(a) + \frac{error(x)}{x - a}$$

Taking the limit of the left hand side as $x \to a$, we obtain an alternative definition:

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

In this definition, the right hand side is the slope of the line connecting $(x, f(x))$ to $(a, f(a))$. In other words, the derivative is the limit of the slopes of those secant lines as we move one end closer and closer to a.

D. Derivative rules.

These are rules for calculating derivatives of functions that are built out of familiar functions. The main idea is that the derivative is associated to a "best linear approximation" (as the slope) and so combining functions by addition/multiplication/composition should result in a new derivative (a slope) which looks like what one would expect from combining two lines in the corresponding way.

This is an early example of the simplifying power of calculus: something which seems complicated is converted into a problem about lines, which are among the nicest of all functions.

Scaling: given some constant c, the derivative of $cf(x)$ is $f'(x)$. This corresponds to the fact that scaling a line by the constant c multiplies the slope by a factor of c.

Sum: given functions $f(x), g(x)$, the derivative of $f(x) + g(x)$ is $f'(x) + g'(x)$, corresponding to the fact that adding lines adds their slopes.

Product: given functions $f(x), g(x)$, the derivative of the product $f(x)g(x)$ is

$$f'(x)g(x) + f(x)g'(x)$$

Let's work this one out in terms of lines. From the definition of the derivative at a point a,

$$f(x) = f(a) + f'(a)(x - a) + (error)$$

$$g(x) = g(a) + g'(a)(x - a) + (error)$$

We can multiply these together. It does no harm to ignore the error

$$f(x)g(x) \approx (f(a) + f'(a)(x - a))(g(a) + g'(a)(x - a))$$

So the right hand side is a product of lines! The cross terms come naturally from expanding this.

$$= f(a)g(a) + f(a)g'(a)(x - a) + f'(a)g(a)(x - a) + f'(a)g'(a)(x - a)^2$$

Remember that $(x - a)^2$ is *very small* compared to $x - a$, and so it can safely be merged into the error. In total, we obtain

$$f(x)g(x) = f(a)g(a) + (f(a)g'(a) + f'(a)g(a))(x - a) + (error)$$

and hence by the definition of the derivative, the derivative of $f(x)g(x)$ at a is $f(a)g'(a)+f'(a)g(a)$. Viewing this as a function of x instead of a, we obtain the claimed result!

Quotient: we don't usually divide lines, but the quotient rule follows formally from the product rule - this is because $\frac{1}{g(x)}$ is defined by the special product $g(x)\frac{1}{g(x)} = 1$. Let $h(x) = \frac{1}{g(x)}$ and then differentiate $g(x)h(x) = 1$ with the product rule:

$$g'(x)h(x) + g(x)h'(x) = 0$$

Rearrange to get $h'(x)$ alone:

$$h'(x) = -\frac{g'(x)h(x)}{g(x)} = \frac{g'(x)}{g(x)^2}$$

(in the last step, remembering that $h(x) = \frac{1}{g(x)}$.

Combine this fact with $f(x)/g(x) = f(x)\frac{1}{g(x)}$ and the product rule to arrive at the quotient rule.

Composition: this is often the trickiest rule. The main idea is that when lines are composed, their slopes multiply. This might suggest that the derivative of $f(g(x))$ is $f'(x)g'(x)$, ***BUT*** we need to make a small adjustment to make sure that the *correct* slopes are multiplied.

This is related to the fact that the derivative is defined with a line in point-slope form, and to compose just lines and get another line in point-slope form requires a certain compatibility.

The actual composition rule for the derivative of $f(g(x))$ is

$$f'(g(x))g'(x)$$

Using $f'(g(x))$ instead of $f(x)$ gives us the right slope to multiply: the point is that $g(x)$ is what's plugged into $f(x)$ during composition, so that is the relevant slope, not the slope of $f(x)$ at x.

We can work this out from the definitions too. Let's expand $f(x)$ and $g(x)$ in terms of their derivatives at $g(a)$ and a, respectively:

$$f(x) = f(g(a)) + f'(g(a))(x - g(a)) + (error)$$

$$g(x) = g(a) + g'(a)(x - a) + (error)$$

Then simply compose the two sides. As usual, we will drop the errors for simplicity

$$\begin{aligned} f(g(x)) &\approx f(g(a)) + f'(g(a))(g(a) + g'(a)(x - a) - g(a)) \\ &= f(g(a)) + f'(g(a))(g'(a)(x - a)) \\ &= f(g(a)) + f'(g(a))g'(a)(x - a) \end{aligned}$$

This matches the definition of the derivative of $f(g(x))$ at a exactly, and has the slope that matches the composition rule, $f'(g(a))g'(a)$.

The final important fact related to the derivative is the Fundamental Theorem of the Derivative. This says that if $f'(x) = 0$ everywhere, then $f(x)$ is a constant. This makes lots of intuitive sense: $f'(x)$ is the growth rate of $f(x)$, and if it's always zero then $f(x)$ isn't increasing or decreasing.

This is a crucial tool for verifying certain identities and properties of functions. The idea is that if we want to show there is some relationship $f(x) = g(x)$, then that is equivalent to $f(x) - g(x) = 0$ or $f(x)/g(x) = 1$. In both cases, we have a function that we want to equal a special constant. It often happens that this is hard to check directly, but if we show the derivative is zero, then at least we know it is constant, and often it is straightforward to find out which particular constant!

E. Rate equations.

A rate equation consists of two pieces of information: an equation involving the derivative of some function, and an initial value. As an example, e^x is defined by the rate equation

$$\begin{cases} f'(x) & = f(x) \\ f(0) & = 1 \end{cases}$$

While we defined the logarithm as the inverse of the exponential function, we could also have defined it with the following rate equation:

$$\begin{cases} f'(x) & = \frac{1}{x} \\ f(1) & = 0 \end{cases}$$

The first part of a rate equation provides information about how the function grows, an dthe initial value tells us "where it starts". In principle, if we know enough about how fast a function grows, and we know where it starts, we should know the function exactly. This is a form of the "racetrack principle": if you know where a horse starts, and its speed at any given moment, then you know also exactly where it is at any time.

Functions defined by rate equations are usually amenable to study by approximation, which is discussed in the next section.

Rate equations arise naturally when modeling real-world problems. We spent a lot of time studying the rate equations associated to continuous growth/interest. In this case, we have something like a population of animals. They should reproduce at a rate proportional to their total population, which means that the derivative (growth rate) of the population is a constant multiple of the current population. In symbols,

$$P'(t) = rP(t)$$

where r is the growth rate. For instance, if a population of weasels reproduces at a rate of 4 baby weasels per weasel per 3 months, then we would have

$$P'(t) = 4P(t)$$

with t in units of 3 months.

The only other piece of information we need is the population at some time. For instance, we might start with a population of 200 weasels. The complete rate equation for this situation, then is

$$\begin{cases} P'(t) & = 4P(t) \\ P(0) & = 200 \end{cases}$$

An **exact solution** to a rate equation is a **function** which satisfies the conditions of the rate equation: if we plug the function into the first condition, the equality is true, and the function has the initial condition as a point. For example, the exact solution to our weasel rate equation is

$$P(t) = 200e^{4t}$$

F. Approximation.

Approximation is ***ESSENTIAL and FUNDAMENTAL*** to calculus, and especially to actually using calculus in the "real world". Here our error perspective is particularly valuable: any experiment is subject to errors which are beyond our control, but calculus gives us a systematic approach for handling error.

For instance, if you go on in any science, and probably lots of other fields (I just don't know about them) then you will eventually have to learn about "error propogation" when you take experimental data and its intrinsic error and need to understand how that error changes when the data is fed into functions. The way error transforms in the course of a calculation is ***entirely*** described by derivatives and calculus.

Also, for many functions it is not possible to compute exact values (e.g. the square root of a non-square) but we can use methods of approximation to work out very precise estimates. Calculators and computers ultimately use such approximations to find the decimal expansions they give you for things like $\sqrt{2}$ and ln(8).

Right now, we know two types of approximation: discrete and polynomial.

The core idea of **discrete approximation** is that the definition of the derivative

$$f(x) = f(a) + f'(a)(x - a) + (error)$$

gives us a good linear approximation to $f(x)$ near $x = a$. We can use this to extrapolate approximations to values of $f(x)$ near a simply by evaluating the linear part at them (and hoping the error is small, which it should be!). This process can be iterated to estimate some points on a graph for the purpose of drawing a sketch, for instance.

The main idea of **polynomial approximation** is that we can use rate-equation constraints and derivatives to find a polynomial which approximates our function near $x = a$. Polynomials are great because evaluating them (at least with a computer) is straightforward. In fact, computers evaluate functions like e^x by using a very good polynomial approximation, because e^x can't be directly computed!

Both kinds of approximation are useful. Discrete approximations are less algebraically complicated, but the calculation has to be redone every time you want an estimate at a new point, or if you want to use smaller increments to increase accuracy. Finding a polynomial approximation is more algebraically difficult, but once you have it you can use it to approximate as many values as you want. A downside of polynomial approximations is that they are not always well-behaved; for instance, polynomial approximations to the logarithm will always exhibit poor behavior related to the fact that $ln(x)$ goes to $-\infty$ near zero.

Let's work out an example of **discrete approximation**. Suppose we have a rate equation like

$$\begin{cases} f'(t) &= x^2 + f(x) \\ f(0) &= 2 \end{cases}$$

We form an empty table

x	$f(x)$	$f'(x)$
0		
.5		
1		
1.5		
2		

The idea is that given a complete row, we are able to fill in the $f(x)$ value on the next row, then use the rate equation to fill in $f'(x)$. That gives us another complete row, so we can go on to the next row!!

We start off knowing $f(0) = 2$. From the rate equation, we get

$$f'(0) = 0^2 + f(0) = 0 + 2 = 2$$

Fill these in on the table above:

x	$f(x)$	$f'(x)$
0	2	2
.5		
1		
1.5		
2		

To get to the next value at $x = 0.5$, we approximate $f(x)$ with the line that comes from the definition of the derivative:

$$f(x) = f(a) + f'(a)(x - a) + (error)$$

The information we have from the first row is all about $a = 0$, which is close to our target ($x = 0.5$). Putting the values from the table into the derivative definition, we get

$$f(x) = f(0) + f'(0)(x - 0) + (error) = 2 + 2x + (error)$$

We can simply plug 0.5 in and ignore the (small) error to get an approximation.

$$f(0.5) \approx 2 + 2(0.5) = 2 + 1 = 3$$

With the rate equation, we then can find $f'(0.5)$:

$$f'(0.5) = 0.5^2 + f(0.5) = .25 + 3 = 3.25$$

Now we can fill these into our table

x	$f(x)$	$f'(x)$
0	2	2
.5	3	3.25
1		
1.5		
2		

This is another complete row, and so we can use the definition of the derivative again, but this time based at $a = 0.5$ (because that's closer to $x = 1$ than the other row).

$$f(x) = f(0.5) + f'(0.5)(x - 0.5) + (error) = 3 + 3.25(x - 0.5)$$

and that lets us estimate $f(1)$:

$$f(1) \approx 3 + 3.25(1 - 0.5) = 3 + 1.875 = 4.875$$

From the rate equation, we are then able to find $f'(1)$:

$$f'(1) = 1^2 + 4.875 = 1 + 4.875 = 5.875$$

This completes another row of the table:

x	$f(x)$	$f'(x)$
0	2	2
.5	3	3.25
1	4.875	5.875
1.5		
2		

I encourage you to work out the next two rows on your own. You'll probably need a calculator to multiply some of the decimal numbers you get (on the exam, you will have numbers that do not require a calculator)

Next, let's work out a **polynomial approximation** to the same rate equation:

$$\begin{cases} f'(x) &= x^2 + f(x) \\ f(0) &= 2 \end{cases}$$

Let's say we want a degree 3 approximation (on the exam I won't ask for you to work out anything higher than degree 2 because the algebra becomes time consuming quickly). So our goal is to find constants A, B, C, D such that

$$f(x) = A + Bx + Cx^2 + Dx^3 + (error)$$

To find A, B, C, D, we will use the rate equation to determine constraints on their values - a system of equations that we can solve for A, B, C, D.

The reason this works is that to say $f(x)$ is well-approximated by a polynomial means that the growth rate of $f(x)$ must behave like that of the polynomial. The growth behavior of a polynomial is determined precisely by its coefficients, and this interaction is what gives us something to work with.

So to start, we will plug

$$f(x) = A + Bx + Cx^2 + Dx^3 + (error)$$

into the rate equation. For this we need

$$f'(x) = B + 2Cx + 3Cx^2 + (error)$$

Starting now I will neglect the small error terms. But it's important to remember they are still secretly present!!

The first part of the rate equation says

$$f'(x) = x^2 + f(x)$$

hence

$$B + 2Cx + 3Dx^2 = x^2 + A + Bx + Cx^2 + Dx^3 = A + Bx + (C+1)x^2 + Dx^3 \approx A + Bx + (C+1)x^2$$

[the Dx^3 term must be absorbed into the error]

Both sides are polynomials. To be equal their coefficients need to all match. So

constant	$B = A$
degree one	$2C = B$
degree two	$3D = C + 1$

Lastly, the initial value gives us one more equation:

$$2 = f(0) = A + B(0) + C(0)^2 + D(0)^3 = A.$$

So we already know $A = 2$, and from the constant terms, $B = 2$. Using B in the linear term, $2C = 2$ and hence $C = 1$. Lastly, using C in the degree 2 part, we get $3D = 1 + 1 = 2$, so $D = \frac{2}{3}$.

We can put these values for A, B, C, D together to get our approximation!!

$$f(x) = 2 + 2x + x^2 + \frac{2}{3}x^3 + (error)$$

This should be a good approximation *near zero*.