1 Goals

The goals of this lesson are to understand the following:

- Higher order derivatives
- Exponential function
- Factorial function
- Definition of Taylor series

2 Higher order derivatives

We have already discussed higher order derivatives in previous lectures. The main point I want to remind you of is that derivatives in a way represent the \textit{neighborhood} of a point:

- Beginning with a function $f(x)$, we can evaluate it at a point $x_0$ and get a single number. The function itself can be considered its \textit{“0th” derivative}, and it gives information about $f(x)$ at the point $x_0$. However, it does not give any information about the points surrounding $x_0$.

- According to the the definition of a derivative, the first derivative $df/dx$ compares the function’s value at $x_0$ to a nearby neighbor at $f(x_0 + h)$. We saw before that the first derivative gives a good description of how the function $f(x)$ changes near $x_0$. Thus, the first derivative encodes some information about the neighborhood of $f(x_0)$.

- Similarly, the second derivative incorporates another nearby point that we can compare to $f(x_0)$. We can calculate the limit formula for the second derivative by taking the derivative of first derivatives:

\[
\frac{d^2 f(x)}{dx^2} = \lim_{h \to 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}
\]

so the second derivative at a point $x_0$ incorporates not only the value $f(x_0)$, but also its two neighbors $f(x_0 + h)$ and $f(x_0 - h)$. Thus second derivative gives us even more information about the neighborhood around $x_0$. 


• The third derivative can be written as a difference of second derivatives:

$$\frac{d^2 f(x)}{dx^2} = \lim_{h \to 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2} = \lim_{h \to 0} \frac{f(x + 2h) - 3f(x + h) + 3f(x) - f(x - h)}{h^3}$$

so it incorporates three nearby points in the neighborhood of $x_0$, namely $f(x_0 + 2h), f(x_0 + h)$, and $f(x_0 - h)$.

• The pattern continues. The fourth derivative incorporates four nearby points in the neighborhood of $x_0$, the fifth derivative uses five points in the neighborhood, and so on. Each higher order derivative incorporates a bit more information from nearby points, and tells us an increasingly more complete picture about the function surrounding the point $x_0$.

### 3 Exponential function

Consider for a moment the function $h^n$. For $h > 1$, the function grows extremely quickly — each subsequent value of $n$ multiplies the previous value of the function by a constant factor. For example: $2^n$ is twice as large as $2^{n-1}$, which is twice as large as $2^{n-2}$, etc.

For values of $h$ between 0 and 1, the function actually shrinks extremely quickly — each subsequent value of the function is only a fraction of the previous value: $0.5^n$ is only half as large as $0.5^{n-1}$, which is only half as large as $0.5^{n-2}$, etc. For small values of $h$ like 1/10 or 1/100, this function is practically zero for even small values of $n$.

### 4 Factorial function

Next look at the factorial function $n!$, which is defined as the product of all positive integers less than or equal to $n$.

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$$

This function grows even more quickly than the exponential function, because each number in the sequence is multiplied by an even larger number than the one before it. For example, $n!$ is quite a big number, but it is dwarfed by the value $(n + 1)!$, which is $(n + 1)$ times larger. However, this large number $(n + 1)!$ is an even smaller fraction of $(n + 2)!$

Given that the factorial function generates large numbers so quickly, dividing by it must produce extremely tiny numbers. The function $\frac{1}{n!}$ shrinks so quickly that $\frac{1}{n!}$ is practically zero for even small values of $n$. 

2
5 Taylor series

The following formula will be essential for the rest of this class:

\[
f(x_0 + h) = \sum_{n=0}^{\infty} \frac{d^n f(x_0)}{dx^n} \frac{h^n}{n!}
\]  

(1)

The large \(\sum\) symbol means that we have to add all of these things together. Writing out the first few terms gives us:

\[
f(x_0 + h) = f(x_0) + \frac{df(x_0)}{dx} h + \frac{d^2 f(x_0)}{dx^2} \frac{h^2}{2} + \frac{d^3 f(x_0)}{dx^3} \frac{h^3}{6} + \frac{d^4 f(x_0)}{dx^4} \frac{h^4}{24} + \ldots
\]  

(2)

This is called the Taylor Series, and it is used to transform any function into an infinitely long polynomial of the following form:

\[
f(x_0 + h) = c_0 + c_1 h + c_2 h^2 + c_3 h^3 + c_4 h^4 + c_5 h^5 + c_6 h^6 + \ldots
\]  

(3)

We have seen each of these pieces before: higher order derivatives, the exponential function \(h^n\), and the factorial function. What does this mean though?

- The Taylor series is a transformation. Regardless of what \(f(x)\) looks like (as long as it is continuous and you can take derivatives), you can turn it into a handy polynomial.

- Each successive term evaluates an increasingly higher order derivative at the position \(x_0\). We know that higher order derivatives represent the shape of the surrounding neighborhood. The higher the derivative, the larger the surrounding neighborhood is. Each successive term in the Taylor series represents the behavior of the function \(f(x)\) in a wider and wider area surrounding \(x_0\).

- We know that \(h^n\) drops very quickly to zero for small values of \(h\), and we know that \(1/n!\) drops to zero even faster. Consequently, the coefficient \(h^n/n!\) becomes insignificantly small for small values of \(h\). Larger values of \(h\) still become insignificant at some point, because \(n!\) will eventually outgrow \(h^n\). For small values of \(h\), each term is an order of magnitude smaller (less important) than the previous term in the series. This means that, even though the Taylor series goes on forever, we can actually ignore all of the terms after a certain point and still get an accurate guess about how \(f(x)\) behaves.

- If we only look at the first few terms and ignore the rest, then we approximate \(f(x)\) with a finite polynomial:
  - If we only keep the first term, then we are guessing that all points near \(f(x_0)\) have the same value.
  - If we only keep the first two terms, then we approximate \(f(x_0 + h)\) with the linear function \(f(x_0) + \frac{df(x_0)}{dx} h\). This assumes that all nearby points behave like a line with slope \(\frac{df(x_0)}{dx}\).
– Keeping the first three terms approximates $f(x_0 + h)$ with the quadratic function $f(x_0) + \frac{df(x_0)}{dx} h + \frac{d^2f(x_0)}{dx^2} \frac{h^2}{2}$. Thus we guess that all nearby points look like a parabola.

– Keeping the first four terms approximates $f(x_0 + h)$ with a cubic function.

– In general, keeping the first $n$ terms makes the approximation that $f(x_0 + h)$ looks like an $n^{th}$ order polynomial function. Because each successive term is far less significant than the ones before it, we very quickly zero in on the correct behavior of $f(x)$.

Visualization

Consider the example $f(x) = \sin(x) + 2$. We can use the Taylor series to expand the function about some input value $x_0$. Let’s choose the value $x_0 = 0$, so $f(x_0 + h) = \sin(x_0 + h) + 2$ becomes $f(h) = \sin(h) + 2$. The series of figures below provide a visualization of the first few terms of the Taylor series for the function $f(h) = \sin(h) + 2$. The black curve is the original function $\sin(h)$, and the red curve is the first few terms of the Taylor series polynomial.

![Figure 1: The original function sin(h) + 2](image-url)
Figure 2: Increasing the number of terms in the Taylor series (red curve) gives an increasingly accurate approximation to the function.