9

THE DERIVATIVE

9.1 Context and motivation

When reasoning about the relationships between observable magnitudes in nature, functions come as a first-class citizens, for many ways of reasoning are based on this very notion.

We will define *functions* as abstract machines that compute an output from a input, so we can think of them as programs that receive some input of type A and returns an output of type B, denoted

$$f : A \to B.$$

In this and subsequent chapters we will focus on numeric functions, i.e., functions for which both the inputs and outputs can be real numbers (single variable case $f : \mathbb{R} \to \mathbb{R}$) or tuples of real numbers (multivariable case $f : \mathbb{R}^n \to \mathbb{R}$).

The output of a function is completely determined by its input. The *domain* of a function f, denoted $\mathrm{Dom}(f)$, is the set of input values for which the function can return an output. It is customary to represent functions graphically using the *graph* of the function: a graphical representation using coordinates both for the input and ouput space in which the geometric place where inputs and outputs meet is drawn. Formally, the graph of a function $f : \mathbb{R} \to \mathbb{R}$ is given by the following set of points in the plane:

$$\Gamma(f) = \{(x, f(x)) \mid x \in \mathrm{Dom}(f)\}$$

9.2 Review on limits

We say that a function $f : \mathbb{R} \to \mathbb{R}$ has a limit L about some point $a \in \mathbb{R}$ if for any given radius $\epsilon > 0$ we can always find a $\delta > 0$ such that if $x \neq a$ satisfies $|x - a| < \delta$, then $|f(x) - L| < \epsilon$. We denote it

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

This somehow technical definition has a very simple intuition behind: when a function has a limit about some point a, the function cannot fluctuate much in the vicinity of a; moreover, given some distance ϵ, no matter how small, there is always an interval about the point a such that the outputs get closer from L than ϵ.

The limit of a function at a point is a very handy idea that we can use to our advantage in order to state some intuitive concepts in a more formal way and also to make them operational.

9.3 Continuity

Functions in general can be very wild: e.g. we can think of functions for which close inputs (in some sense that we can make precise) give rise to arbitrarily far apart outputs. We will consider only functions that ensure that gradual changes in the input yield gradual changes in the output.

We will say that a function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* at some point $a \in \mathbb{R}$ if the limit of f about a exists and it is $f(a)$.

An alternative, more graphical way to put it goes as follows:

Imagine that you are considering all the possible numbers in the output space that are closer than ϵ (no matter how small) from $f(a)$. That f is continous at a means that you can always find a small enough interval $I = [a - \delta, a + \delta]$ such that all the inputs in this interval will yield outputs in the interval $[f(a) - \epsilon, f(a) + \epsilon]$.

We say that the function is *continuous* if it is continuous at every point in its domain.

9.4 Derivative

We can define informally the *derivative* of a function at a point a as the slope or growth rate of the graph of the function when we zoom in closer and closer on a. This is usually formulated using some sort of limit expression:

$$\lim_{\Delta x \to 0} \frac{f(a + \Delta x) - f(a)}{\Delta x}$$

In other words, the derivative tells us how to attach a slope or growth rate at a given point.

More formally, we define the *derivative* of $f : \mathbb{R} \to \mathbb{R}$ at a, whenever it exists, as the following limit:

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

We will say that f is *differentiable at a* if $f'(a)$ exists and we will say f is *differentiable* if it is differentiable at all points of its domain.

Not all continuous functions are differentiable: there are continuous functions that are still a bit rough, in the sense that there may be points for which we lack a sense of growth rate. A prototypical example is the function $f(x) = |x|$: what happens around the point $a = 0$?

Even though there are many examples of non-differentiable functions that are useful for many scientific use cases, the class of differentiable functions is by far the most important from a practical point of view.

9.5 Local affine approximations

Beyond the formal definition of the derivative, it is important to understand what it means geometrically for a function to be differentiable: a function is differentiable at a point a if it can be approaximated by an affine function. This means that we can draw a straight line through the point $P = (a, f(a))$ so that the more we zoom in into the P, the more the graph of f resembles the straight line.

In order to discuss local affine approximations, we will introduce a formal way to say that functions whose values converge towards zero (as x tends to some point of interest a) can converge towards zero at different velocities. For two functions f and g such that

$$lim_{x \to a} f(x) = 0, \quad \lim_{x \to a} g(x) = 0$$

we will say that that the function f is *little-o* of g, denoted $f = \mathfrak{o}(g)$, if the following condition is satisfied:

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 0.$$

Intuitively, f converges faster than g towards zero as x tends to a.

Proposition 33. *Let f, g, h be functions that tend towards zero about the point a. If $f = \mathfrak{o}(h)$ and $g = \mathfrak{o}(h)$ then the following holds:*

1. $f + g = \mathfrak{o}(h)$
2. $\lambda f = \mathfrak{o}(h)$
3. $fg = \mathfrak{o}(h^2)$
4. *If $f = \mathfrak{o}(h^2)$ then $f = \mathfrak{o}(h)$*

We say that a function f has $g(x) = f(a) + k(x - a)$ as a local affine approximation if

$$f(x) = g(x) + \mathfrak{o}(x - a).$$

Observe that there are two important conditions satified:

- $f(a) = g(a)$
- For values $x \neq a$, the difference $f(x) - g(x)$ converges faster towards zero than $x - a$ as x tends to a.

If a function has a local affine approximation, it is unique:

Suppose we have two local affine approximations:

$$f(x) = f(a) + k_1(x-a) + \mathfrak{o}(x-a) = f(a) + k_2(x-a) + \mathfrak{o}(x-a)$$

Then the following identity must hold:

$$k_1(x-a) + \mathfrak{o}(x-a) = k_2(x-a) + \mathfrak{o}(x-a).$$

But if we transform this equation dividing both members by $x - a$ and taking limits $x \longrightarrow a$ we get:

$$k_1 + \lim_{x \longrightarrow a} \frac{\mathfrak{o}(x-a)}{x-a} = k_2 + \lim_{x \longrightarrow a} \frac{\mathfrak{o}(x-a)}{x-a}$$

By definition the limits vanish, then we conclude that $k_1 = k_2$.

Proposition 34. *If f is differentiable at a, then it has a local affine approximation at a. Moreover, the derivative of f at a is the slope of this approximation:*

$$f(x) = f(a) + f'(a)(x-a) + \mathfrak{o}(x-a).$$

9.6 Master rules to compute derivatives

In this section we will prove three basic rules to compute the derivative of a function that results from combining other functions via sum, product and composition. With these important rules we will be able to compute the derivative of more complex functions.

In all cases we will strongly rely on the fact that the derivative at some point must be the slope of the local affine approximation at that point.

9.6.1 Additivity

Let f and g be two differentiable functions, which can be expressed locally at a via their affine approximations as:

$$f(x) = f(a) + f'(a)(x-a) + \mathfrak{o}(x-a)$$

$$g(x) = g(a) + g'(a)(x-a) + \mathfrak{o}(x-a)$$

respectively. Summing both expressions we can get a local affine approximation for the function $h(x) = f(x) + g(x)$:

$$h(x) = f(a) + g(a) + (f'(a) + g'(a))(x-a) + \mathfrak{o}(x-a).$$

We can conclude that

$$h'(a) = f'(a) + g'(a),$$

i.e. the derivative of the sum of functions if the sum of their derivatives.

9.6.2 Leibniz rule

Let's now apply the same idea to the product of functions $p(x) = f(x)g(x)$:

$$p(x) = f(x)g(x) =$$

$$(f(a)+f'(a)(x-a)+\mathfrak{o}(x-a))(g(a)+g'(a)(x-a)+\mathfrak{o}(x-a)) =$$

$$= f(a)g(a) + (f'(a)g(a) + g'(a)f(a))(x-a) + \mathfrak{o}(x-a).$$

We can the conclude that

$$p'(a) = (fg)'(a) = f'(a)g(a) + g'(a)f(a)$$

which is commonly known as *Leibniz rule.*

9.6.3 Chain rule

Let $c(x) = g(f(x))$ which is also commonly expressed as $(g \circ f)$. If we want to infer a rule to compute the derivative of the function $c(x)$ we must first understand what is the result of performing the composition of two affine functions.

Consider the composition $g \circ f$ assuming that f and g are affine functions, say

$$f(x) = m_1 x + n_1, \quad \text{and} \quad g(x) = m_2 x + n_2.$$

Since

$$(g \circ f)(x) = g(f(x)) = m_2 f(x) + n_2 = m_2 m_1 x + (m_2 n_1 + n_2)$$

we can conclude that the composition $g \circ f$ is also an affine function with slope $m_2 m_1$.

If we apply this observation to the composition of two differentiable functions, the local affine approximation of the composition $g \circ f$ at the point a will have a slope that is the product of the two slopes of the local affine approximations of f at a, and of g at $f(a)$, respectively: these slopes are $f'(a)$ and $g'(f(a))$, respectively. We can conclude that:

$$(g \circ f)'(a) = g'(f(a))f'(a)$$

which is commonly known as the *chain rule*.

9.7 TL;DR

The derivative of a function at a point a provides the local trend of the function about a. We may think of it as the limit growth rate of the function when you compare the value of the function at the point a with values at other points that can be as close to a as wanted. This simple intuition requires, however, a bunch of somewhat technical ideas like the limit of a function. Not all functions admit derivatives, but when they do we can define a local affine approximation with the derivative as its slope.