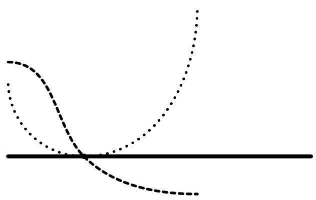
Chapter 6

The Derivative

We have reached the limit of accuracy available if we restrict ourselves to approximating continuous functions using constant polynomial functions.

While this gives a reasonable approximation of the "state" of the system/process we are modelling "near a", it is a static picture, for it gives no insight into how the system "evolves" "near a. Does the dependent variable increase with the independent one? If so, how rapidly?

For example, the best approximation using a polynomial function of degree 0, whose graph is the horizontal line in the following diagram, does not distinguish between the two functions whose graphs are the other two curves.



We turn to examining these questions, beginning with a more precise formulation.

Since we cannot do more with polynomial functions of degree 0, we consider polynomial functions of degree 1,

$$p : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto c_1 x + c_0,$$

where c_0, c_1 are fixed real numbers.

The discrepancy arising from using p to approximate f is

$$|f(x) - p(x)| = |f(x) - c_1 x - c_0|.$$

In order for this to be at least as accurate as utilising a polynomial function of degree 0, we must have

$$f(a) = p(a) = c_1 a + c_0,$$

or

$$c_0 = f(a) - c_1 a,$$

whence

$$p(x) = c_1(x - a) + f(a).$$

Then the discrepancy we need to minimise is

$$|f(x) - p(x)| = |f(x) - f(a) - c_1(x - a)|.$$

Plainly, the discrepancy always converges to 0 as x converges to a, since

$$\lim_{x \to a} |f(x) - p(x)| = \lim_{x \to a} |f(x) - f(a) - c_1(x - a)| = 0,$$

choice of c_1.

So we sharpen our stipulations by requiring the discrepancy to converge to 0 "much faster" than x converges to a. We formulate this as requiring the discrepancy arising from using p instead of f is negligible in comparison with the deviation of x from a. In other words, the ratio of the discrepancy to the deviation should converge to 0 as x converges to a. Formally, we seek c_1 so that

$$\lim_{x \to a} \frac{|f(x) - p(x)|}{|x - a|} = 0.$$

Using Equation **??**, this becomes

$$\lim_{x \to a} \left| \frac{f(x) - f(a)}{x - a} - c_1 \right| = 0,$$

or, equivalently,

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = c_1.$$

Thus, in order to be able to optimise our approximation to f near a by means of a polynomial function of degree at most 1, the ratio

$$\frac{f(x) - f(a)}{x - a}$$

must converge as x converges to a. We formalise this in our next definition.

Definition 6.1. The function $f : \mathbb{R} \to \mathbb{R}$ is *differentiable at* $a \in \mathbb{R}$ with *derivative* ℓ if and only if

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

$$f'(a) = \ell \quad \text{or} \quad \left.\frac{df}{dx}\right|_{x=a} = \ell \quad \text{or} \quad \left.\frac{d}{dx} f(x)\right|_{x=a} = \ell.$$

Then our polynomial function approximating f is $f(a) + f'(a)(x - a)$, so that near a

$$f(x) \approx f(a) - f'(a)(x - a).$$

f is said to be *differentiable on* X if it is differentiable at every $a \in X$, and *differentiable everywhere*, or simply *differentiable* if it is differentiable on its domain.

Note that the graph of $y = f(a) - f'(a)(x - a)$ is a straight line passing through $(a, f(a))$. This is why the derivative at a can be viewed as providing the "best linear approximation to f near a".

If we denote by h the deviation of x from a and by k the variation of $f(x)$ from $f(a)$, we can reformulate the above as

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

and, for h small,

$$k \approx f'(a)h.$$

You will see later that this essentially renders $f'(a)$ a *linear transformation.*

Example 6.2. Let r be any real number. Then

$$f_r : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto r$$

is differentiable and for every $a \in \mathbb{R}$

$$f'(a) = 0.$$

This is a trivial consequence of the definition of f_r, for given any $a, x \in \mathbb{R}$,

$$f_r(x) - f_r(a) = r - r = 0,$$

whence

$$\frac{f_r(x) - f_r(a)}{x - a} - 0.$$

Example 6.3. The function

$$f = id_{\mathbb{R}} : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto x$$

is differentiable and for every $a \in \mathbb{R}$

$$f'(a) = 1.$$

This is a trivial consequence of the definition of $id_{\mathbb{R}}$, for given any $a, x \in \mathbb{R}, x \neq a$,

$$\frac{id_{\mathbb{R}}(x) - id_{\mathbb{R}}(a)}{x - a} = \frac{x - a}{x - a} = 1,$$

whence

$$\lim_{x \to a} \frac{id_{\mathbb{R}}(x) - id_{\mathbb{R}}(a)}{x - a} = \lim_{x \to a} 1 = 1.$$

Example 6.4. The function

$$f : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \sin(x)$$

is differentiable and for every $a \in \mathbb{R}$

$$f'(a) = \cos(a).$$

To see this, note that

$$\sin(x) - \sin(a) = 2\cos(\frac{x+a}{2})\sin(\frac{x-a}{2}),$$

whence
$$\frac{\sin(x)-\sin(a)}{x-a} = \cos(\frac{x+a}{2})\frac{sin(\frac{x-a}{2})}{\frac{x-a}{2}}.$$
Put $u := \frac{x-a}{2}$. Then $u \to 0$ as $x \to a$, and conversely, so that
$$\lim_{x\to a}\frac{\sin(\frac{x-a}{2})}{\frac{x-a}{2}} = \lim_{u\to 0}\frac{\sin(u)}{u} = 1.$$
Since cos is a continuous function,
$$\lim_{x\to a}\cos(\frac{x+a}{2}) = \cos(a).$$
Thus,
$$\lim_{x\to a}\frac{\sin(x)-\sin(a)}{x-a} = \cos(a).$$

While we have defined differentiability only for functions $\mathbb{R} \to \mathbb{R}$, the definition applies also to functions $f : X \to \mathbb{R}$, with $X \subseteq \mathbb{R}$, requiring only the same slight modification as in the case of the definition of continuity.

Example 6.5.
$$f : \mathbb{R}\setminus\{0\} \longrightarrow \mathbb{R}, \quad x \longmapsto \frac{1}{x}$$
is differentiable and
$$f'(a) = -\frac{1}{a^2}$$
for all $a \in \mathbb{R}\setminus\{0\}$.

To see this, note that for $h \neq 0$
$$\frac{\frac{1}{a+h}-\frac{1}{a}}{h} = \frac{a-(a+h)}{ha(a+h)} = \frac{-1}{a(a+h)}.$$

Since $a+h \to a$ as $h \to 0$, the conclusion follows.

Example 6.6. Our final example is that of the exponential function
$$\mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto e^x.$$

Its derivative is, as will be shown in MATH102,
$$f'(a) = e^a = f(a)$$

We investigate the properties of differentiable functions.

Theorem 6.7. *If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is differentiable at a, then it is continuous at a.*

Proof. Take $x \neq a$. Then
$$f(x) = f(x) - f(a) + f(a) = \frac{f(x)-f(a)}{x-a}(x-a) + f(a),$$

whence

$$\begin{aligned}
\lim_{x\to a} f(x) &= \lim_{x\to a}\frac{f(x)-f(a)}{x-a}\lim_{x\to a}(x-a)+\lim_{x\to a} f(a)\\
&= f'(a)0+f(a)\\
&= f(a).
\end{aligned}$$

□

We next investigate the relationship between differentiation and the algebraic operations we introduced on $\mathcal{F}(\mathbb{R})$. This will enable the calculation of derivatives for a large class of functions.

Theorem 6.8. *Take functions $f, g, h : \mathbb{R} \longrightarrow \mathbb{R}$. Suppose that f and g are differentiable at a, and that h is differentiable at $f(a)$. Then*

(i) *$f+g$ is differentiable at a and*

$$(f+g)'(a) = f'(a)+g'(a).$$

(ii) *fg is differentiable at a and*

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

(iii) *$h\circ f$ is differentiable at a and*

$$(h\circ f)'(a) = h'\big(f(a)\big)f'(a).$$

Proof. (i) The conclusion follows immediately from

$$\left|\frac{(f+g)(x)-(f+g)(a)}{x-a}-(f'(a)+g'(a))\right| \le \left|\frac{f(x)-f(a)}{x-a}-f'(a)\right|+\left|\frac{g(x)-g(a)}{x-a}-g'(a)\right|$$

(ii) Since

$$\begin{aligned}
(fg)(x)-(fg)(a) &:= f(x)g(x)-f(a)g(a)\\
&= f(x)g(x)-f(a)g(x)+f(a)g(x)-f(a)g(a),\\
\frac{(fg)(x)-(fg)(a)}{x-a} &= \frac{f(x)-f(a)}{x-a}g(x)+f(a)\frac{g(x)-g(a)}{x-a}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\lim_{x\to a}\frac{(fg)(x)-(fg)(a)}{x-a} &= \lim_{x\to a}\frac{f(x)-f(a)}{x-a}\lim_{x\to a} g(x)+\lim_{x\to a} f(a)\lim_{x\to a}\frac{g(x)-g(a)}{x-a}\\
&= f'(a)g(a)+f(a)g'(a)
\end{aligned}$$

(iii) Since h is differentiable at $f(a)$, there is a real number, $h'(f(a))$, with

$$\lim_{y\to f(a)}\frac{h(y)-h(f(a))}{y-f(a)} = h'(f(a)).$$

Now

$$\begin{aligned}
\lim_{x\to a}\frac{(h\circ f)(x)-(h\circ f)(a)}{x-a} &= \lim_{x\to a}\frac{h\left(f(x)\right)-h\left(f(a)\right)}{x-a} \\
&= \lim_{x\to a}\left(\frac{h\left(f(x)\right)-h\left(f(a)\right)}{f(x)-f(a)}\frac{f(x)-f(a)}{x-a}\right) \\
&= \lim_{x\to a}\frac{h\left(f(x)\right)-h\left(f(a)\right)}{f(x)-f(a)}\lim_{x\to a}\frac{f(x)-f(a)}{x-a} \\
&= \lim_{y\to f(a)}\frac{h(y)-h\left(f(a)\right)}{y-f(a)}\lim_{x\to a}\frac{f(x)-f(a)}{x-a} \\
&=: h'\left(f(a)\right)f'(a)
\end{aligned}$$

□

Definition 6.9. Theorem 6.8 (ii) is the *Leibniz Rule* or *Product Rule* and Theorem 6.8 (iii) is the *Chain Rule*.

Example 6.10. The polynomial function

$$p:\mathbb{R}\longrightarrow\mathbb{R},\quad x\longmapsto\sum_{j=0}^{n}c_jx^j$$

is differentiable and

$$p'(a)=\sum_{j=0}^{n}jc_ja^{j-1}.$$

To prove this, we first use mathematical induction to show that, for any $n\in\mathbb{N}$,

$$\frac{d}{dx}x^n=nx^{n-1}.$$

Examples 6.2 and 6.3 establish that the claim is true for $n=0,1$ respectively.
We make the inductive hypothesis that for some $n\in\mathbb{N}$, $\dfrac{d}{dx}x^n=nx^{n-1}$. Then,

$$\begin{aligned}
\frac{d}{dx}x^{n+1} &= \frac{d}{dx}(xx^n) \\
&= \left(\frac{d}{dx}x\right)x^n+x\frac{d}{dx}x^n && \text{by the Leibniz (Product) Rule} \\
&= 1x^n+x(nx^{n-1}) && \text{by Example 6.2 and the inductive hypothesis} \\
&= (n+1)x^{n+1}.
\end{aligned}$$

completing the proof by induction.

We next observe that by the Leibniz (Product) Rule, if c_n is a fixed real number,

$$\frac{d}{dx}(c_nx^n)=nc_nx^{n-1}.$$

The result now follows from Theorem 6.8(i).

Example 6.11. The derivative of $\cos : \mathbb{R} \longrightarrow \mathbb{R}$ at a is $-\sin(a)$. This follows since

$$\cos(x) = \sin\big(\frac{\pi}{2} - x\big)$$

expresses the cosine function as the composite of two differentiable functions and since, by Example 6.10, $\dfrac{d}{dx}(\frac{\pi}{2} - x) = -1$.

Example 6.12. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a differentiable function with $f(x) \neq 0$ for all $x \neq 0$.

Define

$$g : \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \frac{1}{f(x)}.$$

Then g is differentiable at a and

$$g'(a) = -\frac{f'(a)}{(f(a))^2}$$

.

To see this, note g can be expressed as the composite $h \circ \tilde{f}$, where

$$\tilde{f} : \mathbb{R} \longrightarrow \mathbb{R} \setminus \{0\}, \quad x \longmapsto f(x)$$

and

$$g : \mathbb{R} \setminus \{0\} \longrightarrow \mathbb{R}, \quad y \longmapsto \frac{1}{y}.$$

Plainly, $\tilde{f}'(a) = f'(a)$, and, by Example 6.5, $h'(b) = \dfrac{-1}{b^2}$.

The Chain Rule completes the argument.

We next look at the relationship between the derivative of a function and the derivative of its inverse.

Theorem 6.13. *Let $f, g \colon \mathbb{R} \to \mathbb{R}$ be (mutually) inverse differentiable functions. Suppose that $f'(a) \neq 0$. Let $b = f(a)$. Then*

$$g'(b) = \frac{1}{f'(a)}$$

Proof. Since g and f are mutually inverse, we have $g \circ f = id_{\mathbb{R}}$, that is $(g \circ f)(x) = x$ for all $x \in \mathbb{R}$.

By the Chain Rule, and the fact that $\dfrac{d}{dx}(x) = 1$,

$$(g \circ f)'(a) = g'(f(a))f'(a) = 1,$$

or, since $b = f(a)$,

$$g'(b)f'(a) = 1.$$

□

Definition 6.14. The function $f \colon X \to \mathbb{R}$ with $X \subseteq \mathbb{R}$ has a *local maximum* (resp. *local minimum*) at $a \in X$ if and only if there is an interval I such that $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$ for all $x \in I \cap X$.

An *extremum* is either a maximum or a minimum.

The extremum is an *absolute* or *global extremum* if I can be chosen to be X.

Lemma 6.15. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be differentiable. Suppose that $f'(a) \neq 0$. Then f cannot have an extremum at a.*

Proof. Suppose that $f'(a) > 0$.

Put $\varepsilon := \dfrac{f'(a)}{2} > 0$.

Then there is a $\delta > 0$ such that if $|x - a| < \delta$, then

$$\left|\frac{f(x) - f(a)}{x - a} - f'(a)\right| < \varepsilon,$$

Thus

$$0 < \frac{f'(a)}{2} < \frac{f(x) - f(a)}{x - a}.$$

If we choose x with $a - \delta < x < a$, then $x - a < 0$, and thus $f(x) - f(a)$ must also be negative, that is, $f(x) < f(a)$.

If, on the other hand, we choose x with $a < x < a+\delta$, then $x-a > 0$, and thus $f(x)-f(a)$ must also be positive, that is, $f(x) > f(a)$.

The case $f'(a) < 0$ is left to the reader as an exercise. □

Several important theorems are corollaries.

Theorem 6.16. *The function $f\colon X \to \mathbb{R}$ has an extremum at a only if*

(i) a is a boundary point of X, or

(ii) f is not differentiable at A, or

(iii) $f'(a) = 0$.

Proof. These are the only alternatives to $f'(a) \neq 0$. □

Theorem 6.17 (Rolle's Theorem). *Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function which is differentiable on $]a, b[$. If $f(b) = f(a)$, then there is a $c \in]a, b[$ with $f'(c) = 0$.*

Proof. By the Extreme Value Theorem, f has both a maximum and a minimum, on $[a, b]$.

These can only coincide if f is constant, in which case, by Example 6.2, $f'(c) = 0$ for every $c \in]a, b[$.

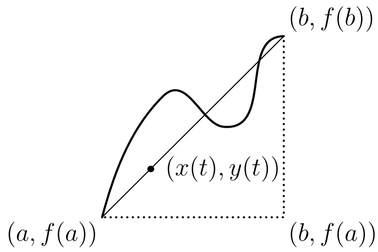
If f is not constant, then, since $f(a) = f(b)$ at least one of the extrema must occur at some $c \in]a, b[$. Since f is differentiable throughout $]a, b[$, it follows from Theorem 6.16 that $f'(c) = 0$. □

Theorem 6.18 (Mean Value Theorem of Differential Calculus). *Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function which is differentiable on $]a, b[$. Then there is a $c \in]a, b[$ with $f(b) - f(a) = f'(c)(b - a)$.*

We make some comments before turning to the proof of the Mean Value Theorem.

Rolle's Theorem is obviously a special case of the Mean Value Theorem. This suggests an approach to proving the theorem by reducing it to Rolle's Theorem.

To do so, we seek a suitable continuous function $g : [a, b] \to \mathbb{R}$, differentiable on $]a, b[$, for which $g(b) = g(a)$, letting geometric intuition guide us.



The graph of f can be regarded as a curve in the Cartesian plane passing through the points with co-ordinates $(a, f(a))$ and $(b, f(b))$.

We can trace the line segment from $(a, f(a))$ to $(b, f(b))$ in one unit of time by specifying that the point with co-ordinates $(x(t), y(t))$ satisfy $x(0) = a, x(1) = b, y(0) = f(a), y(1) = b$.

One way to achieve this is to define

$$\begin{aligned} x &= a + (b - a)t \\ y &= f(a) + \big(f(b) - f(a)\big)t \end{aligned}$$

Solving the first equation for t in terms of x, we obtain

$$t = \frac{x - a}{b - a}.$$

Substituting in the second equation we get

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

Putting $x = b$,

$$f(b) = f(a) + \frac{f(b) - f(a)}{b - a}(b - a),$$

or, equivalently,

$$f(b) - \frac{f(b) - f(a)}{b - a}b = f(a) - \frac{f(b) - f(a)}{b - a}a$$

Note that both sides of this equation are of the form $f(x) - \frac{f(b)-f(a)}{b-a}x$. We use this to prove the Mean Value Theorem.

Proof. Consider the function

$$g\colon [a, b] \longrightarrow \mathbb{R}, \quad x \longmapsto f(x) - \frac{f(b) - f(a)}{b - a}x.$$

Then g is continuous on $[a, b]$ and differentiable on $]a, b[$ with

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

Moreover, $g(b) = g(a)$. Hence, by Rolle's Theorem, there is a $c \in]a, b[$ with $g'(c) = 0$, that is

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

□

Corollary 6.19. *If $f\colon [a,b] \to \mathbb{R}$ is continuous and if $f'(c) > 0$ (resp. < 0) for all $c \in]a,b[$, then f is monotonically strictly increasing (resp. decreasing) on $[a,b]$*

Proof. Suppose that $f'(c) > 0$ for all $c \in]a,b[$.

Take $a \leq x < y \leq b$. Then, by the Meant Value Theorem, there is a $c \in]x,y[$ with

$$f(y) - f(x) = f'(c)(y - x).$$

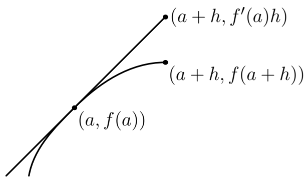
Since $f'(c)(y-x) > 0$, we have $f(y) > f(x)$, so that f is monotonically strictly increasing.

The case $f'(c) < 0$ for all $c \in]a,b[$ is left to the reader. □

Observation 6.20. Recall that we were led to the concept of differentiation by trying to approximate the curve representing the graph of a function using the nearest straight line.

In particular, we saw that if f is differentiable at a, then $f(a+h)$ can be approximated by $f(a) + f'(a)h$ for h sufficiently small, by which we mean that given $\varepsilon > 0$, there is some $\delta > 0$ such that $|f(a+h) - f(a) - f'(a)h| < \varepsilon$ as long as $|h| < \delta$.

If we draw the graph of the approximating function together with the function itself we obtain



From this we recognise the graph of $y = f(a) + f'(a)h$ as being the tangent line at $(a, f(a))$ to the curve which is the graph of $y = f(x)$.

Thus we one application of the derivative of a real-valued function of a real variable is the determination of the *slope of the tangent to the graph of the function.* However the reader is strongly warned against as regarding the derivative primarily in these terms for several reasons, including the following.

1. In order for the graph of f, a real-valued function of a real variable, to possess a tangent at $(a, f(a))$, f must be differentiable at a. So to seek to define the derivative in terms of the slope of the tangent to the graph of the function is circular.

2. The notion of derivative applies to a much broader range of functions. For example, it can be defined for functions $f\colon \mathbb{R}^n \to \mathbb{R}^m$, where, if $m > 1$ there is no single tangent to the graph at any point. In such a case, as the reader will later learn, the derivative is actually a *linear transformation* and appears as an *$m \times n$ matrix*, namely the *Jacobean matrix*.

With this in mind, we can interpret the Mean Value Theorem geometrically. It's effect is to assert that there is a c between a and b such that the tangent at $(c, f(c))$ to the graph of f is parallel to the chord joining $(a, f(a))$ and $(b, f(b))$.