Chapter 5

Derivatives

5.1 Definition

In this chapter we define and study the notion of ***differentiability of functions***. Derivatives were historically introduced in order to answer the need of measuring the "rate of change" of a function. We will actually adopt this approach as a prelude to the formal definition of the derivative.

Before we start, one technical clarification. In the previous chapter, we introduced the limit of a function $f : A \to B$ at an (interior) point a,

$$\lim_a f.$$

Consider the function g defined by

$$g(x) = f(a + x),$$

on the range,

$$\{x : \ a + x \in A\}.$$

In particular, zero is an interior point of this set. We claim that

$$\lim_0 g = \lim_a f,$$

provided, of course, that the right hand side exists. It is a good exercise to prove it formally. Suppose that the right hand side equals ℓ. This means that

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in B^\circ(a, \delta))(|f(x) - \ell| < \varepsilon).$$

Setting $x - a = y$,

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall y \in B^\circ(0,\delta))(|f(a+y) - \ell| < \varepsilon),$$

or equivalently,

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall y \in B^\circ(0,\delta))(|g(y) - \ell| < \varepsilon),$$

i.e., $\lim_0 g = \ell$.

Let now f be a function defined on some interval I, and let a and x be two points inside this interval. The variation in the value of f between these two points is $f(x) - f(a)$. We define the ***mean rate of change*** (קצב שינוי ממוצע) of f between the points a and x to be the ratio

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

This quantity can be attributed with a number of interpretations. First, if we consider the graph of f as a geometry entity, then the mean rate of change is the slope of the secant line (מיתר) that intersects the graph at the points a and x (see Figure 5.1). Second, it has a meaning in many physical situations. For example, f can be the position along a line (say, in meters relative to the origin) as function of time (say, in seconds relative to an origin of time). Thus, $f(a)$ is the distance from the origin in meters a seconds after the time origin, and $f(x)$ is the distance from the origin in meters x seconds after the time origin. Then $f(x) - f(a)$ is the ***displacement*** (העתק) between time a and time x, and $(f(x) - f(a))/(x - a)$ is the mean displacement per unit time, or the ***mean velocity***.

This physical example is a good preliminary toward the definition of the derivative. The mean velocity, or mean rate of displacement, is an average quantity between two instants, but there is nothing to guarantee that within this time interval, the body was in a "fixed state". For example, we could intersect this time interval into two equal sub-intervals, and measure the mean velocity in each half. Nothing guarantees that these mean velocities will equal the mean velocity over the whole interval. Physicists aimed to define an "instantaneous velocity", and the way to do it was to make the length of the interval very small. Of course, having $(x - a)$ small (what does small mean?) does not change the average nature of the measured velocity.

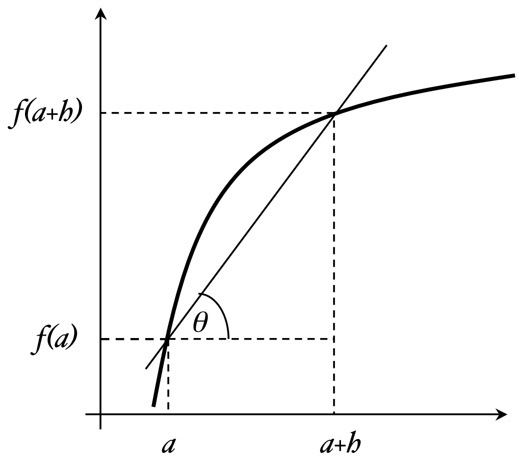

Figure 5.1: Relation between the mean rate of change of a function and the slope of the corresponding secant.

An instantaneous rate of change can be defined by using limits. Fixing the point a, we define a function,

$$\Delta_{f,a} : A \setminus \{a\} \to \mathbb{R},$$

by

$$\Delta_{f,a}(x) \frac{f(x) - f(a)}{x - a}.$$

Definition 5.1 *f **is differentiable at** a (גזירה או דיפרנציאבילית) if $\Delta_{f,a}$ has a limit at a. We write*

$$\lim_{a} \Delta_{f,a} = f'(a).$$

*We call the limit $f'(a)$ the **derivative** (נגזרת) of f at a.*

Comment: In the more traditional notation,

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

At this stage, $f'(a)$ is nothing but a notation. It is not (yet!) a function evaluated at the point a. Another standard notation, due to Leibniz, is

$$\frac{df}{dx}(a),$$

and another popular notation (especially when it comes to multivariate functions) is

$$Df(a).$$

Having identified the mean rate of change as the slope of the secant line, the derivative has a simple geometrical interpretation. As x tends to a, the corresponding family of secants tends to a line which is ***tangent*** to f at the point a. For us, these are only hand-waving arguments, as we have never assigned any meaning to the limit of a family of lines.

Example: Consider the constant function, $f : \mathbb{R} \to \mathbb{R}$, $f : x \mapsto c$ and let $a \in \mathbb{R}$. We construct the function

$$\Delta_{f,a} : \mathbb{R} \setminus \{a\} \to \mathbb{R}, \qquad \Delta_{f,a}(y) = \frac{f(y) - f(a)}{y - a} = \frac{c - c}{y - a} = 0.$$

The limit of this function at a is zero, hence $f'(a) = 0$. ▲▲▲

In the above example, we could have computed the derivative at *any* point. In other words, we can define a *function* that given a point x, returns the derivative of f at that point, i.e., returns $f'(x)$. A function $f : A \to B$ is called ***differentiable*** in a subset $U \subseteq A$ of its domain if it has a derivative at every point $x \in U$. We then define the derivative function, $f' : U \to \mathbb{R}$, as the function,

$$f'(x) = \lim_{x} \Delta_{f,x}.$$

The more traditional notation is,

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

Example: Consider the function $f : x \mapsto x^2$. We calculate its derivative at a point x by first observing that

$$\Delta_{f,x}(y) = \frac{y^2 - x^2}{y - x} = x + y \qquad \text{or} \qquad \Delta_{f,x} = \mathrm{Id} + x,$$

so that

$$f'(x) = \lim_{x} \Delta_{f,x} = 2x.$$

▲▲▲

Example: Consider now the function $f : x \mapsto |x|$. Let's first calculate its derivative at a point $x > 0$. For $x > 0$,

$$\Delta_{f,x}(y) = \frac{|y| - x}{y - x}.$$

Since we are interested in the limit of $\Delta_{f,x}$ at $x > 0$ we may well assume that $y > 0$, in which case $\Delta_{f,x}(y) = 1$. Hence

$$f'(x) = \lim_x \Delta_{f,x} = 1.$$

For negative x, $|x| = -x$ and we may consider $y < 0$ as well, so that

$$\Delta_{f,x}(y) = \frac{(-y) - (-x)}{y - a} = -1,$$

so that

$$f'(x) = \lim_x \Delta_{f,x} = -1.$$

Remains the point 0 itself,

$$\Delta_{f,0}(y) = \frac{|y| - |0|}{y - 0} = \operatorname{sgn}(y).$$

The limit at zero does not exist since every neighborhood of zero has points where this function equals one and points where this function equals minus one. Thus f is differentiable everywhere except for the origin[1]. ▲▲▲

Comment: If f is differentiable at a, then $\Delta_{f,a}$ has a limit at a, hence it has a removable discontinuity at that point. The function

$$x \mapsto \begin{cases} \Delta_{f,a}(x) & x \neq a \\ f'(a) & x = a \end{cases}$$

is continuous at a.

This last example shows a case where a limit does not exist, but one-sided limits do exist. This motivates the following definitions:

[1] We use here a general principle, whereby $\lim_a f$ does not exist if there exist $\ell_1 \neq \ell_2$, such that every neighborhood of a has points x, y, such that $f(x) = \ell_1$ and $f(y) = \ell_2$.

Definition 5.2 A function f is said to be ***differentiable on the right*** *(גזירה מימין) at a if the one-sided limit*

$$\lim_{a^+} \Delta_{f,a}$$

exists. We denote the ***right-hand derivative*** *(נגזרת ימנית) by $f'(a^+)$. A similar definition holds for left-hand derivatives.*

Example: Here is one more example that practices the calculation of the derivative directly from the definition. Consider first the function

$$f : x \mapsto \begin{cases} x \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

We have already seen that that this function is continuous at zero, but is it differentiable at zero? We construct the function

$$\Delta_{f,0}(y) = \frac{f(y) - f(0)}{y - 0} = \sin \frac{1}{y}.$$

The function $\Delta_{f,0}$ does not have a limit at 0, because every neighborhood of zero has both points where $\Delta_{f,0} = 1$ and points where $\Delta_{f,0} = -1$. ▲▲▲

Example: In contrast, consider the function

$$f : x \mapsto \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

We construct

$$\Delta_{f,0}(y) = \frac{f(y) - f(0)}{y - 0} = y \sin \frac{1}{y}.$$

Since $\lim_0 \Delta_{f,0} = 0$, f is differentiable at zero and $f'(0) = 0$. ▲▲▲

5.1.1 Differentiability and continuity

The following theorem shows that differentiable functions are a subclass ("better behaved") of the continuous functions.

Theorem 5.3 If f is differentiable at a then it is continuous at a.

Proof: For $y \neq a$,

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

Viewing $f(a)$ as a constant, we have a functional identity,

$$f = f(a) + (\mathrm{Id} - a)\,\Delta_{f,a}$$

valid in a punctured neighborhood of a. By limit arithmetic,

$$\lim_a f = f(a) + \lim_a (\mathrm{Id} - a)\ \lim_a \Delta_{f,a} = f(a) + 0 \cdot f'(a) = f(a),$$

which proves that f is indeed continuous at a ∎

5.1.2 Higher order derivatives

If a function f is differentiable on an interval A, we can construct a new function—its derivative, f'. The derivative f' can have various properties, For example, it may be continuous, or not, and in particular, it may be differentiable on A, or on a subset of A. Then, we can define a new function—the derivative of the derivative, or the ***second derivative*** (נגזרת שניה), which we denote by f''. By definition

$$f''(x) = \lim_x \Delta_{f',x}.$$

(Note that this is a limit of limits.) Likewise, the second derivative may be differentiable, in which case we may define the third derivative f''', and so on. For derivatives higher than the third, it is customary to use, for example, the notation $f^{(4)}$ rather than f''''. The k-th derivative, $f^{(k)}$, is defined recursively,

$$f^{(k)}(x) = \lim_x \Delta_{f^{(k-1)},x}.$$

For the recursion to hold from $k = 1$, we also set $f^{(0)} = f$.

5.2 Rules of differentiation

Recall that when we studied limits, we first calculated limits by using the definition of the limit, but very soon this became impractical, and we proved a number of theorems (limit arithmetic), with which we were able to easily calculate a large variety of limits. The exact same holds for derivatives. After having calculated the derivatives of a small number of functions, we develop tools enabling us to (easily) compute derivatives without having to go back to the definitions.

Theorem 5.4 If $f\mathbb{R} \to \mathbb{R}$ *is a constant function, then* $f' : \mathbb{R} \to \mathbb{R}$ *is* $f'(x) = 0$.

Proof: We proved this is in the previous section. ■

Theorem 5.5 If $f = \mathrm{Id}$ *then* $f' = 1$ *(i.e.,* $f' : x \mapsto 1$*).*

Proof: Immediate from the definition. ■

Theorem 5.6 Let f, g *be functions defined on the same domain (it suffices that the domains have a non-empty intersection). If both* f *and* g *are differentiable at* a *then* $f + g$ *is differentiable at* a*, and*

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

Proof: We consider the function

$$\begin{aligned}\Delta_{f+g,a} &= \frac{(f+g) - (f+g)(a)}{\mathrm{Id} - a} \\ &= \frac{f + g - f(a) - g(a)}{\mathrm{Id} - a} \\ &= \Delta_{f,a} + \Delta_{g,a},\end{aligned}$$

and by limits arithmetic,

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

■

Theorem 5.7 [Leibniz rule] Let f, g be functions defined in a neighborhood of a. If both f and g are differentiable at a then $f \cdot g$ is differentiable at a, and

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

Proof: We consider the function,

$$\Delta_{fg,a}(x) = \frac{f(x)g(x) - f(a)g(a)}{x-a} = f(x)\frac{g(x)-g(a)}{x-a} + \frac{f(x)-f(a)}{x-a}g(a),$$

i.e.,

$$\Delta_{fg,a} = f\,\Delta_{g,a} + g(a)\,\Delta_{f,a},$$

and it remains to apply limits arithmetic. ■

Corollary 5.8 The derivative is a linear operator,

$$(\alpha f + \beta g)' = \alpha f' + \beta g'.$$

Proof: We only need to verify that if f is differentiable at a then so is αf and $(\alpha f)' = \alpha f'$. This follows from the derivative of the product. ■

Example: Let $n \in \mathbb{N}$ and consider the functions $f_n : x \mapsto x^n$. Then,

$$f_n'(x) = n\, x^{n-1} \qquad \text{or equivalently} \qquad f_n' = n\, f_{n-1}.$$

We can show this inductively. We know already that this is true for $n = 0, 1$. Suppose this were true for $n = k$. Then, $f_{k+1} = f_k \cdot \mathrm{Id}$, and by the differentiation rule for products,

$$f_{k+1}' = f_k' \cdot \mathrm{Id} + f_k \cdot \mathrm{Id}',$$

i.e.,

$$f_{k+1}'(x) = k\, x^{k-1} \cdot x + x_k \cdot 1 = (k+1)\, x^k.$$

▲▲▲

Theorem 5.9 If g is differentiable at a and $g(a) \neq 0$, then $1/g$ is differentiable at a and

$$(1/g)'(a) = -\frac{g'(a)}{g^2(a)}.$$

Proof: Since g is continuous at a (it is continuous since it is differentiable) and it is non-zero, then there exists a neighborhood U of a in which g does not vanish (by Theorem 4.27). In $U \setminus \{a\}$ we consider the function

$$\Delta_{1/g,a(x)} = \frac{1/g(x) - 1/g(a)}{x - a} = \frac{-1}{g(x) \cdot g(a)} \cdot \frac{g(x) - g(a)}{x - a},$$

i.e.,

$$\Delta_{1/g,a} = \frac{-\Delta_{g,a}}{g \cdot g(a)}.$$

It remains to take the limit at a and apply the arithmetic laws of limits. ■

Theorem 5.10 If f and g are differentiable at a and $g(a) \neq 0$, then the function f/g is differentiable at a and

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

Proof: Apply the last two theorems. ■

With these theorems in hand, we can differentiate many other functions. In particular there is no difficulty in differentiating a product of more than two functions. For example,

$$(fgh)' = ((fg)h)' = (fg)'h + (fg)h' = (f'g + fg')h = (fg)h' = f'gh + fg'h + fgh'.$$

Suppose we take for granted that

$$\sin' = \cos \qquad \text{and} \qquad \cos' = -\sin.$$

Then we have no problem calculating the derivative of, say, $x \mapsto \sin^3 x$, as a product of three functions. But what about the derivative of $x \mapsto \sin x^3$? Here we need a rule for how to differentiate compositions.

Consider the composite function $g \circ f$, i.e.,

$$(g \circ f)(x) = g(f(x)).$$

Let's try to calculate its derivative at a point a, assuming for the moment that both f and g are differentiable everywhere. We then need to look at the function

$$\Delta_{g\circ f,a}(y) = \frac{(g \circ f)(y) - (g \circ f)(a)}{y-a} = \frac{g(f(y)) - g(f(a))}{y-a},$$

and calculate its limit at a. When y is close to a, we expect the arguments $f(y)$ and $f(a)$ of g to be very close. This suggest the following treatment,

$$\Delta_{g\circ f,a}(y) = \frac{g(f(y)) - g(f(a))}{f(y)-f(a)} \cdot \frac{f(y)-f(a)}{y-a} = \frac{g(f(y)) - g(f(a))}{f(y)-f(a)}\,\Delta_{f,a}(y).$$

It looks that as $y \to a$, since $f(y) - f(a) \to 0$, this product tends to $g'(f(a)) \cdot f'(a)$. The problem is that while the limit $y \to a$ means that the case $y = a$ is not to be considered, there is nothing to prevent the denominator $f(y) - f(a)$ from vanishing, rendering this expression meaningless. Yet, the result is correct, and it only takes a little more subtlety to prove it.

Theorem 5.11 Let $f : A \to B$ and $g : B \to \mathbb{R}$. If f is differentiable at $a \in A$, and g is differentiable at $f(a)$, then

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

Proof: We introduce the following function, defined in a neighborhood of $f(a)$,

$$\psi : z \mapsto \begin{cases} \Delta_{g,f(a)}(z) & z \neq f(a) \\ g'(f(a)) & z = f(a). \end{cases}$$

The fact that g is differentiable at $f(a)$ implies that ψ is continuous at $f(a)$.

We next claim that for $y \neq a$

$$\Delta_{g\circ f,a} = (\psi \circ f) \cdot \Delta_{f,a}.$$

Why that? If $f(y) \neq f(a)$, then this equation reads

$$\frac{g(f(y)) - g(f(a))}{y - a} = \frac{g(f(y)) - g(f(a))}{f(y) - f(a)} \cdot \frac{f(y) - f(a)}{y - a},$$

which holds indeed, whereas if $f(y) = f(a)$ it reads,

$$\frac{g(f(y)) - g(f(a))}{y - a} = g'(f(a)) \cdot \frac{f(y) - f(a)}{y - a},$$

and both sides are zero. Consider now the limit of the right hand side at a. By limits arithmetic,

$$\lim_a [(\psi \circ f) \cdot \Delta_{f,a}] = \psi(f(a)) \cdot \lim_a \Delta_{f,a} = g'(f(a)) \cdot f'(a).$$

■

5.3 Another look at derivatives

In this short section we provide another (equivalent) characterization of differentiability and the derivative (due to Constantin Carathéodory, 1873–1950). Its purpose is to offer a slightly different angle of view on the subject, and show how clever definitions can sometimes greatly simplify proofs.

Our definition of derivatives states that a function f is differentiable at an interior point a, if the function $\Delta_{f,a}$ has a limit at a, and we denote this limit by $f'(a)$. The function $\Delta_{f,a}$ is not defined at a, hence not continuous at a, but this is a removable discontinuity. We could say that f is differentiable at a if there exists a real number, ℓ, such that the function

$$\mathcal{S}_{f,x}(y) = \begin{cases} \Delta_{f,a}(y) & y \neq a \\ \ell & y = a \end{cases}$$

is continuous at a, and $\mathcal{S}_{f,a}(a) = \ell$ is called the derivative of f at a. For $y \neq a$,

$$\mathcal{S}_{f,a}(y) = \Delta_{f,a}(y) = \frac{f(x) - f(a)}{x - a},$$

or equivalently,

$$f(y) = f(a) + \mathcal{S}_{f,a}(y)(y-a).$$

This equation holds also for $y = a$. This suggests the following alternative definition of the derivative:

Definition 5.12 *A function f defined in on open neighborhood of a point a is said to be differentiable at a if there exists a function $\mathcal{S}_{f,a}$ continuous at a, such that*

$$f(y) = f(a) + \mathcal{S}_{f,a}(y)(y-a).$$

$\mathcal{S}_{f,a}(a)$ is called the derivative of f at a and is denoted by $f'(a)$.

Of course, $\mathcal{S}_{f,a}$ coincides with $\Delta_{f,a}$ in some punctured neighborhood of a.

Example: Take the function $f : x \to x^2$. Then

$$f(y) - f(a) = (y+a)(y-a),$$

or

$$f = f(a) + \underbrace{(\mathrm{Id}+a)}_{\mathcal{S}_{f,a}}(\mathrm{Id}-a).$$

Since the function $\mathrm{Id}+a$ is continuous at a, it follows that f is differentiable at a and

$$f'(a) = \mathcal{S}_{f,a}(a) = 2a.$$

▲ ▲ ▲

Example: Take the function $f : x \to 1/x$ and $a \neq 0$. Then

$$f(x) - f(a) = \frac{1}{x} - \frac{1}{a} = -\frac{1}{xa}a - x,$$

or

$$f = f(a) \underbrace{-\frac{1}{a\,\mathrm{Id}}}_{\mathcal{S}_{f,a}}(\mathrm{Id}-a).$$

Since the function $-1/a\,\mathrm{Id}$ is continuous at a, it follows that f is differentiable at a and

$$f'(a) = \mathcal{S}_{f,a}(a) = -\frac{1}{a^2}.$$

▲ ▲ ▲

5.3.1 Leibniz' rule

Let's now see how this alternative definition simplifies certain proofs. Suppose for example that both f and g are differentiable at a. This implies the existence of two functions, $\mathcal{S}_{f,a}$ and $\mathcal{S}_{g,a}$, both continuous at a, such that

$$\begin{aligned} f(y) &= f(a) + \mathcal{S}_{f,a}(y)(y-a) \\ g(y) &= g(a) + \mathcal{S}_{g,a}(y)(y-a) \end{aligned}$$

in some neighborhood of a and $f'(a) = \mathcal{S}_{f,a}(a)$ and $g'(a) = \mathcal{S}_{g,a}(a)$. Then,

$$\begin{aligned} f(y)g(y) &= (f(a) + \mathcal{S}_{f,a}(y)(y-a))\,(g(a) + \mathcal{S}_{g,a}(y)(y-a)) \\ &= f(a)g(a) + (f(a)\mathcal{S}_{g,a}(y) + g(a)\,\mathcal{S}_{f,a}(y) + \mathcal{S}_{f,a}(y)\mathcal{S}_{g,a}(y)(y-a))\,(y-a). \end{aligned}$$

By limit arithmetic, the function in the brackets,

$$F = f(a)\mathcal{S}_{g,a} + g(a)\,\mathcal{S}_{f,a} + \mathcal{S}_{f,a}\mathcal{S}_{g,a}(\mathrm{Id} - a)$$

is continuous at a, hence fg is differentiable at a and

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

5.3.2 Derivative of a composition

Suppose now that f is differentiable at a and g is differentiable at $f(a)$. Then, there exist functions, $\mathcal{S}_{f,a}$ and $\mathcal{S}_{g,f(a)}$, continuous at a and $f(a)$, such that

$$\begin{aligned} f(y) &= f(a) + \mathcal{S}_{f,a}(y)(y-a) \\ g(z) &= g(f(a)) + \mathcal{S}_{g,f(a)}(z)(z - f(a)). \end{aligned}$$

Now,

$$\begin{aligned} g(f(y)) &= g(f(a)) + \mathcal{S}_{g,f(a)}(f(y))(f(y) - f(a)) \\ &= g(f(a)) + \mathcal{S}_{g,f(a)}(f(y))\mathcal{S}_{f,a}(y)(y-a), \end{aligned}$$

or,

$$(g\circ f)(y) = (g\circ f)(a) + \underbrace{\mathcal{S}_{g,f(a)}(f(y))\mathcal{S}_{f,a}(y)}_{F}(y-a).$$

By the properties of continuous functions, the function in the square brackets is continuous at a, and

$$(g\circ f)'(a) = F(a) = g'(f(a))\,f'(a).$$

5.4 The derivative and extrema

In high-school calculus, one of the main uses of differential calculus is to find extrema of functions. We are going to put this practice on solid grounds. First, recall the definition,

Definition 5.13 Let $f : A \to B$. A point $a \in A$ (not necessarily an internal point) is said to be a ***maximum point*** *of f in A, if*

$$f(x) \le f(a) \qquad \forall x \in A.$$

We define similarly a minimum point.

Comment: By no means a maximum point has to be unique, nor to exist. For example, in a constant function all points are minima and maxima. On the other hand, the function $f : (0,1) \to \mathbb{R}$, $f : x \mapsto x^2$ does not have a maximum in $(0,1)$. Also, we proved that a continuous function defined on a closed interval always has a maximum (and a minumum, both, not necessarily unique).

Here is a first connection between maximum (and minimum) points and derivatives:

Theorem 5.14 Let $f : (a,b) \to \mathbb{R}$. If $x \in (a,b)$ is a maximum point of f and f is differentiable at x, then $f'(x) = 0$.

Proof: Since x is, by assumption, a maximum point, then for every $y \in (a,b)$,

$$f(y) - f(x) \le 0.$$

In particular, for $y \ne x$,

$$\Delta_{f,x}(y) = \frac{f(y) - f(x)}{y - x},$$

satisfies

$$\begin{cases} \Delta_{f,x}(y) \le 0 & y > x \\ \Delta_{f,x}(y) \ge 0 & y < x. \end{cases}$$

It is given that $\Delta_{f,x}$ has a limit at x (it is $f'(x)$). We will show that this limit is necessarily zero.

Indeed, by the relation between limits and order,

$$\lim_{x+} \Delta_{f,x} \le 0 \qquad \text{and} \qquad \lim_{x-} \Delta_{f,x} \ge 0,$$

and since both one-sided limits are equal, we conclude that $f'(x) = 0$.

∎

Comment: This is a uni-directional theorem. It does not imply that if $f'(x) = 0$ then x is a maximum point (nor a minimum point).

Comment: The open interval cannot be replaced by the closed interval $[a,b]$, because at the end points we can only consider one-sided limits.

Definition 5.15 Let $f : A \to B$. An interior point $a \in A$ is called a ***local maximum*** *of f (מקסימום מקומי), if there exists a $\delta > 0$ such that a is a maximum of f in $(a-\delta, a+\delta)$ ("local" always means "in a sufficiently small neighborhood").*

Theorem 5.16 [Pierre de Fermat] If a is a local maximum (or minimum) of f in some open interval and f is differentiable at a then $f'(a) = 0$.

Proof: There is actually nothing to prove, as the previous theorem applies verbatim for f restricted to the interval $(a-\delta, a+\delta)$. ∎

Comment: The converse is not true. Take for example the function $f : x \to x^3$. If is differentiable in $\mathbb{R}$ and $f'(0) = 0$, but zero is not a local minimum, nor a local maximum of f.

Definition 5.17 Let $f : A \to B$ be differentiable. A point $a \in A$ is called a ***critical point*** *of f if $f'(a) = 0$. The value $f(a)$ is called a* ***critical value****.*

5.4.1 Locating extremal points

Let $f : [a, b] \to \mathbb{R}$ and suppose we want to find a maximum point of f. If f is continuous, then a maximum is guaranteed to exist. There are three types of candidates: (1) critical points of f in (a, b), (ii) the end points a and b, and (iii) points where f is not differentiable. If f is differentiable, then the task reduces to finding all the critical points and comparing all the critical values,

$$\{f(x) : f'(x) = 0\}.$$

Finally, the largest critical value has to be compared to $f(a)$ and $f(b)$.

Example: Consider the function $f : [-1, 2] \to \mathbb{R}$, $f(x) = x^3 - x$. This function is differentiable everywhere, and its critical points satisfy

$$f'(x) = 3x^2 - 1 = 0,$$

i.e., $x = \pm 1/\sqrt{3}$, which are both in the domain. It is easily checked that

$$f(1/\sqrt{3}) = -\frac{2}{3\sqrt{3}} \qquad \text{and} \qquad f(-1/\sqrt{3}) = \frac{2}{3\sqrt{3}}.$$

Finally, $f(-1) = 0$ and $f(2) = 6$, hence 2 is the (unique) maximum point. ▲▲▲

5.5 The mean-value theorem

So far, we have derived properties of the derivative given the function. What about the reverse direction? Take the following example: we know that the derivative of a constant function is zero. Is it also true that if the derivative is zero then the function is a constant? A priori, it is not clear how to show it. How can we go from the knowledge that

$$\lim_{y \to x} \frac{f(y) - f(x)}{y - x} = 0,$$

to showing that f is a constant function. The following two theorems will provide us with the necessary tools:

Theorem 5.18 [Michel Rolle, 1691] If f is continuous on $[a,b]$, differentiable in (a,b), and $f(a) = f(b)$, then there exists a point $c \in (a,b)$ where $f'(c) = 0$.

Proof: It follows from the continuity of f that it has a maximum and a minimum point in $[a,b]$. If the maximum occurs at some $c \in (a,b)$ then $f'(c) = 0$ and we are done. If the minimum occurs at some $c \in (a,b)$ then $f'(c) = 0$ and we are done. The only remaining alternative is that a and b are both minima and maxima, in which case f is a constant and its derivative vanishes at some interior point (well, at all of them). ■

Comment: The requirement that f be differentiable everywhere in (a,b) is imperative, for consider the function

$$f(x) = \begin{cases} x & 0 \le x \le \frac{1}{2} \\ 1 - x & \frac{1}{2} < x \le 1. \end{cases}$$

Even though f is continuous in $[0,1]$ and $f(0) = f(1) = 0$, there is no interior point where $f'(x) = 0$.

*Theorem 5.19 [Mean-value theorem (**משפט הערך הממוצע**)] If f is continuous on $[a,b]$ and differentiable in (a,b), then there exists a point $c \in (a,b)$ where*

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

that is, a point at which the derivative equals to the mean rate of change of f between a and b.

Comment: Rolle's theorem is a particular case.

Proof: This is almost a direct consequence of Rolle's theorem. Define the function $g : [a,b] \to \mathbb{R}$,

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

g is continuous on $[a,b]$ and differentiable in (a,b). Moreover, $g(a) = g(b) = f(a)$. Hence by Rolle's theorem there exists a point $c \in (a,b)$, such that

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}.$$

■

Corollary 5.20 Let $f : [a,b] \to \mathbb{R}$. If $f'(x) = 0$ for all $x \in (a,b)$ then f is a constant.

Proof: Let $c, d \in (a,b)$, $c < d$. By the mean-value theorem there exists some $e \in (c,d)$, such that

$$f'(e) = \frac{f(d) - f(c)}{d - c},$$

however $f'(e) = 0$, hence $f(d) = f(c)$. Since this holds for all pair of points, then f is a constant. ■

Comment: This is only true if f is defined on an interval. Take a domain of definition which is the union of two disjoint sets, and this is no longer true (f is only constant in every "connected component" (רכיב קשירות)).

Corollary 5.21 If f and g are differentiable on an interval with $f'(x) = g'(x)$, then there exists a real number c such that $f = g + c$ on that interval.

Proof: Apply the previous corollary for $f - g$. ■

Example: Suppose we are given that a function $f : \mathbb{R} \to \mathbb{R}$ satisfies the ***differential equation***

$$f'(x) = a\, f(x), \qquad \forall x \in \mathbb{R},$$

and the ***initial condition***

$$f(0) = c.$$

We will show that this function is $f(x) = c\, e^{ax}$.

Indeed, we are given that

$$f'(x) - a f(x) = 0,$$

or,

$$e^{-ax} f'(x) - a e^{-ax} f(x) = 0,$$

which we can rewrite as

$$F'(x) = 0,$$

where $F(x) = e^{-ax} f(x)$. Thus $F(x)$ is a constant, however $F(0) = f(0) = c$, which implies, that

$$F(x) = e^{-ax} f(x) = c,$$

which concludes the proof. ▲▲▲

Corollary 5.22 If f is continuous on a closed interval $[a, b]$ and $f'(x) > 0$ in (a, b), then f is increasing on that interval.

Proof: Let $x < y$ belong to that interval. By the mean-value theorem there exists a $c \in (x, y)$ such that

$$f'(c) = \frac{f(y) - f(x)}{y - x},$$

however $f'(c) > 0$, hence $f(y) > f(x)$. ■

Comment: The converse is not true. If f is increasing and differentiable then $f'(x) \geq 0$, but equality may hold, as in $f(x) = x^3$ at zero.

Comment: Suppose now that $f'(a) > 0$. Does it imply that f is increasing in a neighborhood of a. If f' is continuous at a, then $f'(x) > 0$ is some neighborhood of a, and f is increasing in that neighborhood. By f' may fail to be continuous at a. What then? The answer is negative, for consider the function

$$f(x) = \begin{cases} x + 2x^2 \sin(1/x) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Then,

$$f'(x) = \begin{cases} 1 + 4x\sin(1/x) - 2\sin(1/x) & x \neq 0 \\ 1 & x = 0. \end{cases}$$

Thus, any any neighborhood of 0, f' assumes values close to -1 and close to 3. Even though $f'(0) = 1$, f is not increasing in a neighborhood of 0.

We have seen that at a local minimum (or maximum) the derivative (if it exists) vanishes, but that the opposite is not true. The following theorem gives a sufficient condition for a point to be a local minimum (with a corresponding theorem for a local maximum).

Theorem 5.23 If $f'(a) = 0$ and $f''(a) > 0$ then a is a local minimum of f.

Comment: The fact that f'' exists at a implies:

1. f' exists in some neighborhood of a.
2. f' is continuous at a.
3. f is continuous in some neighborhood of a.

Proof: By definition,

$$f''(a) = \lim_{a} \Delta_{f',a} > 0 \qquad \text{where} \qquad \Delta_{f',a}(y) = \frac{f'(y) - f'(a)}{y - a} = \frac{f'(y)}{y - a}.$$

Thus there exists a $\delta > 0$ such that

$$\frac{f'(y)}{y - a} > 0 \qquad \text{whenever} \qquad 0 < |y - a| < \delta,$$

or,

$$\begin{aligned} f'(y) > 0 \qquad & \text{whenever} \qquad a < y < a + \delta \\ f'(y) < 0 \qquad & \text{whenever} \qquad a - \delta < y < a. \end{aligned}$$

It follows that f is increasing in a right-neighborhood of a and decreasing in a left-neighborhood of a, i.e., a is a local minimum.

Another way of completing the argument is as follows: for every $y \in (a, a+\delta)$:

$$\frac{f(y) - f(a)}{y - a} = f'(c_y) > 0 \qquad \text{i.e.,} \qquad f(y) > f(a),$$

where $c_y \in (a, y)$. Similarly, for every $y \in (a - \delta, a)$:

$$\frac{f(y) - f(a)}{y - a} = f'(c_y) < 0 \qquad \text{i.e.,} \qquad f(y) > f(a),$$

where here $c_y \in (y, a)$. ■

Example: Consider the function $f(x) = x^3 - x$. There are three critical points $0, \pm 1/\sqrt{3}$: one local maximum, one local minimum, and one which is neither. ▲▲▲

Example: Consider the function $f(x) = x^4$. Zero is a minimum point, even though $f''(0) = 0$. ▲▲▲

Theorem 5.24 If f has a local minimum at a and $f''(a)$ exists, then $f''(a) \geq 0$.

Proof: Suppose, by contradiction, that $f''(a) < 0$. By the previous theorem this would imply that a is both a local minimum and a local maximum. This in turn implies that there exists a neighborhood of a in which f is constant, i.e., $f'(a) = f''(a) = 0$, a contradiction. ■

The following theorem states that the derivative of a continuous function cannot have a removable discontinuity.

Theorem 5.25 Suppose that f is continuous at a and

$$\lim_a f' \text{ exists,}$$

then $f'(a)$ exists and f' is continuous at a.

Proof: The assumption that f' has a limit at a implies that f' exists in some punctured neighborhood U of a. Thus, for $y \in U \setminus \{a\}$ the function f is continuous on the closed segment that connects y and a and differentiable on the corresponding open segment. By the mean-value theorem, there exists a point c_y between y and a, such that

$$f'(c_y) = \frac{f(y) - f(a)}{y - a} = \Delta_{f,a}(y).$$

Since f' has a limit at a,

$$(\forall \varepsilon > 0)(\exists \delta > 0) : (\forall y \in B^\circ(a, \delta)) \left(|f'(y) - \lim_a f'| < \varepsilon \right).$$

Since $y \in B^\circ(a, \delta)$ implies that $c_y \in B^\circ(a, \delta)$,

$$(\forall \varepsilon > 0)(\exists \delta > 0) : (\forall y \in B^\circ(a, \delta))(|\underbrace{f'(c_y)}_{\Delta_{f,a}(y)} - \lim_a f'| < \varepsilon),$$

which precisely means that

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

■

Example: The function

$$f(x) = \begin{cases} x^2 & x < 0 \\ x^2 + 5 & x \geq 0, \end{cases}$$

is differentiable in a neighborhood of 0 and

$$\lim_0 f' \text{ exists},$$

but it is not continuous, and therefore not differentiable at zero. ▲▲▲

Example: The function

$$f(x) = \begin{cases} x^2 \sin 1/x & x \neq 0 \\ 0 & x = 0, \end{cases}$$

is continuous and differentiable in a neighborhood of zero. However, the derivative does not have a limit at zero, so that the theorem does not apply. Note, however, that f is differentiable at zero. ▲▲▲

Theorem 5.26 [Augustin Louis Cauchy; intermediate value theorem] Suppose that f, g are continuous on $[a,b]$ and differentiable in (a,b). Then, there exists a $c \in (a,b)$, such that

$$[f(b) - f(a)]g'(c) = [g(b) - g(a)]f'(c).$$

Comment: If $g(b) \neq g(a)$ then this theorem states that

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

(provided also that $g'(c) \neq 0$). This may seem like a corollary of the mean-value theorem. We know that there exist c,d, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a} \qquad \text{and} \qquad g'(d) = \frac{g(b) - g(a)}{b - a}.$$

The problem is that there is no reason for the points c and d to coincide.

Proof: Define

$$h(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)],$$

and apply Rolle's theorem. Namely, we verify that h is continuous on $[a,b]$ and differentiable in (a,b), and that $h(a) = h(b)$, hence there exists a point $c \in (a,b)$, such that

$$h'(c) = f'(c)[g(b) - g(a)] - g'(c)[f(b) - f(a)] = 0.$$

■

Theorem 5.27 [L'Hôpital's rule] Suppose that

$$\lim_a f = \lim_a g = 0,$$

and that

$$\lim_a \frac{f'}{g'} \text{ exists.}$$

Then the limit $\lim_a (f/g)$ *exists and*

$$\lim_a \frac{f}{g} = \lim_a \frac{f'}{g'}.$$

Comment: This is a very convenient tool for obtaining a limit of a fraction, when both numerator and denominator vanish in the limit[2]. —

54h(2017)—

Proof: It is implicitly assumed that f and g are differentiable in a neighborhood of a (except perhaps at a) and that g' is non-zero in a neighborhood of a (except perhaps at a); otherwise, the limit $\lim_a (f'/g')$ would not have existed. Note that f and g are not necessarily defined at a; since they have a limit there, there will be no harm assuming that they are continuous at a, i.e., $f(a) = g(a) = 0$[3].

First, we claim that g does not vanish in some neighborhood U of a, for by Rolle's theorem, it would imply that g' vanishes somewhere in this neighborhood. More formally, if

$$g'(x) \neq 0 \qquad \text{whenever} \qquad 0 < |x - a| < \delta,$$

then

$$g(x) \neq 0 \qquad \text{whenever} \qquad 0 < |x - a| < \delta,$$

for if $g(x) = 0$, with, say, $0 < x < \delta$, then there exists a $0 < c_x < x < \delta$, where $g'(c_x) = 0$.

Both f and g are continuous and differentiable in some neighborhood U that contains a, hence by Cauchy's mean-value theorem, there exists for every

[2] L'Hôpital's rule was in fact derived by Johann Bernoulli who was giving to the French nobleman, le Marquis de l'Hôpital, lessons in calculus. L'Hôpital was the one to publish this theorem, acknowledging the help of Bernoulli.

[3] What we really do is to define continuous functions

$$\tilde{f}(x) = \begin{cases} f(x) & x \neq a \\ 0 & x = a \end{cases} \qquad \text{and} \qquad \tilde{g}(x) = \begin{cases} g(x) & x \neq a \\ 0 & x = a \end{cases},$$

and prove the theorem for the "corrected" functions $\tilde{f}$ and $\tilde{g}$. At the end, we observe that the result must apply for f and g as well.

$x \in U$, a point c_x between x and a, such that

$$\frac{f(x) - 0}{g(x) - 0} = \frac{f(x)}{g(x)} = \frac{f'(c_x)}{g'(c_x)}. \tag{5.1}$$

Since the limit of f'/g' at a exists,

$$(\forall \varepsilon > 0)(\exists \delta > 0) x \in B^\circ(a, \delta) \quad \text{implies} \quad \left(\left| \frac{f'(x)}{g'(x)} - \lim_a \frac{f'}{g'} \right| < \varepsilon \right),$$

and since $x \in B^\circ(a, \delta)$ implies $c_x \in B^\circ(a, \delta)$,

$$(\forall \varepsilon > 0)(\exists \delta > 0) x \in B^\circ(a, \delta) \quad \text{implies} \quad \left(\left| \underbrace{\frac{f'(c_x)}{g'(c_x)}}_{f(x)/g(x)} - \lim_a \frac{f'}{g'} \right| < \varepsilon \right),$$

which means that

$$\lim_a \frac{f}{g} = \lim_a \frac{f'}{g'}.$$

■

Example:

1. The limit
$$\lim_{x \to 0} \frac{\sin x}{x} = 1.$$
2. The limit
$$\lim_{x \to 0} \frac{\sin^2 x}{x^2} = 1,$$
with two consecutive applications of l'Hôpital's rule.

▲▲▲

Comment: Purposely, we only prove one variant among many other of l'Hôpital's rule. Another variants is: Suppose that

$$\lim_\infty f = \lim_\infty g = 0,$$

and that

$$\lim_{\infty} \frac{f'}{g'} \text{ exists.}$$

Then

$$\lim_{\infty} \frac{f}{g} = \lim_{\infty} \frac{f'}{g'}.$$

Yet another one is: Suppose that

$$\lim_{a} f = \lim_{a} g = \infty,$$

and that

$$\lim_{a} \frac{f'}{g'} \text{ exists.}$$

Then

$$\lim_{a} \frac{f}{g} = \lim_{a} \frac{f'}{g'}.$$

The following theorem is very reminiscent of l'Hôpital's rule, but note how different it is:

Theorem 5.28 Suppose that f and g are differentiable at a, with

$$f(a) = g(a) = 0 \qquad \textit{and} \qquad g'(a) \neq 0.$$

Then,

$$\lim_{a} \frac{f}{g} = \frac{f'(a)}{g'(a)}.$$

Proof: Without loss of generality, assume $g'(a) > 0$. That is,

$$\lim_{y \to a} \frac{g(y) - g(a)}{y - a} = \lim_{y \to a} \frac{g(y)}{y - a} = g'(a) > 0,$$

hence there exists a punctured neighborhood of a where $g(x)/(x-a) > 0$, and in particular, in which the numerator $g(x)$ is non-zero. Thus, we can divide by $g(x)$ in a punctured neighborhood of a, and

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{\frac{f(x)-f(a)}{x-a}}{\frac{g(x)-g(a)}{x-a}} = \frac{\Delta_{f,a}(x)}{\Delta_{g,a}(x)}.$$

Using the arithmetic of limits, we obtain the desired result. ■

5.6 The derivative of the exponential

Let $f : \mathbb{R} \to \mathbb{R}$ be the exponential function,

$$f(x) = e^x.$$

We have seen that

$$f(x) = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n.$$

We are going to prove that the exponential function has the very special property

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

Lemma 5.29 For every $-1/2 < x < 1/2$,

$$1 + y - 2y^2 \le f(x) \le 1 + y + 2y^2.$$

Proof: For every $-1/2 < x < 1/2$ and every $n \in \mathbb{N}$,

$$(1 + x/n)^n = 1 + x + \frac{n(n-1)}{2n^2} x^2 + \frac{n(n-1)}{3!\, n^3} x^3 + \cdots \frac{1}{n^n} x^n,$$

i.e.,

$$\left| \left(1 + \frac{x}{n}\right)^n - (1 + x) \right| \le x^2 \left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{2^{n-2}}\right) < 2x^2.$$

It follows that for all n,

$$1 + x - 2x^2 < \left(1 + \frac{x}{n}\right)^n < 1 + x + 2x^2,$$

and the statement follows from the properties of limits and order. ∎

Proposition 5.30 f is differentiable at zero and

$$f'(0) = 1.$$

Proof: We have

$$\Delta_{f,0}(x) = \frac{e^x - 1}{x},$$

hence, for $-1/2 < x < 1/2$,

$$1 - 2x \le \Delta_{f,0}(x) \le 1 + 2x.$$

By the sandwich theorem,

$$\lim_{0} \Delta_{f,0} = 1.$$

■

Proposition 5.31 f is everywhere differentiable and

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

Proof: Note that

$$\Delta_{f,a}(x) = \frac{e^x - e^a}{x - a} = e^a \frac{e^{x-a} - 1}{x - a} = e^a \, \Delta_{f,0}(x - a).$$

Now,

$$\lim_{x \to a} \Delta_{f,0}(x - a) = \lim_{x \to 0} \Delta_{f,0}(x) = f'(0) = 1,$$

hence by limit arithmetic,

$$\lim_{a} \Delta_{f,a} = e^a = f(a).$$

■

Comment: The only functions satisfying $f' = f$ are of the form $f(x) = c\,e^x$. We've seen how to prove it. —

56h(2017)—

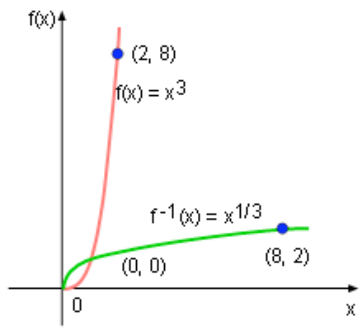

Figure 5.2: The derivative of the inverse function

5.7 Derivatives of inverse functions

We have already shown that if $f : A \to B$ is one-to-one and onto, then it has an inverse $f^{-1} : B \to A$. We then say that f is ***invertible*** (הפיכה). We have seen that an invertible function defined on an interval must be monotonic, and that the inverse of a continuous function is continuous. In this section, we study under what conditions is the inverse of a differentiable function differentiable.

Recall also that

$$f^{-1} \circ f = \text{Id},$$

with $\text{Id} : A \to A$. If both f and f^{-1} were differentiable at a and $f(a)$, respectively, then by the composition rule,

$$(f^{-1})'(f(a))\, f'(a) = 1.$$

Setting $f(a) = b$, or equivalently, $a = f^{-1}(b)$, we get

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))}.$$

There is one little flaw with this argument. We have assumed $(f^{-1})'$ to exist. If this were indeed the case, then we have just derived an expression for the derivative of the inverse.

Corollary 5.32 If $f'(f^{-1}(b)) = 0$ then f^{-1} is not differentiable at b.

Example: The function $f(x) = x^3 + 2$ is continuous and invertible, with $f^{-1}(x) = \sqrt[3]{x-2}$. Since, $f'(f^{-1}(2)) = 0$, then f^{-1} is not differentiable at 2. ▲▲▲

Theorem 5.33 Let $f : I \to \mathbb{R}$ be continuous and invertible. If f is differentiable at $f^{-1}(a)$ and $f'(f^{-1}(a)) \neq 0$, then f^{-1} is differentiable at b.

Proof: We have

$$\Delta_{f^{-1},a}(x) = \frac{f^{-1}(x) - f^{-1}(a)}{x - a} = \frac{f^{-1}(x) - f^{-1}(a)}{f(f^{-1}(x)) - f(f^{-1}(a))} = \frac{1}{\Delta_{f,f^{-1}(a)}(f^{-1}(x))},$$

i.e.,

$$\Delta_{f^{-1},a} = \frac{1}{\Delta_{f,f^{-1}(a)} \circ f^{-1}}.$$

Since f is differentiable at $f^{-1}(a)$, it is in particular continuous at that point, hence f^{-1} is continuous at a. Since $\Delta_{f,f^{-1}(a)}$ is continuous at $f^{-1}(a)$ (the limit is $f'(f^{-1}(a))$, the theorem follows by limit arithmetic,

$$\begin{aligned} \lim_a \Delta_{f^{-1},a} &= \frac{1}{\lim_a \Delta_{f,f^{-1}(a)} \circ f^{-1}} \\ &= \frac{1}{\lim_{f^{-1}(a)} \Delta_{f,f^{-1}(a)}} \\ &= \frac{1}{f'(f^{-1}(a))}. \end{aligned}$$

■

Example: Consider the function $f(x) = \tan x$ defined on $(-\pi/2, \pi/2)$. We denote its inverse by $\arctan x$. Then[4],

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))} = \cos^2(\arctan x) = \frac{1}{1 + \tan^2(\arctan x)} = \frac{1}{1 + x^2}.$$

▲▲▲

[4] We used the identity $\cos^2 = 1/(1 + \tan^2)$.

Example: Consider the function $f(x) = x^n$, with n integer. This proves that the derivative of $f^{-1}(x) = x^{1/n}$ is

$$(f^{-1})'(x) = \frac{1}{n} x^{1/n-1},$$

and together with the composition rule gives the derivative for all rational powers. ▲▲▲

Example: The logarithm is the function inverse to the exponential,

$$\log : (0, \infty) \to \mathbb{R},$$

where

$$\log(e^x) = x.$$

Then,

$$\log'(x) = \frac{1}{\exp'(\log(x))} = \frac{1}{\exp(\log(x))} = \frac{1}{x}.$$

▲▲▲

5.8 Complements

Theorem 5.34 Suppose that $f : A \to B$ is monotonic (say, increasing), then it has one-sided limits at every point that has a one-sided neighborhood that belongs to A. (In particular, if A is a connected set then f has one-sided limits everywhere.)

Proof: Consider a segment $(b, a] \in A$. We need to show that f has a left-limit at a. The set

$$K = \{f(x) : b < x < a\}$$

is non-empty and upper bounded by $f(a)$ (since f is increasing). Hence we can define $\ell = \sup K$. We are going to show that

$$\ell = \lim_{a^-} f.$$

Let $\varepsilon > 0$. By the definition of the supremum, there exists an $m \in K$ such that $m > \ell - \varepsilon$. Let y be such $f(y) = m$. By the monotonicity of f,

$$\ell - \varepsilon < f(x) \le \ell \qquad \text{whenever} \qquad y < x < a.$$

This concludes the proof. We proceed similarly for right-limits. ■

Theorem 5.35 Let $f : [a, b] \to \mathbb{R}$ be monotonic (say, increasing). Then the image of f is a segment if and only if f is continuous.

Proof: (i) Suppose first that f is continuous. By monotonicity,

$$\text{Image}(f) \subseteq [f(a), f(b)].$$

That every point of $[f(a), f(b)]$ is in the image of f follows from the intermediate-value theorem.

(ii) Suppose then that $\text{Image}(f) = [f(a), f(b)]$. Suppose, by contradiction, that f is not continuous. This implies the existence of a point $c \in [a, b]$, where

$$f(c) > \lim_{c^-} f \qquad \text{and/or} \qquad f(c) < \lim_{c^+} f.$$

(We rely on the fact that one-sided derivatives exist.)

Say, for example, that $f(c) < \lim_{c^+} f$. Then,

$$f(c) < \frac{f(c) + \lim_{c^+} f}{2} < \lim_{c^+} f$$

is not in the image of f, which contradicts the fact that $\text{Image}(f) = [f(a), f(b)]$. ■

Theorem 5.36 [Jean-Gaston Darboux] Let $f : A \to B$ be differentiable on (a, b), including a right-derivative at a and a left-derivative at b. Then, f' has the "intermediate-value property", whereby for every t between $f'(a^+)$ and $f'(b^-)$ there exists an $x \in [a, b]$, such that $f'(x) = t$.

Comment: The statement is trivial when f is continuously differentiable (by the intermediate-value theorem). It is nevertheless correct even if f' has discontinuities. This theorem thus limits the functions that are derivatives of other functions—not every function can be a derivative.

Proof: Without loss of generality, let us assume that $f'(a^+) > f'(b^-)$. Let $f'(a^+) > t > f'(b^-)$, and consider the function

$$g(x) = f(x) - tx.$$

(Here t is a fixed parameter; for every t we can construct such a function.) We have

$$g'(a^+) = f'(a^+) - t > 0 \qquad \text{and} \qquad g'(b^-) = f'(b^-) - t < 0,$$

and we wish to show the existence of an $x \in [a, b]$ such that $g'(x) = 0$.

Since g is continuous on $[a, b]$ then it must attain a maximum. The point a cannot be a maximum point because g is increasing at a. Similarly, b cannot be a maximum point because g is decreasing at b. Thus, there exists an interior point c which is a maximum, and by Fermat's theorem, $g'(c) = f'(c) - t = 0$. ■

Example: The function

$$f(x) = \begin{cases} x^2 \sin 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is differentiable everywhere, and its derivative is

$$f'(x) = \begin{cases} 2x \sin 1/x - \cos 1/x & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

The derivative is not continuous at zero, but nevertheless satisfes the Darboux theorem. ▲▲▲

Example: How crazy can a continuous function be? It was Weierstrass who shocked the world by constructing a continuous function that it nowhere differentiable! As instance of his class of examples is

$$f(x) = \sum_{k=0}^{\infty} \frac{\cos(21^k \pi x)}{3^k}.$$

▲▲▲

5.9 Taylor's theorem

Recall the mean-value theorem. If $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and differentiable in (a,b), then for every point $x \in (a,b)$ there exists a point c_x, such that

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

We can write it alternatively as

$$f(x) = f(a) + f'(c_x)(x - a).$$

Suppose, for example, that we knew that $|f'(y)| < M$ for all y. Then, we would deduce that $f(x)$ cannot differ from $f(a)$ by more than M times the separation $|x - a|$. In particular,

$$\lim_a [f - f(a)] = 0.$$

It turns out that if we have more information, about higher derivatives of f, then we can refine a lot the estimates we have on the variation of f as we get away from the point a.

Definition 5.37 Let f be n-times differentiable at a. We define the ***Taylor polynomial of f of order n at a*** *by*

$$P_{f,n,a} = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(\mathrm{Id} - a)^k,$$

or equivalently,

$$P_{f,n,a}(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x - a)^k.$$

Explicitly,

$$P_{f,n,a}(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \cdots + \ldots \frac{f^{(n)}(a)}{n!}(x - a)^n.$$

Lemma 5.38 Let f be n times differentiable at a. Then,

$$P'_{f,n,a} = P_{f',n-1,a}.$$

Proof: This is an immediate consequence of the definition of the Taylor polynomial. ■

Theorem 5.39 Let f be n times differentiable at a. Then,

$$\lim_a \frac{f - P_{f,n,a}}{(\mathrm{Id} - a)^n} = 0.$$

Comment: Loosely speaking, this theorem states that, as x approaches a, $P_{f,n,a}(x)$ is closer to $f(x)$ than $(x-a)^n$ (or that f and $P_{f,n,a}$ are "equal up to order n"; see below). For $n = 0$ it states that

$$\lim_a [f - f(a)] = 0,$$

whereas for $n = 1$,

$$\lim_a \left(\frac{f - f(a)}{(\mathrm{Id} - a)} - f'(a) \right) = \lim_a \Delta_{f,a} - f'(a) = 0,$$

which holds by definition.

Proof: We have seen above that the theorem holds for $n = 1$. Suppose that it holds for $k < n$, i.e., for every function g that is k times differentiable at a,

$$\lim_a \frac{g - P_{g,k,a}}{(\mathrm{Id} - a)^k} = 0.$$

Since f' is k times differentiable at a,

$$\lim_a \frac{f' - P_{f',k,a}}{(\mathrm{Id} - a)^k} = 0,$$

which we may rewrite as follows,

$$\lim_a \frac{(f - P_{f,k+1,a})'}{[(\mathrm{Id} - a)^{k+1}]'} = 0.$$

Since

$$\lim_a (f - P_{f,k+1,a}) = 0 \qquad \text{and} \qquad \lim_a (\mathrm{Id} - a)^{k+1} = 0,$$

it follows by L'hôpital's rule that

$$\lim_a \frac{f - P_{f,k+1,a}}{(\mathrm{Id} - a)^{k+1}} = 0.$$

This completes the proof. ■

If we denote by Π_n the set of polynomials of degree at most n, then the defining property of $P_{f,n,a}$ is:

1. $P_{f,n,a} \in \Pi_n$.
2. $P^{(k)}_{f,n,a}(a) = f^{(k)}(a)$ for $k = 0, 1, \dots, n$.

Definition 5.40 Let f, g be defined in a neighborhood of a. We say that f and g are ***equal up to order*** n ***at*** a *if*

$$\lim_a \frac{f - g}{(\mathrm{Id} - a)^n} = 0.$$

It is easy to see that "equal up to order n" is an ***equivalence relation***, which defines an ***equivalence class***.

Thus, the above theorem proves that $f(x)$ and $P_{f,n,a}$ (provided that the latter exists) are equal up to order n at a.

The next task is to obtain an expression for the difference between the function f and its Taylor polynomial. We define the ***remainder*** (שארית) of order n, $R_n(x)$, by 60h(2017)

$$f(x) = P_{f,n,a}(x) + R_{f,n,a}(x).$$

Proposition 5.41 Let $P, Q \in \Pi_n$ be equal up to order n at a. Then $P = Q$.

Proof: We can always express these polynomials as

$$\begin{aligned} P(x) &= a_0 + a_1(x - a) + \cdots + a_n(x - a)^n \\ Q(x) &= b_0 + b_1(x - a) + \cdots + b_n(x - a)^n. \end{aligned}$$

We know that

$$\lim_a \frac{P - Q}{(\mathrm{Id} - a)^k} = 0 \qquad \text{for } k = 0, 1, \dots, n.$$

For $k = 0$

$$0 = \lim_a (P - Q) = a_0 - b_0,$$

i.e., $a_0 = b_0$. We proceed similarly for each k to show that $a_k = b_k$ for $k = 0, 1, \dots, n$. ■

Corollary 5.42 If f is n times differentiable at a and f is equal to $Q \in \Pi_n$ up to order n at a, then

$$P_{f,n,a} = Q.$$

Example: We know from high-school math that for $|x| < 1$

$$f(x) = \frac{1}{1-x} = \sum_{k=0}^{n} x^k + \frac{x^{n+1}}{1-x}.$$

Thus,

$$\frac{f(x) - \sum_{k=0}^{n} x^k}{x^n} = \frac{x}{1-x}.$$

Since the right hand side tends to zero as $x \to 0$ it follows that f and $\sum_{k=0}^{n} x^k$ are equal up to order n at 0, which means that

$$P_{f,n,0}(x) = \sum_{k=0}^{n} x^k.$$

▲ ▲ ▲

Example: Similarly for $|x| < 1$

$$f(x) = \frac{1}{1+x^2} = \sum_{k=0}^{n} (-1)^k x^{2k} + (-1)^{n+1} \frac{x^{2n+2}}{1+x^2}.$$

Thus,

$$\frac{f(x) - \sum_{k=0}^{n} (-1)^k x^{2k}}{x^{2n}} = \frac{x^2}{1+x^2}.$$

Since the right hand side tends to zero as $x \to 0$ it follows that

$$P_{f,2n,0}(x) = \sum_{k=0}^{n} (-1)^k x^{2k}.$$

▲ ▲ ▲

Example: Let $f = \arctan$. Then,

$$f(x) = \int_0^x \frac{dt}{1+t^2} = \int_0^x \left(\sum_{k=0}^n (-1)^k t^{2k} + (-1)^{n+1} \frac{t^{2n+2}}{1+t^2} \right) dt$$
$$= \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{2k+1} + \int_0^x (-1)^{n+1} \frac{t^{2n+2}}{1+t^2}\, dt.$$

Since

$$\left| f(x) - \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{2k+1} \right| \le |x|^{2n+3},$$

it follows that

$$P_{f,2n+2,0} = \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{2k+1}.$$

▲▲▲

Theorem 5.43 Let f, g be n times differentiable at a. Then,

$$P_{f+g,n,a} = P_{f,n,a} + P_{g,n,a},$$

and

$$P_{fg,n,a} = [P_{f,n,a} P_{g,n,a}]_n,$$

where $[]_n$ stands for a truncation of the polynomial in $(\mathrm{Id} - a)$at the n-th power.

Proof: We use again the fact that if a polynomial "looks" like the Taylor polynomial then it is. The sum $P_{f,n,a} + P_{g,n,a}$ belongs to Π_n and

$$\lim_a \frac{(f+g) - (P_{f,n,a} + P_{g,n,a})}{(\mathrm{Id} - a)^n} = \lim_a \frac{f - P_{f,n,a}}{(\mathrm{Id} - a)^n} + \lim_a \frac{g - P_{g,n,a}}{(\mathrm{Id} - a)^n} = 0,$$

which proves the first statement.

Second, we note that $P_{f,n,a} P_{g,n,a} \in \Pi_{2n}$, and that

$$\frac{fg - P_{f,n,a} P_{g,n,a}}{(\mathrm{Id} - a)^n} = \frac{fg - P_{f,n,a} g + P_{f,n,a} g - P_{f,n,a} P_{g,n,a}}{(\mathrm{Id} - a)^n}$$
$$= g \frac{f - P_{f,n,a}}{(\mathrm{Id} - a)^n} + P_{f,n,a} \frac{g - P_{g,n,a}}{(\mathrm{Id} - a)^n},$$

whose limit at a is zero. This proves that $P_{f,n,a}P_{g,n,a}$ and fg are equivalent up to order n at a. All that remains it to truncate the product at the n-th power of $(\mathrm{Id}-a)$. ■

Theorem 5.44 [Taylor] Suppose that f is $(n+1)$-times differentiable on the interval $[a,x]$. Then,

$$f(x) = P_{f,n,a}(x) + R_{n,f,a}(x),$$

where the remainder $R_{f,n,a}(x)$ can be represented in the ***Lagrange form****,*

$$R_{f,n,a}(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1},$$

for some $c \in (a,x)$. It can also be represented in the ***Cauchy form****,*

$$R_{f,n,a}(x) = \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x-a),$$

for some $\xi \in (a,x)$.

Comment: For $n=0$ this is simply the mean-value theorem.

Proof: Fix x, and consider the function $\phi : [a,x] \to \mathbb{R}$,

$$\begin{aligned}\phi(z) &= f(x) - P_{f,n,z}(x)\\ &= f(x) - f(z) - f'(z)(x-z) - \frac{f''(z)}{2!}(x-z)^2 - \cdots - \frac{f^{(n)}(z)}{n!}(x-z)^n.\end{aligned}$$

We note that

$$\phi(a) = R_{f,n,a}(x) \qquad \text{and} \qquad \phi(x) = 0.$$

Moreover, ϕ is differentiable, with

$$\begin{aligned}\phi'(z) &= -f'(z) - [f''(z)(x-z) - f'(z)] - \left(\frac{f'''(z)}{2!}(x-z)^2 - f''(z)(x-z)\right)\\ &\quad - \cdots - \left(\frac{f^{(n+1)}(z)}{n!}(x-z)^n - \frac{f^{(n)}(z)}{(n-1)!}(x-z)^{n-1}\right)\\ &= -\frac{f^{(n+1)}(z)}{n!}(x-z)^n.\end{aligned}$$

Let now ψ be any function defined on $[a,x]$, differentiable on (a,x) with non-vanishing derivative. By Cauchy's mean-value theorem, there exists a point $a < \xi < x$, such that

$$\frac{\phi(x) - \phi(a)}{\psi(x) - \psi(a)} = \frac{\phi'(\xi)}{\psi'(\xi)}.$$

That is,

$$R_{f,n,a}(x) = \frac{\psi(x) - \psi(a)}{\psi'(\xi)} \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n$$

Set for example $\psi(z) = (x-z)^p$ for some p. Then,

$$R_{f,n,a}(x) = (x-a)^p \frac{f^{(n+1)}(\xi)}{p\,n!}(x-\xi)^{n-p+1}$$

For $p = 1$ we retrieve the Cauchy form. For $p = n+1$ we retrieve the Lagrange form. ∎

Comment: Later in this course we will study series and ask questions like "does the Taylor polynomial $P_{f,n,a}(x)$ tend to $f(x)$ as $n \to \infty$?". This is the notion of ***converging series***. Here we deal with a different beast: the degree of the polynomial n is fixed and we consider how does $P_{f,n,a}(x)$ approach $f(x)$ as $x \to a$. The fact that $P_{f,n,a}$ approaches f faster than $(x-a)^n$ as $x \to a$ means that $P_{f,n,a}$ is an ***asymptotic series*** of f. —

62h(2017)—

Proof:[Alternative proof of the remainder formula] Note that

$$R_{f,n,a} = f - P_{f,n,a}$$

satisfies,

$$R^{(k)}_{f,n,a}(a) = 0 \quad \text{for all } 0 \le k \le n \qquad \text{and} \qquad R^{(n+1)}_{f,n,a} = f^{(n+1)}.$$

The Lagrange representation is that for every x there exists a mid-point c, such that

$$R_{f,n,a}(x) = \frac{R^{(n+1)}_{f,n,a}(c)}{(n+1)!}(x-a)^{n+1}.$$

Consider the following function defined in a punctured neighborhood of a,

$$\frac{R_{f,n,a}(x)}{(x-a)^{n+1}} = \frac{R_{f,n,a}(x) - R_{f,n,a}(a)}{(x-a)^{n+1} - (a-a)^{n+1}}.$$

By the Cauchy mean value theorem, there exists a mid-point c_1, such that

$$\frac{R_{f,n,a}(x)}{(x-a)^{n+1}} = \frac{R'_{f,n,a}(c_1)}{(n+1)(c_1-a)^n}.$$

We rewrite the right hand side as

$$\frac{R'_{f,n,a}(c_1)}{(n+1)(c_1-a)^n} = \frac{1}{n+1}\frac{R'_{f,n,a}(c_1)-R'_{f,n,a}(a)}{(c_1-a)^n-(a-a)^n}.$$

Applying the Cauchy mean value theorem once again, there exists a mid-point c_2, such that

$$\frac{R_{f,n,a}(x)}{(x-a)^{n+1}} = \frac{1}{n+1}\frac{R''_{f,n,a}(c_2)}{n(c_1-a)^{n-1}}.$$

We proceed n times, until we obtain

$$\frac{R_{f,n,a}(x)}{(x-a)^{n+1}} = \frac{1}{(n+1)!}R^{(n+1)}_{f,n,a}(c_{n+1}).$$

■

5.10 Convex functions

Definition 5.45 Let I be an interval. A function $f : I \to \mathbb{R}$ is called ***convex*** *(קמורה), if for every $a, b \in I$ (assume $a < b$) and every $x \in (a, b)$,*

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

it is called ***strictly convex*** *(קמורה ממש) if the inequality is strong.*

Let's first understand the meaning of this definition. For fixed a, b, define the function $\ell : (a, b) \to \mathbb{R}$,

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

Its graph is a straight line connecting the points $(a, f(a))$ and $(b, f(b))$. Convexity of f means that for every two points in that interval, the graph of

f is under the graph of the secant (מיתר) connecting the graph at those two points.

Note that any $x \in (a, b)$ can be written in the form

$$x = ta + (1 - t)b,$$

for some $t \in (0, 1)$. The inverse relation is

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

Note also that

$$1 - t = 1 - \frac{b - x}{b - a} = \frac{x - a}{b - a}.$$

Thus, f is convex if for every $a, b \in I$ and every $t \in (0, 1)$,

$$f(ta + (1 - t)b) \le tf(a) + (1 - t)f(b). \tag{5.2}$$

Comment: For $t = 1/2$ we get that convexity implies

$$f\left(\frac{a + b}{2}\right) \le \frac{f(a) + f(b)}{2},$$

i.e., "*the function of the mean is less than the mean of the function.*"

Example: Every affine function,

$$f(x) = mx + n,$$

is convex. Indeed, for every $a, b \in \mathbb{R}$ and $t \in (0, 1)$,

$$\begin{aligned} f(ta + (1 - t)b) &= m(ta + (1 - t)b) + n \\ &= t(ma + n) + (1 - t)(mb + n) \\ &= t\, f(a) + (1 - t)\, f(b). \end{aligned}$$

Note that an affine function is not strictly convex. ▲▲▲

Example: The function $f(x) = |x|$ is convex. Indeed, by the triangle inequality,

$$\begin{aligned} |ta + (1 - t)b| &\le |ta| + |(1 - t)b| \\ &= t\,|a| + (1 - t)\,|b| \\ &= t\, f(a) + (1 - t)\, f(b). \end{aligned}$$

▲▲▲

Lemma 5.46 For every function f defined on an interval, and every $x < y < z$

$$\Delta_{f,x}(y) \le \Delta_{f,y}(z)$$

if and only if

$$\Delta_{f,x}(y) \le \Delta_{f,x}(z)$$

if and only if

$$\Delta_{f,x}(z) \le \Delta_{f,y}(z).$$

Proof: It is easy to convince ourselves graphically that this is the case. For a formal proof, note that

$$\begin{aligned}
\Delta_{f,x}(z) &= \frac{f(z) - f(x)}{z - x} \\
&= \frac{f(z) - f(y) + f(y) - f(x)}{z - x} \\
&= \frac{y - x}{y - x}\frac{f(z) - f(y)}{z - x} + \frac{z - y}{z - y}\frac{f(y) - f(x)}{z - x} \\
&= \underbrace{\frac{y - x}{z - x}}_{\equiv t} \Delta_{f,y}(z) + \underbrace{\frac{z - y}{z - x}}_{\equiv 1-t} \Delta_{f,x}(y).
\end{aligned}$$

Thus, $\Delta_{f,x}(z)$ is a weighted average of $\Delta_{f,y}(z)$ and $\Delta_{f,x}(y)$. It is therefore greater than the smallest of the two and greater than the largest of the two.
■

Proposition 5.47 f is convex on I if and only if for every $x < y < z$,

$$\Delta_{f,x}(y) \le \Delta_{f,x}(z) \le \Delta_{f,y}(z).$$

Proof: By the previous lemma,

$$\Delta_{f,x}(y) \le \Delta_{f,x}(z)$$

and

$$\Delta_{f,x}(z) \le \Delta_{f,y}(z)$$

are equivalent statements, hence it suffices to show that $\Delta_{f,x}(y) \le \Delta_{f,x}(z)$. Differently stated, we need to show that for every $x, z \in I$ and every $t \in (0,1)$,

$$\Delta_{f,x}(tx + (1-t)z) \le \Delta_{f,x}(z).$$

Let's write this condition explicitly,

$$\frac{f(tx + (1-t)z) - f(x)}{(tx + (1-t)z) - x} \le \frac{f(z) - f(x)}{z - x},$$

which amounts to

$$\frac{f(tx + (1-t)z) - f(x)}{(1-t)z - (1-t)x} \le \frac{f(z) - f(x)}{z - x},$$

i.e.,

$$f(tx + (1-t)z) - f(x) \le (1-t)(f(z) - f(x)),$$

which further reduces to

$$f(tx + (1-t)z) \le t\, f(x) + (1-t)\, f(z),$$

which is precisely the definition of convexity. ■

Corollary 5.48 If f is convex on an interval I, then it has one-sided derivatives at every point, and for all $a \in I$,

$$f'(a^-) \le f'(a^+).$$

Furthermore, if $a < b$, then

$$f'(a^+) \le f'(b^-).$$

Proof: Proposition 5.47 asserts that $\Delta_{f,a}$ is monotonically increasing, hence it has one-sided derivatives. Moreover,

$$\lim_{a^-} \Delta_{f,a} \le \lim_{a^+} \Delta_{f,a}.$$

As for the second claim, note that for $a < b$ and every $a < x < b$,

$$f'(a^+) \le \Delta_{f,a}(x) \le \Delta_{f,x}(b) = \Delta_{f,b}(x) \le f'(b^-).$$

■

Proposition 5.49 If f has a right-derivative at a then it is right-continuous.

Proof: This is a one-sided version of the differentiable implies continuity property. Indeed, for every $x \ne a$,

$$f(x) = f(a) + \Delta_{f,a}(x)\,(x - a).$$

Hence

$$\lim_{a^+} f = f(a),$$

i.e., f is right-continuous. ■

Corollary 5.50 If f is convex on an interval I then it is continuous on that interval.

Proof: Since f has two one-sided derivatives, it is continuous on both sides, hence continuous. ■

Proposition 5.51 Let I be an open segment. If $f : I \to \mathbb{R}$ is differentiable, then it is convex if and only if f' is increasing. If, in addition, f is convex and twice differentiable, then $f'' \ge 0$.

Proof: Suppose that f is differentiable and convex. We know that for $a < b$,

$$f'(a) = f'(a^+) \le f'(b^-) = f'(b),$$

i.e., f' is monotonically increasing. Conversely, if f is differentiable and f' is increasing, then for every $x < y < z$ there exist $c \in (x, y)$ and $d \in (y, z)$, such that

$$\Delta_{f,x}(y) = f'(c) \qquad \text{and} \qquad \Delta_{f,y}(z) = f'(d).$$

It follows that

$$\Delta_{f,x}(y) \leq \Delta_{f,y}(z),$$

which by Proposition 5.47 implies that f is convex.

Finally, if f is twice differentiable and convex, then the monotonicity of f' implies that $f'' \geq 0$. ■

Convexity is a property relating the value of f at a weighted average of x and y to the weighted average of $f(x)$ and $f(y)$. This property can be generalized:

Definition 5.52 Let $x_1, \dots, x_n$ be a set of numbers. A ***weighted average*** *(ממוצע משוקלל) of those numbers is a* ***convex combination****,*

$$\sum_{i=1}^{n} \lambda_i x_i,$$

where $\lambda_i \leq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$. The λ_i are called ***convex weights****.*

Proposition 5.53 (Jensen's inequality) If f is convex on an interval I, then for every $x_1, \dots, x_n \in I$ and every set of convex weights $\lambda_1, \dots, \lambda_n$,

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i).$$

Proof: This is proved by induction on n. For $n = 1$ this is obvious as the only convex weight is $\lambda_1 = 1$. Suppose this holds for $n - 1$, then

$$\begin{aligned}
f\left(\sum_{i=1}^{n} \lambda_i x_i\right) &= f\left(\sum_{i=1}^{n-1} \lambda_i x_i + \lambda_n x_n\right) \\
&= f\left((1-\lambda_n)\frac{\sum_{i=1}^{n-1} \lambda_i x_i}{1-\lambda_n} + \lambda_n x_n\right) \\
&\leq (1-\lambda_n) f\left(\sum_{i=1}^{n-1} \frac{\lambda_i}{1-\lambda_n} x_i\right) + \lambda_n f(x_n) \\
&\leq (1-\lambda_n) \sum_{i=1}^{n-1} \frac{\lambda_i}{1-\lambda_n} f(x_i) + \lambda_n f(x_n) \\
&= \sum_{i=1}^{n} \lambda_i f(x_i),
\end{aligned}$$

where we used the fact that $\sum_{i=1}^{n-1} \lambda_i = 1 - \lambda_n$. ■

Example: In particular, for $\lambda_i = 1/n$,

$$f\left(\frac{x_1 + \cdots + x_n}{n}\right) \le \frac{f(x_1) + \cdots + f(x_n)}{n}.$$

▲▲▲

Example: Consider the function

$$-\log : (0, \infty) \to \mathbb{R}.$$

It it differentiable. Since

$$(-\log)'(x) = -\frac{1}{x}$$

is increasing, it follows that $(-\log)$ is convex. It follows that

$$-\log\left(\frac{x_1 + \cdots + x_n}{n}\right) \le -\frac{\log(x_1) + \cdots + \log(x_n)}{n}.$$

We can rewrite this as

$$\log\left(\frac{x_1 + \cdots + x_n}{n}\right) \ge \log(x_1 \dots x_n)^{1/n}.$$

Since the logarithm is monotonically increasing, we recover the ***inequality of arithmetic and geometric means***,

$$\left(\frac{x_1 + \cdots + x_n}{n}\right) \ge (x_1 \dots x_n)^{1/n}.$$

▲▲▲