8 Lectures 15–16

- In mechanics on a geometric line, one studies the position coordinate x as a real-valued function of a time parameter t. The velocity is the "rate-of-change of the position with respect to time". This is actually a tricky concept which took thousands of years to develop.

  At a given time $t_0 \in \mathbb{R}$, we can consider a small finite change Δt in time, then the corresponding change in position is

  $$\Delta x = x(t_0 + \Delta t) - x(t_0).$$

  The ratio $\frac{\Delta x}{\Delta t}$ is the *average* velocity over the time interval $[t_0, t_0 + \Delta t]$. This average velocity is a concrete concept involving quantities that are routinely measured by scientific instruments (rulers, clocks, and their modern versions). Generally, it depends on Δt as well as t_0.

  In general, motion is not uniform, in the sense that its average velocities vary with t_0 (and Δt). In the 14th century, a group of scholars, called the Oxford Calculators, considered a concept of *instantaneous* velocity at a single time t_0. They imagined that at time t_0, the non-uniformity of the motion $x \equiv x(t)$ could be "switched off", so that only a uniform motion $x_{\mathrm{u}} \equiv x_{\mathrm{u}}(t)$ remains. Then the instantaneous velocity at t_0 is considered to be the average velocity for x_{u} (which is independent of Δt).

  In the 17th century, Galileo made the idea of instantaneous velocity at t_0 empirically accessible. He considered a ball rolling down an inclined plane, so its speed (|velocity|) increases. When it arrives at the floor, the motion becomes horizontal, and it is *uniform* according to his famous inertia principle. The constant velocity of this uniform horizontal motion gives the instantaneous velocity at time that the ball reaches the floor.

- Newton and Leibniz formalized the idea of instantaneous velocity at *any* time t_0 of a motion. Semi-rigorously, they considered the limit of the ratio $\frac{\Delta x}{\Delta t}$ as Δt is taken to 0. Because Δx also goes to 0 as Δt goes to 0, they had to make sense of the apparent ill-definedness of $\frac{0}{0}$. In today's language, they introduced the *derivative* concept, with "velocity = derivative of position" being the basic example. The only thing that was lacking in their definition is the rigorous concept of limit, which we now have.

Definition 8. Let x be a real-valued function defined on a neighbourhood of $t_0 \in \mathbb{R}$. We say that x is *differentiable at* t_0, if

$$\lim_{t \to t_0} \frac{x(t) - x(t_0)}{t - t_0} \tag{8.1}$$

exists, in which case, (8.1) is the *derivative* of x at t_0.

If x is defined on $[t_0, t_0 + \delta_0)$ for some $\delta_0 > 0$, then we say that it is *right-differentiable* at t_0, if

$$\lim_{t \to t_0^+} \frac{x(t) - x(t_0)}{t - t_0} \tag{8.2}$$

exists, in which case, (8.2) is the *right-sided derivative* of x at t_0; similarly for the left-sided derivative.

Comments:

- In (8.1) the function $t \mapsto \frac{x(t)-x(t_0)}{t-t_0}$ is regarded as being defined on a deleted neighbourhood of t_0.
- By convention, if x is defined on an interval $[a, b]$, differentiability at a or b means one-sided differentiability. At an interior point, differentiability is equivalent to existence and equality of the left-sided and right-sided derivatives (Recall Exercise 5.3).
- By changing variables to $\Delta t = t - t_0$, an equivalent version of (8.1) is
$$\lim_{\Delta t \to 0} \frac{x(t_0 + \Delta t) - x(t_0)}{\Delta t} \equiv \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}\Big|_{t_0}.$$
- The derivative (8.1) is often denoted by
$$\frac{dx}{dt}\Big|_{t_0}, \qquad \frac{dx}{dt}(t_0), \qquad x'(t_0), \qquad \dot{x}(t_0).$$
- In modern Calculus textbooks, one typically considers the derivative of a function $y \equiv y(x)$ or $f \equiv f(x)$, so the typical derivative would be $\frac{dy}{dx}\Big|_{x=x_0}$ or $f'(x_0)$. We deliberately used $x \equiv x(t)$ in Definition 8 to emphasize the historical and geometrical-mechanical aspect of the derivative.

You should get used to all these different notations. In differential geometry and modern physics, the concept of derivative and velocity is studied in much greater depth, and there, one studies the velocities of parametrized curves $x \equiv x(t)$ in general geometric spaces.

– Suppose $\frac{dx}{dt}\Big|_{t_0} = v$ exists. So, given $\epsilon > 0$, there exists $\delta > 0$ such that

$$0 < |t - t_0| < \delta \Rightarrow \left| \frac{x(t) - x(t_0)}{t - t_0} - v \right| < \epsilon$$
$$\text{i.e.,} \qquad |t - t_0| < \delta \Rightarrow \big|x(t) - \big(x(t_0) + v(t - t_0)\big)\big| < \epsilon |t - t_0|.$$

Note that $x_{\mathrm{u}} = x(t_0) + v(t - t_0)$ is precisely the uniform motion with velocity v, starting from position $x(t_0)$ at time t_0. So x is linearly approximated by x_{u}, with controlled error rate. Furthermore, uniform motion with any other velocity $\tilde{v} \neq v$ will be a worse approximation to x than x_{u} is.

- It is straightforward to see that differentiability of a general abstract function $f \equiv f(x)$ at x_0 implies continuity at x_0, since

$$\begin{aligned} \lim_{x \to x_0} f(x) &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}(x - x_0) + f(x_0) \\ &= f'(x_0)(x - x_0) \cdot \lim_{x \to x_0} (x - x_0) + f(x_0) \\ &= f(x_0). \end{aligned} \tag{8.3}$$

The converse is sometimes false. For example, the function

$$x \mapsto |x| = \begin{cases} x, & x \geq 0, \\ -x, & x \leq 0, \end{cases}$$

is continuous at 0, but not differentiable at 0 (Exercise).

Exercise 8.1. Consider the functions

$$f_1 : x \mapsto \begin{cases} x, & x \in \mathbb{Q} \\ -x, & x \in \mathbb{R} \setminus \mathbb{Q}, \end{cases} \qquad f_2 : x \mapsto \begin{cases} x^2, & x \in \mathbb{Q} \\ -x^2, & x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Find their points of differentiability (if any). You may need (8.4).

- Unsurprisingly, linear functions $x \mapsto ax + b$ are differentiable at every $x \in \mathbb{R}$, with constant derivative a. (The $a = 0$ case is the constant function $x \mapsto b$.)

  By an abuse of notation, for positive integers $n \geq 1$, we usually write x^n for the monomial function $x \mapsto x^n$. Its derivative at $x_0 \in \mathbb{R}$ can be computed with the telescopic trick,

$$\begin{aligned}\frac{d(x^n)}{dx}\Big|_{x_0} = \lim_{x\to x_0} \frac{x^n - x_0^n}{x - x_0} &= \lim_{x\to x_0} \frac{(x - x_0)(x^{n-1} + x^{n-2}x_0 + \dots x_0^{n-1})}{x - x_0} \\ &= \sum_{j=0}^{n-1} \lim_{x\to x_0} x^{n-1-j}x_0^j \\ &= nx_0^{n-1}\end{aligned} \tag{8.4}$$

  For integer degree $n \leq -1$, a similar argument (exercise) leads to the same formula

$$\frac{d(x^n)}{dx}\Big|_{x_0} = nx_0^{n-1}, \qquad x_0 \in \mathbb{R} \setminus \{0\}.$$

  The limits calculated in Exercise 7.6 can be used to calculate the limits of the power functions and the exponential functions.

  - For general exponent $\alpha \in \mathbb{R}$, and $x_0 > 0$,

$$\begin{aligned}\frac{d(x^\alpha)}{dx}\Big|_{x_0} &= \lim_{\Delta x\to 0} \frac{(x_0 + \Delta x)^\alpha - x_0^\alpha}{\Delta x} \\ &= \lim_{\Delta x\to 0} \frac{\left(1 + \frac{\Delta x}{x_0}\right)^\alpha - 1}{\frac{\Delta x}{x_0}} \frac{x_0^\alpha}{x_0} \\ &= \lim_{h\to 0} \frac{(1+h)^\alpha - 1}{h} x_0^{\alpha-1} \qquad \text{(change variables)} \\ &= \alpha x_0^{\alpha-1}.\end{aligned} \tag{8.5}$$

  - For $x_0 \in \mathbb{R}$,

$$\begin{aligned}\frac{d(e^x)}{dx}\Big|_{x_0} &= \lim_{\Delta x\to 0} \frac{e^{x_0+\Delta x} - e^{x_0}}{\Delta x} \\ &= e^{x_0} \lim_{\Delta x\to 0} \frac{e^{\Delta x} - 1}{\Delta x} \\ &= e^{x_0}.\end{aligned} \tag{8.6}$$

Remark. Some authors define e by first defining the exponential function $x \mapsto e^x$ to be the (unique) solution to the differential equation $\frac{d(e^x)}{dx}\Big|_{x_0} = e^{x_0}, x_0 \in \mathbb{R}$ with $e^0 = 1$.

– For base $a > 0$, we have $a = e^{\ln a}$. So for $x \in \mathbb{R}$,

$$a^x = e^{x \ln a},$$

thus by the chain rule and (8.6),

$$\frac{d(a^x)}{dx}\Big|_{x_0} = e^{x_0 \ln a} \cdot \frac{d(x \ln a)}{dx}\Big|_{x_0} = a^{x_0} \ln a. \tag{8.7}$$

- Suppose f, g are real-valued functions which are differentiable at x_0. The, for any $\lambda, \mu \in \mathbb{R}$, the algebraic properties of limits immediately give

$$(\lambda f + \mu g)'(x_0) = \lambda f'(x_0) + \mu g'(x_0).$$

By induction, this linearity property holds for general linear combinations, $\sum_{k=1}^{n} \lambda_n f_n$, of functions differentiable at x_0. In particular, polynomial functions are differentiable everywhere.

There are also differentiation rules for the pointwise product and pointwise quotient of functions.

Exercise 8.2. Let f, g be real-valued functions which are differentiable at x_0. Show that

– The *Leibniz rule*

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0). \tag{8.8}$$

holds.

– Suppose $g(x_0) \neq 0$. Use the Leibniz rule to deduce that

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{(g(x_0))^2}.$$

Recall the *hyperbolic trigonometric* functions,

$$\cosh x = \frac{e^x + e^{-x}}{2}, \qquad \sinh x = \frac{e^x - e^{-x}}{2}.$$

By the quotient rule (or changing variables from x to $-x$), it is straightforward to see that $\frac{d(e^{-x})}{dx}\Big|_{x_0} = -e^{-x_0}$, thus

$$\cosh'(x_0) = \sinh(x_0), \qquad \sinh'(x_0) = \cosh(x_0), \qquad x_0 \in \mathbb{R}. \tag{8.9}$$

- Besides the Leibniz rule (8.8), another key property of differentiation is the *chain rule*.

  Proposition 8.1 (Chain rule)**.** *Let f, g be real-valued functions which are differentiable at x_0 and at $f(x_0)$ respectively. Then $g \circ f$ is differentiable at x_0, with*

  $$(g \circ f)'(x_0) = g'(f(x_0)) \cdot f'(x_0).$$

  Proof. Note that f is continuous at x_0 and g is continuous at $f(x_0)$, and as mentioned in Prop. 7.1, $g \circ f$ is defined on some neighbourhood of x_0, and we can consider whether its derivative exists there.

  For x in a deleted neighbourhood of x_0, we can write

  The equation

  $$\frac{g \circ f(x) - g \circ f(x_0)}{x - x_0} = \frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} \cdot \frac{f(x) - f(x_0)}{x - x_0}, \tag{8.10}$$

  holds for all x in some deleted neighbourhood of x_0, *except* for those x such that $f(x) = f(x_0)$. To extend the identity (8.10) to the $f(x) = f(x_0)$ situation, we define the function

  $$\psi(y) = \begin{cases} \frac{g(y) - g(f(x_0))}{y - f(x_0)}, & y \neq f(x_0), \\ g'(f(x_0)), & y = f(x_0). \end{cases}$$

  Then

  $$\frac{g \circ f(x) - g \circ f(x_0)}{x - x_0} = \psi(f(x)) \cdot \frac{f(x) - f(x_0)}{x - x_0} \tag{8.11}$$

  holds for *all* x in some deleted neighbourhood of x_0, because it is the same as (8.10) if $f(x) = f(x_0)$, otherwise it is "$0 = 0$" if $f(x) = f(x_0)$. Therefore, it makes sense to take the $x \to x_0$ limit of (8.11).

  Note that ψ is continuous at $f(x_0)$, because

  $$\lim_{y \to f(x_0)} \psi(y) = \lim_{y \to f(x_0)} \frac{g(y) - g(f(x_0))}{y - f(x_0)} \equiv g'(f(x_0)) = \psi(f(x_0)).$$

So the composition $\psi \circ f$ is continuous at x_0 (Prop. 7.1). Thus, the $x \to x_0$ limit of the identity (8.11) becomes

$$\begin{aligned}(g \circ f)'(x_0) = \lim_{x \to x_0} \frac{g \circ f(x) - g \circ f(x_0)}{x - x_0} &= \lim_{x \to x_0} \psi(f(x)) \cdot \frac{f(x) - f(x_0)}{x - x_0} \\ &= \lim_{x \to x_0} \psi(f(x)) \cdot \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \\ &= \psi(f(x_0)) \cdot f'(x_0) \\ &= g'(f(x_0)) \cdot f'(x_0).\end{aligned}$$

□

Exercise 8.3. Let f be an injective real-valued function on some neighbourhood W of x_0, which is differentiable at x_0 with $f'(x_0) \neq 0$.

– The function $f \circ f^{-1}$ is the identity function $y \mapsto y$ on $f(W)$. So, apparently, $(f \circ f^{-1})'(y) = 1, y \in f(W)$. Write $y_0 = f(x_0)$. By the chain rule, we should have

$$1 = (f \circ f^{-1})'(y_0) = f'(f^{-1}(y_0)) \cdot (f^{-1})'(y_0) = f'(x_0) \cdot (f^{-1})'(y_0).$$

Thus, since $f'(x_0) \neq 0$, we have the formula

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}. \tag{8.12}$$

Some problems with the above argument are:

∗ The assumptions of injectivity and differentiability at x_0, do not guarantee that f^{-1} is differentiable at $y_0 = f(x_0)$. The classic example is:

$$\begin{aligned} f(x) &= x^3, && x \in \mathbb{R}, \\ f^{-1}(y) &= \begin{cases} y^{1/3}, & y \geq 0, \\ -(-y)^{1/3}, & y \leq 0, \end{cases} && y \in \mathbb{R}. \end{aligned}$$

Check that formula (8.12) holds for nonzero y_0, but f^{-1} is not differentiable at $y_0 = f(0) = 0$. (What happens there?)

∗ In the $f(x) = x^3$ example, we have $f'(0) = 0$. There are examples of injective f which have $f'(0) \neq 0$, but f^{-1} is not differentiable (or even continuous) at $f(0)$. (See Exercises.)

– Now make the extra assumption that f^{-1} is continuous at $f(x_0)$. Show that f^{-1} is now differentiable at $f(x_0)$, and that the formula (8.12) holds.

Example 8.1. We saw that the exponential function $x \mapsto e^x$ is strictly monotone on the interval $\mathbb{R}$ with range $(0, \infty)$. By Theorem 7.4 (or Corollary 7.6), its inverse function, the natural logarithm ln, is continuous on $(0, \infty)$. We also found the derivative of e^x to be e^x itself (Eq. (8.6)), which is always non-zero. By Exercise 8.3, we conclude that ln is differentiable at every $y_0 \in (0, \infty)$, with derivative

$$\ln'(y_0) = \frac{1}{e^{\ln(x_0)}} = \frac{1}{x_0}, \qquad x_0 \in (0, \infty). \tag{8.13}$$

More generally, for logarithms with general base $a > 0, a \neq 1$, we have

$$\ln x = \ln(a^{\log_a x}) = (\log_a x) \ln a.$$

So, for all $x_0 \in \mathbb{R}$

$$\frac{d(\log_a(x))}{dx}\Big|_{x_0} = (\ln a)^{-1} \ln'(x_0) = \frac{1}{x_0 \ln a}.$$

- Let f be a *positive* function, then $\ln \circ f$ is defined. If f is differentiable at x_0, then the chain rule and (8.13) give

$$(\ln \circ f)'(x_0) = \frac{f'(x_0)}{f(x_0)}. \tag{8.14}$$

Eq. (8.14) is sometimes called the *logarithmic derivative* of f at x_0.

The *logarithmic derivative* helps to simplify derivative calculations involving (products of) exponentials.

Example 8.2. Consider

$$\begin{aligned} f(x) &= x^x, && x \in (0, \infty), \\ g(x) &= (\ln x)^{\ln x}, && x > 1, \\ h(x) &= \prod_{j=1}^{n} (p_j(x))^{\alpha_j}, && \text{positive functions } p_j, \alpha_j \in \mathbb{R}, \text{ suitable domain.} \end{aligned}$$

We have, by Leibniz product rule,

$$(\ln \circ f)'(x_0) = (x \ln x)'(x_0) = x_0 \cdot \frac{1}{x_0} + \ln x_0 = 1 + \ln x_0.$$

Put this calculation into (8.14) to get

$$f'(x_0) = f(x_0) \cdot (\ln \circ f)'(x_0) = x_0^{x_0}(1 + \ln x_0).$$

Similarly (Exercise), we can compute

$$g'(x_0) = \frac{1}{x_0}(\ln x_0)^{\ln x_0}(1 + \ln(\ln x_0)).$$

If the positive functions p_j are differentiable at x_0, then

$$\ln \circ h = \sum_{j=1}^{n} \alpha_j (\ln \circ p_j)$$

is differentiable at x_0. The logarithmic derivative $(\ln \circ h)'(x_0)$ is just a linear combination of the logarithmic derivatives $(\ln \circ p_j)'(x_0)$, so it is easier than a direct computation of $h'(x_0)$.

- Some of you were concerned that we have been lacking some "practice examples" for computing limits. In semi-rigorous calculus textbooks, it is common to study the non-algebraic functions $e^x, \ln(x), \sin(x), \cos(x)$, and provide non-rigourous presentations of their definitions and properties. In a rigorous analysis course, it is more important to understand the *reasons* why such functions are sensible and well-behaved, before playing with limits/derivative of various combinations of them. To establish these properties, we need *theory* (guided by geometric intuition/physical application), then we can reliably carry out *computation*.

  For example, the derivatives of $\sin, \cos$, and the Pythagoras identity

  $$\sin^2(x) + \cos^2(x) = 1,$$

  are not so easy to rigorously establish, compared to their hyperbolic versions (8.9). We can "cheat" and use a geometric "definition" of $\sin, \cos$, but this needs some calculus *theory* to make precise. At the moment, we have the power series definition of $\sin, \cos$, and to obtain their derivatives, we need two further results: the derivative of certain power series, and the Mean Value Theorem.

Proposition 8.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be given by a power series,*

$$f(x) = \sum_{k=0}^{\infty} a_k x^k, \qquad x \in \mathbb{R}, a_k \in \mathbb{R},$$

assumed to be absolutely convergent, i.e., $\sum_{k=0}^{\infty} |a_k||x^k|$ exists. Then f is differentiable at every $x \in \mathbb{R}$, with derivative being the power series of term-by-term derivatives,

$$f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}, \qquad x \in \mathbb{R}. \tag{8.15}$$

Proof. [Not examinable[17]] In the first place, we should check that the power series

$$g(y) = \sum_{k=1}^{\infty} k a_k y^{k-1} \tag{8.16}$$

converges.

– By assumption, $\sum_{k=0}^{\infty} |a_k||3y|^k$ converges[18] for every $y \in \mathbb{R}$. So $|a_k||3y|^k$, thus also $|a_k||3y|^{k-1}$ converges to 0 as $k \to \infty$ (why?). So boundedness holds: for some $M \equiv M(y) > 0$, we have

$$|a_k||3y|^{k-1} < M, \qquad \forall k \geq 1.$$

– The k-th term of (8.16) satisfies the inequality

$$|k a_k y^{k-1}| = k \underbrace{|a_k||3y|^{k-1}}_{<M} (\tfrac{1}{3})^{k-1} < 3M \cdot \frac{k}{3^k}, \qquad k \geq 1.$$

The "larger" series $\sum_{k=1}^{\infty} \frac{k}{3^k}$ converges because:

∗ The ratio of its consecutive terms is $\frac{k+1}{3k}$, which is less than $\frac{1}{2}$ for all $k > 2$. So for some $\lambda > 0$, we have

$$\frac{k}{3^k} < \lambda \cdot \frac{1}{2^k}, \qquad \forall k > 2.$$

Since $\sum_{k=1}^{\infty} \frac{1}{2^k}$ converges, $\sum_{k=1}^{\infty} \frac{k}{3^k}$ also converges.

[17] You will encounter similar power series analysis in Analysis II.

[18] The choice of 3 is for convenience, any $r > 1$ could also be used.

Thus, $\sum_{k=1}^{\infty} |ka_k y^{k-1}| < \infty$, i.e., (8.16) is absolutely convergent, for every $x \in \mathbb{R}$.

For any distinct $x_0, x \in \mathbb{R}$, the power series for $f(x)$ and $f(x_0)$ are both convergent, so we may write

$$\begin{aligned} f(x) - f(x_0) &= \sum_{k=1}^{\infty} a_k (x^k - x_0^k) \\ &= \sum_{k=1}^{\infty} a_k (x - x_0)\big(x^{k-1} + x^{k-2}x_0 + \ldots + x_0^{k-1}\big), \end{aligned}$$

using the "telescopic sum trick". Then

$$\frac{f(x) - f(x_0)}{x - x_0} = \sum_{k=1}^{\infty} a_k \sum_{j=0}^{k-1} x^{k-1-j} x_0^j, \qquad \forall x \neq x_0. \tag{8.17}$$

Therefore,

$$\begin{aligned} \frac{f(x) - f(x_0)}{x - x_0} - g(x_0) &= \sum_{k=1}^{\infty} a_k \left(\Big(\sum_{j=0}^{k-1} x^{k-1-j} x_0^j \Big) - k x_0^{k-1} \right) \\ &= \sum_{k=2}^{\infty} a_k \sum_{j=0}^{k-1} x_0^j \big(x^{k-1-j} - x_0^{k-1-j}\big) \qquad (k = 1 \text{ term vanishes}) \\ &= \sum_{k=2}^{\infty} a_k \sum_{j=0}^{k-1} x_0^j (x - x_0)\big(x^{k-2-j} + x^{k-3-j}x_0 + \ldots + x_0^{k-2-j}\big). \end{aligned}$$

Now restrict x to lie within $(x_0 - 1, x_0 + 1)$. So $|x_0| < |x_0| + 1$ and $|x| < |x_0| + 1$, as well as $|x - x_0| < 1$. Then, by triangle inequality,

$$\begin{aligned} \left| \frac{f(x) - f(x_0)}{x - x_0} - g(x_0) \right| &< \sum_{k=2}^{\infty} |a_k| \sum_{j=0}^{k-1} (|x_0| + 1)^j |x - x_0| (k - 1 - j)(|x_0| + 1)^{k-2-j} \\ &< \frac{|x - x_0|}{2} \sum_{k=2}^{\infty} |a_k| (|x_0| + 1)^{k-2} k(k-1). \end{aligned} \tag{8.18}$$

The infinite sum on the right side of (8.18) converges, because:

– The power series

$$h(y) = \sum_{k=2}^{\infty} k(k-1)a_k y^{k-2}$$

is obtained by differentiating the everywhere absolutely convergent series (8.16) term-by-term. So $h(y)$ is also absolutely convergent, for every $y \in \mathbb{R}$.

So the inequality (8.18) reduces to

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - g(x_0) \right| < |x - x_0| C, \qquad 0 < |x - x_0| < 1,$$

where $C \equiv C(x_0)$ is some real constant.

Finally, by the squeezing principle,

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = g(x_0) = \sum_{k=1}^{\infty} k a_k x_0^{k-1}, \qquad \forall x_0 \in \mathbb{R},$$

which is formula (8.15).

□

Example 8.3. Applying Prop. 8.2 to the everywhere-convergent power series for the functions $\cos, \sin$, we obtain

$$\sin'(x) = \cos(x), \qquad \cos'(x) = -\sin(x), \qquad x \in \mathbb{R}. \tag{8.19}$$

- As mentioned, we also need a "Mean Value Theorem" to establish the Pythagoras formula for $\sin, \cos$. Let us first provide some preparatory definitions.

  For a real-valued function $f : E \to \mathbb{R}$, we say that $x_0 \in E$ is a *global maximum* of f if $f(x_0) \geq f(x)$ for all $x \in E$. A *global minumum* is similarly defined.

  We say that x_0 is a *local maximum*, if there exists $\delta > 0$ such that

  $$\forall x \in E, \quad |x - x_0| < \delta \Rightarrow f(x) \leq f(x_0).$$

  Similarly, x_0 is a *local minimum* at $x_0 \in E$, if there exists $\delta > 0$ such that

  $$\forall x \in E, \quad |x - x_0| < \delta \Rightarrow f(x) \geq f(x_0).$$

Lemma 8.3 (Fermat's stationary point theorem)**.** *Let f be a real-valued function which is differentiable*[19] *at $x_0 \in \mathbb{R}$. If x_0 is a local minimum or local maximum of f, then $f'(x_0) = 0$.*

Proof. Exercise. Note that f has left-sided and right-sided derivatives at x_0, both equal to $f'(x_0)$. □

Theorem 8.4 (Rolle's theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists $x \in (a, b)$ such that $f'(x) = 0$.*

Proof. If f is constant on $[a, b]$, then $f'(x) = 0$ holds for any $x \in (a, b)$.

So we shall assume that f is non-constant. We may further assume that $f(a)$ is not the maximum value of f on $[a, b]$. (Otherwise, $f(a)$ is not the minimum, and a similar argument as the one below will apply.) By the Extreme Value Theorem 6.5, and non-maximality of $f(a) = f(b)$, it attains its maximum value at some interior point $x_0 \in (a, b)$. This x_0 is also a local maximum, and f is differentiable at x_0. So $f'(x_0) = 0$ by Lemma 8.3. □

Theorem 8.5 (Lagrange's Mean Value Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). There exists $x \in (a, b)$ such that*

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

Proof. By choosing a suitable constant c, we can arrange for the function $g(x) = f(x) - cx$ to satisfy $g(a) = g(b)$. Specifically, take $c = \frac{f(b)-f(a)}{b-a}$. This g satisfies the conditions of Rolle's Theorem 8.4, so there exists $x \in (a, b)$ such that

$$0 = g'(x) = f'(x) - c.$$

So $f'(x) = c$ as claimed. □

Remark. The MVT for the position function $x \equiv x(t)$ has the following interpretation. Over a time interval $[t_0, t_1]$, the average velocity is $c = \frac{x(t_1)-x(t_0)}{t_1-t_0}$. Subtracting the uniform motion $t \mapsto ct$ from $x(t)$, we are left

[19] Recall that this requires f to be defined on some neighbourhood of x_0.

with the non-uniform part, which has the same start and end points, $x(t_0) = x(t_1)$. Then Rolle's theorem is applied to this non-uniform part of the motion.

We conclude that no matter how the instantaneous velocities $v(t) = x'(t)$ vary with $t \in [t_0, t_1]$, there will be some time at which the instantaneous velocity equals the average velocity.

Exercise 8.4. Prove the order completeness property of $\mathbb{R}$ (Theorem 3.8) from the MVT.

Hint: For a non-empty set A which has an upper bound, consider the function

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} 1, & x \text{ is upper bound for } A, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Argue that if A is no least upper bound, then $f'(x) = 0$ holds for every $x \in \mathbb{R}$, and derive a contradiction.

Remark. Rolle's Theorem uses the EVT for intervals, which can be proved directly from the completeness property. Then MVT is proved from Rolle's Theorem, while Exercise 8.4 proves the completeness property from MVT. So Rolle's Theorem and MVT are equivalent to the completeness property.

Corollary 8.6. *If f is a real-valued function such that $f'(x) = 0$ holds for all x in an open interval I, then f is constant on I.*

Proof. Note that f is differentiable, thus continuous on I (Eq. (8.3)). Let a, b be any two distinct points in I; we may assume that $a < b$. Since $[a, b] \subseteq I$, we know that f is continuous on $[a, b]$ and differentiable on (a, b). By MVT, there exists $x \in (a, b)$ such that $f'(x) = \frac{f(b)-f(a)}{b-a}$. By assumption, $f'(x) = 0$, thus $f(b) = f(a)$. As $a, b \in I$ were arbitrary, this shows that f has a constant value on I. □

Here is another similar consequence of the MVT.

Exercise 8.5. Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Show that if $f'(x) > 0$ for all $x \in (a, b)$, then f is strictly monotone increasing on $[a, b]$.