6. The mean-value theorem and applications

The mean-value theorem is one of the most important theorems of analysis. It is the key to deducing information about a function from information about its derivative. The statement is as follows.

6.1 THEOREM (the mean-value theorem). If \( f \) is differentiable on \([a, b]\), then there is a point \( x_0 \) in \((a, b)\) such that

\[
f(b) - f(a) = (b - a)f'(x_0).\]

\[\Box\]

In other words,

\[
f'(x_0) = \frac{f(b) - f(a)}{b - a}.
\]

This is the slope of the chord from \([a, f(a)]\) to \([b, f(b)]\), and the theorem says there is a point in \((a, b)\) where the gradient is the same as the slope of this chord.

The statement seems highly convincing from the picture, but of course it must be proved properly. This can be done quite easily using our earlier theorems (in particular, 5.9). We start with the special case where \(f(b) = f(a)\).

6.2 (Rolle’s theorem). Suppose that \( f \) is differentiable on \([a, b]\) and \( f(a) = f(b) \). Then there is a point \( x_0 \) in \((a, b)\) such that \( f'(x_0) = 0 \).

Proof. By 5.9, \( f \) is bounded and attains its bounds at certain points of \([a, b]\). As the diagram suggests, these are the points to consider. Let \( M \) and the supremum and infimum of \( \{ f(x) : a \leq x \leq b \} \).

If \( M = m \), then \( f \) is constant on \([a, b]\), so \( f'(x) = 0 \) for all \( x \) in \((a, b)\). So assume that \( M > m \). Then either \( M > f(a) \) or \( m < f(a) \) (or both). If \( M > f(a) \), then there exists \( x_0 \) in \([a, b]\) such that \( f(x_0) = M \), and \( x_0 \) is not \( a \) or \( b \), so in fact \( x_0 \in (a, b) \). Hence \( f \) has a local maximum at \( x_0 \), so by 4.7, \( f'(x_0) = 0 \). If, instead, \( m < f(a) \), we consider instead a point where \( f(x_0) = m \).

We now derive the mean-value theorem itself by a simple and elegant argument.

Proof of 6.1. Let \( g(x) = f(x) - kx \), with \( k \) chosen so that \( g(a) = g(b) \). In other
words, \( f(a) - ka = f(b) - kb \), so

\[
k = \frac{f(b) - f(a)}{b - a}.
\]

Then \( g \) is differentiable on \([a, b]\), with \( g'(x) = f'(x) - k \). By Rolle’s theorem, there is a point \( x_0 \) in \((a, b)\) such that \( g'(x_0) = 0 \), so that

\[\text{Notes.} \]

(1) The result may fail if \( f \) fails to be differentiable, even at one point. This is shown by \( f(x) = |x| \), for which \( f(-1) = f(1) = 1 \), but there is no point where \( f'(x) = 0 \).

(2) As the proof shows, it is actually sufficient if \( f \) is differentiable on \((a, b)\) and right-continuous at \( a \), left-continuous at \( b \).

(3) Note that the point \( x_0 \) is in \((a, b)\); in some applications, it is important that it is not an end-point.

6.3 COROLLARY. Suppose that \( f \) is differentiable on \([a, b]\). Then:

(i) if \( f'(x) \geq 0 \) for all \( x \) in \((a, b)\), then \( f \) is increasing on \([a, b]\);

(ii) if \( f'(x) > 0 \) for all \( x \) in \((a, b)\), then \( f \) is strictly increasing on \([a, b]\);

(iii) similarly for decreasing or strictly decreasing;

(iv) if \( f'(x) = 0 \) for all \( x \) in \((a, b)\), then \( f \) is constant on \([a, b]\).

(v) if \( |f'(x)| \leq M \) for all \( x \) in \([a, b]\), then \( |f(x_2) - f(x_1)| \leq M|x_2 - x_1| \) for all \( x_1, x_2 \) in \([a, b]\).

Proof. Choose points \( x_1, x_2 \) such that \( a \leq x_1 < x_2 \leq b \). By the mean-value theorem (applied with end points \( x_1 \) and \( x_2 \)), there exists \( x_0 \) in \((x_1, x_2)\) such that

\[
f(x_2) - f(x_1) = (x_2 - x_1)f'(x_0).
\]

In case (i), we have \( f'(x_0) \geq 0 \), hence \( f(x_2) \geq f(x_1) \). In case (ii), \( f'(x_0) > 0 \), so \( f(x_2) > f(x_1) \). Similarly with \( \leq \) or \(< \) in case (iii).

In case (iv), \( f(x_0) = 0 \), so \( f(x_2) = f(x_1) \): this holds for all \( x_1 \) and \( x_2 \) in \([a, b]\), so \( f \) is constant on \([a, b]\). Case (v) is immediate.

Note the following obvious extension of (iv): if \( f'(x) = g'(x) \) on \([a, b]\), then there is a constant \( c \) such that \( f(x) = g(x) + c \) on \([a, b]\) (just apply (iv) to \( f - g \)).

The mean-value theorem has numerous applications. We shall describe some of them.
Applications to standard functions

We show how the mean-value theorem (mostly in the form of 6.3(iv)) can be used to establish some of the basic properties of the exponential and trigonometric functions. Recall that we are working from the series definitions of these functions.

6.4. For any $a$ and $b$, we have $e^{a+b} = e^a e^b$.

Proof. Fix $a$ and $b$, and let $f(x) = e^{a+x} e^{b-x}$. Then

$$f'(x) = e^{a+x} e^{b-x} + e^{a+x} (-e^{b-x}) = 0$$

for all $x$. Hence $f(x)$ is constant. In particular, $f(0) = f(b)$. But

$$f(0) = e^a e^b, \quad f(b) = e^{a+b} e^0 = e^{a+b},$$

so the statement follows. \square

(Compare the earlier proof by multiplication of series.)

Of course, it follows that $e^x e^{-x} = 1$, so that $e^{-x} = 1/e^x$ and $e^x > 0$ for all $x$. Also, $(e^x)^n = e^{nx}$ for integers $n$.

It also follows that $\log xy = \log x + \log y$, $\log(1/x) = -\log x$ and $\log x^n = n \log x$ for integers $n$ and $x, y > 0$.

6.5. If $f$ is a function such that $f'(x) = af(x)$ for all $x$, then there is a constant $c$ such that $f(x) = ce^{ax}$.

Proof.

6.6. For any $a, b$,

$$\cos(a + b) = \cos a \cos b - \sin a \sin b,$$

$$\sin(a + b) = \sin a \cos b + \cos a \sin b.$$ 

In particular,

$$\cos 2a = \cos^2 a - \sin^2 a = 2 \cos^2 a - 1, \quad \sin 2a = 2 \sin a \cos a.$$
Proof. Let

6.7 COROLLARY. For all \( x \), we have \( \cos^2 x + \sin^2 x = 1 \). Hence \( \cos x \) and \( \sin x \) are between \(-1\) and \(1\).

Proof.

Applications to inequalities; greatest and least values

These are largely deductions from (i)–(iii) of 6.3, or directly from the mean-value theorem itself.

Example 1. The function \( x - \sin x \) is increasing for all \( x \), since its derivative is \( 1 - \cos x \geq 0 \) for all \( x \).

Example 2. By finding the greatest value, show that \( 4x^3 - 3x^4 \leq 1 \) for all \( x \).

Note that it is not enough just to discuss points where \( f'(x) = 0 \)!

Example 3. Let \( f(x) = x^3 - 6x^2 + 9x \). Find the greatest and least values of \( f \) on \([0,5]\).

How many solutions are there in this interval of the equation \( f(x) = 2 \)?
By the *intermediate* value theorem, there is a solution of $f(x) = 2$ in the interval $(0, 1)$, another in $(1, 3)$ and another in $(3, 5)$. In each case, there is *only* one solution, since $f'(x) \neq 0$ on the open interval in question. Hence there are three solutions in $[0, 5]$ (and in fact no others, since $f(x) < 0$ for $x < 0$ and $f(x) > 20$ for $x > 5$).

*Example 4.* To show that \( \tan^{-1} x \geq x - \frac{1}{3} x^3 \) for all \( x > 0 \).

*Example 5.* Show that \( \tan x \geq x \) for \( 0 < x < \pi/2 \). Using this fact, apply the method a second time to show that \( \tan x \geq x + \frac{1}{3} x^3 \) on this interval.

*Example 6.* Given that \( f'(x) \geq m > 0 \) for all \( x \geq a \), show that \( f(x) \geq f(a) + m(x-a) \) for all \( x > a \) (so that \( f(x) \to \infty \) as \( x \to \infty \)).

*Example 7.* To show that, for \( a > 0 \),

\[
\frac{a}{1+a} \leq \log(1+a) \leq a.
\]

*Note.* This implies that \( \frac{1}{x} \log(1+x) \to 1 \) as \( x \to 0^+ \). Hence, for any \( a \neq 0 \), we have \( n \log(1+a/n) \to a \), so that \( (1+a/n)^n \to e^a \) as \( n \to \infty \) (proved quite differently in 2.18).
Zeros of polynomials

6.8 PROPOSITION. A polynomial of degree \( n \) has at most \( n \) (real) zeros.

Proof. By induction on \( n \). The statement is true for \( n = 1 \), since \( ax + b \) has one zero, \( x = -b/a \). Assume that the statement is true for a certain \( n \), and let \( p \) be a polynomial of degree \( n + 1 \). Then its derivative \( p' \) is a polynomial of degree \( n \), so, by assumption, \( p' \) has at most \( n \) zeros. Suppose that \( p \) is zero at \( k \) points \( x_1, x_2, \ldots, x_k \) (in order). By Rolle's theorem, \( p' \) has a zeros in each open interval \( (x_{r-1}, x_r) \), a total of at least \( k - 1 \) zeros. So \( k - 1 \leq n \), hence \( k \leq n + 1 \), so the statement is true for \( n + 1 \). \( \square \)

Increasing or decreasing derivative; convex functions

6.9. Suppose that \( f'(x) \) is increasing on \([a, b]\). Then for any \( x \) and \( x_0 \) in \([a, b]\),

\[
f(x) \geq f(x_0) + (x - x_0)f'(x_0).
\]

In other words, the graph is above the tangent at \( x_0 \). The opposite inequality holds if \( f'(x) \) is decreasing.

Proof. First suppose that \( x > x_0 \). By 6.1, there exists \( \xi \) such that \( x_0 < \xi < x \) and

\[
\begin{align*}
f(x) - f(x_0) & = (x - x_0)f'(\xi) \\
& \geq (x - x_0)f'(x_0),
\end{align*}
\]

since \( f'(\xi) \geq f'(x_0) \). Now suppose that \( x < x_0 \).

Example 8. Take \( f(x) = \log x \). Then \( f'(x) = 1/x \), which is decreasing for \( x > 0 \).

Taking \( x_0 = 1 \) in 6.9, we obtain at once

\[
\log x \leq \log x_0
\]

Example 9. Note first that \( x^p (= e^{p \log x}) \) is increasing if \( p > 0 \) and decreasing if \( p < 0 \).

Hence the derivative \( px^{p-1} \) is

increasing \hspace{1cm} \text{if} \hspace{1cm} \text{decreasing} \hspace{1cm} \text{if} \.

Take \( f(x) = x^p \) in 6.9 and substitute \( 1 + x \) for \( x \) and 1 for \( x_0 \): if \( p > 1 \) or \( p < 0 \), we obtain
\[(1 + x)^p = \]

and the opposite inequality when 0 < p < 1.

**6.10 PROPOSITION** (Jensen’s inequality). Suppose that \( f'(x) \) is increasing on an interval \( I \). Suppose that \( x_j \in I \) and \( \lambda_j \geq 0 \) for \( 1 \leq j \leq n \), and \( \sum_{j=1}^n \lambda_j = 1 \). Let \( y = \sum_{j=1}^n \lambda_j x_j \). Then

\[ f(y) \leq \sum_{j=1}^n \lambda_j f(x_j). \]

The opposite inequality holds if \( f'(x) \) is decreasing.

**Proof.** By 6.9, we have \( f(x_j) - f(y) \geq (x_j - y)f'(y) \) for each \( j \). Hence

\[ \sum_{j=1}^n \lambda_j f(x_j) - f(y) = \]

In the case \( n = 2 \), this says

\[ f[(1 - \lambda)x_1 + \lambda x_2] \leq (1 - \lambda)f(x_1) + \lambda f(x_2) \]

for 0 < \( \lambda < 1 \). Any function satisfying this property (whether or not differentiable) is said to be **convex**. It amounts to saying that the curve is below the straight line joining two of its points (which, of course, is what you would expect if the gradient is increasing).

A very nice application of Jensen’s inequality is the following generalization of the famous inequality of geometric and arithmetic means:

**6.11 PROPOSITION.** Let \( a_j, w_j \ (1 \leq j \leq n) \) be positive numbers, with \( \sum_{j=1}^n w_j = 1 \). Then

\[ a_1^w_1 a_2^w_2 \ldots a_n^w_n \leq \sum_{j=1}^n w_j a_j. \]

In particular, the geometric mean \( (a_1 a_2 \ldots a_n)^{1/n} \) is not greater than the arithmetic mean \( \frac{1}{n} \sum_{j=1}^n a_j \).

**Proof.** Let \( \log a_j = x_j \), so that \( a_j^w = \exp(w_j x_j) \). Since \( e^x \) has increasing derivative (\( e^x \) itself),
The periodic nature of $\cos$ and $\sin$

6.12. There is a unique number (denoted by $\pi/2$) in $(0, 2)$ such that $\cos \pi/2 = 0$. We have $\sin \pi/2 = 1$. On the interval $[0, \pi/2]$, $\cos$ is strictly decreasing and $\sin$ is strictly increasing.

Proof. The existence of such a number was shown in 5.8. By 6.7, $\sin^2 \pi/2 = 1$. For $0 < x < 2$,

$$\sin x =$$

since each bracket is greater than 0. In particular, $\sin \pi/2 > 0$, so $\sin \pi/2 = 1$. Since $\frac{d}{dx} \cos x = -\sin x$, it follows that $\cos x$ is strictly decreasing on $[0, 2]$. Hence $\pi/2$ is its only zero in this interval, and $\cos x > 0$ for $0 < x < \pi/2$. Since $\frac{d}{dx} \sin x = \cos x$, this shows that $\sin x$ is strictly increasing on $[0, \pi/2]$. □

6.13. We have: $\cos \pi = -1$, $\cos(\pi/2 - x) = \sin x$, and if $f(x)$ is either $\cos x$ or $\sin x$, then $f(x + \pi) = -f(x)$, $f(x + 2\pi) = f(x)$. Also, $\cos \pi/4 = \sin \pi/4 = 1/\sqrt{2}$.

Proof.

The series for $\log(1 + x)$ and $\tan^{-1} x$

The derivative of $\log(1 + x)$ is $1/(1 + x)$, which equals $1 - x + x^2 - \cdots$. This idea leads us to the following series expression for $\log(1 + x)$:

6.14 PROPOSITION. For $|x| < 1$, we have

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}.$$ 

Proof. Let $f(x) = \log(1 + x)$ and let $g(x)$ be the sum of the stated series. Then, differentiating term by term, we have

$$f'(x) = \frac{1}{1 + x} = 1 - x + x^2 - \cdots = g'(x)$$

for $-1 < x < 1$. Hence $f(x) - g(x)$ is constant (say equal to $c$) on this interval. Since $f(0) = g(0) = 0$, we have $c = 0$, hence $f(x) = g(x)$ on the interval. □
It follows at once from the series that $\frac{1}{x} \log(1 + x) \to 1$ as $x \to 0$ (compare Example 7). We derive further:

$$\log \frac{1 + x}{1 - x} = \log(1 + x) - \log(1 - x) = 2 \left( x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots \right).$$

(To evaluate $\log 2$, put $x = \frac{1}{3}$ in this.)

In exactly the same way, from the series

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1},$$

we obtain:

**6.15 PROPOSITION.** For $|x| < 1$, we have

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}.$$

This leads to various series for $\pi$, but we defer this topic until we have given a second proof, based on integration.

**Taylor’s theorem**

6.16 (Lagrange’s form of Taylor’s theorem). Suppose that the $n$th derivative of $f$ exists throughout an open interval $I$, and that $a, b \in I$ (with $a < b$ or $a > b$). Let

$$f(b) = f(a) + (b - a)f'(a) + \cdots + \frac{(b - a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + R_n.$$  

Then $R_n = \frac{(b - a)^n}{n!}f^{(n)}(\xi)$ for some $\xi$ between $a$ and $b$.

**Note.** The case $n = 1$ is the mean-value theorem.

**Proof.** Define $g(x)$ by:

$$f(b) = f(x) + (b - x)f'(x) + \cdots + \frac{(b - x)^{n-1}}{(n-1)!}f^{(n-1)}(x) + g(x).$$

Clearly, $g(b) = 0$ and $g(a) = R_n$. Differentiate: we get

Now let $h(x) = g(x) - k\frac{(b - x)^n}{n!}$, with $k$ chosen so that $h(a) = 0 (= h(b))$, that is,

$$k\frac{(b - a)^n}{n!} = g(a) = R_n.$$
Then

\[ h'(x) = \frac{(b - x)^{n-1}}{(n - 1)!} (k - f^{(n)}(x)) \].

By Rolle’s theorem, there exists \( \xi \) between \( a \) and \( b \) such that \( h'(\xi) = 0 \), hence \( k = f^{(n)}(\xi) \).

If we can show, in a particular case, that \( R_n \to 0 \) as \( n \to \infty \), then it follows that \( \sum_{n=0}^{\infty} \frac{(b - a)^n}{n!} f^{(n)}(a) \) converges to \( f(b) \). This is often applied with \( a = 0 \).

6.17 PROPOSITION (the binomial series). For \( 0 \leq b < 1 \) and \( \alpha \) not a positive integer,

\[ (1 + b)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} b^n, \]

where \( \binom{\alpha}{n} = \frac{\alpha(\alpha - 1) \ldots (\alpha - n + 1)}{n!} \).

Proof. Let \( f(x) = (1 + x)^\alpha \). Then \( f^{(n)}(x) = \alpha(\alpha - 1) \ldots (\alpha - n + 1)(1 + x)^{\alpha-n} \), so \( f^{(n)}(0)/n! = \binom{\alpha}{n} \), and

\[ R_n = b^n \binom{\alpha}{n} (1 + \xi_n)^{\alpha-n} \]

for some \( \xi_n \) in \( (0, b) \). For \( n > \alpha \), \( (1 + \xi_n)^{\alpha-n} < 1 \), so it is enough to show that \( b^n \binom{\alpha}{n} \to 0 \) as \( n \to \infty \). Denote this by \( u_n \). Then for \( n > \alpha \),

\[ \left| \frac{u_{n+1}}{u_n} \right| = b \left| \frac{\alpha - n}{n + 1} \right| = b \frac{n - \alpha}{n + 1} \to b \quad \text{as} \quad n \to \infty, \]

hence, by the ratio test, \( |u_n| \to 0 \) as \( n \to \infty \), as required. \( \Box \)