Mean value theorems, the continuity of derivatives, L'Hospital's rule, derivatives of higher order, Taylor's theorem, differentiation of vector-valued functions

Theorem *If f and g are continuous real functions on $[a, b]$ which are differentiable in (a, b), then there is a point $x \in (a, b)$ at which*

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

Proof Put

$$h(t) = [f(b) - f(a)]g(t) - [g(b) - g(a)]f(t) \qquad (a \leq t \leq b).$$

Then h is continuous on $[a, b]$, h is differentiable in (a, b), and

$$h(a) = f(b)g(a) - f(a)g(b) = h(b).$$

To prove the theorem, we have to show that $h'(x) = 0$ for some $x \in (a, b)$.

If h is constant, this holds for every $x \in (a, b)$. If $h(t) > h(a)$ for some $t \in (a, b)$, let x be a point on $[a, b]$ at which h attains its maximum.

It follows that $h'(x) = 0$.

If $h(t) < h(a)$ for some $t \in (a, b)$, the same argument applies if we choose for x a point on $[a, b]$ where h attains its minimum.

Taking g(x)=x, we obtain the first mean value theorem:

Theorem *If f is a real continuous function on $[a, b]$ which is differentiable in (a, b), then there is a point $x \in (a, b)$ at which*

$$f(b) - f(a) = (b - a)f'(x).$$

Corollary *Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b) with*

$$\mu \leq f'(x) \leq m \quad \text{for all } x \in (a,b).$$

We then have for $a \leq x_1 \leq x_2 \leq b$

$$\mu(x_2 - x_1) \leq f(x_2) - f(x_1) \leq m(x_2 - x_1).$$

In particular if $M := \max(|\mu|, |m|)$ then

$$|f(x_2) - f(x_1)| \leq M|x_2 - x_1| \quad \text{for all } x_1, x_2 \in (a,b).$$

Therefore if $f'(x) \equiv 0$ then f is constant.

Corollary *Let $f : [a,b] \to \mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b). Suppose*

$$f'(x) \equiv \gamma$$

for some constant γ.

Then

$$f(x) = \gamma x + c,$$

with some constant c, for all $x \in [a,b]$.

Theorem *Let $\gamma \in \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be a function which satisfies the differential equation $f' = \gamma f$, that is, $f'(x) = \gamma f(x)$ for all $x \in \mathbb{R}$. Then*

$$f(x) = f(0)e^{\gamma x} \quad \textit{for all } x \in \mathbb{R}.$$

Proof We consider

$$F(x) := f(x)e^{-\gamma x}.$$

Now

$$F'(x) = f'(x)e^{-\gamma x} - \gamma f(x)e^{-\gamma x} = 0,$$

Thus

$$F \equiv \text{ const. } = F(0) = f(0),$$

and therefore

$$f(x) = f(0)e^{\gamma x}.$$

$\square$

Theorem *Let $f : [a,b] \to \mathbb{R}$ be a differentiable function that satisfies for all $x \in [a,b]$ $|f'(x)| \le \gamma |f(x)|$, γ a constant. If $f(x_0) = 0$ for some $x_0 \in [a,b]$ then $f \equiv 0$ on $[a,b]$.*

Proof. We may assume that $\gamma > 0$, otherwise there is nothing to prove. Set $\delta := \frac{1}{2\gamma}$ and choose $x_1 \in [x_0 - \delta, x_0 + \delta] \cap [a,b] =: I$ such that 属于[a,b]的小邻域

$$|f(x_1)| = \sup_{x \in I} |f(x)|$$

We have

$$|f(x_1)| = |f(x_1) - f(x_0)| \le |x_1 - x_0| \sup_{\xi \in I} |f'(\xi)|$$

$$\le \gamma |x_1 - x_0| \sup_{\xi \in I} |f(\xi)| \le \gamma\delta |f(x_1)| = \frac{1}{2} |f(x_1)|,$$

and therefore $f(x_1) = 0$. It follows that

$$f(x) = 0 \quad \text{for all } x \in I.$$

We have therefore shown that there exists a $\delta > 0$ with the following property: If $f(x_0) = 0$ then $f(x) = 0$ for all $x \in [x_0 - \delta, x_0 + \delta] \cap [a,b]$. If f is not identically zero, there exists a smallest ξ_1 with $a < \xi_1 \le b$ and $f(\xi_1) = 0$, or a greatest ξ_2 with $a \le \xi_2 < b$ and $f(\xi_2) = 0$. However, this is not compatible with the statement which we just proved. □

Corollary *Let $\phi : \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous, $c \in \mathbb{R}$ and $[a,b] \subset \mathbb{R}$. There exists at most one solution $f : [a,b] \to \mathbb{R}$ of the differential equation*

$$f'(x) = \phi(f(x)) \text{ for all } x \in [a,b]$$

with

$$f(a) = c.$$

Proof. Let f_1 and f_2 be solutions with $f_1(a) = f_2(a) = c$. The function $F = f_1 - f_2$ satisfies

$$F(a) = 0$$

and

$$\begin{aligned} |F'(x)| &= |\phi(f_1(x)) - \phi(f_2(x))| \\ &\leq L|f_1(x) - f_2(x)| = L|F(x)| \end{aligned}$$

for a suitable constant L, as f_1 and f_2, being continuous, map the bounded interval $[a,b]$ onto a bounded interval and ϕ is Lipschitz continuous. Theorem above implies that $F \equiv 0$, that is $f_1 \equiv f_2$, whence the uniqueness of the solutions. □

Theorem *Suppose f is differentiable in (a, b).*

(a) *If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotonically increasing.*

(b) *If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.*

(c) *If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.*

Proof All conclusions can be read off from the equation

$$f(x_2) - f(x_1) = (x_2 - x_1)f'(x),$$

which is valid, for each pair of numbers x_1, x_2 in (a, b), for *some* x between x_1 and x_2.

Remark If the inequality about the derivative of f is replaced by the strict inequality, then the function is strictly monotone.

Definition If f has a derivative f' on an interval, and if f' is itself differentiable, we denote the derivative of f' by f'' and call f'' the second derivative of f. Continuing in this manner, we obtain functions

$$f, f', f'', f^{(3)}, \ldots, f^{(n)},$$

each of which is the derivative of the preceding one. $f^{(n)}$ is called the nth derivative, or the derivative of order n, of f.

In order for $f^{(n)}(x)$ to exist at a point x, $f^{(n-1)}(t)$ must exist in a neighborhood of x (or in a one-sided neighborhood, if x is an endpoint of the interval on which f is defined), and $f^{(n-1)}$ must be differentiable at x. Since $f^{(n-1)}$ must exist in a neighborhood of x, $f^{(n-2)}$ must be differentiable in that neighborhood.

Theorem *Let $f : (a, b) \to \mathbb{R}$ be twice differentiable, and let $x_0 \in (a, b)$, with*

$$f'(x_0) = 0, f''(x_0) > 0. \tag{8}$$

Then f has a strict local minimum at x_0. If we have $f''(x_0) < 0$ instead, it has a strict local maximum at x_0. Conversely, if f has a local minimum at $x_0 \in (a, b)$, and if it is twice differentiable there, then

$$f''(x_0) \geq 0. \tag{9}$$

Proof We only treat the case of a local minimum. If

6

$$\lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{f'(x) - f'(x_0)}{x - x_0} = f''(x_0) > 0,$$

then there exists $\delta > 0$ with

$$f'(x) < f'(x_0) = 0 \quad \text{for } x_0 - \delta < x < x_0$$

and

$$f'(x) > f'(x_0) = 0 \quad \text{for } x_0 < x < x_0 + \delta.$$

Thus, f is strictly monotonically decreasing on $(x_0 - \delta, x_0)$, and strictly monotonically increasing on $(x_0, x_0 + \delta)$. This implies that

$$f(x) > f(x_0) \quad \text{for } 0 < |x - x_0| < \delta,$$

and consequently f has a strict local minimum at x_0.
The second half of the theorem follows from the first half. $\square$

Continuity of derivatives

Theorem *Suppose f is a real differentiable function on $[a, b]$ and suppose $f'(a) < \lambda < f'(b)$. Then there is a point $x \in (a, b)$ such that $f'(x) = \lambda$.*

Proof Put $g(t) = f(t) - \lambda t$. Then $g'(a) < 0$, so that $g(t_1) < g(a)$ for some $t_1 \in (a, b)$, and $g'(b) > 0$, so that $g(t_2) < g(b)$ for some $t_2 \in (a, b)$. Hence g attains its minimum on $[a, b]$ at some point x such that $a < x < b$.

Thus $g'(x) = 0$ and so $f'(x) = \lambda$.

Remark A similar result holds if $f'(a) > f'(b)$.

Theorem *Suppose f and g are real and differentiable in (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$, where $-\infty \leq a < b \leq +\infty$. Suppose*

$$\frac{f'(x)}{g'(x)} \to A \text{ as } x \to a. \tag{13}$$

If

$$f(x) \to 0 \text{ and } g(x) \to 0 \text{ as } x \to a, \tag{14}$$

or if

$$g(x) \to +\infty \text{ as } x \to a, \tag{15}$$

then

$$\frac{f(x)}{g(x)} \to A \text{ as } x \to a. \tag{16}$$

Remark The analogous statement is also true if $x \to \mathrm{b}$ or $\mathrm{g}(x) \to -\infty$ in (15).

Example

$$\lim_{x\to+\infty}\frac{\ln x}{x^\alpha}=\lim_{x\to+\infty}\frac{\left(\frac{1}{x}\right)}{\alpha x^{\alpha-1}}=\lim_{x\to+\infty}\frac{1}{\alpha x^\alpha}=0 \text{ for } \alpha>0\,.$$

Example

$$\lim_{x\to+\infty}\frac{x^\alpha}{a^x}=\lim_{x\to+\infty}\frac{\alpha x^{\alpha-1}}{a^x\ln a}=\cdots=\lim_{x\to+\infty}\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)x^{\alpha-n}}{a^x(\ln a)^n}=0$$

for $a>1$, since for $n>\alpha$ and $a>1$ it is obvious that $\frac{x^{\alpha-n}}{a^x}\to 0$ if $x\to+\infty$.

Theorem *Suppose f is a real function on $[a, b]$, n is a positive integer, $f^{(n-1)}$ is continuous on $[a, b]$, $f^{(n)}(t)$ exists for every $t \in (a, b)$. Let α, β be distinct points of $[a, b]$, and define*

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(t-\alpha)^k. \tag{23}$$

Then there exists a point x between α and β such that

$$f(\beta) = P(\beta) + \frac{f^{(n)}(x)}{n!}(\beta-\alpha)^n. \tag{24}$$

Remark For $n = 1$, this is just the mean value theorem. In general, the theorem shows that f can be approximated by a polynomial of degree $n - 1$, and that (24) allows us to estimate the error, if we know bounds on $|f^{(n)}(x)|$.

Proof Let M be the number defined by

$$f(\beta) = P(\beta) + M(\beta - \alpha)^n \tag{25}$$

and put

$$g(t) = f(t) - P(t) - M(t - \alpha)^n \qquad (a \le t \le b). \tag{26}$$

We have to show that $n!M = f^{(n)}(x)$ for some x between α and β. By (23) and (26),

$$g^{(n)}(t) = f^{(n)}(t) - n!M \qquad (a < t < b). \tag{27}$$

Hence the proof will be complete if we can show that $g^{(n)}(x) = 0$ for some x between α and β.

Since $P^{(k)}(\alpha) = f^{(k)}(\alpha)$ for $k = 0, \ldots, n-1$, we have

$$g(\alpha) = g'(\alpha) = \cdots = g^{(n-1)}(\alpha) = 0. \tag{28}$$

Our choice of M shows that $g(\beta) = 0$, so that $g'(x_1) = 0$ for some x_1 between α and β, by the mean value theorem. Since $g'(\alpha) = 0$, we conclude similarly that $g''(x_2) = 0$ for some x_2 between α and x_1. After n steps we arrive at the conclusion that $g^{(n)}(x_n) = 0$ for some x_n between α and x_{n-1}, that is, between α and β.

Definition Suppose that E is a real or complex vector space, and that $u, v \in E$. Let $\sigma : [0,1] \to E$ be defined by

$$\sigma(t) = u + (v-u)t = (1-t)u + tv \text{ for } 0 \le t \le 1.$$

Then $\sigma([0,1])$ is the *straight line segment* $[u,v]$ between u and v. A subset C of E is *convex* if $[u,v] \subseteq C$, for each u, v in C. Thus a subset of $\mathbf{R}$ is convex if and only if it is an interval.

Suppose that f is a function on an interval I. f is said to be *convex* if the subset $\{(x,y) \in \mathbf{R}^2 : x \in I, y \ge f(x)\}$ of $\mathbf{R}^2$ is a convex set. Equivalently, if $x_0, x_1 \in I$, then the straight line segment $[(x_0, f(x_0)), (x_1, f(x_1))]$ in $\mathbf{R}^2$ lies above the graph $G_f = \{(x, f(x)) \in \mathbf{R}^2 : x \in I\}$. Since

$$\begin{aligned}&[(x_0, f(x_0)), (x_1, f(x_1))] = \\ &\quad \{((1-t)x_0 + tx_1, (1-t)f(x_0) + tf(x_1)) : 0 \le t \le 1\},\end{aligned}$$

this says that

$$(1-t)f(x_0) + tf(x_1) \ge f((1-t)x_0 + tx_1)$$

for all $x_0, x_1 \in I$ and all $0 \le t \le 1$.

We say that f is *strictly convex* if

$$(1-t)f(x_0) + tf(x_1) > f((1-t)x_0 + tx_1)$$

for distinct $x_0, x_1 \in I$ and all $0 < t < 1$. f is *concave* if $-f$ is convex; that is,

$(1-t)f(x_0)+tf(x_1) \leq f((1-t)x_0+tx_1)$ for all $x_0, x_1 \in I$ and all $0 \leq t \leq 1$.

Strict concavity is defined similarly.

Proposition *(i) If f and g are convex functions on an interval I and $a \geq 0$ then $f+g$ and af are convex.*

(ii) If $(f_n)_{n=1}^{\infty}$ is a sequence of convex functions on an interval I, and if $f_n(x) \to f(x)$ as $n \to \infty$, for each $x \in I$, then f is convex.

(iii) If $\{f : f \in F\}$ is a family of convex functions on an interval I for which $g(x) = \sup\{f(x) : f \in F\}$ is finite for each $x \in I$ then g is convex.

(iv) If f, g are convex, non-negative increasing functions on an interval I then fg is convex.

(v) If f is a convex function on an interval I, and if ϕ is an increasing convex function on an interval J which contains $f(I)$, then $\phi \circ f$ is a convex function on I.

Proof (i) and (ii) follow immediately from the definitions.

We suppose that $x_0, x_1 \in I$ and that $0 < t < 1$, and we set $x_t = (1-t)x_0 + tx_1$.

(iii) Suppose that $\epsilon > 0$. There exists a function f in F such that $f(x_t) \geq g(x_t) - \epsilon$. Then

$$g(x_t) - \epsilon \leq f(x_t) \leq (1-t)f(x_0) + tf(x_1) \leq (1-t)g(x_0) + tg(x_1).$$

Since this holds for all $\epsilon > 0$, $g(x_t) \leq (1-t)g(x_0) + tg(x_1)$.

(iv) Since f and g are increasing,

$$(g(x_1) - g(x_0))(f(x_1) - f(x_0)) \geq 0.$$

Expanding and rearranging,

$$f(x_0)g(x_1) + f(x_1)g(x_0) \leq f(x_0)g(x_0) + f(x_1)g(x_1),$$

and so

$$\begin{aligned}
&f(x_t)g(x_t) \leq \\
&\leq ((1-t)f(x_0) + tf(x_1))((1-t)g(x_0) + tg(x_1)) \\
&= (1-t)^2 f(x_0)g(x_0) + t(1-t)(f(x_0)g(x_1) + f(x_1)g(x_0)) + t^2(f(x_1)g(x_1) \\
&\leq (1-t)^2 f(x_0)g(x_0) + t(1-t)(f(x_0)g(x_0) + f(x_1)g(x_1)) + t^2(f(x_1)g(x_1) \\
&= (1-t)f(x_0)g(x_0) + tf(x_1)g(x_1).
\end{aligned}$$

(v) Since ϕ is convex and increasing,

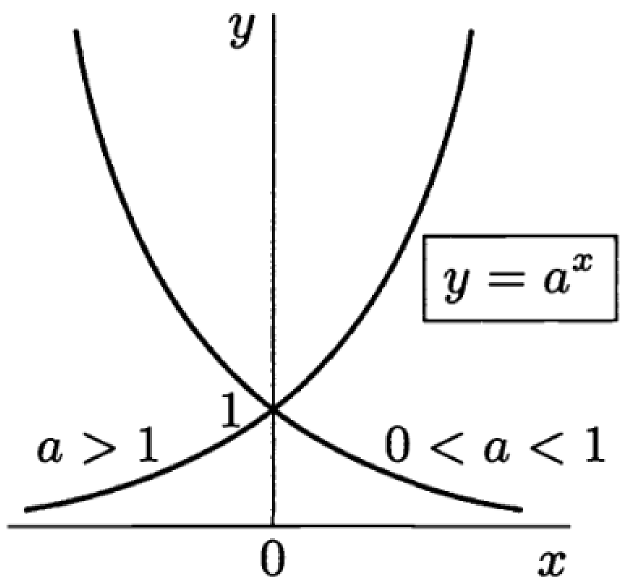
$$\phi(f(x_t)) \leq \phi((1-t)f(x_0) + tf(x_1)) \leq (1-t)\phi(f(x_0)) + t\phi(f(x_1)). \quad \square$$

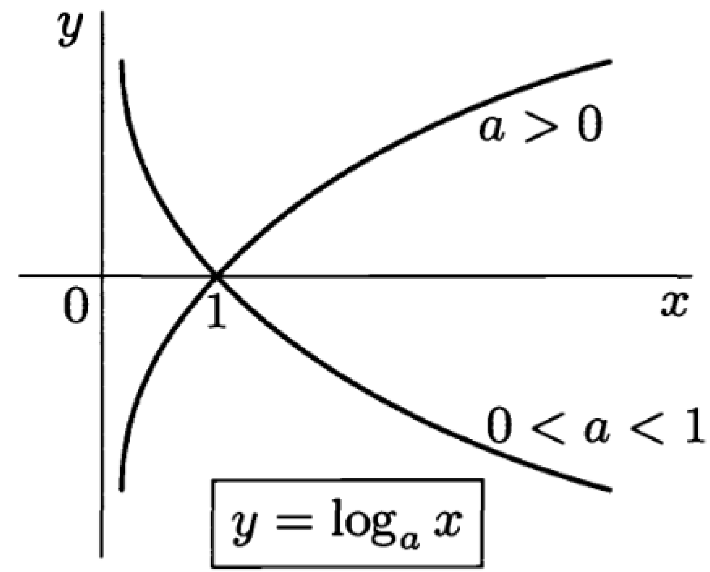
Proposition *A necessary and sufficient condition for a function* $f : (a, b) \to \mathbb{R}$ *that is differentiable on the open interval* (a, b) *to be convex on that interval is that its derivative* f' *be nondecreasing on* (a, b). *A strictly increasing* f' *corresponds to a strictly convex function.*

Corollary. *A necessary and sufficient condition for a function* $f : (a, b) \to \mathbb{R}$ *having a second derivative on the open interval* (a, b) *to be convex on* (a, b) *is that* $f''(x) \geq 0$ *on that interval. The condition* $f''(x) > 0$ *on* (a, b) *is sufficient to guarantee that* f *is strictly convex.*

Example Let $f(x) = a^x$, $0 < a$, $a \neq 1$. Since $f''(x) = a^x \ln^2 a > 0$, the exponential function a^x is strictly convex.

Example For the function $f(x) = \log_a x$ we have $f''(x) = -\frac{1}{x^2 \ln a}$, so that the function is strictly convex if $0 < a < 1$, and strictly concave if $1 < a$.





Proposition (Jensen's inequality). *If $f: (a,b) \to \mathbb{R}$ is a convex function, $x_1, \ldots, x_n$ are points of (a,b), and $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers such that $\alpha_1 + \cdots + \alpha_n = 1$, then*

$$f(\alpha_1 x_1 + \cdots + \alpha_n x_n) \le \alpha_1 f(x_1) + \cdots + \alpha_n f(x_n)\,. \quad \heartsuit$$

Proof We show that if $\heartsuit$ is valid for n = m- 1, it is also valid for n = m. We can assume $\alpha_n \neq 0$.
Then $\beta = \alpha_2 + \cdots + \alpha_n > 0$ and $\frac{\alpha_2}{\beta} + \cdots + \frac{\alpha_n}{\beta} = 1$. Using the convexity of the function, we find

$$f(\alpha_1 x_1 + \cdots + \alpha_n x_n) = f\left(\alpha_1 x_1 + \beta\left(\frac{\alpha_2}{\beta} x_2 + \cdots + \frac{\alpha_n}{\beta} x_n\right)\right) \le$$
$$\le \alpha_1 f(x_1) + \beta f\left(\frac{\alpha_2}{\beta} x_2 + \cdots + \frac{\alpha_n}{\beta} x_n\right),$$

since $\alpha_1 + \beta = 1$ and $\left(\frac{\alpha_2}{\beta} x_1 + \cdots + \frac{\alpha_n}{\beta} x_n\right) \in$ (a, b).

By the induction hypothesis, we now have

$$f\left(\frac{\alpha_2}{\beta} x_2 + \cdots + \frac{\alpha_n}{\beta} x_n\right) \le \frac{\alpha_2}{\beta} f(x_2) + \cdots + \frac{\alpha_n}{\beta} f(x_n) \quad \text{and so}$$

$$f(\alpha_1 x_1 + \cdots + \alpha_n x_n) \le \alpha_1 f(x_1) + \beta f\left(\frac{\alpha_2}{\beta} x_2 + \cdots + \frac{\alpha_n}{\beta} x_n\right) \le$$
$$\le \alpha_1 f(x_1) + \alpha_2 f(x_2) + \cdots + \alpha_n f(x_n)\,.$$

Example The function $f(x) =-\ln x$ is strictly convex on the set of positive numbers, and so

$$\alpha_1 \ln x_1 + \cdots + \alpha_n \ln x_n \leq \ln(\alpha_1 x_1 + \cdots + \alpha_n x_n)$$

or,

$$x_1^{\alpha_1} \cdots x_n^{\alpha_n} \leq \alpha_1 x_1 + \cdots + \alpha_n x_n$$

for $x_i \geq 0$, $\alpha_i \geq 0$, $i = 1, \ldots, n$, and $\sum_{i=1}^{n} \alpha_i = 1$.

In particular, if $\alpha_1 = \cdots = \alpha_n = \frac{1}{n}$, we obtain the classical inequality

$$\sqrt[n]{x_1 \cdots x_n} \leq \frac{x_1 + \cdots + x_n}{n} .$$

Example Let $f(x) = x^p$, $x \geq 0$, $p > 1$. Since such a function is convex, we have

$$\Big(\sum_{i=1}^{n} \alpha_i x_i\Big)^p \leq \sum_{i=1}^{n} \alpha_i x_i^p \ .$$

Setting $q = \frac{p}{p-1}$, $\alpha_i = b_i^q \Big(\sum_{i=1}^{n} b_i^q\Big)^{-1}$, and $x_i = a_i b_i^{-1/(p_i-1)} \sum_{i=1}^{n} b_i^q$ here, we obtain Hölder's inequality

$$\sum_{i=1}^{n} a_i b_i \leq \Big(\sum_{i=1}^{n} a_i^p\Big)^{1/p} \Big(\sum_{i=1}^{n} b_i^q\Big)^{1/q} \ ,$$

where $\frac{1}{p} + \frac{1}{q} = 1$ and $p > 1$.

Definition Suppose $U \subseteq \mathbb{R}^n$ is an open set. A function $f : U \longrightarrow \mathbb{R}^m$ is *differentiable at* $x \in U$ if there is a linear map $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Th\|}{\|h\|} = 0.$$

First, note that $h \to 0$ in $\mathbb{R}^n$. Notice also that the norm sign in the numerator denotes the Euclidean norm in $\mathbb{R}^m$ while the norm sign in the denominator denotes the Euclidean norm in $\mathbb{R}^n$. If we use the norm sign for an element of $\mathbb{R}$, it indicates the usual absolute value on $\mathbb{R}$. We write $T = Df(x)$ and we call this the *derivative* of f at x. We say that f is *differentiable on* U if f is differentiable at each point in U.

THEOREM *Suppose U is an open set in $\mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}^m$ is differentiable at a point $x_0 \in U$. Then f is continuous at x_0.*

Proof. Take $\varepsilon = 1$. Then there exists a $\delta > 0$ such that

$$\|f(x_0 + h) - f(x_0) - Df(x_0)h\| < \varepsilon\|h\| = \|h\|$$

whenever $\|h\| < \delta$. It follows from the triangle inequality that

$$\begin{aligned}\|f(x_0 + h) - f(x_0)\| &< \|h\| + \|Df(x_0)h\| \\ &\leq \|h\| + \|Df(x_0)\| \cdot \|h\| = (1 + \|Df(x_0)\|)\|h\|\end{aligned}$$

when $\|h\| < \delta$.

Theorem *Suppose* $\mathbf{f}$ *is a continuous mapping of* $[a, b]$ *into* R^k *and* $\mathbf{f}$ *is differentiable in* (a, b). *Then there exists* $x \in (a, b)$ *such that*

$$|\mathbf{f}(b) - \mathbf{f}(a)| \leq (b - a)|\mathbf{f}'(x)|.$$

Proof Put $\mathbf{z} = \mathbf{f}(b) - \mathbf{f}(a)$, and define

$$\varphi(t) = \mathbf{z} \cdot \mathbf{f}(t) \qquad (a \leq t \leq b).$$

Then φ is a real-valued continuous function on $[a, b]$ which is differentiable in (a, b). The mean value theorem shows therefore that

$$\varphi(b) - \varphi(a) = (b - a)\varphi'(x) = (b - a)\mathbf{z} \cdot \mathbf{f}'(x)$$

for some $x \in (a, b)$. On the other hand,

$$\varphi(b) - \varphi(a) = \mathbf{z} \cdot \mathbf{f}(b) - \mathbf{z} \cdot \mathbf{f}(a) = \mathbf{z} \cdot \mathbf{z} = |\mathbf{z}|^2.$$

The Schwarz inequality now gives

$$|\mathbf{z}|^2 = (b - a)|\mathbf{z} \cdot \mathbf{f}'(x)| \leq (b - a)|\mathbf{z}|\,|\mathbf{f}'(x)|.$$

Hence $|\mathbf{z}| \leq (b - a)|\mathbf{f}'(x)|$, which is the desired conclusion.