Mathematical Analysis: Supplementary notes I

0 FIELDS

The real numbers, $\mathbb{R}$, form a *field.* This means that we have a *set*, here $\mathbb{R}$, and two binary operations

$$\textit{addition}, + : \mathbb{R}\times\mathbb{R}\to\mathbb{R}, \quad \text{and } \textit{multiplication}, \cdot : \mathbb{R}\times\mathbb{R}\to\mathbb{R},$$

for which the axioms **F1-F7** below hold true. Note that multiplication can also be denoted by $\times$, or simply by juxtaposition.

(F1) Associativity of addition: $(r+s)+t=(r+s)+t$ for all $r,s,t\in\mathbb{R}$;

(F2)(i) Existence of additive identity: There exists $0\in\mathbb{R}$ such that

$$r+0=r \qquad \text{for all } r\in\mathbb{R};$$

(F2)(ii) Existence of additive inverse: Given $r\in\mathbb{R}$ there exists $s\in\mathbb{R}$ such that

$$r+s=0 \qquad (\text{we write } s=-r);$$

(F3) Commutativity of addition: $r+s=s+r$ for all $r,s\in\mathbb{R}$;

(F4) Associativity of multiplication: $(rs)t=r(st)$ for all $r,s,t\in\mathbb{R}$;

(F5)(i) Existence of multiplicative identity: There exists $1\in\mathbb{R}$ with $1\neq 0$ such that

$$r\cdot 1=r \qquad \text{for all } r\in\mathbb{R};$$

(F5)(ii) Existence of multiplicative inverse: Given $r\in\mathbb{R}\setminus\{0\}$ there exists $t\in\mathbb{R}$ such that

$$rt=1 \qquad (\text{we write } t=r^{-1});$$

(F6) Commutativity of multiplication: $rs=sr$ for all $r,s\in\mathbb{R}$;

(F7) Distributive Law: $r(s+t)=rs+rt$ for all $r,s,t\in\mathbb{R}$.

1 SEQUENCES

DEFINITION Let $f : \mathbb{N} \to \mathbb{R}$ (or $\mathbb{N}_0 \to \mathbb{R}$). Then f is called a *sequence*.

If $f(n) = a_n$, then a_n is called the nth term. It is customary to write such a sequence as $\{a_n\}$ (or $\{a_1, a_2, \ldots\}$).

DEFINITION A sequence $\{a_n\}$ is said to converge to a limit a if for every $\varepsilon > 0$ there is an integer N such that $|a_n - a| < \varepsilon$ whenever $n \geq N$. (The number N may depend on ε; in particular, smaller ε may (and often does) require larger N). In this case we write $\lim_{n\to\infty} a_n = a$ or $a_n \to a$ as $n \to \infty$.

A sequence that does not converge is said to diverge. This can be further qualified (e.g. "the sequence diverges to $+\infty$").

LEAST UPPER BOUND AND GREATEST LOWER BOUND

DEFINITION Let $E \subset \mathbb{R}$, $E \neq \emptyset$. A number $b \in \mathbb{R}$ is called an upper (resp. lower) bound for E if $x \leq b$ (resp. $b \leq x$) for every $x \in E$. In the case that such a b exists, we say that E is bounded above (resp. below).
If E has both an upper bound and a lower bound, then we say that E is bounded.

DEFINITION By a **supremum** (or least upper bound) of E we mean a number $s \in \mathbb{R}$ such that

(i) s is an upper bound for E and

(ii) if b is an upper bound for E, then $s \leq b$.

If E has a supremum s, we write

$$s = \sup E.$$

(The notation $s = \text{lub } E$ is used in some textbooks)

The **infimum** (or greatest lower bound) of E is defined similarly. It is a lower bound for E which is larger than every other lower bound for E. We write $\inf E$ (or glb E).

Axiom 5 (**Supremum principle**) Every nonempty set of real number that is bounded above has a supremum in $\mathbb{R}$.
Every nonempty set of real numbers that is bounded below has an infimum in $\mathbb{R}$.

MONOTONE SEQUENCES

DEFINITION A sequence $\{a_n\}$ is said to be :

nondecreasing if $a_n \leq a_{n+1}$;

nonincreasing if $a_n \geq a_{n+1}$;

strictly increasing if $a_n < a_{n+1}$;

$$\textbf{strictly decreasing} \text{ if } a_n > a_{n+1},$$

for every $n \in \mathbb{N}$. All such sequences are said to be **monotone**, in the latter two cases **strictly monotone**.

EXAMPLE The number e.
For every $n \in \mathbb{N}$, let

$$a_n = \big(1 + \frac{1}{n}\big)^n, \quad b_n = \big(1 + \frac{1}{n}\big)^{n+1}.$$

Then

(i) $\{a_n\}$ is strictly increasing,

(ii) $\{b_n\}$ is strictly decreasing,

(iii) $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$.

The common limit of these two sequences is denoted by e. There holds $2 < e < 4$.

LIMSUP and LIMINF

Even though a sequence need not have a limit, there are two important real numbers which can be associated with every sequence of real numbers.

DEFINITION Let $\{a_n\} \subset \mathbb{R}$. For every $k \in \mathbb{N}$ define

$$y_k = \inf\{a_n;\, n \in \mathbb{N},\, n \geq k\}$$

$$z_k = \sup\{a_n;\, n \in \mathbb{N},\, n \geq k\}.$$

We have

$$y_1 \leq y_2 \leq y_3 \leq \dots \text{ and } z_1 \geq z_2 \geq z_3 \geq \dots$$

Let

$$y = \lim_{k\to\infty} y_k = \sup\{y_k;\, k \in \mathbb{N}\}$$

$$z = \lim_{k\to\infty} z_k = \inf\{z_k;\, k \in \mathbb{N}\}.$$

The numbers y and z are called the **limit inferior** of $\{a_n\}$ and the **limit superior** of $\{a_n\}$, respectively. Note: y and/or z may be $+\infty$ or $-\infty$.

We write

$$y = \liminf_{n\to\infty} a_n = \sup_k \inf\{a_n;\, n \in \mathbb{N},\, n \geq k\},$$

$$z = \limsup_{n\to\infty} a_n = \inf_k \sup\{a_n;\, n \in \mathbb{N},\, n \geq k\}.$$

A useful tool in the study of sequences is the notion of a **cluster point**.

DEFINITION. A point x is called a cluster point of the sequence $\{x_n\}$ if for **every** $\varepsilon > 0$ there are infinitely many values of n with $|x_n - x| < \varepsilon$.

Let $\{x_n\}$ be a sequence in $\mathbb{R}$ that is bounded above. The limit superior of x_n ($\limsup_{n\to\infty} x_n$) is the greatest cluster point of $\{x_n\}$; equivalently, it is the supremum of the set of cluster points. If the sequence is not bounded above, $\limsup_{n\to\infty} x_n = \infty$.
Similarly, if $\{x_n\}$ is bounded below, the limit inferior of x_n ($\liminf_{n\to\infty} x_n$) is the infimum of the set of cluster points. If $\{x_n\}$ is not bounded below, $\liminf_{n\to\infty} x_n = -\infty$.

CAUCHY SEQUENCES

DEFINITION A sequence $\{x_n\}$ of real numbers is called a **Cauchy sequence** if for every number $\varepsilon > 0$, there is an integer N (depending on ε) such that $|x_n - x_m| < \varepsilon$ whenever $n \geq N$ and $m \geq N$.

2 SERIES

DEFINITION Let $\{a_n\} \subset \mathbb{R}$ and define

$$s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \ldots + a_n$$

for each $n \in \mathbb{N}$.
The symbol $\sum_{n=1}^{\infty} a_n$ or $a_1 + a_2 + \ldots$ is called an **infinite series** having nth term a_n and nth partial sum s_n.

DEFINITION A series $\sum_{n=1}^{\infty} a_n$ is said to converge to $a \in \mathbb{R}$ if the sequence of **partial sums** $s_n = \sum_{k=1}^{n} a_k$ converges to a, and if so we write

$$a = \sum_{n=1}^{\infty} a_n.$$

Otherwise, we say that $\sum_{n=1}^{\infty} a_n$ diverges.

Cauchy criterion *Let $\{a_n\} \subset \mathbb{R}$. A series $\sum_{n=1}^{\infty} a_n$ converges if and only if for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that*

$$\left| \sum_{n=p+1}^{q} a_n \right| < \varepsilon \quad \textit{whenever} \quad q > p \geq n_0.$$

Geometric series. Let $a \in \mathbb{R}$ and $r \in \mathbb{R}$. Then $\displaystyle\sum_{n=0}^{\infty} ar^n$ converges and its sum is $\frac{a}{1-r}$ if $|r| < 1$. If $a \neq 0$ and $|r| \geq 1$, then this series diverges.
By the formula for geometric progressions we have ($r \neq 1$):

$$s_n = \sum_{k=0}^{n} ar^k = a\frac{1 - r^{n+1}}{1 - r}.$$

Assuming that $|r| < 1$, $\lim_{n\to\infty} s_n = \frac{a}{1-r}$. If $a \neq 0$ and $|r| \geq 1$, then $|ar^n| \geq |a| \neq 0$ and this shows that the series cannot converge.

Comparison test *Suppose that $0 \leq a_n \leq b_n$ for all but finitely many $n \in \mathbb{N}$.*

(1) If $\sum_n b_n$ converges, then $\sum_n a_n$ converges.

(2) If $\sum_n a_n$ diverges, then $\sum_n b_n$ diverges.

ABSOLUTE CONVERGENCE

DEFINITION We say that $\sum_{n=1}^{\infty} a_n$ *converges absolutely* if $\sum_{n=1}^{\infty} |a_n|$ converges.

Leibniz's alternating test *Let $\{a_n\}$ be a nonincreasing sequence of positive numbers such that $\lim_{n\to\infty} a_n = 0$. Then $\sum_{n=1}^{\infty}(-1)^n a_n$ is convergent.*

Root test (Cauchy) *Let $\sum_{n=1}^{\infty} a_n$ and*

$$\rho = \limsup_{n\to\infty} \sqrt[n]{|a_n|}.$$

(i) If $\rho < 1$, the series converges absolutely,

(ii) If $\rho > 1$, the series diverges.

(iii) If $\rho = 1$, then the test is inconclusive.

Ratio test (d'Alembert) *Let $\sum_{n=1}^{\infty} a_n$ be a series with $a_n \neq 0$ for every n.*

(i) If $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$, then the series converges absolutely.

(ii) If $\liminf_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| > 1$, then the series diverges.

(iii) If $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = 1$, then the test is inconclusive.

3 LIMITS AND CONTINUITY

DEFINITION Let $X \subset \mathbb{R}$ and $a \in \mathbb{R}$. We call a a **limit point** (or **accumulation point**) of X if every interval $(-\delta + a, a + \delta)$, $\delta > 0$, contains at least one point of X other than a (a need **not** be in X).
All points of X that are not limit points of X are called **isolated points** of X.

DEFINITION Let $f : X \to \mathbb{R}$, $X \subset \mathbb{R}$ and let a be a limit point of X. We say that f converges to the limit ℓ as x approaches a if for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$|f(x) - \ell| < \varepsilon \quad \text{whenever} \quad 0 < |x - a| < \delta \quad \text{and} \quad x \in X.$$

Theorem 1 *Let $\lim\limits_{x\to a} f(x) = \ell$ and $\lim\limits_{x\to a} g(x) = m$. Then* *(i)* $\lim_{x\to a}(f(x) \pm g(x)) = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x) = \ell \pm m$,

(ii) $\lim_{x\to a} \alpha f(x) = \alpha \lim_{x\to a} f(x) = \alpha\ell$ *for every* $\alpha \in \mathbb{R}$,

(iii) if $m \neq 0$, *then* $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \dfrac{\ell}{m}$,

(iv) $\lim_{x\to a} f(x)g(x) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = \ell m$.

Theorem 2 (Squeeze principle) *Let* $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}$, $h : X \to \mathbb{R}$ *be such that* $f(x) \leq g(x) \leq h(x)$ *for all* $x \in X$. *If* $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = \ell$, *then* $\lim_{x\to a} g(x) = \ell$.

LIMITS AT INFINITY

DEFINITION Let $f : [a,\infty) \to \mathbb{R}$. We say that $\lim_{x\to\infty} f(x) = \ell$ if and only if for every $\varepsilon > 0$ there exists $M > 0$ (think of M as needing to be large if ε is small) such that

$$|f(x) - \ell| < \varepsilon \quad \text{whenever } \ x \geq M.$$

We say that $\lim_{x\to\infty} f(x) = \infty$ if and only if for every $N > 0$ there exists $M > 0$ such that $f(x) \geq N$ whenever $x > M$.
In a similar manner we define limits at $-\infty$.

The squeeze principle remains valid for limits at ∞ $(-\infty)$.

EXAMPLE Let $f(x) = \frac{\sin x}{1+|x|}$. For every $x \in \mathbb{R}$ we have

$$-\frac{1}{1+|x|} \leq \frac{\sin x}{1+|x|} \leq \frac{1}{1+|x|}.$$

Since $\lim_{x\to\infty} \dfrac{1}{1+|x|} = 0$, we also have $\lim_{x\to\infty} f(x) = 0$.

CONTINUOUS FUNCTIONS

DEFINITION Let $f : (a,b) \to \mathbb{R}$ and $x_0 \in (a,b)$. We say that f is continuous at x_0 if $\lim_{x\to x_0} f(x) = f(x_0)$.
This means: f is **continuous** at x_0 if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - f(x_0)| < \varepsilon \quad \text{whenever } \ |x - x_0| < \delta.$$

The function f is continuous on (a,b) if f is continuous at each point of (a,b).

Theorem 3 *Let* f *and* g *be defined on* (a,b) *and continuous at the point* $x_0 \in (a,b)$, *and let* λ *be a constant. Then*

(i) $\lambda \cdot f(x)$ *is continuous at* x_0,

(ii) $f(x) \pm g(x)$ *is continuous at* x_0,

(iii) $f(x) \cdot g(x)$ *is continuous at* x_0,

(iv) $\frac{f(x)}{g(x)}$ *is continuous at* x_0, *provided* $g(x_0) \neq 0$.

Theorem 4 (**composition of functions**) *Let* $\lim_{x \to x_0} g(x) = \ell$ *and let* f *be continuous at* ℓ*: then there holds* $\lim_{x \to x_0} f((g(x)) = f(\ell)$.
If g *is continuous at* x_0 *and* f *is continuous at* $g(x_0)$, *then* $f \circ g(x) = f(g(x))$ *is continuous at* x_0.

DEFINITION Let I be an interval, $f : I \to \mathbb{R}$. We say that f is *uniformly continuous* on I if for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$x, y \in I \text{ and } |x - y| < \delta \text{ imply } |f(x) - f(y)| < \varepsilon.$$

Theorem 5 Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is uniformly continuous.

PROPERTIES OF CONTINUOUS FUNCTIONS

Intermediate Value Theorem *Let* λ *be a constant and* $f : [a, b] \to \mathbb{R}$ *continuous on* $[a, b]$.
If $f(a) < \lambda < f(b)$ *or* $f(a) > \lambda > f(b)$, *then there is a* $c \in (a, b)$ *such that* $f(c) = \lambda$.

DEFINITION We say that a function f attains its minimum (maximum) value on a set $\Omega \subset \mathbb{R}$ at $a \in \Omega$ if $f(x) \geq f(a)$ $(f(x) \leq f(a))$ for all $x \in \Omega$. We say that a function f has an **extremum** at a on Ω if it attains its maximum or its minimum value on Ω at a.

Extreme value theorem *A continuous function* f *on a closed interval* $[a, b]$ *attains its maximum and minimum values on* $[a, b]$.

4 SEQUENCES AND SERIES OF FUNCTIONS

An important way to construct nontrivial functions is to obtain them as limits of sequences or series of given functions.

DEFINITION (**pointwise convergence**) Let $X \subset \mathbb{R}$ and $f_n : X \to \mathbb{R}$ for each $n \in \mathbb{N}$. The sequence $\{f_n\}$ is said to be *pointwise convergent on* X if there exists a function $f : X \to \mathbb{R}$ such that $f(x) = \lim_{n\to\infty} f_n(x)$ for every $x \in X$. In this case, f is called the *pointwise limit* on X of the sequence $\{f_n\}$.

DEFINITION (**uniform convergence**) The sequence $\{f_n\}$ is said to be uniformly convergent on X if there exists a function $f : X \to \mathbb{R}$ such that for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \varepsilon \quad \text{whenever} \quad n \geq N \quad \text{and} \quad x \in X.$$

The number N may depend on ε but not on x.

Theorem 1 *Let $f_n : [a, b] \to \mathbb{R}$ be continuous for each $n \in \mathbb{N}$. If $f_n \to f$ uniformly on $[a, b]$, then f is continuous on $[a, b]$.*

Theorem 2 (Cauchy condition) *Let $X \subset \mathbb{R}$ and $f_n : X \to \mathbb{R}$. Then the following two assertions are equivalent:*

(i) $\{f_n\}$ is uniformly convergent,

(ii) for every $\varepsilon > 0$ there exists some $N \in \mathbb{N}$ such that for every $x \in X$

$$|f_n(x) - f_m(x)| < \varepsilon \quad \text{whenever } n, m \geq N.$$

SERIES OF FUNCTIONS

Let $f_n : X \to \mathbb{R}$, $X \subset \mathbb{R}$ and

$$s_n(x) = f_1(x) + \ldots + f_n(x).$$

DEFINITION If $\{s_n(x)\}$ is pointwise convergent on X to $s(x)$, then we say that the series $\sum_{k=1}^{\infty} f_k(x)$ is *pointwise convergent* on X and that s is its *pointwise sum* on X. If $\{s_n(x)\}$ is uniformly convergent on X to $s(x)$, then we say that the series $\sum_{k=1}^{\infty} f_k(x)$ is *uniformly convergent* on X and that s is its *uniform sum* on X.

Theorem 3 (Weierstrass M-test) *Let $X \subset \mathbb{R}$ and $f_n : X \to \mathbb{R}$. Suppose that there exists a sequence $\{M_n\}$ of nonnegative numbers such that $\sum_{n=1}^{\infty} M_n < \infty$ and $|f_n(x)| \leq M_n$ for $x \in X$ and $n \in \mathbb{N}$. Then $\sum_{n=1}^{\infty} f_n(x)$ is uniformly convergent.*

Theorem 4 (Abel's test) *Let $A \subset \mathbb{R}$ and $\phi_n : A \to \mathbb{R}$ be a decreasing sequence of functions, that is, $\phi_{n+1}(x) \leq \phi_n(x)$ for every $x \in A$. Suppose that there is a constant M such that $|\phi_n(x)| \leq M$ for every $x \in A$ and every n. If $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on A, then so does $\sum_{n=1}^{\infty} \phi_n(x) f_n(x)$.*

Theorem 5 (Dirichlet's test) *Let $s_n(x) = \sum_{k=1}^{n} f_k(x)$ for a sequence $f_k : A \to \mathbb{R}$. Assume that there is a constant M such that $|s_n(x)| \leq M$ for every $x \in A$ and every n. Let $g_n : A \to \mathbb{R}$ be such that $g_n \to 0$ uniformly, $g_n \geq 0$ and $g_{n+1}(x) \leq g_n(x)$. Then $\sum_{n=1}^{\infty} f_n(x) g_n(x)$ converges uniformly on A.*

5 DIFFERENTIATION OF FUNCTIONS OF ONE VARIABLE

DEFINITION Let a function f be defined on some open interval containing $x_0 \in \mathbb{R}$. We say that f is **differentiable** at x_0 if

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

exists. We call $f'(x_0)$ the derivative of f at x_0. Rewriting this condition as

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

we see that the straight line $y = f(x_0) + f'(x_0)(x - x_0)$, called the tangent line to the graph of f at x_0, is a good approximation to f near x_0, and rewriting it as

$$\lim_{\Delta x \to 0} \left[\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - f'(x_0) \right] = 0,$$

we see that $f'(x_0)$, being the limit of slopes of the secant lines, can be interpreted as the slope of the tangent line to the graph of f at $(x_0, f(x_0))$.

Proposition 1 *If f is differentiable at x_0, then f is continuous at x_0.*

Theorem 2 *Let $f, g : I \to \mathbb{R}$ be defined on an open interval I and differentiable at $x \in I$. Let $\alpha \in \mathbb{R}$. Then the functions αf, $f + g$, $f \cdot g$ and $\frac{f}{g}$ (provided $g \neq 0$) are differentiable at x. Moreover,*

(i) $(\alpha f)'(x) = \alpha f'(x)$,

(ii) $(f + g)'(x) = f'(x) + g'(x)$,

(iii) $(f \cdot g)'(x) = f(x)g'(x) + g(x)f'(x)$,

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

Theorem 3 (Chain rule) *Let $f : I \to J$ and $g : J \to \mathbb{R}$, where I and J are open intervals. Suppose that f is differentiable at $c \in I$ and that g is differentiable at $f(c)$. Then the composite function $g \circ f : I \to \mathbb{R}$ defined by $g \circ f(x) = g(f(x))$ is differentiable at c and*

$$(g \circ f)'(c) = g'(f(c))f'(c).$$

LOCAL EXTREMA

DEFINITION Let $f : E \to \mathbb{R}$, $E \subset \mathbb{R}$. We say that f has a **local maximum** (**local minimum**) at c if there exists a neighbourhood U of c such that $f(x) \leq f(c)$ ($f(x) \geq f(c)$) for every $x \in U$. If f has either a local maximum or a local minimum at c, we say that f has a **local extremum** at c.

The next theorem gives a necessary, but not sufficient, condition that a local extremum exists at a given point.

Theorem 4 *Let $a < c < b$ and $f : [a, b] \to \mathbb{R}$ be given. If f has a local extremum at c and $f'(c)$ exists, then $f'(c) = 0$.*

Remarks

(a) The restriction that c is not an endpoint of $[a, b]$ is necessary. For instance, the function $f(x) = \sqrt{x}$ on $[0, 1]$ has a local minimum at 0 and a local maximum at 1, $f'_+(0) = \infty$ and $f'_-(1) = \frac{1}{2}$.

(b) The function $f(x) = x^3$ on $(-1, 1)$ satisfies $f'(0) = 0$ but does not have a local extremum at 0.

(c) The theorem assures us that if we are seeking all local extrema of a differentiable function on an open interval, then we need only consider, as candidates, those c for which $f'(c) = 0$.

(d) If $f(x) = |x|$ for $x \in \mathbb{R}$, then f has a local minimum at $c = 0$, but $f'(0)$ does not exist.

MEAN VALUE THEOREMS

Theorem 5 (Rolle's theorem) *Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, differentiable on (a, b) and $f(a) = f(b)$. Then there exists a number $\xi \in (a, b)$ such that $f'(\xi) = 0$.*

Theorem 6 (Lagrange) *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$, and differentiable on (a, b). Then there exists a point $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$.*

Theorem 7 *Suppose that f is continuous on $[a, b]$ and differentiable on (a, b).*

(i) If $f'(x) \geq 0$ for every $x \in (a, b)$, then f is nondecreasing on $[a, b]$.

(ii) If $f'(x) \leq 0$ for every $x \in (a, b)$, then f is nonincreasing on $[a, b]$.

(iii) If $f'(x) > 0$ for every $x \in (a, b)$, then f is strictly increasing on $[a, b]$.

(iv) If $f'(x) < 0$ for every $x \in (a, b)$, then f is strictly decreasing on $[a, b]$.

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

Theorem 8 *Suppose that f is continuous on $[a, b]$ and is twice differentiable on (a, b), and that $x_0 \in (a, b)$.*

(i) If $f'(x_0) = 0$ and $f''(x_0) > 0$, then x_0 is a strict local minimum of f.

(ii) If $f'(x_0) = 0$ and $f''(x_0) < 0$, then x_0 is a strict local maximum of f.

6 INTEGRALS OF FUNCTIONS OF ONE VARIABLE

DEFINITION Let $f : [a, b] \to \mathbb{R}$ be a bounded function. We **partition** $[a, b]$, which means we choose an integer n and points $x_0, x_1, \ldots, x_{n-1}, x_n$ in such a way that $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$. Denote such a partition by P, that is, let $P = \{x_0, x_1, \ldots, x_n\}$. Then form two sums

$$U(f, P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}), \quad \text{where } M_i = \sup_{x \in [x_{i-1}, x_i]} f(x)$$

and

$$L(f, P) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}), \quad \text{where } m_i = \inf_{x \in [x_{i-1}, x_i]} f(x),$$

called the upper and lower Riemann sum (with respect to P), respectively.

Since f is bounded, say $-M \le f(x) \le M$ for every $x \in [a, b]$, we see that

$$-(b-a)M \le L(f, P) \le U(f, P) \le (b-a)M \tag{6.1}$$

for every partition P of $[a, b]$.

It seems reasonable to expect that as the size of the intervals in P gets smaller, $U(f, P)$ decreases while $L(f, P)$ increases.

DEFINITION If P and P' are partitions of $[a, b]$ with $P \subset P'$, then P' is called a *refinement* of P.

Lemma 1 *If P' is a refinement of P, then $L(f, P) \le L(f, P')$ and $U(f, P) \ge U(f, P')$.*

According to the inequality (6.1) Riemann sums are bounded: therefore we can introduce the following notation:

$$\overline{\int_a^b} f(x)\, dx = \inf\{U(f, P);\ P \text{ is any partition of } [a, b]\},$$

the upper Riemann integral, and

$$\underline{\int_a^b} f(x)\, dx = \sup\{L(f, P);\ P \text{ is any partition of } [a, b]\},$$

the lower Riemann integral.

Lemma 2 *Let P_1 and P_2 be any partitions of $[a, b]$. Then $L(f, P_1) \le U(f, P_2)$.*

Corollary 3

$$\underline{\int_a^b} f(x)\, dx \le \overline{\int_a^b} f(x)\, dx.$$

DEFINITION We say that $f : [a,b] \to \mathbb{R}$ is **Riemann integrable** or that the Riemann integral exists, if

$$\underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx.$$

The common value is denoted by $\int_a^b f(x)\,dx$.

Theorem 4 *A function $f : [a,b] \to \mathbb{R}$ is integrable on $[a,b]$ if given any $\varepsilon > 0$ there is a partition P such that*

$$U(f,P) - L(f,P) < \varepsilon.$$

Theorem 5

(i) If $f : [a,b] \to \mathbb{R}$ is bounded and continuous at all but finitely many points of $[a,b]$, then f is integrable on $[a,b]$.

(ii) Any increasing (decreasing) function on $[a,b]$ is integrable on $[a,b]$.

PROPERTIES OF INTEGRALS

Theorem 6

(i) If f is bounded and integrable on $[a,b]$ and $k \in \mathbb{R}$, then $k \cdot f$ is integrable on $[a,b]$ and

$$\int_a^b k \cdot f(x)\,dx = k \int_a^b f(x)\,dx.$$

(ii) If f and g are bounded and integrable on $[a,b]$, then $f+g$ is integrable on $[a,b]$ and

$$\int_a^b (f+g)\,dx = \int_a^b f\,dx + \int_a^b g\,dx.$$

(iii) If f and g are bounded and integrable on $[a,b]$ and $f(x) \le g(x)$ for every $x \in [a,b]$, then

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx.$$

(iv) If f is bounded and integrable on $[a,b]$ and $[b,c]$, then f is integrable on $[a,c]$ and

$$\int_a^c f\,dx = \int_a^b f\,dx + \int_b^c f\,dx.$$

Theorem 7 (Mean value theorem) *If f is continuous on $[a,b]$, then*

$$\int_a^b f(x)\,dx = f(c)(b-a)$$

for some $c \in [a, b]$.

Let $f : [a, b] \to \mathbb{R}$ be a continuous function.

DEFINITION An antiderivative of f is a continuous function $F : [a, b] \to \mathbb{R}$ such that F is differentiable on (a, b) and $F'(x) = f(x)$ for $a < x < b$.

Theorem 8 *Let* f *be bounded and integrable on* $[a, b]$. *Then* $|f|$ *is integrable on* $[a, b]$, *and*

$$\Big| \int_a^b f \, dx \Big| \leq \int_a^b |f| \, dx \, .$$

Theorem 9 (The fundamental theorem of Calculus) *Let* $f : [a, b] \to \mathbb{R}$ *be a continuous function. Then* f *has an antiderivative* F *and*

$$\int_a^b f(x) \, dx = F(b) - F(a).$$

If G *is any other antiderivative of* f, *we also have*

$$\int_a^b f(x) \, dx = G(b) - G(a).$$

Theorem 10 (Integration by parts) *If* $\frac{du}{dx}$ *and* $\frac{dv}{dx}$ *are continuous on* $[a, b]$, *then*

$$\int_a^b u \frac{dv}{dx} \, dx = u(b)v(b) - u(a)v(a) - \int_a^b \frac{du}{dx} v \, dx.$$

We often need to integrate unbounded functions or to integrate over unbounded regions. The resulting improper integrals lead to convergence problems analogous to those for an infinite series.

DEFINITION (**improper integrals - first kind**) Let $f : (a, b] \to \mathbb{R}$ and suppose that f is not necessarily bounded at a (near a) but f is integrable on $[a + \varepsilon, b]$ for every $\varepsilon > 0$ sufficiently small. We say that f is improperly integrable (or $\int_a^b f(x) \, dx$ exists) if

$$\lim_{\varepsilon \to 0^+} \int_{a+\varepsilon}^b f(x) \, dx \;\; \text{exists.}$$

If this limit exists it is denoted by $\int_a^b f(x) \, dx$ (that is, $\displaystyle\int_a^b f(x) \, dx = \lim_{\varepsilon \to 0^+} \int_{a+\varepsilon}^b f(x) \, dx$).

In a similar manner we define improper integrals $f : [a, b) \to \mathbb{R}$ (if f is unbounded near b).

DEFINITION (**improper integrals - second kind**) Let $f : [a, \infty) \to \mathbb{R}$ and suppose that $\int_a^b f(x) \, dx$ exists for every $b > a$. We say that f is improperly integrable if

$$\lim_{b \to \infty} \int_a^b f(x) \, dx \;\; \text{exists.}$$

If this limit exists, it is denoted by $\int_a^\infty f(x)\,dx$.

Similarly we define improper integrals for $f:\,(-\infty,a]\to\mathbb{R}$,

$$\int_{-\infty}^{a} f(x)\,dx = \lim_{b\to-\infty}\int_b^a f(x)\,dx.$$

If $f:\,(-\infty,\infty)\to\mathbb{R}$, we set

$$\int_{-\infty}^{\infty} f(x)\,dx = \lim_{a\to-\infty}\int_a^0 f(x)\,dx + \lim_{b\to\infty}\int_0^b f(x)\,dx.$$

The integral $\int_{-\infty}^{\infty} f(x)\,dx$ diverges if one of these limits does not exist.

Theorem 11 (Comparison test for improper integrals) *Suppose* $f(x)\geq 0$ *and* $g(x)\geq 0$ *for* $x\geq a$.

(i) If $g(x)\leq f(x)$, *then the convergence of* $\int_a^\infty f(x)\,dx$ *implies the convergence of* $\int_a^\infty g(x)\,dx$.

(ii) If $g(x)\geq f(x)$, *then the divergence of* $\int_a^\infty f(x)\,dx$ *implies the divergence of* $\int_a^\infty g(x)\,dx$.

The analogy between positive - term series and improper integrals of positive functions is the key to the **integral test**.

Theorem 12 *If* f *is continuous, nonnegative and nonincreasing on* $[1,\infty)$, *then* $\sum_{n=1}^{\infty} f(n)$ *and* $\int_1^\infty f(x)\,dx$ *converge or diverge together.*

7 TAYLOR SERIES

TAYLOR'S FORMULA

Theorem 1 (**Taylor's formula**) *Suppose that the first $(n+1)$ derivatives of the function f exist on an interval containing points a and b. Then*

$$\begin{aligned} f(b) &= f(a) + f'(a)(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \frac{f^{(3)}(a)}{3!}(b-a)^3 \\ &+ \ldots + \frac{f^{(n)}(a)}{n!}(b-a)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(b-a)^{n+1} \end{aligned} \tag{7.1}$$

for some number ξ between a and b.

REMARK *Taylor's formula with the Cauchy form of the remainder:*

$$\begin{aligned} f(b) &= f(a) + f'(a)(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(b-a)^n \\ &+ \frac{1}{n!}\int_a^b (b-t)^n f^{(n+1)}(t)\, dt. \end{aligned}$$

for some number t between a and b.

TAYLOR SERIES

Suppose that f is a function with continuous derivatives of all orders in an interval (c, d). Let $a \in (c, d)$ and let n be an arbitrary positive integer. We know by Taylor's formula that

$$f(x) = P_n(x) + R_n(x),$$

where

$$P_n(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2!} + \ldots + f^{(n)}(a)\frac{(x-a)^n}{n!}$$

and R_n is either the Lagrange or Cauchy remainder.
Now suppose that, for some particular fixed value of x, we can show that

$$\lim_{n\to\infty} R_n(x) = 0.$$

Then it follows from (7.1) that

$$\begin{aligned} f(x) &= \lim_{n\to\infty} P_n(x) = \lim_{n\to\infty}\left(\sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k\right) \\ &= \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k. \end{aligned}$$

The infinite series in this equation is called the **Taylor series** (of f at a).

EXAMPLES We have the following Taylor formulae for the exponential and trigonometric functions:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \frac{e^\xi}{(n+1)!} x^{n+1},$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + (-1)^n \frac{x^{2n}}{(2n)!} + (-1)^{n+1} \frac{\cos \xi}{(2n+2)!} x^{2n+2},$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + (-1)^{n+1} \frac{\cos \xi}{(2n+3)!} x^{2n+3},$$

In each case ξ is some number between 0 and x. Since ξ is between 0 and x, it follows that $0 < e^\xi \leq e^{|x|}$ in Taylor's formula for e^x. In the formulas for the sine and cosine functions, $0 \leq |\cos \xi| \leq 1$. Therefore the fact that

$$\lim_{n \to \infty} \frac{x^n}{n!} = 0 \quad \text{for all } x$$

implies that $\lim_{n \to \infty} R_n(x) = 0$ in all three cases above. This gives the following Taylor series:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots \quad ,$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots \quad ,$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots \quad .$$

8 VECTOR FUNCTIONS - FUNCTIONS OF SEVERAL VARIABLES

Addition and scalar multiplication of n-tuples are defined by

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and

$$a(x_1, \ldots, x_n) = (ax_1, \ldots, ax_n) \text{ for } a \in \mathbb{R}.$$

The length or norm of a vector x in $\mathbb{R}^n$ is defined by

$$|x| = \|x\| = \left\{ \sum_{i=1}^{n} x_i^2 \right\}^{\frac{1}{2}} .$$

The distance between two vectors x and y is defined by

$$|x-y| = \|x-y\| = \left\{ \sum_{i=1}^{n} (x_i - y_i)^2 \right\}^{\frac{1}{2}}.$$

The inner (scalar) product of x and y is defined by

$$(x, y) = \sum_{i=1}^{n} x_i y_i.$$

We have:

Theorem 1 *For x, y, $z \in \mathbb{R}^n$ there holds:*

(i) $(x, x) = |x|^2$,

(ii) $|(x, y)| \leq |x| \cdot |y|$ *(Cauchy - Schwarz inequality),*

(iii) $|x + y| \leq |x| + |y|$

(iv) $|x - y| \leq |x - z| + |z - y|$ *((iii) and (iv) are called triangle inequalities)*

DEFINITION Let S and T be given sets. A function $f : S \to T$ consists of two sets S and T together with a "rule" that assigns to each $x \in S$ a specific element of T, denoted by $f(x)$. One often writes $x \mapsto f(x)$ to denote that x is mapped to the element $f(x)$.

For a function $f : S \to T$, the set S is called the domain of F. The range, or image, of f is the subset of T defined by $f(S) = \{f(x) \in T;\ x \in S\}$.

If $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, we write

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)).$$

The f_i are called coordinate functions, or components of f.

Composite functions If two functions f and g are so related that the range space of f is the same as the domain space of g, we may form the composite function $g \circ f$ by first applying f and then g. Thus

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

for every vector x in the domain of f.

The operation of addition and multiplication of vector functions

Let f and g be functions with the same domain and having the same range space. Then the function $f + g$ is the sum of f and g defined by

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

for all x in the domain of both f and g.

Similarly, if $r \in \mathbb{R}$, then rf is the numerical multiple of f by r and is defined by $rf(x) = r \cdot f(x)$.

LIMITS AND CONTINUITY OF VECTOR FUNCTIONS

Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$. We use $|x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$, $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$.

DEFINITION Let $\Omega \subset \mathbb{R}^n$. Then a is a **limit point** of Ω if, for every $\varepsilon > 0$, there exists a point $y \in \Omega$ such that $0 < |a - y| < \varepsilon$.

In other words, the definition says that a is a limit point (or accumulation point) of Ω if there are points in Ω other than a that are contained in a ball of arbitrarily small radius with centre at a.

We come now to the definition of a limit for a function $f : \Omega \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$.

DEFINITION Let $y_0 \in \mathbb{R}^m$, and let $x_0 \in \mathbb{R}^n$ be a limit point of Ω. Then y_0 is the limit of f at x_0 if, for every $\varepsilon > 0$, there is a $\delta > 0$ such that $|f(x) - y_0| < \varepsilon$ whenever x satisfies $0 < |x - x_0| < \delta$ and $x \in \Omega$. (We write $\lim\limits_{x \to x_0} f(x) = y_0$)

Less formally, the definition says that for $x \neq x_0$, $f(x)$ can be made arbitrarily close to y_0 by choosing x sufficiently close to x_0.

Geometrically, the idea is this: given any ball $B(y_0, \varepsilon)$ in $\mathbb{R}^m$, there exists a ball $B(x_0, \delta) \subset \mathbb{R}^n$ whose intersection with Ω (the domain of f), except possibly for x_0 itself, is sent by f into $B(y_0, \varepsilon)$.

Theorem 2 *Let $f : \Omega \to \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ and let x^0 be a limit point of Ω. Then $\lim\limits_{x \to x^0} f(x) = y^0$ if and only if $\lim\limits_{x \to x^0} f_i(x) = y_i^0$, $i = 1, \ldots, m$.*

DEFINITION A function f is **continuous** at x_0 if x_0 is in the domain of f and $\lim\limits_{x \to x_0} f(x) = f(x_0)$.

At a nonlimit or isolated point of the domain, we cannot ask for a limit; instead we simply define f to be automatically continuous at such a point.

Theorem 3 *A vector function is continuous at a point x_0 if and only if its coordinate functions are continuous there.*

Theorem 4 *Every linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ is continuous on $\mathbb{R}^n$, and for such an L there is a number k such that*

$$|L(x)| \leq k|x| \quad \textit{for every} \quad x \in \mathbb{R}^n.$$

The continuity of many functions can be deduced from repeated applications of the following theorem:

Theorem 5

(1) The functions $P_k : \mathbb{R}^n \to \mathbb{R}$, where $P_k(x) = x_k$, (i.e. $P_k : (x_1, \ldots, x_n) \mapsto x_k$) are continuous for $k = 1, \ldots, n$.

(2) The functions $S : \mathbb{R}^2 \to \mathbb{R}$ and $M : \mathbb{R}^2 \to \mathbb{R}$, defined by $S(x, y) = x + y$ and $M(x, y) = xy$ are continuous.

(3) If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ are continuous, then the composition $g \circ f$ given by $g \circ f(x) = g(f(x))$ is continuous wherever it is defined.

9 DIFFERENTIABILITY OF VECTOR FUNCTIONS

DEFINITION Let $\Omega \subset \mathbb{R}^n$. A point $x_0 \in \Omega$ is an **i**nterior point of a set S if there exists a positive number δ such that $\{x : |x - x_0| < \delta\} \subset \Omega$ (equivalently $B(x_0, \delta) \subset \Omega$)
A subset of $\mathbb{R}^n$, all of whose points are interior, is called **o**pen.

Many of the techniques of calculus have as their foundation the idea of approximating a vector function by a linear function or by an affine function. Recall that a function $A : \mathbb{R}^n \to \mathbb{R}^m$ is affine if there exists a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ and a vector $y_0 \in \mathbb{R}^m$ such that

$$A(x) = L(x) + y_0 \quad \text{for every} \quad x \in \mathbb{R}^n.$$

EXAMPLE Consider a point $y_0 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and a linear function $L : \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$y = L(x) = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

equivalently,

$$\begin{aligned} y_1 &= x_1 + 2x_2 + x_3 \\ y_2 &= x_1 + 4x_2 + 5x_3. \end{aligned}$$

The affine function $A(x) = L(x) + y_0$ is defined by the equations

$$\begin{aligned} y_1 &= x_1 + 2x_2 + x_3 + 1 \\ y_2 &= 3x_1 + 4x_2 + 5x_3 + 3. \end{aligned}$$

We shall now study the possibility of approximating an arbitrary vector function $f : \Omega \to \mathbb{R}^m$, with $\Omega \subseteq \mathbb{R}^n$, near a point x_0 of Ω by an affine function A.

We begin by requiring that $f(x_0) = A(x_0)$. Since $A(x) = L(x) + y_0$, where L is linear, we obtain $f(x_0) = L(x_0) + y_0$ and so

(7.1) $$A(x) = L(x - x_0) + f(x_0).$$

A natural requirement is that

(7.2) $$\lim_{x \to x_0} \big(f(x) - A(x)\big) = 0.$$

We observe that from (7.1) we have

$$f(x) - A(x) = f(x) - f(x_0) - L(x - x_0).$$

Since L is continuous, (7.2) says that

$$0 = \lim_{x \to x_0} \big(f(x) - A(x)\big) = \lim_{x \to x_0} \big(f(x) - f(x_0)\big),$$

which is precisely the statement that f is continuous at x_0. This is significant, but it says nothing about L. Thus, in order for our notion of approximation to distinguish one affine function from another or to measure how well A approximates f, some additional requirement is necessary. We require that $f(x) - A(x)$ approaches 0 faster than x approaches x_0. That is, we demand that

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - L(x - x_0)}{|x - x_0|} = 0.$$

Equivalently, we can ask that f be representable in the form

$$f(x) = f(x_0) + L(x - x_0) + |x - x_0| z(x - x_0),$$

where $z(y)$ is some function that tends to 0 as y tends to 0.

DEFINITION A function $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is said to be **differentiable** at x_0, if

(1) x_0 is an interior point of the domain D of f,

(2) there is an affine function that approximates f near x_0. That is, there exists a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - L(x - x_0)}{|x - x_0|} = 0.$$

The linear function L is called the *differential* of f at x_0.

The function f is said to be differentiable if it is differentiable at every point of its domain.

If $n = m = 1$, an affine function has the form $ax+b$. Hence $f : \mathbb{R} \to \mathbb{R}$ that is differentiable at x_0 can be approximated near x_0 by a function $A(x) = ax+b$. Since $f(x_0) = A(x_0) = ax_0+b$, we obtain $b = f(x_0) - ax_0$ and

$$A(x) = ax + b = a(x - x_0) + f(x_0).$$

The linear part of A, denoted earlier by L, is $L(x) = ax$. The condition (2) of the definition becomes

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - a(x - x_0)}{|x - x_0|} = 0.$$

This is equivalent to

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = a.$$

The number a is commonly denoted by $f'(x_0)$ and it is the derivative of f at x_0. The affine function A is therefore given by

$$A(x) = f(x_0) + f'(x_0)(x - x_0).$$

Its graph is the tangent line to the graph of f at x_0.

General case: $n, m \geq 1$. A linear function $L : \mathbb{R}^n \to \mathbb{R}^m$ must be representable by an m - by - n matrix, that is, $L(x) = Ax$. We shall show that the matrix A of L satisfying (1) and (2) of the definition can be computed in terms of partial derivatives of f. To find the matrix A, we consider the standard basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$. If x_0 is an interior point of the domain of f, the vectors

$$x^j = x_0 + te_j, \quad j = 1, \ldots, m,$$

are all in the domain of f for sufficiently small t. By condition (2) of the definition we have

$$\lim_{t \to 0} \frac{f(x^j) - f(x_0) - L(te_j)}{t} = 0.$$

Since L is linear, we deduce from this

$$\lim_{t \to 0} \frac{f(x^j) - f(x_0)}{t} = L(e_j).$$

$L(e_j)$ is the jth column of the matrix A of L. On the other hand the vector x^j differs from x_0 only in the jth coordinate, that is,

$$\lim_{t \to 0} \frac{f(x^j) - f(x_0)}{t} = \begin{pmatrix} \frac{\partial f_1(x_0)}{\partial x^j} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f_m(x_0)}{\partial x^j} \end{pmatrix}$$

and the entire matrix of L has the form

$$\begin{bmatrix} \frac{\partial f_1(x_0)}{\partial x^1}, & \frac{\partial f_1(x_0)}{\partial x^2}, & \ldots, & \frac{\partial f_1(x_0)}{\partial x^n} \\ \frac{\partial f_2(x_0)}{\partial x^1}, & \frac{\partial f_2(x_0)}{\partial x^2}, & \ldots, & \frac{\partial f_2(x_0)}{\partial x^n} \\ \ldots & & & \\ \frac{\partial f_m(x_0)}{\partial x^1}, & \frac{\partial f_m(x_0)}{\partial x^2}, & \ldots, & \frac{\partial f_m(x_0)}{\partial x^n} \end{bmatrix}, \quad a_{ij} = \frac{\partial f_i(x_0)}{\partial x^j}.$$

This matrix is called the **Jacobian matrix** or the derivative of f at x_0, and it is denoted by $f'(x_0)$. It follows that L is uniquely determined by the partial derivatives $\frac{\partial f_i(x_0)}{\partial x^j}$.

The differential of L at x_0 is also denoted by $d_{x_0}f$ or $D_{x_0}f$.

We summarize what we have just proved as follows.

Theorem 1 *If the function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x_0, then the differential $d_{x_0}f$ is uniquely determined, and its matrix is the Jacobian matrix of f. That is, for every vector y in $\mathbb{R}^n$ we have*

$$D_{x_0}f(y) = f'(x_0)y.$$

We interpret $y = f(x_0) + f'(x_0)(x - x_0)$ as the equation of the tangent plane to the graph of f at $(x_0, f(x_0))$.

EXAMPLE The function

$$\begin{pmatrix} x^2 + e^y \\ x + y \sin z \end{pmatrix}$$

has coordinate functions $f_1(x, y, z) = x^2 + e^y$ and $f_2(x, y, z) = x + y \sin z$. The Jacobian matrix of f at (x, y, z) is

$$f'(x, y, z) = \begin{bmatrix} \frac{\partial f_1}{\partial x}, & \frac{\partial f_1}{\partial y}, & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x}, & \frac{\partial f_2}{\partial y}, & \frac{\partial f_2}{\partial z} \end{bmatrix} = \begin{bmatrix} 2x, & e^y, & 0 \\ 1, & \sin z, & y \cos z \end{bmatrix}.$$

The differential of f at $(1, 1, \pi)$ is

$$d_{(1,1,\pi)} f(x, y, z) = \begin{bmatrix} 2 & e & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2x + ey \\ x - z \end{bmatrix}.$$

The affine mapping that approximates f near $(1, 1, \pi)$ is

$$\begin{aligned} A(x, y, z) &= \begin{bmatrix} 2 & e & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x - 1 \\ y - 1 \\ z - \pi \end{bmatrix} + f(1, 1, \pi) \\ &= \begin{bmatrix} 1 + e \\ 1 \end{bmatrix} + \begin{bmatrix} 2 & e & 0 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x - 1 \\ y - 1 \\ z - \pi \end{bmatrix} \\ &= \begin{bmatrix} 1 + e + 2(x - 1) + e(y - 1) \\ 1 + (x - 1) - (z - \pi) \end{bmatrix} = \begin{bmatrix} 2x + ey - 1 \\ x - z + \pi \end{bmatrix}. \end{aligned}$$

Remark: If $f : \mathbb{R}^n \to \mathbb{R}$, then the Jacobian matrix is reduced to a gradient

$$f'(x) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}).$$

The notation $\nabla f(x)$ is also used. In this case the differential of f at x_0 is given by

$$L(x) = d_{x_0} f(x) = x_1 \frac{\partial f(x_0)}{\partial x_1} + x_2 \frac{\partial f(x_0)}{\partial x_2} + \ldots + x_n \frac{\partial f(x_0)}{\partial x_n}.$$

EXAMPLE Let $f(x) = \sum_{i=1}^{n} x_i^2 = |x|^2$. Show that f is differentiable at every point $x_0 \in \mathbb{R}^n$.
If f is differentiable at x_0, then the differential of f at x_0 must be given by

$$L(x) = d_{x_0} f(x) = \sum_{i=1}^{n} 2x_i^0 x_i = 2x_0 x.$$

Then

$$\begin{aligned} \frac{f(x) - f(x_0) - L(x - x_0)}{|x - x_0|} &= \frac{|x|^2 - |x_0|^2 - 2x_0(x - x_0)}{|x - x_0|} \\ = \frac{|x|^2 - |x_0|^2 - 2x_0 x + 2|x_0|^2}{|x - x_0|} &= \frac{|x|^2 + |x_0|^2 - 2x_0 x}{|x - x_0|} = \frac{|x - x_0|^2}{|x - x_0|} \\ = |x - x_0| \to 0 \;\text{ as }\; x \to x_0. \end{aligned}$$

Therefore f is differentiable at each point $x_0 \in \mathbb{R}^n$.

How one can tell whether or not a vector function is differentiable? We only know that if f is differentiable then the differential is represented by the Jacobian matrix.

Theorem 2 *If the domain of $f : \mathbb{R}^n \to \mathbb{R}^m$ is an open set $D \subset \mathbb{R}^n$ on which all partial derivatives $\frac{\partial f_i}{\partial x_j}$ are continuous, then f is differentiable at every point of D.*

For example

$$f(x_1, x_2, x_3) = \left(x_1^2 + x_1x_2x_3, e^{x_1+x_2+x_3}, \sin(x_1 + x_2 + x_3)\right)$$

is differentiable at every point of $\mathbb{R}^3$.

It follows from the definition of a differentiable vector function that:

Theorem 3 *If f is differentiable at x_0, then f is continuous at x_0.*

The converse is not true.

EXAMPLE Consider the function

$$f(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & \text{for } (x,y) \neq (0,0) \\ 0 & \text{for } (x,y) = (0,0). \end{cases}$$

Since $\lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{x^2+y^2}} = 0 = f(0,0)$, f is continuous at $(0,0)$. If f were differentiable at $(0,0)$, then

$$d_{(0,0)}f(x,y) = (f_x(0,0), f_y(0,0)).$$

Since

$$f_x(0,0) = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = 0$$

and

$$f_y(0,0) = \lim_{h\to 0} \frac{f(0,h) - f(0,0)}{h} = 0,$$

$d_{(0,0)}f(x,y) = 0$ for all $(x,y) \in \mathbb{R}^2$. On the other hand

$$\lim_{(x,y)\to(0,0)} \frac{f(x,y) - f(0,0) - d_{(0,0)}f(x,y)}{\sqrt{x^2+y^2}} = \lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2} = 0,$$

which is impossible as the function $\frac{xy}{x^2+y^2}$ does not have a limit at the origin.

Notice that f has partial derivatives at $(0,0)$: $\frac{\partial f(0,0)}{\partial x} = 0$ and $\frac{\partial f(0,0)}{\partial y} = 0$. This means that the gradient (the Jacobian matrix) exists. However, this does not guarantee the differentiability of f at $(0,0)$.
Let $f : \mathbb{R}^n \to \mathbb{R}$, $x_0 \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$ be a unit vector. This means $|u| = 1$, where

$$u = (u_1, \ldots, u_n), \quad |u| = \left(\sum_{i=1}^{n} u_i^2\right)^{\frac{1}{2}}.$$

DEFINITION The directional derivative of f at x_0 in the direction u, denoted by $\frac{\partial f(x_0)}{\partial u}$ or $D_u f(x_0)$, is defined by

$$\frac{\partial f(x_0)}{\partial u} = \lim_{t \to 0} \frac{f(x_0 + tu) - f(x_0)}{t}.$$

From this definition we see that the directional derivative is the rate of change of f in the direction u.

Theorem 4 *If f is differentiable at x, then*

$$\frac{\partial f(x)}{\partial u} = f'(x)u$$

for every unit vector in $\mathbb{R}^n$.

EXAMPLE Let $f(x_1, x_2, x_3) = x_1^2 + x_3^2 + e^{x_2}$, $u = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$.
Then

$$\begin{aligned} \frac{\partial f(1,1,1)}{\partial u} &= \frac{\partial f(1,1,1)}{\partial x_1}u_1 + \frac{\partial f(1,1,1)}{\partial x_2}u_2 + \frac{\partial f(1,1,1)}{\partial x_3}u_3 \\ &= 2 \cdot \frac{1}{2} + e \cdot \frac{1}{2} + 2 \cdot \frac{1}{\sqrt{2}} = 1 + \frac{e}{2} + \sqrt{2}. \end{aligned}$$

EXAMPLE Show that the existence of all directional derivatives at a point x does not imply the differentiability at this point.
Let

$$f(x, y) = \begin{cases} \frac{x|y|}{\sqrt{x^2+y^2}} & \text{for } (x, y) \neq (0, 0), \\ 0 & \text{for } (x, y) = (0, 0). \end{cases}$$

We know that f is not differentiable at $(0, 0)$. However, $\frac{\partial f(0,0)}{\partial u}$ exists at each direction $u = (u_1, u_2)$. Indeed,

$$\frac{\partial f(0,0)}{\partial u} = \lim_{t \to 0} \frac{tu_1|tu_2|}{t\sqrt{t^2u_1^2 + t^2u_2^2}} = \lim_{t \to 0} \frac{t|t|u_1|u_2|}{t|t|} = u_1|u_2|.$$

We have that $\frac{\partial f(x_0)}{\partial u}$ is the slope of the tangent line at $(x_0, f(x_0))$ to the curve formed by the intersection of the graph of f with the plane that contains $(x_0 + tu)$ and x_0, and is parallel to the z-axis.

We always have

$$\frac{\partial f(x_0)}{\partial u} = \nabla f(x_0)u \leq |\nabla f(x_0)||u| = |\nabla f(x_0)|.$$

For $u = \frac{\nabla f(x_0)}{|\nabla f(x_0)|}$, we have

$$\frac{\partial f(x_0)}{\partial u} = \frac{\nabla f(x_0) \cdot \nabla f(x_0)}{|\nabla f(x_0)|} = \frac{|\nabla f(x_0)|^2}{|\nabla f(x_0)|} = |\nabla f(x_0)|.$$

This shows that the rate of change of $\frac{\partial f(x_0)}{\partial u}$ is never greater than $|\nabla f(x_0)|$ and is equal to it in the direction of the gradient.

CHAIN RULE

As in the one-dimensional case, the chain rule is a rule for differentiating composite functions.

Theorem 5 *Let g be a continuously differentiable function on an open set $\Omega \subset \mathbb{R}^n$ and let f be defined and differentiable for $a < t < b$, taking its values in Ω. Then the composite function $F(t) = g(f(t))$ is differentiable for $a < t < b$ and*

$$F'(t) = \nabla g(f(t)) \cdot f'(t).$$

EXAMPLE Let $g(x, y) = x^2 y + e^{x+y}$ for $(x, y) \in \mathbb{R}^2$ and let $f(t) = (t, t^2)$. Then

$$f'(t) = (1, 2t), \quad \nabla g(x, y) = (2xy + e^{x+y}, x^2 + e^{x+y})$$

and

$$\begin{aligned} F'(t) &= (2t^3 + e^{t+t^2}, t^2 + e^{t+t^2}) \cdot (1, 2t) \\ &= 2t^3 + e^{t+t^2} + 2t^3 + 2te^{t+t^2} \\ &= 4t^3 + e^{t+t^2} + 2te^{t+t^2}. \end{aligned}$$

The following theorem gives the extension to any dimension for the domain and range of g and f.

Theorem 6 (the Chain Rule) *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable at x and let $g : \mathbb{R}^m \to \mathbb{R}^p$ be continuously differentiable at $f(x)$. If $g \circ f$ is defined on an open set containing x, then $g \circ f$ is continuously differentiable at x, and*

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

Proof The matrices here have the form

$$\begin{bmatrix} \frac{\partial g_1(f(x))}{\partial y_1} & ,\ldots, & \frac{\partial g_1(f(x))}{\partial y_m} \\ & \cdots & \\ \frac{\partial g_p(f(x))}{\partial y_1} & ,\ldots, & \frac{\partial g_p(f(x))}{\partial y_m} \end{bmatrix} \text{ and } \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & ,\ldots, & \frac{\partial f_1(x)}{\partial x_n} \\ & \cdots & \\ \frac{\partial f_m(x)}{\partial x_1} & ,\ldots, & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

The product of the matrices has as its ij th entry the sum of products

$$\sum_{k=1}^{m} \frac{\partial g_i(f(x))}{\partial y_k} \frac{\partial f_k(x)}{\partial x_j}.$$

This expression is the scalar product of two vectors $\nabla g_i(f(x))$ and $\frac{\partial f(x)}{\partial x_j}$. It follows from Theorem 5 that

$$\nabla g_i(f(x)) \frac{\partial f(x)}{\partial x_j} = \frac{\partial (g_i \circ f)(x)}{\partial x_j},$$

because we differentiate with respect to the single variable x_j. □

EXAMPLES

(1) Let $f(x,y) = (x^2+y^2, x^2-y^2)$ and $g(u,v) = (uv, u+v)$. Find $(g\circ f)'(2,1)$. First we compute g' and f'

$$g'(u,v) = \begin{bmatrix} v, & u \\ 1, & 1 \end{bmatrix}, \quad f'(x,y) = \begin{bmatrix} 2x, & 2y \\ 2x, & -2y \end{bmatrix}.$$

To find $(g\circ f)'(2,1)$, we note that $f(2,1) = (5,3)$,

$$g'(5,3) = \begin{bmatrix} 3, & 5 \\ 1, & 1 \end{bmatrix} \quad f'(2,1) = \begin{bmatrix} 4, & 2 \\ 4, & -2 \end{bmatrix}.$$

Then the product of the matrices $g'(5,3)$ and $f'(2,1)$ gives

$$(g\circ f)(2,1) = \begin{bmatrix} 3, & 5 \\ 1, & 1 \end{bmatrix} \cdot \begin{bmatrix} 4, & 2 \\ 4, & -2 \end{bmatrix} = \begin{bmatrix} 12+20, & 6-10 \\ 4+4, & 2-2 \end{bmatrix} = \begin{bmatrix} 32, & -4 \\ 8, & 0 \end{bmatrix}.$$

(2) Let $w = g(x,y,z) : \mathbb{R}^3 \to \mathbb{R}$ and

$$(x,y,z) = f(s,t) = \big(f_1(s,t), f_2(s,t), f_3(s,t)\big) : \mathbb{R}^2 \to \mathbb{R}^3.$$

Compute the partial derivatives $\frac{\partial w}{\partial s}$ and $\frac{\partial w}{\partial t}$ of the composite function

$$w = g\big(f_1(s,t), f_2(s,t), f_3(s,t)\big).$$

By the chain rule we have

$$(\frac{\partial w}{\partial s}, \frac{\partial w}{\partial t}) = (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}) \cdot \begin{bmatrix} \frac{\partial x}{\partial s}, & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s}, & \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial s}, & \frac{\partial z}{\partial t} \end{bmatrix}.$$

Matrix multiplication yields

$$\begin{aligned} \frac{\partial w}{\partial s} &= \frac{\partial g}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial s} \\ \frac{\partial w}{\partial t} &= \frac{\partial g}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial t}. \end{aligned}$$

(3) Let $g : \mathbb{R}^3 \to \mathbb{R}^3$, $f : \mathbb{R}^2 \to \mathbb{R}^3$. We introduce notation

$$w = (w_1, w_2, w_3) = \big(g_1(x,y,z), g_2(x,y,z), g_3(x,y,z)\big),$$
$$(x,y,z) = \big(f_1(s,t), f_2(s,t), f_3(s,t)\big).$$

Compute $\frac{\partial w_1}{\partial s}$, $\frac{\partial w_1}{\partial t}$, $\frac{\partial w_2}{\partial s}$, $\frac{\partial w_2}{\partial t}$, $\frac{\partial w_3}{\partial s}$ and $\frac{\partial w_3}{\partial t}$. We have

$$\begin{bmatrix} \frac{\partial w_1}{\partial s}, & \frac{\partial w_1}{\partial t} \\ \frac{\partial w_2}{\partial s}, & \frac{\partial w_2}{\partial t} \\ \frac{\partial w_3}{\partial s}, & \frac{\partial w_3}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial x}, & \frac{\partial g_1}{\partial y}, & \frac{\partial g_1}{\partial z} \\ \frac{\partial g_2}{\partial x}, & \frac{\partial g_2}{\partial y}, & \frac{\partial g_2}{\partial z} \\ \frac{\partial g_3}{\partial x}, & \frac{\partial g_3}{\partial y}, & \frac{\partial g_3}{\partial z} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial f_1}{\partial s}, & \frac{\partial f_1}{\partial t} \\ \frac{\partial f_2}{\partial s}, & \frac{\partial f_2}{\partial t} \\ \frac{\partial f_3}{\partial s}, & \frac{\partial f_3}{\partial t} \end{bmatrix}.$$

Matrix multiplication yields, for $j = 1,2,3$:

$$\begin{aligned} \frac{\partial w_j}{\partial s} &= \frac{\partial g_j}{\partial x}\frac{\partial f_1}{\partial s} + \frac{\partial g_j}{\partial y}\frac{\partial f_2}{\partial s} + \frac{\partial g_j}{\partial z}\frac{\partial f_3}{\partial s}, \\ \frac{\partial w_j}{\partial t} &= \frac{\partial g_j}{\partial x}\frac{\partial f_1}{\partial t} + \frac{\partial g_j}{\partial y}\frac{\partial f_2}{\partial t} + \frac{\partial g_j}{\partial z}\frac{\partial f_3}{\partial t}. \end{aligned}$$

10 THE IMPLICIT AND INVERSE FUNCTION THEOREMS

THE IMPLICIT FUNCTION THEOREM

An equation in two variables x and y may have one or more solutions for y in terms of x or for x in terms of y. We say that these solutions are functions implicitly defined by the equation.
For example the equation of the unit circle , $x^2 + y^2 = 1$, implicitly defines four function (among others):

$$\begin{aligned} y &= \sqrt{1-x^2} \text{ for } x \in [-1,1], \\ y &= -\sqrt{1-x^2} \text{ for } x \in [-1,1], \\ x &= \sqrt{1-y^2} \text{ for } y \in [-1,1], \\ x &= -\sqrt{1-y^2} \text{ for } y \in [-1,1]. \end{aligned}$$

In general case, we consider a function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, and study the relation $F(x, y) = 0$, or, written out,

$$F_1(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0$$

$$\ldots$$

$$F_m(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0.$$

The goal is to solve for the m unknowns $y_1, \ldots, y_m$ from the m equations in terms of $x_1, \ldots, x_n$.

Theorem 1 (The Implicit Function Theorem) *Let $\Omega \subset \mathbb{R}^n \times \mathbb{R}^m$ be an open set, and let $F : \Omega \to \mathbb{R}^m$ be function of class C^1. Suppose $(x_0, y_0) \in \Omega$ and $F(x_0, y_0) = 0$. Assume that*

$$\det \begin{bmatrix} \frac{\partial F_1}{\partial y_1}, & \ldots, & \frac{\partial F_1}{\partial y_m} \\ & \ldots & \\ \frac{\partial F_m}{\partial y_1}, & \ldots, & \frac{\partial F_m}{\partial y_m} \end{bmatrix} \neq 0 \text{ evaluated at } (x_0, y_0),$$

where $F = (F_1, \ldots, F_m)$. Then there are open sets $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$, with $x_0 \in U$ and $y_0 \in V$, and a unique function $f : U \to V$ such that

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

for all $x \in U$. Furthermore, f is of class C^1.

THE INVERSE FUNCTION THEOREM

If a function f is thought of as sending vectors x into vectors y in the range of f, then it is natural to start with y and ask what vector or vectors x are sent by f into y.

More particularly, we may ask if there is a function that reverses the action of f. If there is a function f^{-1} with the property

$$f^{-1}(y) = x \text{ if and only if } f(x) = y,$$

then f^{-1} is called the inverse function of f. It follows that the domain of f^{-1} is the range of f and that the range of f^{-1} is the domain of f.

Given a function $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$, where $\Omega \subseteq \mathbb{R}^n$, one may ask:

(1) Does it have an inverse?

(2) If it does, what are its properties?

In general it is not easy to answer these questions just by looking at the function. However, in certain circumstances we get a useful result.

Theorem 2 (The inverse function theorem) *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable function such that $f'(x_0)$ has an inverse. Then there is an open set Ω containing x_0 such that f, when restricted to Ω, has a continuously differentiable inverse. The image set $f(\Omega)$ is open. In addition,*

$$\left[f^{-1}(y_0)\right]' = \left[f'(x_0)\right]^{-1},$$

where $y_0 = f(x_0)$. That is, the differential of the inverse function at y_0 is the inverse of the differential of f at x_0.