

# COMP40018 Calculus

## Imperial College London

Boxuan Tang

Autumn 2021

### Contents

<b>1 Sequences</b>	<b>2</b>
1.1 Convergence Tests for Sequences . . . . .	4
<b>2 Continuous Functions</b>	<b>5</b>
<b>3 Integration</b>	<b>7</b>
<b>4 Series</b>	<b>9</b>
4.1 Convergence Tests for Series . . . . .	10
<b>5 Differentiation</b>	<b>11</b>
<b>6 Power Series</b>	<b>13</b>
<b>7 Numerical Methods</b>	<b>14</b>
<b>8 Metric Spaces</b>	<b>15</b>
<b>9 Deep Learning and Multivariate Chain Rule</b>	<b>18</b>

## 1 Sequences

**Definition 1.1 (Sequence)** A sequence is a function  $f : \mathbb{N}^+ \rightarrow \mathbb{R}$  that maps positive natural numbers to real numbers, written as  $(a_n)_{n \geq 1}$  where  $a_n = f(n)$ .

**Definition 1.2 (Arithmetic Sequence)** An arithmetic sequence is the sequence  $f : \mathbb{N}^+ \rightarrow \mathbb{R}$  defined by

$$f : n \mapsto \begin{cases} a_1, & n = 1 \\ a_1 + (n - 1)d, & \text{otherwise} \end{cases}$$

$$S_n = \frac{n}{2}(a_1 + a_n)$$

**Definition 1.3 (Geometric Sequence)** A geometric sequence is the sequence  $f : \mathbb{N}^+ \rightarrow \mathbb{R}$  defined by

$$f : n \mapsto ar^{n-1}$$

$$S_n = \frac{a(1 - r^n)}{1 - r}$$

**Definition 1.4 (Fibonacci Sequence)** A fibonacci sequence is the sequence  $f : \mathbb{N}^+ \rightarrow \mathbb{R}$  defined by

$$f : n \mapsto \begin{cases} 0, & n = 1 \\ 1, & n = 2 \\ f(n - 1) + f(n - 2) & n \geq 3 \end{cases}$$

**Definition 1.5 (Monotonic)** A sequence is increasing if  $a_{n+1} \geq a_n$  for  $n \geq 1$  and decreasing if  $a_{n+1} \leq a_n$  for  $n \geq 1$ . A sequence is monotonic if it is either increasing or decreasing.

**Theorem 1.6 (Triangle Inequality and Reverse)**

$$|a + b| \leq |a| + |b| \text{ AND } |a - b| \geq ||a| - |b||$$

**Definition 1.7 (Cauchy Sequences)** A sequence is a Cauchy sequence iff for all  $\epsilon > 0$ , there exists  $N \in \mathbb{N}$  such that for all  $n, m > N$  we have  $|a_n - a_m| < \epsilon$ .

- Every sequence that converges to a real number is a Cauchy sequence.

**Definition 1.8 (Completeness)** A subset  $A \subseteq \mathbb{R}$  is complete iff any Cauchy sequence in  $A$  converges to a limit in  $A$ .

- $\mathbb{Q}$  is not complete but  $\mathbb{R}$  is complete.

**Definition 1.9 (Subsequence)** If  $f : \mathbb{N} \rightarrow \mathbb{R}$  is a sequence  $(a_n)_{n \geq 1}$  and  $M$  is an infinite subset of  $\mathbb{N}$ , then  $g : M \rightarrow \mathbb{R}$  is a subsequence  $(a_{n_i})_{i \geq 1}$  of  $f$ .

- Any subsequence converges to the limit of the sequence.
- Any sequence of real numbers has a monotonic subsequence.

**Theorem 1.10 (Order Theory)** Let  $X \subseteq \mathbb{R}$  and  $l, u, s$  and  $i \in \mathbb{R}$

1.  $u$  is an **upper bound** of  $X$  if  $x \leq u$  for all  $x \in X$
2.  $l$  is an **lower bound** of  $X$  if  $l \leq x$  for all  $x \in X$
3.  $s$  is the **supremum** (least upper bound) of  $X$  if  $s \leq u$  for all  $u$  of  $X$
4.  $i$  is the **infimum** (greatest lower bound) of  $X$  if  $l \leq i$  for all  $l$  of  $X$
5.  $X$  is **bounded above** if  $X$  has an upper bound
6.  $X$  is **bounded below** if  $X$  has a lower bound
7.  $X$  is **bounded** if  $X$  has an upper and lower bound

**Theorem 1.11 (Dedekind-completeness of  $\mathbb{R}$ )** Every nonempty subset of  $\mathbb{R}$  that is bounded above has a supremum (least upper bound).

**Theorem 1.12 (Fundamental Theorem of Analysis)**

If  $(a_n)_{n \geq 1}$  is increasing and bounded above, then  $s = \sup\{a_n | n \geq 1\}$  exists and is the limit of  $(a_n)_{n \geq 1}$ .

If  $(a_n)_{n \geq 1}$  is decreasing and bounded below, then  $i = \inf\{a_n | n \geq 1\}$  exists and is the limit of  $(a_n)_{n \geq 1}$ .

**Theorem 1.13 (Bolzano-Weierstrass Theorem)** Every bounded sequence in  $\mathbb{R}$  has a convergent subsequence.

### 1.1 Convergence Tests for Sequences

**Definition 1.14 (Convergence to a limit)** Let  $(a_n)_{n \geq 1}$  be a sequence which **converges to a limit  $l$  in  $\mathbb{R}$**  iff for all  $\epsilon > 0$ , there exists an  $N \in \mathbb{N}$  such that for all  $n > N$  we have  $|a_n - l| < \epsilon$

**converges to  $+\infty$**  iff for all  $r$  in  $\mathbb{R}$  there exists an  $N \in \mathbb{N}$  such that for all  $n \geq N$  we have  $a_n > r$

**converges to  $-\infty$**  iff for all  $r$  in  $\mathbb{R}$  there exists an  $N \in \mathbb{N}$  such that for all  $n \geq N$  we have  $a_n < r$

written as  $\lim_{n \rightarrow \infty} a_n = x$  or  $(a_n)_{n \geq 1} \rightarrow x$ , where  $x$  is the unique limit

**diverges** if it does not converge to a real number,  $\infty$  or  $-\infty$

**Corollary 1.15 (Common convergent sequences)**

$$\left(\frac{1}{n^c}\right)_{n \geq 1} \rightarrow 0, \text{ when } c > 0$$

$$\left(\frac{1}{c^n}\right)_{n \geq 1} \rightarrow 0, \text{ when } |c| > 1 \text{ OR } (c^n)_{n \geq 1} \rightarrow 0, \text{ when } |c| < 1$$

$$\left(\frac{1}{n!}\right)_{n \geq 1} \rightarrow 0$$

$$\left(\frac{1}{\log n}\right)_{n \geq 1} \rightarrow 0$$

**Theorem 1.16 (Limits of combination of sequences)**

Given  $(a_n)_{n \geq 1} \rightarrow a$  and  $(b_n)_{n \geq 1} \rightarrow b$  with a real constant  $\lambda$

$$(\lambda a_n)_{n \geq 1} \rightarrow \lambda a$$

$$(a_n + b_n)_{n \geq 1} \rightarrow a + b$$

$$(a_n b_n)_{n \geq 1} \rightarrow ab$$

$$\left(\frac{a_n}{b_n}\right)_{n \geq 1} \rightarrow \frac{a}{b} \text{ provided } b \neq 0$$

**Theorem 1.17 (Sandwich Theorem)** Let  $(l_n)_{n \geq 1} \rightarrow l$  and  $(u_n)_{n \geq 1} \rightarrow l$  for some real number  $l$ . If for  $(a_n)_{n \geq 1}$  we have some  $N \in \mathbb{N}$  such that  $l_n \leq a_n \leq u_n$  for all  $n \geq N$ , then  $(a_n)_{n \geq 1} \rightarrow l$  as well.

**Theorem 1.18 (Ratio Test for Sequences)** Let  $c \in \mathbb{R}$  such that  $0 \leq c \leq 1$ . If there exists  $N \in \mathbb{N}$  such that for all  $n \geq N$  we have  $\left|\frac{a_{n+1}}{a_n}\right| \leq c$ , then  $(a_n)_{n \geq 1} \rightarrow 0$ .

**Theorem 1.19 (Limit Ratio Test for Sequences)** Let  $c \in \mathbb{R}$  such that  $0 \leq c \leq 1$ . If there exists  $N \in \mathbb{N}$  such that for all  $n \geq N$  we have  $\left|\frac{a_{n+1}}{a_n}\right| \leq c$ , then  $(a_n)_{n \geq 1} \rightarrow 0$ .

## 2 Continuous Functions

**Definition 2.1 (Neighbourhood)** A set  $A \subseteq \mathbb{R}$  is called a neighbourhood of  $a$  if there exists an open interval  $I$  where  $a \in I \subseteq A$

- An open interval is a neighbourhood of each of its points

**Definition 2.2 (Accumulation Point)** A real number  $\xi$  is an accumulation point of a set  $A \subseteq \mathbb{R}$  if every neighbourhood of  $\xi$  contains an infinite number of members of  $A$ .

**Definition 2.3 (Limit of a Function)**  $f : A \rightarrow \mathbb{R}$ ,  $A \subseteq \mathbb{R}$ , has a limit  $l \in \mathbb{R}$

at the accumulation point  $x_0$  of  $A$  if for all  $\epsilon > 0$  there exists  $\delta > 0$  such that if  $x \in A$  and  $|x - x_0| < \delta$ , then  $|f(x) - l| < \epsilon$

as  $x$  approaches  $+\infty$  if for all  $\epsilon > 0$  there exists  $c$  such that if  $x > c$  then  $|f(x) - l| < \epsilon$

as  $x$  approaches  $-\infty$  if for all  $\epsilon > 0$  there exists  $c$  such that if  $x < c$  then  $|f(x) - l| < \epsilon$

written as  $\lim_{x \rightarrow x_0} f(x) = l$  or  $f(x) \rightarrow l$  as  $x \rightarrow x_0$ , where  $l$  is the unique limit

- Let  $f : I \rightarrow \mathbb{R}$  where  $I \subseteq \mathbb{R}$  is an open interval and  $x_0$  is an accumulation point of  $I$ , then  $\lim_{x \rightarrow x_0} f(x) = l \in \mathbb{R}$  iff for all sequences of points of  $I$  with  $(y_n)_{n \geq 1} \rightarrow x_0$  we have  $\lim_{n \rightarrow \infty} f(y_n) = l$

**Theorem 2.4 (Limits of combination of sequences)**

Given  $f, g : A \rightarrow \mathbb{R}$  have limits  $k, l \in \mathbb{R}$  at accumulation point  $x_0$  of  $A$ ,

$f \pm g$  has limit  $k \pm l$  at  $x_0$

$fg$  has limit  $kl$  at  $x_0$

$\frac{f}{g}$  has limit  $\frac{k}{l}$  at  $x_0$  if  $l \neq 0$

**Lemma 2.5 (Axiom of Choice)** For any collection  $\chi$  of nonempty sets, there exists a choice function  $f$  that maps each set of  $\chi$  to an element of that set.

**Lemma 2.6 (Axiom of Countable Choice)** Let  $(S_n)_{n \in \mathbb{N}}$  be a sequence of nonempty sets, then there exists a sequence  $(x_n)_{n \in \mathbb{N}}$  such that  $x_n \in S_n$  for all  $n \in \mathbb{N}$ .

**Definition 2.7 (Continuity of functions)**  $f : [a, b] \rightarrow \mathbb{R}$ , where  $x \in \mathbb{R}$ , is

continuous at  $x_0 \in [A, B]$  iff  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$  OR for every  $\epsilon > 0$  there exists  $\delta > 0$  such that for all  $x$ , if  $|x - x_0| < \delta$

continuous in  $[A, B]$  iff  $f$  is continuous at all  $x_0 \in [A, B]$

**Theorem 2.8 (Combination of continuous functions)**

Given  $f, g : A \rightarrow \mathbb{R}$  are continuous at  $x_0$ ,

$f \pm g$  is continuous at  $x_0$

$fg$  is continuous at  $x_0$

$\frac{f}{g}$  is continuous at  $x_0$  if  $g(x_0) \neq 0$

**Theorem 2.9 (Composition of continuous functions)** If  $g$  is continuous at  $x_0$  and  $f$  is continuous at  $g(x_0)$ , then  $f \circ g$  is continuous at  $x_0$ . Note that  $f$  need not be continuous at  $x_0$ .

**Theorem 2.10 (Maxima and Minima)** If  $f : [a, b] \rightarrow \mathbb{R}$  with  $a, b \in \mathbb{R}$ , then there exists  $r, s \in [a, b]$  such that  $f(r) = \sup_{x \in [a, b]} f(x)$  and  $f(s) = \inf_{x \in [a, b]} f(x)$ .

- A continuous function on a closed bounded interval is bounded and attains their supremum and infimum

**Theorem 2.11 (Intermediate Value Theorem)** If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous with  $s \in \mathbb{R}$  such that  $f(a) < s < f(b)$ , then there exists  $c \in (a, b)$  such that  $f(c) = s$ .

**Definition 2.12 (Uniform Continuity)**  $f : A \rightarrow \mathbb{R}$  is uniformly continuous if for all  $\epsilon > 0$  there exists  $\delta > 0$  such that for all  $x, x_0 \in A$  we have  $|f(x) - f(x_0)| < \epsilon$  if  $|x - x_0| < \delta$ . In other words,  $\delta$  is independent of  $x_0$ .

**Theorem 2.13** If  $f : [a, b] \rightarrow \mathbb{R}$ , for  $a, b \in \mathbb{R}$ , is continuous then it is uniformly continuous on  $[a, b]$ .

### 3 Integration

**Definition 3.1 (Partition)** A partition  $P$  of  $[a, b]$  is given by the finite set

$$P = \{r_i : 0 \leq i \leq n-1, a = r_0, b = r_n, r_i < r_{i+1}\}$$

**subinterval** of  $P$  is a closed interval  $[r_i, r_{i+1}]$  for  $0 \leq i \leq n-1$

**norm** of  $P$  is the largest length of the subintervals in  $P$ , or

$$\|P\| = \max\{r_{i+1} - r_i : 0 \leq i \leq n-1\}$$

$P_2$  **refines**  $P_1$  if  $P_1 \subset P_2$

**Definition 3.2 (Sums)** Given  $f : [a, b] \rightarrow \mathbb{R}$  and a partition  $P$  of  $[a, b]$ , the **Lower Sum**  $L(f, P)$  and **Upper Sum**  $U(f, P)$  are defined as

$$L(f, P) = \sum_{i=0}^{n-1} (r_{i+1} - r_i) \times \inf_{x \in [r_i, r_{i+1}]} f(x), \quad U(f, P) = \sum_{i=0}^{n-1} (r_{i+1} - r_i) \times \sup_{x \in [r_i, r_{i+1}]} f(x)$$

**Riemann Sum** for  $P$  for any choice of  $s_i \in [r_i, r_{i+1}]$  for  $0 \leq i \leq n-1$  is

$$S(f, P, (s_i)_{0 \leq i \leq n-1}) = \sum_{i=0}^{n-1} (r_{i+1} - r_i) \times f(s_i)$$

- $L(f, P) \leq S(f, P, (s_i)_{0 \leq i \leq n-1}) \leq U(f, P)$
- If  $P_1 \subset P_2$ , then  $L(f, P_1) \leq L(f, P_2) \leq U(f, P_2) \leq U(f, P_1)$

**Definition 3.3 (Integrals)** **Lower** and **Upper** integrals of  $f : [a, b] \rightarrow \mathbb{R}$  are

$$\int_a^b f(x) dx = \sup_P L(f, P), \quad \overline{\int_a^b} f(x) dx = \inf_P U(f, P)$$

**Riemann Integral**  $\int_a^b f(x) dx$  exists if  $\int_a^b f(x) dx = \overline{\int_a^b} f(x) dx$ .

$f : [a, b] \rightarrow \mathbb{R}$  is Riemann integrable with Riemann integral  $c \in \mathbb{R}$  iff

- for each  $\epsilon > 0$  there exists a partition  $P$  of  $[a, b]$  with  $c - L(f, P) < \epsilon$  and  $U(f, P) - c < \epsilon$
- for each  $\epsilon > 0$  there exists a  $\delta > 0$  such that for all partitions  $P$  of  $[a, b]$  with  $\|P\| < \delta$  we have  $|S(f, P, (s_i)_{0 \leq i \leq n-1}) - c| < \epsilon$

**Theorem 3.4** A bounded function  $f : [a, b] \rightarrow \mathbb{R}$  with a countable set of discontinuities on  $[a, b]$  is Riemann integrable.

- If  $f$  is continuous on  $[a, b]$  then the Riemann integral  $\int_a^b f(x) dx$  exists

**Theorem 3.5 (Properties of Riemann Integrals)**

$$\int_a^b s f(x) + t g(x) dx = s \int_a^b f(x) dx + t \int_a^b g(x) dx, \text{ if } f \text{ and } g \text{ are integrable}$$

$$\int_a^b c dx = c(b - a), \text{ where } c \in \mathbb{R}$$

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx, \text{ if the integrals exists for } a < b < c$$

$$\text{If } f(x) \geq 0 \text{ for } x \in [a, b], \text{ then } \int_a^b f(x) dx \geq 0 \text{ if the integral exists}$$

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx \text{ if the integral exists}$$

**Definition 3.6 (Improper Riemann Integral)**  $f : [a, b) \rightarrow \mathbb{R}$  has improper Riemann integral (integral converges) if  $\lim_{x \rightarrow b} \int_a^x f(x) dx \in \mathbb{R}$  exists. If the limit does not exists or is  $\pm\infty$ , the integral diverges.



## 4 Series

**Definition 4.1 (Series)** Series are formal infinite sums of real numbers  $\sum_{i=1}^{\infty} a_i$ .

**Remark 4.1.1** We can associate  $\sum_{i=1}^{\infty} a_i$  to  $(S_n)_{n \geq 1}$  where for each  $n \geq 1$ ,  $S_n$  is defined as the partial sum  $\sum_{i=1}^n a_i$ .

**Definition 4.2 (Convergence)** The series  $\sum_{i=1}^{\infty} a_i$

**converges** iff  $(S_n)_{n \geq 1}$  has limit  $l \in \mathbb{R}$  OR  $\sum_{i=1}^{\infty} a_i$  converges for any  $N \in \mathbb{N}$

**diverges** if it does not converges to some  $l \in \mathbb{R}$

**Theorem 4.3 (Increasing and bounded above)** When  $a_i \geq 0$  for all  $i \geq 1$ ,  $(S_n)_{n \geq 1}$  is increasing. If  $(S_n)_{n \geq 1}$  is also bounded above, by the Fundamental Theorem of Analysis,  $\sum_{i=1}^{\infty} a_i$  converges to a limit.

**Definition 4.4 (Permutation)** A permutation  $\pi$  over the natural numbers  $\mathbb{N}$  is a function  $\pi : \mathbb{N} \rightarrow \mathbb{N}$  that has an inverse (injective & surjective).

**Definition 4.5 (Unconditional Convergence)** A series  $\sum_{i=1}^{\infty} a_i$  is unconditionally convergent iff it converges and the permuted series  $\sum_{i=1}^{\infty} a_{\pi(i)}$  converges to the same limit for all permutations  $\pi$ .

**Definition 4.6 (Absolute Convergence)** A series  $\sum_{i=1}^{\infty} a_i$  is absolutely convergent iff the corresponding series  $\sum_{i=1}^{\infty} |a_i|$  converges. Absolute convergence implies unconditional convergence.

**Definition 4.7 (Limit Superior)** The limit superior of  $(a_n)_{n \geq 1}$  is the limit of  $(b_n)_{n \geq 1} \in \mathbb{R}$ , where  $b_n = \sup\{a_m | m \geq n\}$ , denoted  $\limsup_{n \rightarrow \infty} a_n$

**Definition 4.8 (Limit Inferior)** The limit inferior of  $(a_n)_{n \geq 1}$  is the limit of  $(c_n)_{n \geq 1} \in \mathbb{R}$ , where  $c_n = \inf\{a_m | m \geq n\}$ , denoted  $\liminf_{n \rightarrow \infty} a_n$

**Lemma 4.9 (Known Divergent & Convergent Series)**

**Geometric series**  $\sum_{n=1}^{\infty} x^n \rightarrow \frac{x}{1-x}$  for all  $x \in \mathbb{R}$  with  $|x| < 1$

**Inverse squares series**  $\sum_{n=1}^{\infty} \frac{1}{n^2} \rightarrow \frac{\pi^2}{6}$

$\frac{1}{n^c}$  **series**  $\sum_{n=1}^{\infty} \frac{1}{n^c}$  converges for all  $c \in \mathbb{R}$  with  $c > 1$

**Harmonic series**  $\sum_{n=1}^{\infty} \frac{1}{n}$  diverges

**Harmonic primes**  $\sum_{p:\text{prime}}^{\infty} \frac{1}{p}$  diverges

**Geometric series**  $S = \sum_{n=1}^{\infty} x^n$  diverges for  $|x| \geq 1$

### 4.1 Convergence Tests for Series

**Theorem 4.10** If  $\lim_{n \rightarrow \infty} a_n \neq 0$ ,  $\sum_{i=1}^{\infty} a_i$  diverges.

**Theorem 4.11 (Comparison Test)** Let  $\lambda > 0$  and  $N \in \mathbb{N}$ .

If  $a_i \leq \lambda c_i$  for all  $i > N$  for some convergent series  $\sum_{i=1}^{\infty} c_i$ ,  $\sum_{i=1}^{\infty} a_i$  converges.

If  $a_i \geq \lambda d_i$  for all  $i > N$  for some divergent series  $\sum_{i=1}^{\infty} d_i$ ,  $\sum_{i=1}^{\infty} a_i$  diverges.

**Theorem 4.12 (Limit Comparison Test)**

If  $\lim_{i \rightarrow \infty} \frac{a_i}{c_i} \in \mathbb{R}$  exists for some convergent series  $\sum_{i=1}^{\infty} c_i$ ,  $\sum_{i=1}^{\infty} a_i$  converges.

If  $\lim_{i \rightarrow \infty} \frac{d_i}{a_i} \in \mathbb{R}$  exists for some divergent series  $\sum_{i=1}^{\infty} d_i$ ,  $\sum_{i=1}^{\infty} a_i$  diverges.

**Theorem 4.13 (D'Alembert's Ratio Test)** Let  $N \in \mathbb{N}$

If there exists  $k \in \mathbb{R}$  with  $k < 1$  such that  $\frac{a_{i+1}}{a_i} \leq k$  for all  $i \geq N$ , then  $\sum_{i=1}^{\infty} a_i$  converges.

If  $\frac{a_{i+1}}{a_i} \geq 1$  for all  $i \geq N$ , then  $\sum_{i=1}^{\infty} a_i$  diverges.

**Theorem 4.14 (D'Alembert's Limit Ratio Test)** If  $\lim_{i \rightarrow \infty} \frac{a_{i+1}}{a_i}$  exists,

if  $\lim_{i \rightarrow \infty} \frac{a_{i+1}}{a_i} < 1$ , then  $\sum_{i=1}^{\infty} a_i$  converges.

if  $\lim_{i \rightarrow \infty} \frac{a_{i+1}}{a_i} > 1$ , then  $\sum_{i=1}^{\infty} a_i$  diverges.

else  $\lim_{i \rightarrow \infty} \frac{a_{i+1}}{a_i} = 1$ , the test is inconclusive.

**Theorem 4.15 (Integral Test)** Let  $f : \mathbb{R} \rightarrow \mathbb{R}^+$  be continuous, decreasing and positive on  $[N, \infty)$ , where  $N \in \mathbb{Z}$ , with  $a_n = f(n)$  for all  $n \in \mathbb{N}$ .

If  $\int_N^{\infty} f(x) dx$  converges, then  $\sum_{i=1}^{\infty} a_i$  converges.

If  $\int_N^{\infty} f(x) dx$  diverges, then  $\sum_{i=1}^{\infty} a_i$  diverges.

**Theorem 4.16 (Absolute Value Comparison Test)** Let  $(b_n)_{n \geq 1}$  be a non-negative sequence such that  $\sum_{i=1}^{\infty} b_i$  converges, and  $(a_n)_{n \geq 1}$  be a sequence such that  $|a_i| \leq b_i$  for all  $i \geq 1$ . Then  $\sum_{i=1}^{\infty} a_i$  converges.

**Theorem 4.17 (Limit Absolute Value Ratio Test)**

Let  $\sum_{i=1}^{\infty} a_i$  with  $a_i \neq 0$  for  $i \geq 1$ ,

if  $\lim_{i \rightarrow \infty} \left| \frac{a_{i+1}}{a_i} \right| < 1$ , then  $\sum_{i=1}^{\infty} a_i$  converges and absolutely converges.

if  $\lim_{i \rightarrow \infty} \left| \frac{a_{i+1}}{a_i} \right| > 1$ , then  $\sum_{i=1}^{\infty} a_i$  diverges.

else  $\lim_{i \rightarrow \infty} \left| \frac{a_{i+1}}{a_i} \right| = 1$ , the test is inconclusive.

**Theorem 4.18 ( $n^{\text{th}}$  Root Test)** Consider  $\sum_{i=1}^{\infty} a_i$ ,

if  $\limsup_{n \rightarrow \infty} |a_n|^{\frac{1}{n}} < 1$ ,  $\sum_{i=1}^{\infty} a_i$  converges and absolutely converges.

if  $\limsup_{n \rightarrow \infty} |a_n|^{\frac{1}{n}} > 1$ ,  $\sum_{i=1}^{\infty} a_i$  diverges.

else  $\limsup_{n \rightarrow \infty} |a_n|^{\frac{1}{n}} = 1$ , the test is inconclusive.

## 5 Differentiation

**Definition 5.1** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $x \in \mathbb{R}$  and  $h > 0$ .

**Newton's Difference Quotient at  $x$  for  $f$**  is given by

$$\frac{\Delta f(x)}{\Delta(x)} = \frac{f(x+h) - f(x)}{(x+h) - x} = \frac{f(x+h) - f(x)}{h}$$

**$f$  is differentiable at  $x$**  iff  $\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$  exists as a real number and has the same value for all ways where  $h \rightarrow 0$

**derivative of  $f$  at  $x$**  equals to this limit if it exists, denoted as  $f'(x)$  or  $\frac{dy}{dx}$

### Theorem 5.2 (Properties of derivatives)

Given  $f, g : (a, b) \rightarrow \mathbb{R}$  be two functions,

1. Polynomials have derivatives at all points
2. If  $f$  is differentiable at  $x$ ,  $f$  is continuous at  $x$
3. If  $f$  is differentiable in  $(a, b)$ ,  $f'(x_0) = 0$  for any point  $x_0$  where  $f$  is maximum or minimum
4. If  $f$  and  $g$  are differentiable at  $x$ ,  $f \cdot g$  is differentiable under **product rule**

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

5. If  $f$  and  $g$  are differentiable at  $g(x)$  and  $x$  respectively,  $f \circ g$  is differentiable under **product rule**

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

6. Differentiation is a **linear function**, where for all  $f$  and  $g$  differentiable at  $x$  and for all  $a, b \in \mathbb{R}$

$$(a \cdot f + b \cdot g)'(x) = a \cdot f'(x) + b \cdot g'(x)$$

**Theorem 5.3 (Rolle's Theorem)** If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous and  $f : (a, b) \rightarrow \mathbb{R}$  is differentiable with  $f(a) = f(b)$ , there exists  $c \in (a, b)$  such that  $f'(c) = 0$ .

**Theorem 5.4 (Mean Value Theorem)** If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous and  $f : (a, b) \rightarrow \mathbb{R}$  is differentiable, there exists  $c \in (a, b)$  such that  $\frac{f(b) - f(a)}{b - a} = f'(c)$ .

- If  $x$  is close to  $x_0$ ,  $f(x) = f(x_0) + f'(x_0)(x - x_0) + E$ , where  $E$  is small

**Theorem 5.5 (Taylor's Theorem)** *If  $f$  is  $n$  times differentiable in  $(a, b)$  with  $x_0 \in (a, b)$ , then for any  $x \in (a, b)$  we have*

$$f(x) = f(x_0) + \frac{1}{1!}(x - x_0)f'(x_0) + \dots + \frac{1}{n!}(x - x_0)^n f^{(n)}(x_0) + E_n$$

where  $E_n = \frac{f^{(n+1)}(x^*)}{(n+1)!}(x - x_0)^{n+1}$  is the **Lagrange error term** with  $x^*$  between  $x$  and  $x_0$ .

**Theorem 5.6 (L'Hospital's Rule)** *Suppose  $f, g : (a, b) \rightarrow \mathbb{R}$  have derivatives  $f', g' : (a, b) \rightarrow \mathbb{R}$  that are continuous in  $(a, b)$ .*

- *If  $f(c) = g(c) = 0$  for some  $c \in (a, b)$  and  $g'(c) \neq 0$ ,*

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$$

- *If  $\lim_{x \rightarrow c} |f(x)| = \lim_{x \rightarrow c} |g(x)| = \infty$  for some  $c \in (a, b)$*

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{\frac{1}{g(x)}}{\frac{1}{f(x)}}$$

- *It can be extended to  $x \rightarrow \infty$  by restricting  $f, g : [0, \infty) \rightarrow [0, \infty)$*

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{y \rightarrow 0} \frac{f(\frac{1}{y})}{g(\frac{1}{y})}$$

**Theorem 5.7 (Fundamental Theorem of Calculus)** *If  $f : [a, b] \rightarrow \mathbb{R}$  is continuous and  $F : [a, b] \rightarrow \mathbb{R}$  is defined by  $F(y) = \int_a^y f(x) dx$ , then  $F$  is uniformly continuous on  $[a, b]$  and  $F'(x) = f(x)$  for  $x \in (a, b)$ .*

**Corollary 5.8 (Change of variable)** *Let  $g : [a, b] \rightarrow [c, d]$  be a differentiable function with  $g' : [a, b] \rightarrow \mathbb{R}$  and  $f : [c, d] \rightarrow \mathbb{R}$  be a continuous function with  $y = g(x)$ .*

$$\int_a^b f(g(x))g'(x) dx = \int_{g(a)}^{g(b)} f(y) dy$$

## 6 Power Series

**Definition 6.1** A power series is a series of the form  $\sum_{i=0}^{\infty} a_i \cdot (x - c)^i$  where  $x$  is a variable  $\in \mathbb{R}$ ,  $c$  is a constant  $\in \mathbb{R}$  and  $(a_n)_{n \geq 0} \subseteq \mathbb{R}$ .

- Polynomials are power series where  $c = 0$  and there exists  $N$  such that  $a_n = 0$  for all  $n \geq N$ . They converge for all  $x \in \mathbb{R}$ .

**Definition 6.2 (Radius of Convergence)** Let  $c \in \mathbb{R}$  and  $(a_n)_{n \geq 0} \subseteq \mathbb{R}$ .  $\sum_{i=0}^{\infty} a_i \cdot (x - c)^i$  has a radius of convergence  $r \in [0, \infty) \cup \{\infty\}$  such that:

1. If  $r \neq \infty$ , then  $\sum_{i=0}^{\infty} a_i \cdot (x - c)^i$  converges for all  $x \in \mathbb{R}$  when  $|x - c| < r$  and diverges for all  $x \in \mathbb{R}$  when  $|x - c| > r$ .
2. If  $r = \infty$ , then  $\sum_{i=0}^{\infty} a_i \cdot (x - c)^i$  converges for all  $x \in \mathbb{R}$

given by  $r^{-1} = \limsup_{n \rightarrow \infty} |a_n|^{\frac{1}{n}}$ .

**Theorem 6.3 (Ratio Test)** Suppose  $(\frac{|a_{n+1}|}{|a_n|})_{n \geq 1}$  has a limit  $l \in \mathbb{R}$ , then setting  $l < 1$  gives  $l^{-1}$  the radius of convergence of any  $\sum_{i=0}^{\infty} a_i \cdot (x - c)^i$ .

**Definition 6.4 (Smoothness)**  $f : \mathbb{R} \rightarrow \mathbb{R}$  is smooth at  $x_0$  if for all  $k \geq 1$  the  $k^{\text{th}}$  derivative exists at  $x_0$ .

**Definition 6.5 (Analytical)** Given  $f : \mathbb{R} \rightarrow \mathbb{R}$ , if the power series has the same outputs as  $f$  within the radius of convergence,  $f$  is a real analytical function.

- Not every smooth real function is analytical.

**Definition 6.6 (Maclaurin Series)**  $\sum_{i=0}^{\infty} a_i \cdot (x - c)^i$  is called the Maclaurin Series when  $c = 0$ , or  $f(x) = \sum_{i=0}^{\infty} a_i \cdot x^i$ .

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n, \text{ with } r^{-1} = \limsup_{n \rightarrow \infty} \left| \frac{f^{(n)}(0)}{n!} \right|^{\frac{1}{n}}$$

**Definition 6.7 (Taylor Series)**

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!} (x - c)^n$$

**Theorem 6.8** Within the radius of convergence,  $\sum_{i=0}^{\infty} a_i \cdot (x - c)^i$  is continuous and can be differentiated and integrated term by term.

## 7 Numerical Methods

**Theorem 7.1 (Newton's Method)** *Approximates  $f(x) = 0$*

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

*The rate of convergence is at least quadratic if*

1.  $f'(x) \neq 0$  for  $x \in I$  where  $I = [\alpha - r, \alpha + r]$  for some  $r \geq |\alpha - x_0|$
2.  $f''(x)$  is continuous in  $I$
3.  $x_0$  is sufficiently close to  $\alpha$

**Theorem 7.2 (Relaxed Newton's Method)** *For some  $0 < \gamma \leq 1$ ,*

$$x_{n+1} = x_n - \gamma \frac{f(x_n)}{f'(x_n)}$$

**Theorem 7.3 (Secant Method)** *Reduces calculation of derivatives*

$$x_{n+1} = x_n - f(x_n) \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

**Theorem 7.4 (Gradient Descent)** *Minimises  $f(x)$  using a small enough  $\eta$*

$$x_{n+1} = x_n - \eta f'(x_n)$$

**Theorem 7.5 (Euler's Method)** *Approximates the solution to the initial-value problem in differential equations. Given  $y' = f(x, y)$  and  $y(x_0) = y_0$ ,*

$$x_{n+1} = x_0 + nh, y_{n+1} = y_n + hf(x_n, y_n), \text{ for } n = 0, 1, \dots$$

**Theorem 7.6 (Heun's Method)** *A predictor-corrector method to modify Euler's Method. First approximate  $y(x_{n+1})$  with  $y_{n+1}^* = y_n + hf(x_n, y_n)$ .*

$$y_{n+1} = y_n + \frac{1}{2}h[f(x_n, y_n) + f(x_{n+1}, y_{n+1}^*)]$$

**Theorem 7.7 (Runge-Kutta Method of Order Four)** *Uses Simpson's Rule*

$$y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$k_1 = hf(x_n, y_n), k_2 = hf(x_n + \frac{h}{2}, y_n + \frac{k_1}{2})$$

$$k_3 = hf(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}), k_4 = hf(x_{n+1}, y_n + k_3)$$

## 8 Metric Spaces

**Definition 8.1 (Sequence of Functions)** Let  $I \in \mathbb{R}$ . For each  $n \in \mathbb{N}$ ,  $f_n(x) : I \rightarrow \mathbb{R}$ . Then  $(f_n)$  is a sequence of functions on  $I$ .

**Definition 8.2 (Pointwise Convergence)**  $(f_n)$  converges pointwise to  $f$  iff  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  for every  $x \in I$ , written as  $\lim_{n \rightarrow \infty} f_n = f$  pointwise.

**Definition 8.3 (Uniform Convergence)**  $(f_n)$  converges uniformly to  $f$  iff  $\lim_{n \rightarrow \infty} \sup\{|f_n(x) - f(x)| \mid x \in I\} = 0$ , written as  $\lim_{n \rightarrow \infty} f_n = f$  uniformly.

- Every uniformly convergent sequence is pointwise convergent to the same limiting function.
- The pointwise limit of a sequence of continuous functions may be a discontinuous function but only if the sequence is not uniformly convergent.

**Definition 8.4 (Ring)** A ring is a set  $R$  equipped with 2 binary operators  $+$  (addition) and  $\cdot$  (multiplication) satisfying the ring axioms.

1.  $R$  is an **abelian group** under addition, meaning for all  $a \in R$ 
  - (a)  $+$  is associative
  - (b)  $+$  is commutative
  - (c) There exists  $0 \in R$  which is the additive identity ( $a + 0 = a$ )
  - (d)  $-a$  is the additive inverse of  $a$  ( $a + (-a) = 0$ )
2.  $R$  is a **monoid** under multiplication, meaning for all  $a \in R$ 
  - (a)  $\cdot$  is associative
  - (b) There exists  $1 \in R$  which is the multiplicative identity ( $a \cdot 1 = 1 \cdot a = a$ )
3. Multiplication is **distributive** to addition, meaning for all  $a, b, c \in R$ 
  - (a)  $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$  (left associative)
  - (b)  $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$  (right associative)

**Definition 8.5 (Fields)** Fields are commutative rings with identity ( $1 \neq 0$ ) in which every nonzero element has a multiplicative inverse.

**Definition 8.6 (Distance)**  $d : X \times X \rightarrow \mathbb{R}$  is a distance function or metric on the underlying set  $X$  if

1.  $d(x, y) = 0$  iff  $x = y$  (identity of indiscernibles)
2.  $d(x, y) = d(y, x)$  (symmetry)
3.  $d(x, z) \leq d(x, y) + d(y, z)$  (subadditivity or triangle inequality)

**Definition 8.7 (Metric Space)** An ordered pair  $(X, d)$  consisting of a nonempty set  $X$  and distance function  $d$  on  $X$  is a metric space

- If  $(X, d)$  is a metric space and  $S \subset X$ ,  $(S, d)$  is a metric subspace of  $(X, d)$

**Definition 8.8 (Limit)** Let  $(X, d)$  be a metric space and  $(x_n)_{n \geq 1}$  be a sequence of points in  $X$ , a point  $x \in X$  is the limit of  $(x_n)_{n \geq 1}$  if  $\lim_{n \rightarrow \infty} d(x, x_n) = 0$

**Definition 8.9 (Open Ball)** Let  $(X, d)$  be a metric space,  $a \in X$  and  $\delta > 0$ . The subset of  $X$  containing all points  $x \in X$  such that  $d(a, x) < \delta$  is called the open ball about  $a$  of radius delta, denoted by  $B(a; \delta)$ .

**Definition 8.10 (Neighbourhood)** Let  $(X, d)$  be a metric space and  $a \in X$ . A subset  $N$  of  $X$  is a neighbourhood of  $a$  if there exists  $\delta > 0$  such that  $B(a; \delta) \subseteq N$ . The **complete system of neighbourhoods** of the point  $a$   $\mathcal{N}_a$  is the collection of all neighbourhoods of  $a$ . For all  $a$  and any neighbourhood  $N$  of  $a$ ,

1. there exists at least one neighbourhood of  $a$
2.  $a \in N$
3. if  $N' \supseteq N$ , then  $N'$  is a neighbourhood of  $a$
4. if  $M$  is another neighbourhood of  $a$ ,  $N \cap M$  is also a neighbourhood of  $a$
5. there exists a neighbourhood  $O$  of  $a$  such that  $O \subseteq N$  such that  $O$  is a neighbourhood of each of its points

**Lemma 8.11 (First Axiom of Countability)** Let  $(X, d)$  be a metric space. For every  $a \in X$ , there is a sequence of neighbourhoods of  $a$   $(O_n)_{n \geq 1}$  such that  $N$  contains at least one neighbourhood of this sequence.

**Lemma 8.12 (Hausdorff Axiom)** For every pair of distinct points  $x, y$  of  $(X, d)$ , there is a neighbourhood  $M$  of  $x$  and  $N$  of  $y$  such that  $M \cap N = \emptyset$ .

**Theorem 8.13** Let  $(X, d)$  be a metric space and  $a \in X$ . For each  $\delta > 0$ , the open ball  $B(a; \delta)$  is a neighbourhood of each of its points.

**Definition 8.14 (Open Set)** A subset  $O$  of a metric space  $(X, d)$  is open if  $O$  is a neighbourhood of each of its points.

- $O$  is an open set iff it is a union of open balls

1.  $\emptyset$  and  $X$  is open
2. The union and intersection of open sets is open

**Definition 8.15 (Function Composition)** Let  $(X, d)$ ,  $(Y, d')$  and  $(Z, d'')$  be metric spaces. Also let  $f : X \rightarrow Y$  be continuous at  $a \in X$  and  $g : Y \rightarrow Z$  be continuous at  $f(a) \in Y$ . Then  $g \circ f : X \rightarrow Z$  is continuous at  $a \in X$ .



**Definition 8.16 (Topology)** An ordered pair  $(X, \tau)$  consisting of a set  $X$  and a collection  $\tau$  of subsets of  $X$  satisfying the following axioms:

1.  $\emptyset$  and  $X$  belongs to  $\tau$
2. any arbitrary union of members of  $\tau$  belong to  $\tau$
3. the intersection of any finite members of  $\tau$  belongs to  $\tau$

with the elements of  $\tau$  called **open sets** and  $\tau$  is called a topology on  $X$ . A subset  $C \subseteq X$  is said to be closed in  $(X, \tau)$  iff its complement  $X \setminus C$  is open.

**Definition 8.17 (Homeomorphism)**  $(X, \tau)$  and  $(X', \tau')$  are homeomorphic if there exists mutually inverse continuous functions  $f : X \rightarrow X'$  and  $g : X' \rightarrow X$ .  $f$  and  $g$  then define a homeomorphism between  $(X, \tau)$  and  $(X', \tau')$ .

- Homeomorphisms form an equivalence relation on the class of topological spaces. The resulting equivalence classes are **homeomorphism classes**.

**Definition 8.18 (Embedding)** A map  $f$  between  $(X, \tau)$  and  $(X', \tau')$  is a topological embedding if  $f$  yields a homeomorphism between  $X$  and  $f(X)$ , where  $f(X)$  carries the relative topology inherited from  $X'$ , denoted by  $X \hookrightarrow X'$ .

- $X$  can be treated as a subspace of  $X'$

**Definition 8.19 (Compactness)**  $(X, \tau)$  is compact if each of its open covers has a finite subcover, that is for any collection  $C$  of open subsets of  $X$  such that  $X = \bigcup_{x \in C} x$  there exists a finite subset  $F \subseteq C$  such that  $X = \bigcup_{x \in F} x$ .

**Theorem 8.20 (Heine-Borel Theorem)** For any subset  $A$  of a Euclidean space,  $A$  is compact iff  $A$  is closed and bounded.

- Closed intervals are compact

**Definition 8.21 (Continuity)**

(Metric Spaces)  $f : (X, d) \rightarrow (Y, d')$  is continuous at  $a \in X$  if

- (**Epsilon-Delta**) for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that
  - $d'(f(x), f(a)) < \epsilon$  whenever  $x \in X$  and  $d(x, a) < \delta$
  - $f(B(a; \delta)) \subseteq B(f(a); \epsilon)$  OR  $B(a; \delta) \subseteq f^{-1}(B(f(a); \epsilon))$
- (**Neighbourhood**) for any neighbourhood  $M$  of  $f(a)$ 
  - there exists a corresponding neighbourhood  $N$  of  $a$  such that  $f(N) \subseteq M$  OR  $N \subseteq f^{-1}(M)$
  - $f^{-1}(M)$  is a neighbourhood of  $a$
- (**Open Set**) for any open set  $O$  of  $Y$ , the subset  $f^{-1}(O)$  is an open subset of  $X$

(Topology)  $f : (X, \tau) \rightarrow (X', \tau')$  is continuous at  $x \in X$  if for any neighbourhood  $G$  of  $f(x)$ , where  $G \in \tau'$ ,  $f^{-1}(G)$  is a neighbourhood of  $x$ .

## 9 Deep Learning and Multivariate Chain Rule

**Definition 9.1 (Perceptron)** The function  $f : \mathbf{x} \mapsto \sigma(\mathbf{w}^\top \mathbf{x} + b)$  where  $w$  is the vector of weights,  $b$  the scalar bias (offset), and the activation function  $\sigma$  is the Heaviside step function.

$$\sigma(v) = \begin{cases} 0, & v < 0 \\ 1, & v \geq 0 \end{cases}$$

function	$w_1$	$w_2$	$b$
$\neg$	-1		0
$\wedge$	0.5	0.5	-1
$\vee$	1	1	-1

**Definition 9.2 (Loss Function)** Let  $\hat{y}$  be the output of  $\sigma(\mathbf{w}^\top \mathbf{x} + b)$  and  $y$  the true value (target). The zero-one loss function is defined as

$$l_{0-1}(y, \hat{y}) = \begin{cases} 0, & \hat{y} \neq y \\ 1, & \hat{y} = y \end{cases}$$

which is a piecewise constant function of the weights and bias. Hence we use the surrogate loss function

$$l_{SE}(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$$

**Theorem 9.3 (Gradient Descent)**

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \gamma \frac{\partial l_{SE}}{\partial \mathbf{w}} = \mathbf{w}_n - \gamma(\hat{y} - y)\mathbf{x}$$

$$b_{n+1} = b_n - \gamma \frac{\partial l_{SE}}{\partial b} = b_n - \gamma(\hat{y} - y)$$

**Definition 9.4 (Multilayer Perceptron)** or feedforward network

$$\hat{Y}(X) := F_{\mathbf{w}, b}(X) = \left( \mathbf{f}_{\mathbf{w}^{(L)}, b^{(L)}}^{(L)} \circ \dots \circ \mathbf{f}_{\mathbf{w}^{(1)}, b^{(1)}}^{(1)} \right)(X)$$

- If we take  $\sigma$  to be the identity function in each layer, the MLP becomes a **linear regression**

**Theorem 9.5 (Chain Rule)**

$$(f \circ g)' = (f' \circ g) \cdot g', \text{ in Lagrange's notation}$$

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}, \text{ in Leibniz's notation}$$

**Theorem 9.6 (Multivariate Chain Rule)** Let  $g : \mathbb{R}^p \rightarrow \mathbb{R}^n$  and  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , with  $\mathbf{b} = g(\mathbf{a})$  and  $\mathbf{c} = f(\mathbf{b})$  where  $\mathbf{a} \in \mathbb{R}^p$ ,  $\mathbf{b} \in \mathbb{R}^n$  and  $\mathbf{c} \in \mathbb{R}^m$ .

$$\frac{\partial c_i}{\partial a_j} = \sum_{k=1}^n \frac{\partial c_i}{\partial b_k} \cdot \frac{\partial b_k}{\partial a_j}$$