

RESULTS IN FIRST PART OF METHODS AND CALCULUS

T.W.KÖRNER

Definition 1. Let $\mathbf{a}_n, \mathbf{a} \in \mathbb{R}^k$. We say that $\mathbf{a}_n \rightarrow \mathbf{a}$ as $n \rightarrow \infty$, if given $\epsilon > 0$, we can find $N(\epsilon)$ such that $\|\mathbf{a}_n - \mathbf{a}\| < \epsilon$ for all $n > N(\epsilon)$.

Theorem 2. (i) If $\mathbf{a}_n \rightarrow \mathbf{a}$ and $\mathbf{b}_n \rightarrow \mathbf{b}$ in \mathbb{R}^k then $\mathbf{a}_n + \mathbf{b}_n \rightarrow \mathbf{a} + \mathbf{b}$ as $n \rightarrow \infty$.

(ii) If $a_n \rightarrow a$ and $b_n \rightarrow b$ in \mathbb{R} (or \mathbb{C}) then $a_n b_n \rightarrow ab$ as $n \rightarrow \infty$.

(iii) If $a_n \rightarrow a$ as $n \rightarrow \infty$ in \mathbb{R} (or \mathbb{C}) and $a \neq 0$, $a_n \neq 0$ [$n = 1, 2, \dots$] then $a_n^{-1} \rightarrow a^{-1}$ as $n \rightarrow \infty$.

There are many related definitions.

Definition 3. (i) If $f : \mathbb{R}^l \rightarrow \mathbb{R}^k$ we say that $f(\mathbf{x}) \rightarrow \mathbf{a}$ as $\mathbf{x} \rightarrow \mathbf{y}$ if given $\epsilon > 0$ we can find a $\delta(\epsilon) > 0$ such that, whenever $0 < \|\mathbf{x} - \mathbf{y}\| < \delta(\epsilon)$ it follows that $\|f(\mathbf{x}) - \mathbf{a}\| < \epsilon$.

(ii) Let $a_n \in \mathbb{R}$. We say that $a_n \rightarrow \infty$ if, given any K we can find $N(K)$ such that $a_n > K$ for all $n > N(K)$.

Axiom 4 (Fundamental Axiom of Analysis). If a_1, a_2, \dots is an increasing sequence in \mathbb{R} and there exists an $A \in \mathbb{R}$ such that $a_n \leq A$ for all n , then there exists an $a \in \mathbb{R}$ such that $a_n \rightarrow a$ as $n \rightarrow \infty$.

Definition 5. We say that the series \mathbf{a}_n in \mathbb{R}^k converges to the sum \mathbf{a} if

$$\sum_{n=1}^N \mathbf{a}_n \rightarrow \mathbf{a} \text{ as } N \rightarrow \infty.$$

We write

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

Lemma 6 (Absolute Convergence implies Convergence). If $\mathbf{a}_n \in \mathbb{R}^k$ and $\sum_{n=1}^{\infty} \|\mathbf{a}_n\|$ converges then $\sum_{n=1}^{\infty} \mathbf{a}_n$ converges.

Lemma 7 (Comparison Test). If $a_n, b_n \in \mathbb{R}$ and $0 \leq a_n \leq b_n$ then whenever $\sum_{n=1}^{\infty} b_n$ converges $\sum_{n=1}^{\infty} a_n$ must converge.

Corollary 8 (Ratio Test). *If $a_n \in \mathbb{R}$, $0 < a_n$ and $a_{n+1}/a_n \rightarrow l$ then*

- (i) *If $l < 1$ then $\sum_{n=1}^{\infty} a_n$ converges.*
(ii) *If $l > 1$ then $\sum_{n=1}^{\infty} a_n$ diverges.*

Example 9. *If $a_{2n} = 2^{-2n}$, $a_{2n+1} = 2^{-2n-2}$ then the ratio test fails but comparison with 2^{-n} show that the series a_n is convergent.*

Lemma 10 (Integral Test). *Suppose that $f : [0, \infty) \rightarrow [0, \infty)$ is a decreasing continuous function. Then if one of*

$$\int_0^N f(x)dx \quad \text{and} \quad \sum_{n=0}^N f(n)$$

tends to a (finite) limit as $N \rightarrow \infty$ so does the other.

Corollary 11. $\sum_{n=1}^{\infty} n^{-p}$ *converges if and only if $p > 1$.*

Note the failure of the ratio test for the series n^{-p} .

Example 12. *If $f(x) = 1 - \cos(2\pi x)$ then $\int_0^N f(x)dx$ diverges and $\sum_{n=0}^N f(n)$ converges as $N \rightarrow \infty$.*

Lemma 13 (Alternating Series Test). *(Not in syllabus.) Suppose that a_n is a decreasing sequence of positive terms with $a_n \rightarrow 0$ as $n \rightarrow \infty$. Then $\sum_{n=1}^{\infty} (-1)^n a_n$ converges.*

Example 14. $\sum_{n=1}^{\infty} (-1)^n n^{-p}$ *converges for $p > 0$ but only converges absolutely for $p > 1$.*

Theorem 15 (Rearrangement of Positive Series). *Let $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ be a bijection. If $a_n \geq 0$ then, if the series a_n converges, so does the rearranged series $a_{\sigma(n)}$ and*

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_{\sigma(n)}.$$

Corollary 16 (Rearrangement of Absolutely Convergent Series). *(Not in syllabus.) Let $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ be a bijection. If $\mathbf{a}_n \in \mathbb{R}^k$ then, if the series \mathbf{a}_n is absolutely convergent, so is the rearranged series $\mathbf{a}_{\sigma(n)}$ and*

$$\sum_{n=1}^{\infty} \mathbf{a}_n = \sum_{n=1}^{\infty} \mathbf{a}_{\sigma(n)}.$$

Example 17. *Let $a_{2n-1} = n^{-1}$, $a_{2n} = -n^{-1}$ [$n = 1, 2, \dots$]. Then (by comparison with appropriate integrals)*

$$\sum_{n=1}^{kN} a_{2n-1} + \sum_{n=1}^{lN} a_{2n} = \sum_{n=lN+1}^{kN} n^{-1} \rightarrow \log(k/l)$$

as $N \rightarrow \infty$ whenever k and l are integers with $k \geq l > 0$.

Theorem 18. Let $a_n \in \mathbb{C}$.

(i) If $\sum_{n=0}^{\infty} a_n z_0^n$ converges then $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely for all $z \in \mathbb{C}$ with $|z| < |z_0|$.

(ii) If $\sum_{n=0}^{\infty} a_n z_0^n$ diverges then the sequence $|a_n z^n|$ is unbounded for all $z \in \mathbb{C}$ with $|z| > |z_0|$.

Theorem 19 (Radius of Convergence). (Proof next year.) Let $a_n \in \mathbb{C}$. Then, either $\sum_{n=0}^{\infty} a_n z^n$ converges for all $z \in \mathbb{C}$ (and we say that the power series has radius of convergence infinity) or there exists an R with $R \geq 0$ such that

(i) $\sum_{n=0}^{\infty} a_n z^n$ converges for all $|z| < R$,

(ii) $\sum_{n=0}^{\infty} a_n z^n$ diverges for all $|z| > R$.

We call R the radius of convergence.

Example 20. (i) If $a_n = 1/n!$ then $R = \infty$.

(ii) If $a_n = n!$ then $R = 0$.

(iii) If $a_n = r^{-n}$ then $R = r$.

(iv) If $a_{4n} = 0$, $a_{4n+1} = 1$, $a_{4n+2} = 2$, $a_{4n+3} = 1$ then $R = 1$.

(v) If $a_0 = 1$, $a_n = n^p$ then $R = 1$. If $p < -1$ then we have convergence of $\sum_{n=0}^{\infty} a_n z^n$ for $|z| = 1$, if $p \geq 0$ we have divergence. If $-1 \leq p < 0$ have convergence when $z = -1$ and divergence when $z = 1$.

Definition 21. We say that a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is continuous at $\mathbf{x} \in \mathbb{R}^n$ if given $\epsilon > 0$ we can find a $\delta(\epsilon, \mathbf{x}) > 0$ such that, whenever $\|\mathbf{h}\| < \delta(\epsilon, \mathbf{x})$, we have $\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})\| < \epsilon$.

Theorem 22. (i) If $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is continuous at $\mathbf{x} \in \mathbb{R}^n$ and $g : \mathbb{R}^m \rightarrow \mathbb{R}^p$ is continuous at $f(\mathbf{x})$ then the composed function $g \circ f$ (defined by $g \circ f(\mathbf{y}) = g(f(\mathbf{y}))$) is continuous at \mathbf{x} .

(ii) If $F, G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are continuous at $\mathbf{x} \in \mathbb{R}^n$ then so is $F + G$.

Definition 23. We say that a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable at $\mathbf{x} \in \mathbb{R}^n$ with derivative the linear map $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^m$ if given $\epsilon > 0$ we can find a $\delta(\epsilon, \mathbf{x}) > 0$ such that, whenever $\|\mathbf{h}\| < \delta(\epsilon, \mathbf{x})$, we have

$$\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \alpha(\mathbf{h})\| < \epsilon \|\mathbf{h}\|$$

We shall write $\alpha = (Df)(\mathbf{x})$. If we consider the special case $m = n = 1$ then $(Df)(x)$ is a linear map $\mathbb{R} \rightarrow \mathbb{R}$ and so there exists a $\lambda \in \mathbb{R}$ such that $(Df)(x)h = \lambda h$ for all $h \in \mathbb{R}$. We thus have

$$|f(x+h) - f(x) - \lambda h| < \epsilon |h|$$

for $|h| < \delta(\epsilon, x)$, and so

$$\frac{f(x+h) - f(x)}{h} \rightarrow \lambda.$$

If we set $\lambda = f'(x)$ we recover our old definition of differentiation. (Sometimes people write $f'(\mathbf{x}) = (Df)(\mathbf{x})$ in the general case, but we must then remember that $f'(\mathbf{x})$ is a linear map.)

Lemma 24. *Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable at $\mathbf{x} \in \mathbb{R}^n$ with derivative $Df(\mathbf{x})$. If \mathbf{u} is a vector in \mathbb{R}^n and \mathbf{v} is a vector in \mathbb{R}^m then*

$$g_{\mathbf{u}\mathbf{v}}(t) = \mathbf{v} \cdot f(\mathbf{x} + t\mathbf{u})$$

defines a function $g_{\mathbf{u}\mathbf{v}} : \mathbb{R} \rightarrow \mathbb{R}$ which is differentiable at $\mathbf{0}$ with derivative $\mathbf{v} \cdot (Df(\mathbf{x}))(\mathbf{u})$.

Lemma 25. *Suppose that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ is a basis for \mathbb{R}^n and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ is a basis for \mathbb{R}^m . If write $f_{i,j}(\mathbf{x}) = \mathbf{v}_i \cdot (Df(\mathbf{x}))(\mathbf{u}_j)$ then the linear map $Df(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ has matrix $(f_{i,j}(\mathbf{x}))$ with respect to the given bases.*

There is an older and, in many circumstances, more convenient notation called the partial derivative. Suppose g is a well behaved real function of n real variables x_1, x_2, \dots, x_n . Then we write

$$\frac{\partial g}{\partial x_i} = \lim_{h \rightarrow 0} \frac{g(x_1, x_2, \dots, x_i + h, \dots, x_n) - g(x_1, x_2, \dots, x_i, \dots, x_n)}{h}$$

for the derivative of g with respect to x_i when $x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ are kept fixed. With this notation, if we write f_j for the j th component of the function f of the preceding lemma, we have

$$f_{j,i}(\mathbf{x}) = \frac{\partial f_j}{\partial x_i}$$

However when reading material which uses partial derivatives (as in Classical Thermodynamics) it is important to be aware of possible ambiguities.

Example 26. *Let $f(x, y) = x + y$. If we keep y fixed and allow x to vary then $\frac{\partial f}{\partial x} = 1$, but if we keep $x + y$ fixed and allow x to vary then $\frac{\partial f}{\partial x} = 0$*

Lemma 27. *If $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear there exists a constant A such that*

$$\|\alpha(\mathbf{x})\| \leq A\|\mathbf{x}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Theorem 28 (Chain Rule). *If $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable at $\mathbf{x} \in \mathbb{R}^n$ and $g : \mathbb{R}^m \rightarrow \mathbb{R}^p$ is differentiable at $f(\mathbf{x})$ then the composed function $g \circ f$ (defined by $g \circ f(\mathbf{y}) = g(f(\mathbf{y}))$) is differentiable at \mathbf{x} with*

$$D(g \circ f)(\mathbf{x}) = (D(g)(f(\mathbf{x}))D(f)(\mathbf{x})).$$

Corollary 29. (i) *If $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are differentiable at $\mathbf{x} \in \mathbb{R}^n$ then so is $f + g$ and*

$$D(f + g)(\mathbf{x}) = D(f)(\mathbf{x}) + D(g)(\mathbf{x}).$$

(ii) *If $f, g : \mathbb{R} \rightarrow \mathbb{R}$ are differentiable at x then so is their product and*

$$\frac{d(f(x)g(x))}{dx} = g(x)\frac{d(f(x))}{dx} + f(x)\frac{d(g(x))}{dx}.$$

In the matrix notation of Lemma 25 the chain rule of Theorem 28 becomes

$$(g \circ f)_{i,k}(\mathbf{x}) = \sum_{j=1}^m g_{i,j}(f(\mathbf{x}))f_{j,k}(\mathbf{x}).$$

Still more briefly we may use the summation convention with i ranging from 1 to p , j from 1 to m and k from 1 to n to write

$$(g \circ f)_{i,k}(\mathbf{x}) = g_{i,j}(f(\mathbf{x}))f_{j,k}(\mathbf{x}).$$

If g is a real function of the real variables y_1, y_2, \dots, y_m which are themselves functions of the real variables x_1, x_2, \dots, x_n we recover the traditional form

$$\frac{\partial g}{\partial x_k} = \frac{\partial g}{\partial y_1} \frac{\partial y_1}{\partial x_k} + \frac{\partial g}{\partial y_2} \frac{\partial y_2}{\partial x_k} + \dots + \frac{\partial g}{\partial y_m} \frac{\partial y_m}{\partial x_k}.$$

If $f : \mathbb{C} \rightarrow \mathbb{C}$ then, just as in the real case we may define the derivative $f'(z)$ by

$$f'(z) = \lim_{h \rightarrow 0} \frac{f(z+h) - f(z)}{h}$$

where it exists (a $\mathbb{C} \rightarrow \mathbb{C}$ differentiable function is also called analytic). Since \mathbb{C} and \mathbb{R} are so similar those parts of the real theory which ought to go over to the complex case do.

Example 30. *If P and Q are polynomials and Q has roots at $\omega_1, \omega_2, \dots, \omega_n$ then P/Q is differentiable at all $z \in \mathbb{C} \setminus \{\omega_1, \omega_2, \dots, \omega_n\}$.*

As the next theorem shows analytic functions are rather special.

Theorem 31 (Characterisation of Analytic Functions). *Let $f : \mathbb{C} \rightarrow \mathbb{C}$ be given and write*

$$f(x + iy) = u(x, y) + iv(x, y)$$

with x, y, u, v real. Let $z_0 = x_0 + iy_0$. The following statements are equivalent.

(i) f is differentiable as a function $\mathbb{C} \rightarrow \mathbb{C}$ at z_0 .

(ii) $f(z_0 + h) = f(z_0) + \lambda h + \epsilon(h)|h|$ with $\lambda \in \mathbb{C}$ and $|\epsilon(h)| \rightarrow 0$ as $|h| \rightarrow 0$. (We have $f'(z_0) = \lambda$.)

(iii) Near z_0 , f is the composition of translations, rotations, and (possibly zero) dilatations with an error which decreases faster than linear.

(iv) The function $(u, v) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is differentiable at (x_0, y_0) with derivative

$$D(u, v)(x_0, y_0) = l\alpha$$

where $l \geq 0$ and $\alpha \in SO(2)$ the special orthogonal group.

(v) (Cauchy Riemann Equations) The function $(u, v) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is differentiable at (x_0, y_0) and (at that point)

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}.$$

(vi) The function $(u, v) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is differentiable at (x_0, y_0) and (at that point)

$$\frac{\partial u}{\partial X} = \frac{\partial v}{\partial Y}, \quad \frac{\partial v}{\partial X} = -\frac{\partial u}{\partial Y}.$$

for any orthogonal coordinate system (X, Y) .

If $f'(z_0) \neq 0$ then (iii) tells us that f is locally conformal (i.e. angle preserving).

Example 32. (i) The map $z \mapsto z^2$ is analytic but is not conformal at 0.

(ii) The mapping $z \mapsto z^*$ is a reflection and nowhere analytic. It preserves the magnitude of angles but changes their sense.

Example 33. (This non-examinable example shows the need for a rigorous treatment of the calculus) Define $f : \mathbb{Q} \rightarrow \mathbb{Q}$ by

$$\begin{aligned} f(q) &= q \text{ if } q < 0 \text{ or } q^2 < 2, \\ f(q) &= q - 3 \text{ otherwise.} \end{aligned}$$

Then, considered as a function $\mathbb{Q} \rightarrow \mathbb{Q}$, f is continuous, yet $f(1) = 1$, $f(2) = -1$ and $f(t) \neq 0$ for all $t \in \mathbb{Q}$ with $1 \leq t \leq 2$. Further, considered as a function $\mathbb{Q} \rightarrow \mathbb{Q}$, f is differentiable with derivative 1 everywhere, but $f(1) > f(2)$ so f is not everywhere increasing.

In spite of this, we shall be able next year to use the Fundamental Axiom (Axiom 4) to prove rigorously the following two fundamental theorems.

Theorem 34 (Intermediate Value Theorem). *(Proof next year.)* If $f : [a, b] \rightarrow \mathbb{R}$ is continuous and $f(a) \leq d \leq f(b)$ then there exists a c with $a \leq c \leq b$ and $f(c) = d$.

Theorem 35. *(Proof next year.)* If $f : [a, b] \rightarrow \mathbb{R}$ is continuous then f is bounded and attains its bounds. In other words we can find c_1 and c_2 with $a \leq c_1, c_2 \leq b$ and $f(c_1) \leq f(t) \leq f(c_2)$ for all $t \in [a, b]$.

We use Theorem 35 to give rigorous proofs of the results that follow.

Theorem 36 (Rolle's Theorem). If $f : [a, b] \rightarrow \mathbb{R}$ is continuous, f is differentiable on (a, b) and $f(a) = f(b)$ then there exists a c with $a < c < b$ and $f'(c) = 0$.

Theorem 37 (Mean Value Theorem). If $f : [a, b] \rightarrow \mathbb{R}$ is continuous and f is differentiable on (a, b) then there exists a c with $a < c < b$ and $f(b) - f(a) = (b - a)f'(c)$.

Corollary 38. Suppose $b > a$ and $f : [a, b] \rightarrow \mathbb{R}$ is continuous and f is differentiable on (a, b) .

(i) If $|f'(t)| \leq M$ for all $a < t < b$ then $|f(b) - f(a)| \leq M(b - a)$.

(ii) If $f'(t) = 0$ for all $a < t < b$ then $f(b) = f(a)$.

(iii) If $f'(t) \geq 0$ for all $a < t < b$ then $f(b) \geq f(a)$.

(iv) If $f'(t) > 0$ for all $a < t < b$ then $f(b) > f(a)$.

Corollary 39. If f and g are differentiable functions $(a, b) \rightarrow \mathbb{R}$ with $f'(t) = g'(t)$ for all $a < t < b$ then we can find a constant c such that $f(t) = g(t) + c$ for all $a < t < b$.

Example 40. If $f(t) = t^3$ then f is strictly increasing yet $f'(0) = 0$.

The Mean Value Theorem **does not** go over unchanged to higher dimension.

Example 41. (i) Define $f : \mathbb{R} \rightarrow \mathbb{R}^2$ by $f(t) = (\cos(t), \sin(t))$. Then $f(0) = f(2\pi)$ but $Df(t)$ has matrix $(-\sin(t), \cos(t))$ and so is never zero.

(ii) Define $g : \mathbb{C} \rightarrow \mathbb{C}$ by $g(z) = \exp(iz)$. Then $g(0) = g(2\pi)$ but $g'(z) = i \exp(iz) \neq 0$.

However there is a modified version which is just as useful.

Theorem 42 (Modified Mean Value Theorem). (i) Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable with

$$\|Df((1-t)\mathbf{a} + t\mathbf{b})\mathbf{h}\| \leq M\|\mathbf{h}\|$$

for all $\mathbf{h} \in \mathbb{R}^n$ and all $0 \leq t \leq 1$. Then

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

(ii) Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable with $Df(\mathbf{x}) = \mathbf{0}$ for all \mathbf{x} with $\|\mathbf{x} - \mathbf{a}\| < r$. Then $f(\mathbf{x}) = f(\mathbf{a})$ for all \mathbf{x} with $\|\mathbf{x} - \mathbf{a}\| < r$.

(iii) Suppose $g : \mathbb{C} \rightarrow \mathbb{C}$ is analytic with

$$|g'((1-t)z_1 + tz_2)| \leq M$$

for all $0 \leq t \leq 1$. Then

$$|g(z_2) - g(z_1)| \leq M|z_2 - z_1|.$$

(iv) Suppose $g : \mathbb{C} \rightarrow \mathbb{C}$ is analytic with $g'(z) = 0$ for all $z \in \mathbb{C}$ with $|z - w| < r$. Then $g(z) = g(w)$ for all z with $|z - w| < r$.

Lemma 43 (Operations Within Radius of Convergence). (*Proof next year.*)

(i) If $\sum_{n=0}^{\infty} a_n z^n$ has radius of convergence R and $\sum_{n=0}^{\infty} b_n z^n$ has radius of convergence S then, writing

$$c_n = \sum_{r=0}^n a_{n-r} b_r,$$

$\sum_{n=0}^{\infty} c_n z^n$ has radius of convergence $T \geq \min(R, S)$ and

$$\sum_{n=0}^{\infty} a_n z^n \sum_{n=0}^{\infty} b_n z^n = \sum_{n=0}^{\infty} c_n z^n$$

for all $|z| < \min(R, S)$.

(ii) If $\sum_{n=0}^{\infty} a_n z^n$ has radius of convergence R then $\sum_{n=1}^{\infty} n a_n z^{n-1}$ has the same radius of convergence. Further $\sum_{n=0}^{\infty} a_n z^n$ is differentiable and

$$\frac{d}{dz} \sum_{n=0}^{\infty} a_n z^n = \sum_{n=1}^{\infty} n a_n z^{n-1}$$

for all $|z| < R$.

Theorem 44. *The infinite sum*

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

defines a function $\exp : \mathbb{R} \rightarrow \mathbb{R}$ with the following properties.

(i) \exp is differentiable with $\exp'(x) = \exp(x)$ for all $x \in \mathbb{R}$.

(ii) $\exp(x + y) = \exp(x) \exp(y)$ for all $x, y \in \mathbb{R}$.

(iii) $\exp(x)$ increases strictly from 0 to ∞ as x increases from $-\infty$ to ∞ .

Theorem 45. *There exists a unique function $\log : (0, \infty) \rightarrow \mathbb{R}$ such that $\log(\exp(x)) = x$ for all $x \in \mathbb{R}$ and $\exp(\log(y)) = y$ for all $y \in (0, \infty)$. The function \log has the following properties.*

- (i) \log is differentiable with $\log'(x) = 1/x$ for all $x > 0$.
- (ii) $\log(xy) = \log(x) + \log(y)$ for all $x, y > 0$.
- (iii) $\log(x)$ increases strictly from $-\infty$ to ∞ as x increases from 0 to ∞ .

Corollary 46. *The additive group $(\mathbb{R}, +)$ and the multiplicative group $((0, \infty), \times)$ are isomorphic.*

Theorem 47. *If we set $x^\alpha = \exp(\alpha \log(x))$ for $x > 0$, $\alpha \in \mathbb{R}$ then*

- (i) $x^\alpha y^\alpha = (xy)^\alpha$,
- (ii) $x^\alpha x^\beta = x^{(\alpha+\beta)}$,
- (iii) $(x^\alpha)^\beta = x^{\alpha\beta}$,
- (iv) $x^1 = x$,

for all $x, y > 0$, $\alpha, \beta \in \mathbb{R}$. Further, if $\alpha \in \mathbb{R}$ is fixed, x^α is differentiable with

$$\frac{dx^\alpha}{dx} = \alpha x^{\alpha-1}.$$

Theorem 48. *If we write*

$$\begin{aligned} \sin(x) &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}, \\ \cos(x) &= \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}, \end{aligned}$$

then \sin and \cos are well defined differentiable functions $\mathbb{R} \rightarrow \mathbb{R}$ such that

- (i) $\sin'(x) = \cos(x)$ and $\cos'(x) = -\sin(x)$,
- (ii) $\sin(x+y) = \sin(x)\cos(y) + \cos(x)\sin(y)$, $\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y)$ and $(\cos(x))^2 + (\sin(x))^2 = 1$, for all $x, y \in \mathbb{R}$.

Moreover there exists a real number π such that $2 < \pi < 4$ and (for $x \in \mathbb{R}$)

- (iii) $\cos(x) \geq 0$ for $0 \leq x < \pi/2$, $\cos(\pi/2) = 0$,
- (iv) $\sin(x) = \cos(x - \pi/2)$, $\cos(x + \pi) = -\cos(x)$,
- (v) If $u, v \in \mathbb{R}$ and $u^2 + v^2 = 1$ then the equation $(u, v) = (\cos(\theta), \sin(\theta))$ has exactly one solution θ with $0 \leq \theta < 2\pi$.

Corollary 49. *If $(x, y) \in \mathbb{R}^2$ and $(x, y) \neq (0, 0)$ then there is a unique $r > 0$ and a unique θ with $0 \leq \theta < 2\pi$ with $x = r \cos(\theta)$, $y = r \sin(\theta)$.*

Theorem 50. *We can extend the function \exp in a consistent manner to a function $\mathbb{C} \rightarrow \mathbb{C}$ by setting*

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

With this definition,

- (i) \exp is differentiable with $\exp'(z) = \exp(z)$ for all $z \in \mathbb{C}$.
- (ii) $\exp(z_1 + z_2) = \exp(z_1)\exp(z_2)$ for all $z_1, z_2 \in \mathbb{C}$.
- (iii) If x and y are real, then

$$\exp(x + iy) = \exp(x)(\cos(y) + i \sin(y)).$$

Theorem 51. We can extend the functions \cos and \sin in a consistent manner to functions $\mathbb{C} \rightarrow \mathbb{C}$ by setting

$$\begin{aligned}\sin(z) &= \frac{\exp(iz) - \exp(-iz)}{2i}, \\ \cos(z) &= \frac{\exp(iz) + \exp(-iz)}{2}.\end{aligned}$$

With these definitions,

- (i) \cos and \sin are differentiable with $\sin'(z) = \cos(z)$ and $\cos'(z) = -\sin(z)$,
- (ii) $\sin(z+w) = \sin(z)\cos(w) + \cos(z)\sin(w)$, $\cos(z+w) = \cos(z)\cos(w) - \sin(z)\sin(w)$ and $(\cos(z))^2 + (\sin(z))^2 = 1$, for all $z, w \in \mathbb{C}$.

Lemma 52. If x is real, $\cosh(x) = \cos(ix)$ and $\sinh(x) = -i \sin(ix)$.

Theorem 53. (i) If $w \in \mathbb{C}$ the equation $z^2 = w$ has a solution. However there is no continuous function $f : \mathbb{C} \rightarrow \mathbb{C}$ such that $f(w)^2 = w$ for all $w \in \mathbb{C}$.

(ii) The equation $\exp(z) = 0$ has no solution. If $w \neq 0$ then the set of solutions for $\exp(z) = w$ has the form

$$\{\log(|w|) + i(\theta + 2n\pi) : n \in \mathbb{Z}\}$$

for some real θ . However there is no continuous function $g : \mathbb{C} \setminus \{0\} \rightarrow \mathbb{C}$ such that $\exp(g(w)) = w$ for all $w \in \mathbb{C} \setminus \{0\}$.

Definition 54. A continuous map $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^3$ is called a curve in \mathbb{R}^3 . If $\mathbf{r}(a) = \mathbf{r}(b)$ we say that the curve is closed. We agree to identify (i.e. consider as the same) the two curves $\mathbf{r}_1 : [a_1, b_1] \rightarrow \mathbb{R}^3$ and $\mathbf{r}_2 : [a_2, b_2] \rightarrow \mathbb{R}^3$ if there exists a continuous bijective function $\gamma : [a_1, b_1] \rightarrow [a_2, b_2]$ with $\gamma(a_1) = a_2$ and $\gamma(b_1) = b_2$ such that $\mathbf{r}_1(t) = \mathbf{r}_2(\gamma(t))$.

Example 55. If the curves $\mathbf{r}_j : [0, 1] \rightarrow \mathbb{R}^3$ are defined by

$$\begin{aligned}\mathbf{r}_1(t) &= (\cos(2\pi t), \sin(2\pi t), 0), \\ \mathbf{r}_2(t) &= (\cos(2\pi t^2), \sin(2\pi t^2), 0), \\ \mathbf{r}_3(t) &= (\cos(4\pi t), \sin(4\pi t), 0),\end{aligned}$$

then \mathbf{r}_1 and \mathbf{r}_2 represent the same curve but \mathbf{r}_3 represents a different curve.

We say that \mathbf{r}_1 and \mathbf{r}_2 in the preceding definition are two parameterisations of the same curve.

Definition 56. If $\mathbf{r} : [0, l] \rightarrow \mathbb{R}^3$ is continuous and

$$\left\| \frac{\mathbf{r}(s + \delta s) - \mathbf{r}(s)}{\delta s} \right\| \rightarrow 1$$

as $\delta s \rightarrow 0$ we say that \mathbf{r} is an arc length parameterisation of the curve.

Just as in the case of functions $f : \mathbb{R} \rightarrow \mathbb{R}$ where it is natural to identify the linear map $Df(t) : \mathbb{R} \rightarrow \mathbb{R}$ which takes h to $f'(t)h$ with the real number $f'(t)$ so, in the case of functions $\mathbf{r} : [0, l] \rightarrow \mathbb{R}^3$, it is natural to identify the linear map $D\mathbf{r}(t) : [0, l] \rightarrow \mathbb{R}^3$ which takes h to $(r_1(t)h, r_2(t)h, r_3(t)h)$ with the vector $\dot{\mathbf{r}}(t) = (r_1(t), r_2(t), r_3(t))$. We observe that (when \mathbf{r} is well behaved)

$$\left\| \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t)}{\delta t} - \dot{\mathbf{r}}(t) \right\| \rightarrow 0$$

as $\delta t \rightarrow 0$.

Theorem 57. If $\mathbf{r} : [0, l] \rightarrow \mathbb{R}^3$ is an arc length parameterisation of a well behaved curve then

$$\dot{\mathbf{r}}(s) = \mathbf{t}(s),$$

where $\mathbf{t}(s)$ is a unit vector defining the direction of the tangent. Either $\dot{\mathbf{t}}(s) = \mathbf{0}$, or

$$\dot{\mathbf{t}}(s) = \frac{\mathbf{n}(s)}{\rho(s)},$$

where $\mathbf{n}(s)$ is a unit vector, perpendicular to $\mathbf{t}(s)$, defining the direction of the normal. The scalar $\rho(s) > 0$ is called the radius of curvature (not in the syllabus).

Theorem 58 (Taylor's Theorem With Integral Remainder). Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ is $n + 1$ times continuously differentiable. Then

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \frac{f'''(a)(x-a)^3}{3!} + \dots + \frac{f^{(n)}(a)(x-a)^n}{n!} + R_n(f, x),$$

where

$$R_n(f, x) = \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x-t)^n dt.$$

Corollary 59 (Local Taylor Theorem). If $f : \mathbb{R} \rightarrow \mathbb{R}$ is n times continuously differentiable then, given any $\epsilon > 0$, there exists a $\delta(\epsilon, a)$ such that

$$\left| f(x) - \left(f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} \dots + \frac{f^{(n)}(a)(x-a)^n}{n!} \right) \right| < \epsilon |x-a|^n$$

for all x with $|x - a| < \delta(\epsilon, a)$.

Theorem 60 (Binomial Theorem). *If α and x are real and $|x| < 1$ then*

$$(1 - x)^\alpha = \sum_{n=0}^{\infty} \frac{\alpha(\alpha - 1)(\alpha - 2) \dots (\alpha - (n - 1))}{n!} x^n.$$

If α is a positive integer there are only a finite number of non-zero terms; otherwise the sum diverges if $|x| > 1$.

Example 61 (Cauchy). *Let*

$$\begin{aligned} E(x) &= \exp(-1/x^2) \text{ for } x \neq 0, \\ E(0) &= 0. \end{aligned}$$

Then E is infinitely differentiable with

$$\begin{aligned} E^{(n)}(x) &= Q_n(1/x) \exp(-1/x^2) \text{ for } x \neq 0, \\ E^{(n)}(0) &= 0, \end{aligned}$$

where Q_n is a polynomial. The Taylor expansion

$$E(x) = \sum_{n=0}^{\infty} \frac{E^{(n)}(0)}{n!} x^n$$

is only valid at the single point $x = 0$.

Next year we shall see that analytic functions (i.e. differentiable functions $f : \mathbb{C} \rightarrow \mathbb{C}$) form such a restricted class that a Taylor expansion is always possible. (We shall make no essential use of this result so it serves mainly as a trailer for sensational results to come.)

Theorem 62 (Taylor's Theorem For Analytic Functions). *If $f : \{z \in \mathbb{C} : |z| < r\} \rightarrow \mathbb{C}$ is once complex differentiable it is infinitely complex differentiable and*

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} z^n$$

for all $|z| < r$.

So far in this course we have been able to express the concepts in a direct geometric manner and then give computational rules for them involving coordinates. Our treatment of higher derivatives will, however, be purely computational. (This will be remedied in more advanced courses but the standard geometric approach requires a certain amount of multilinear algebra.)

In what follows we shall consider \mathbb{R}^n and \mathbb{R}^m each with a fixed basis and so fixed coordinate systems (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_m) . Let

f be a function $\mathbb{R}^n \rightarrow \mathbb{R}^m$. Then, if $1 \leq i \leq m$, f_i is a function $\mathbb{R}^n \rightarrow \mathbb{R}$ and (if f is well behaved) $f_{i,j}$ is just another such function $\mathbb{R}^n \rightarrow \mathbb{R}$ [$1 \leq j \leq n$]. We can therefore form $(f_{i,j})_{,k}$ [$1 \leq k \leq n$] and so on. To simplify notation we write

$$f_{i,jk} = (f_{i,j})_{,k}, \quad f_{i,jkl} = (f_{i,jk})_{,l}, \dots$$

and so on [$1 \leq l \leq n$]. In partial derivative notation

$$f_{i,jk} = \frac{\partial^2 f_i}{\partial x_k \partial x_j}, \quad f_{i,jkl} = \frac{\partial^3 f_i}{\partial x_l \partial x_k \partial x_j},$$

Applying the Local Taylor Theorem of Corollary 59 to the function $F : \mathbb{R} \rightarrow \mathbb{R}$ given by $F(t) = f_i(\mathbf{a} + (\mathbf{x} - \mathbf{a})t)$ we obtain the following multidimensional version.

Theorem 63 (Local Multidimensional Taylor Theorem). *Let f be a function $\mathbb{R}^n \rightarrow \mathbb{R}^m$. If f is sufficiently well behaved (more exactly if f has continuous partial derivatives of sufficiently high order) then, given any $\epsilon > 0$, there exists a $\delta(\epsilon, \mathbf{a})$ such that*

$$\left\| f_i(\mathbf{x}) - \left(f_i(\mathbf{a}) + \sum_{1 \leq j \leq n} f_{i,j}(\mathbf{a})(x_j - a_j) + \sum_{1 \leq j, k \leq n} \frac{f_{i,jk}(\mathbf{a})(x_j - a_j)(x_k - a_k)}{2!} + \dots \right. \right. \\ \left. \left. + \sum_{1 \leq j, k, \dots, p \leq n} \frac{f_{i,jk\dots p}(\mathbf{a})(x_j - a_j)(x_k - a_k) \dots (x_p - a_p)}{N!} \right) \right\| < \epsilon \|\mathbf{x} - \mathbf{a}\|^N.$$

By a careful exploitation of these ideas (which we leave to next year) it is possible to prove the following useful result.

Theorem 64 (Partial Differentiation Commutes). *Let f be a function $\mathbb{R} \rightarrow \mathbb{R}^m$. If f is sufficiently well behaved (more exactly if f has continuous second partial derivatives) then $f_{,jk} = f_{,kj}$.*

In other words

$$\frac{\partial^2 f}{\partial x_k \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_k}.$$

Applied to the function of Theorem 63 the result shows that $f_{i,jk} = f_{i,kj}$ and that the higher order terms have similar symmetries. Note that some condition on the behaviour of f is required in Theorem 64 since it is possible to construct pathological functions for which $f_{,jk}$ and $f_{,kj}$ exist but are not equal.

Combining Theorem 62 which says that analytic functions are well behaved, with Theorem 64 which says that, for well behaved functions, the partial derivatives commute and with the Cauchy Riemann equations of Theorem 31 we obtain the following result which turns out to be very important for later work.

Theorem 65. *If $f : \mathbb{C} \rightarrow \mathbb{C}$ is a well behaved analytic function and we write $f(x + iy) = u(x, y) + iv(x, y)$ as in Theorem 31 then u and v both satisfy Laplace's equation*

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0.$$

Functions which satisfy Laplace's equation are called harmonic. In the second half of this course you will see that the converse to Theorem 65 holds and every harmonic function is (at least locally) the real part of an analytic function.

We now return to the local Taylor expansion of Theorem 63 in the special but important case when $m = 1$ concentrating particularly on the cases when $n \leq 3$ which are easiest to visualise. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is well behaved Theorem 63 tells us that

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{1 \leq j \leq n} f_{,j}(\mathbf{a})(x_j - a_j) + \sum_{1 \leq j, k \leq n} \frac{f_{,jk}(\mathbf{a})(x_j - a_j)(x_k - a_k)}{2!} + \text{error},$$

where the error term decreases faster than $\|\mathbf{x} - \mathbf{a}\|^2$ as \mathbf{x} approaches \mathbf{a} . Bearing in mind the symmetry of the second derivative, we may write this in matrix terms as

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \mathbf{b}^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T B \mathbf{h} + \text{error}, \quad (*')$$

where \mathbf{h} is a column vector, \mathbf{b} is a column vector with $b_j = f_{,j}(\mathbf{a})$ and B is an $n \times n$ symmetric matrix with $b_{jk} = f_{,jk}(\mathbf{a})$. We call B the Hessian matrix. We give \mathbf{b} the name $\text{grad}(f)$, ∇f , or $\nabla f(\mathbf{a})$ so that $(*)$ becomes

$$f(\mathbf{a} + \mathbf{h}) = (f(\mathbf{a}) + \nabla f(\mathbf{a})^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T B \mathbf{h} + \text{error}), \quad (*'')$$

Alternatively we may write the equation in terms of linear maps as

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + (Df(\mathbf{a}))(\mathbf{h}) + \frac{1}{2} \mathbf{h} \cdot \beta \mathbf{h} + \text{error}, \quad (**)$$

where the dot denotes inner product and β is the linear map with matrix B . Combining the ideas of $(*'')$ and $(**)$ gives yet another form of the equation.

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{h}) + \frac{1}{2} \mathbf{h} \cdot \beta \mathbf{h} + \text{error}. \quad (*)$$

To help follow the rest of this discussion the reader should draw contour lines (that is lines in \mathbb{R}^2 on which f is constant) for f when $n = 2$ and imagine contour surfaces (that is surfaces in \mathbb{R}^3 on which f is constant) for f when $n = 3$. Observe that $(*)$ tells us that

$$f(\mathbf{a} + \delta \mathbf{a}) - f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot \delta \mathbf{a} + \text{error},$$

or, still more briefly

$$\delta f = \nabla f \cdot \delta \mathbf{a} + \text{error},$$

where the error decreases faster than linearly. If we move along a line (or surface) on which f is constant then $\delta f = 0$ so ∇f must be perpendicular to contour lines (or surfaces). Working in this context it is clear that ∇f is a vector in the direction that f changes most rapidly of length proportional to the rate of change of f in that direction¹

So long as $\nabla f \neq \mathbf{0}$ the linear term $\nabla f \cdot \delta \mathbf{h}$ dominates all the other non-constant terms in (*) (so, in particular, we cannot have a maximum or a minimum at such a point). At, so called, stationary points $\nabla f = \mathbf{0}$ and (*) becomes

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = \frac{1}{2} \mathbf{h} \cdot \beta \mathbf{h} + \text{error}. \quad (*)$$

If we consider the case when $n = 2$ and write (*) in coordinate form we get

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) = \frac{1}{2}(f_{,11}h_1^2 + 2f_{,12}h_1h_2 + f_{,22}h_2^2) + \text{error},$$

with the error decreasing faster than quadratically. Using the ideas of the previous course we see that

(i) If $f_{,11} > 0$ and $\det B = f_{,11}f_{,22} - f_{,12}^2 > 0$ then the Hessian

$$B = \begin{pmatrix} f_{,11} & f_{,12} \\ f_{,21} & f_{,22} \end{pmatrix}$$

is positive definite and \mathbf{a} is a minimum.

(ii) If $f_{,11} < 0$ and $\det B < 0$ then the Hessian is negative definite and \mathbf{a} is a maximum.

(iii) If $f_{,11}$ and $\det B$ are non-zero and of opposite signs then \mathbf{a} is not a maximum nor a minimum. (We have a saddle point.)

(iv) In all other cases the behaviour depends on higher order terms. (But see part (ii) of Remark 66 below.)

No new phenomena emerge in higher dimensions ($n \geq 3$) but the calculations become a bit more complicated.

We make the following remarks which are non-examinable and will be proved by hand waving.

¹WARNING The geometric introduction of ∇f above depends on using a particular inner product. ∇f behaves as a vector so long as we confine ourselves to orthogonal changes of coordinates. If we change our scales of measurement then it behaves in an unexpected manner. The Pure Mathematician explains this by saying that ∇f is indeed a vector but lives in a dual space and the Applied Mathematician explains this by saying that ∇f is indeed a vector but of contravariant type. Fortunately the problem does not arise until Part II, and, possibly, not even then, so you may snopake this footnote out and forget it.

Remark 66. (i) If $f : \mathbb{R} \rightarrow \mathbb{R}$ is a well behaved function and $\epsilon > 0$ there exists a well behaved perturbation $\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}$ such that $|f(t) - \tilde{f}(t)| < \epsilon$ for all $t \in \mathbb{R}$ and all the stationary points of \tilde{f} are maxima and minima.

(ii) If $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a well behaved function and $\epsilon > 0$ there exists a well behaved perturbation $\tilde{g} : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that $|g(\mathbf{t}) - \tilde{g}(\mathbf{t})| < \epsilon$ for all $\mathbf{t} \in \mathbb{R}^2$ and, at each stationary point the matrix

$$\begin{pmatrix} \tilde{g}_{,11} & \tilde{g}_{,12} \\ \tilde{g}_{,21} & \tilde{g}_{,22} \end{pmatrix}$$

is invertible and $\tilde{g}_{,11} \neq 0$.

The following theorem (which is not in the syllabus) indicates why saddle points cannot be perturbed away.

Theorem 67 (Lakes, Peaks and Passes). Consider the sphere

$$S_2 = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\}.$$

If $f : S_2 \rightarrow \mathbb{R}$ is well behaved with L minima, P maxima and S saddle points then $P - S + L = 2$.

More generally the number $P - S + L$ is a ‘topological invariant’ for the associated surface.