

MA3101

# Analysis III

Autumn 2021

Satvik Saha  
19MS154

*Indian Institute of Science Education and Research, Kolkata,  
Mohanpur, West Bengal, 741246, India.*

## Contents

<b>1</b>	<b>Euclidean spaces</b>	<b>1</b>
1.1	$\mathbb{R}^n$ as a vector space . . . . .	1
1.2	$\mathbb{R}^n$ as a metric space . . . . .	2
1.3	$\mathbb{R}^n$ as a topological space . . . . .	4
1.4	Compact sets in $\mathbb{R}^n$ . . . . .	5
1.5	Continuous maps . . . . .	7
1.6	Connectedness . . . . .	8
<b>2</b>	<b>Differential calculus</b>	<b>10</b>
2.1	Differentiability . . . . .	10
2.2	Chain rule . . . . .	11
2.3	Partial derivatives . . . . .	12
2.3.1	Directional derivatives . . . . .	14
2.3.2	Differentiation on manifolds * . . . . .	15
2.4	Mean value theorem . . . . .	16
2.5	Inverse and implicit function theorems . . . . .	18
2.6	Taylor's theorem . . . . .	19
2.7	Critical points and extrema . . . . .	19
<b>3</b>	<b>Integral calculus</b>	<b>20</b>
3.1	Path integrals . . . . .	20
3.2	Multiple integrals . . . . .	22

## 1 Euclidean spaces

### 1.1 $\mathbb{R}^n$ as a vector space

We are familiar with the vector space  $\mathbb{R}^n$ , with the standard inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \cdots + x_n y_n.$$

The standard norm is defined as

$$\|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle = \sum_{k=1}^n (x_k - y_k)^2.$$

**Exercise 1.1.** What are all possible inner products on  $\mathbb{R}^n$ ?

*Solution.* Note that an inner product is a bilinear, symmetric map such that  $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ , and  $\langle \mathbf{x}, \mathbf{x} \rangle = 0$  if and only if  $\mathbf{x} = \mathbf{0}$ . Thus, an product map on  $\mathbb{R}^n$  is completely and uniquely determined by the values  $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = a_{ij}$ . Let  $A$  be the  $n \times n$  matrix with entries  $a_{ij}$ . Note that  $A$  is a real symmetric matrix with positive entries. Now,

$$\langle \mathbf{x}, \mathbf{e}_j \rangle = x_1 a_{1j} + \cdots + x_n a_{nj} = \mathbf{x}^\top \mathbf{a}_j,$$

where  $\mathbf{a}_j$  is the  $j^{\text{th}}$  column of  $A$ . Thus,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{a}_1 y_1 + \cdots + \mathbf{x}^\top \mathbf{a}_n y_n = \mathbf{x}^\top A \mathbf{y}.$$

Furthermore, any choice of real symmetric  $A$  with positive entries produces an inner product.

**Theorem 1.1** (Cauchy-Schwarz). *Given two vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ , we have*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|.$$

*Proof.* This is trivial when  $\mathbf{w} = \mathbf{0}$ . When  $\mathbf{w} \neq \mathbf{0}$ , set  $\lambda = \langle \mathbf{v}, \mathbf{w} \rangle / \|\mathbf{w}\|^2$ . Thus,

$$0 \leq \|\mathbf{v} - \lambda \mathbf{w}\|^2 = \|\mathbf{v}\|^2 - 2\lambda \langle \mathbf{v}, \mathbf{w} \rangle + \lambda^2 \|\mathbf{w}\|^2.$$

Simplifying,

$$0 \leq \|\mathbf{v}\|^2 - \frac{|\langle \mathbf{v}, \mathbf{w} \rangle|^2}{\|\mathbf{w}\|^2}.$$

This gives the desired result. Clearly, equality holds if and only if  $\mathbf{v} = \lambda \mathbf{w}$ . □

**Theorem 1.2** (Triangle inequality). *Given two vectors  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ , we have*

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

*Proof.* Write

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle + \|\mathbf{w}\|^2 \leq \|\mathbf{v}\|^2 + 2|\langle \mathbf{v}, \mathbf{w} \rangle| + \|\mathbf{w}\|^2.$$

Applying Cauchy-Schwarz gives

$$\|\mathbf{v} + \mathbf{w}\|^2 \leq (\|\mathbf{v}\| + \|\mathbf{w}\|)^2.$$

Equality holds if and only if  $\mathbf{v} = \lambda \mathbf{w}$  for  $\lambda \geq 0$ . □

## 1.2 $\mathbb{R}^n$ as a metric space

Our previous observations allow us to define the standard metric on  $\mathbb{R}^n$ , seen as a point set.

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

**Definition 1.1.** For any  $\delta > 0$ , the set

$$B_\delta(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) < \delta\}$$

is called the open ball centred at  $\mathbf{x} \in \mathbb{R}^n$  with radius  $\delta$ . This is also called the  $\delta$  neighbourhood of  $\mathbf{x}$ .

**Definition 1.2.** A set  $U$  is open in  $\mathbb{R}^n$  if for every  $\mathbf{x} \in U$ , there exists an open ball  $B_\delta(\mathbf{x}) \subset U$ .

*Remark.* Every open ball in  $\mathbb{R}^n$  is open.

*Remark.* Both  $\emptyset$  and  $\mathbb{R}^n$  are open.

**Definition 1.3.** A set  $F$  is closed in  $\mathbb{R}^n$  if its complement  $\mathbb{R}^n \setminus F$  is open in  $\mathbb{R}^n$ .

*Remark.* Both  $\emptyset$  and  $\mathbb{R}^n$  are closed.

*Remark.* Finite sets in  $\mathbb{R}^n$  are closed.

**Theorem 1.3.** *Unions and finite intersections of open sets are open.*

**Corollary 1.3.1.** *Intersections and finite unions of closed sets are closed.*

**Definition 1.4.** An interior point  $x$  of a set  $S \subseteq \mathbb{R}^n$  is such that there is a neighbourhood of  $x$  contained within  $S$ .

*Example.* Every point in an open set is an interior point by definition. The interior of a set is the largest open set contained within it.

**Definition 1.5.** An exterior point  $x$  of a set  $S \subseteq \mathbb{R}^n$  is an interior point of the complement  $\mathbb{R}^n \setminus S$ .

**Definition 1.6.** A boundary point of a set is neither an interior point, nor an exterior point.

*Example.* The boundary of the unit open ball  $B_1(0) \subset \mathbb{R}^n$  is the sphere  $S^{n-1}$ .

**Definition 1.7.** A limit point  $x$  of a set  $S \subseteq \mathbb{R}^n$  is such that every neighbourhood of  $x$  contains a point from  $S$  other than itself.

**Definition 1.8.** The closure of a set  $S \subseteq \mathbb{R}^n$  is the union of  $S$  and its limit points.

*Remark.* The closure of a set is the smallest closed set containing it.

**Lemma 1.4.** Every open set in  $\mathbb{R}^n$  is a union of open balls.

*Proof.* Let  $U \subseteq \mathbb{R}^n$  be open. Thus, for every  $\mathbf{x} \in \mathbb{R}^n$ , we can choose  $\delta_{\mathbf{x}} > 0$  such that  $B_{\delta_{\mathbf{x}}}(\mathbf{x}) \subset U$ . The union of all such open balls is precisely the set  $U$ .  $\square$

### 1.3 $\mathbb{R}^n$ as a topological space

**Definition 1.9.** A topology on a set  $X$  is a collection  $\tau$  of subsets of  $X$  such that

1.  $\emptyset \in \tau$
2.  $X \in \tau$
3. Arbitrary union of sets from  $\tau$  belong to  $\tau$ .
4. Finite intersections of sets from  $\tau$  belong to  $\tau$ .

Sets from  $\tau$  are called open sets.

*Example.* The Euclidean metric induces the standard topology on  $\mathbb{R}^n$ .

*Example.* The discrete topology on a set  $X$  is one where every singleton set is open. This is the topology induced by the discrete metric,

$$d_{\text{discrete}}: X \times X \rightarrow \mathbb{R}, \quad (x, y) \mapsto \begin{cases} 0, & \text{if } x = y, \\ 1, & \text{if } x \neq y. \end{cases}$$

*Example.* Let  $X$  be an infinite set. The collection of sets consisting of  $\emptyset$  along with all sets  $A$  such that  $X \setminus A$  is finite is a topology on  $X$ . This is called the Zariski topology.

*Example.* Consider the set of real numbers, and let  $\tau$  be the collection  $\emptyset, \mathbb{R}$ , and all intervals  $(-x, +x)$  for  $x > 0$ . This constitutes a topology on  $\mathbb{R}$ , very different from the usual one.

This topology cannot be induced by a metric; it is not metrizable.

Consider the constant sequence of zeros. In this topology  $(\mathbb{R}, \tau)$ , this sequence converges to *every* point in  $\mathbb{R}$ . Given any  $\ell \in \mathbb{R}$ , the open neighbourhoods of  $\ell$  are precisely the sets  $\mathbb{R}$  and the open intervals  $(-x, +x)$  for  $x > |\ell|$ . The tail of the constant sequence of zeros is contained within every such neighbourhood of  $\ell$ , hence  $0 \rightarrow \ell$ . Indeed, the element zero belongs to every open set apart from  $\emptyset$  in this topology.

**Definition 1.10.** A topological space is called Hausdorff if for every distinct  $x, y \in X$ , there exist disjoint neighbourhoods of  $x$  and  $y$ .

*Example.* Every metric space is Hausdorff. Given distinct  $x, y$  in a metric space  $(X, d)$ , set  $\delta = d(x, y)/3$  and consider the open balls  $B_\delta(x)$  and  $B_\delta(y)$ .

**Lemma 1.5.** Every convergent sequence in a Hausdorff space has exactly one limit.

*Proof.* Consider a sequence  $\{x_n\}_{n \in \mathbb{N}}$ , and suppose that it converges to distinct  $x_1$  and  $x_2$ . Construct disjoint neighbourhoods  $U_1$  and  $U_2$  around  $x_1$  and  $x_2$ . Now, convergence implies that both  $U_1$  and  $U_2$  contain the tail of  $\{x_n\}$ , which is impossible since they are disjoint and hence contain no elements in common.  $\square$

**Definition 1.11.** Given a topological space  $(X, \tau)$  and a subset  $Y \subseteq X$ , the collection of sets  $U \cap Y$  where  $U \in \tau$  is a topology  $\tau_Y$  on  $Y$ . We call this collection the subspace topology on  $Y$ , induced by the topology on  $X$ .

## 1.4 Compact sets in $\mathbb{R}^n$

**Definition 1.12.** A set  $K \subset X$  in a topological space is compact if every open cover of  $K$  has a finite sub-cover. That is, for every collection of open sets  $\{U_\alpha\}_{\alpha \in A}$  such that  $K$  is contained in their union, there exists a finite sub-collection  $U_{\alpha_1}, \dots, U_{\alpha_k}$  such that  $K$  is also contained in their union.

*Example.* All finite sets are compact.

*Example.* Given a convergent sequence of real numbers  $x_n \rightarrow x$ , the collection  $\{x_n\}_{n \in \mathbb{N}} \cup \{x\}$  is compact.

*Example.* In  $\mathbb{R}^n$ , compact sets are precisely those sets which are closed and bounded. This is the Heine-Borel Theorem.

**Theorem 1.6.** The closed intervals  $[a, b] \subset \mathbb{R}$  are compact.

*Remark.* This can be extended to show that any  $k$ -cell  $[a_1, b_1] \times \dots \times [a_n, b_n] \subset \mathbb{R}^n$  is compact.

*Proof.* Let  $\{U_\alpha\}_{\alpha \in A}$  be an open cover of  $[a, b]$ , and suppose that  $I_1 = [a, b]$  has no finite sub-cover. Then, at least one of the intervals  $[a, (a+b)/2]$  and  $[(a+b)/2, b]$  must not have a finite sub-cover; pick one and call it  $I_2$ . Similarly, one of the halves of  $I_2$  must not have a finite

sub-cover; call it  $I_3$ . In this process, we generate a sequence of closed intervals  $I_1 \supset I_2 \supset \dots$ , none of which have a finite sub-cover. The length of each interval is given by

$$|I_n| = 2^{-n+1}\|b - a\| \rightarrow 0.$$

Now, pick a sequence of points  $\{x_n\}$  where each  $x_n \in I_n$ . Then,  $\{x_n\}$  is a Cauchy sequence. To see this, given any  $\epsilon > 0$ , we can find sufficiently large  $n_0$  such that  $2^{-n_0+1}\|b - a\| < \epsilon$ . Thus,  $x_n \in I_n \subset I_{n_0}$  for all  $n \geq n_0$ , which means that for any  $m, n \geq n_0$ , we have  $x_m, x_n \in I_{n_0}$  forcing<sup>1</sup>

$$\|x_m - x_n\| \leq |I_{n_0}| = 2^{-n_0+1}\|b - a\| < \epsilon.$$

From the completeness of  $\mathbb{R}$ , this sequence must converge in  $\mathbb{R}$ , specifically in  $[a, b]$ . Thus,  $x_n \rightarrow x$  for some  $x \in [a, b]$ . It can also be seen that the limit  $x \in I_n$  for all  $n \in \mathbb{N}$ ; if not, say  $x \notin I_{n_0}$ , then  $x \in [a, b] \setminus I_{n_0}$  which is open, hence there is an open interval such that  $(x - \delta, x + \delta) \cap I_{n_0} = \emptyset$ . However,  $I_{n_0}$  contains all  $x_{n \geq n_0}$ , thus this  $\delta$ -neighbourhood of  $x$  would miss out a tail of  $\{x_n\}$ .

Now, pick the open set  $U \in \{U_\alpha\}$  which covers the point  $x$ . Thus,  $x \in U$  so  $U$  contains some non-empty open interval  $(x - \delta, x + \delta)$  around  $x$ . Choose  $n_0$  such that  $2^{-n_0+1}\|b - a\| < \delta$ ; this immediately gives  $I_{n_0} \subseteq (x - \delta, x + \delta) \subset U$ . This contradicts that fact that  $I_{n_0}$  has no finite sub-cover from  $\{U_\alpha\}$ , completing the proof.  $\square$

*Remark.* The fact that Cauchy sequences in  $\mathbb{R}^n$  converge isn't immediately obvious; it is a consequence of the completeness of  $\mathbb{R}^n$ . Start by noting that  $\mathbb{R}$  has the Least Upper Bound property, from which the Monotone Convergence Theorem follows; every monotonic, bounded sequence of reals converges. It can also be shown that any sequence of reals with contain a monotone subsequence, from which it follows that every bounded sequence contains a convergent subsequence (Bolzano-Weierstrass). Finally, it can be shown that if a subsequence of a Cauchy sequence converges, then the entire sequence also converges to the same limit, giving us the desired result for  $\mathbb{R}$ . For sequence in  $\mathbb{R}^n$ , we may apply this coordinate-wise to obtain the result.

**Lemma 1.7.** *Compact sets in  $\mathbb{R}^n$  are closed and bounded.*

*Proof.* Consider a compact set  $K \subset \mathbb{R}^n$ . Let  $x \in \mathbb{R}^n \setminus K$ , and let  $y \in K$ . Since  $x \neq y$ , we choose open balls  $U_y$  around  $y$  and  $V_y$  around  $x$  such that  $U_y \cap V_y = \emptyset$ . Repeating this for all  $y \in K$ , we generate an open cover  $\{U_y\}$  of  $K$  consisting of open balls. The compactness of  $K$  guarantees that this has a finite sub-cover, i.e. there is a finite set  $Y$  such that the collection  $\{U_y\}_{y \in Y}$  covers  $K$ . As a result, the finite intersection of all  $V_y$  for  $y \in Y$  is contained within  $\mathbb{R}^n \setminus K$ . Thus,  $x$  is in the exterior of  $K$ . Since  $x$  was chosen arbitrarily from  $\mathbb{R}^n \setminus K$ , we see that  $K$  is closed.

Now, consider the open cover  $\{B_1(x)\}_{x \in K}$ , and extract a finite sub-cover of unit open balls. The distance between any two points in  $K$  is at most the maximum distance between the centres of any two balls in our sub-cover, plus two.  $\square$

**Lemma 1.8.** *The intersection of a closed set and a compact set is compact.*

<sup>1</sup>If  $x_1, x_2 \in [a, b]$  with  $x_1 < x_2$ , note that  $a \leq x_1 < x_2 \leq b$ , so

$$|x_2 - x_1| = x_2 - x_1 \leq b - a.$$

*Proof.* Let  $F \subseteq \mathbb{R}^n$  be closed and let  $K \subseteq \mathbb{R}^n$  be compact. Suppose that the open cover  $\{U_\alpha\}$  of  $F \cap K$  has no finite sub-cover. Now the complement  $U = F^c$  is open in  $\mathbb{R}^n$ , hence the collection  $\{U_\alpha\} \cup \{U\}$  is an open cover of  $K$ , and hence must admit a finite sub-cover of  $K$ . In particular, this must be a finite sub-cover of  $F \cap K$ . However, we can remove the set  $U$  from this sub-cover since it shares no element with  $F \cap K$ ; as a result, our sub-cover must be a finite sub-collection of sets  $U_\alpha$ , contradicting our assumption. This shows that  $F \cap K$  is compact.  $\square$

**Lemma 1.9** (Finite intersection property). *Let  $\{K_\alpha\}$  be a collection of compact sets in  $\mathbb{R}^n$  which have the property that any finite intersection of them is non-empty. Then,*

$$\bigcap_{\alpha} K_{\alpha} \neq \emptyset.$$

*Proof.* Suppose to the contrary that the intersection of all  $K_\alpha$  is empty. Fix an index  $\beta$ , and note that no element of  $K_\beta$  lies in every  $K_\alpha$ . Set  $J_\alpha = K_\alpha^c$ , whence the collection  $\{J_\alpha : \alpha \neq \beta\}$  is an open cover of  $K_\beta$ . This must admit a finite sub-cover  $\{J_{\alpha_1}, \dots, J_{\alpha_k}\}$  of  $K_\beta$ . Thus, we must have

$$K_\beta^c \cup J_{\alpha_1} \cup \dots \cup J_{\alpha_k} = \mathbb{R}^n.$$

This immediately gives the contradiction

$$K_\beta \cap K_{\alpha_1} \cap \dots \cap K_{\alpha_k} = \emptyset. \quad \square$$

**Theorem 1.10** (Heine-Borel). *Compact sets in  $\mathbb{R}^n$  are precisely those that are closed and bounded.*

*Proof.* Given a compact set in  $\mathbb{R}^n$ , we have already shown that it must be closed and bounded. Next, if  $F \subset \mathbb{R}^n$  is closed and bounded, it can be enclosed within a  $k$ -cell which we know is compact. Thus,  $F$  is the intersection of the closed set  $F$  and the compact  $k$ -cell, proving that  $F$  must be compact.  $\square$

## 1.5 Continuous maps

**Definition 1.13.** A map  $f: X \rightarrow Y$  is continuous if the pre-image of every open set from  $Y$  is open in  $X$ .

**Lemma 1.11.** *A map  $f: X \rightarrow Y$  is continuous if the pre-image of every closed set from  $Y$  is closed in  $X$ .*

**Theorem 1.12.** *The projection maps  $\pi_i: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\mathbf{x} \mapsto x_i$  are continuous.*

*Proof.* Let  $U \subseteq \mathbb{R}$  be open; we claim that  $\pi_i^{-1}(U)$  is open. Pick  $\mathbf{x} \in \pi_i^{-1}(U)$ , and note that  $\pi_i(\mathbf{x}) = x_i \in U$ . Thus, there exists  $\delta > 0$  such that  $(x_i - \delta, x_i + \delta) \subset U$ . Now examine  $B_\delta(\mathbf{x})$ ; for any point  $\mathbf{y}$  within this open ball, we have  $d(\mathbf{x}, \mathbf{y}) < \delta$  hence

$$|x_i - y_i|^2 \leq \sum_{k=1}^n (x_k - y_k)^2 = d(\mathbf{x}, \mathbf{y})^2 < \delta^2.$$

In other words,  $\pi_i(\mathbf{y}) = y_i \in (x_i - \delta, x_i + \delta)$ , hence  $\pi_i B_\delta(\mathbf{x}) \subseteq (x_i - \delta, x_i + \delta) \subset U$ . Thus, given arbitrary  $\mathbf{x} \in \pi_i^{-1}(U)$ , we have found an open ball  $B_\delta(\mathbf{x}) \subset \pi_i^{-1}(U)$ .  $\square$

**Lemma 1.13.** *Finite sums, products, and compositions of continuous functions are continuous.*

**Corollary 1.13.1.** *A function  $f: [a, b] \rightarrow \mathbb{R}^n$  is continuous if and only if the components,  $\pi_i \circ f$ , are continuous.*

**Theorem 1.14.** *All polynomial functions of the coordinates in  $\mathbb{R}^n$  are continuous.*

*Example.* The unit sphere  $S^{n-1} \subset \mathbb{R}^n$  is closed. It is by definition the pre-image of the singleton closed set  $\{1\}$  under the continuous map

$$\mathbf{x} \mapsto x_1^2 + \cdots + x_n^2.$$

**Theorem 1.15.** *The continuous image of a compact set is compact.*

*Proof.* Let  $f: X \rightarrow Y$  be continuous, where  $Y$  is the image of the compact set  $X$ , and let  $\{U_\alpha\}$  be an open cover of  $Y$ . Then, the collection  $\{f^{-1}(U_\alpha)\}$  is an open cover of  $X$ . Using the compactness of  $X$ , extract a finite sub-cover  $f^{-1}(U_{\alpha_1}), \dots, f^{-1}(U_{\alpha_k})$  of  $X$ . It follows that the collection  $U_{\alpha_1}, \dots, U_{\alpha_k}$  is a finite sub-cover of  $Y$ .  $\square$

## 1.6 Connectedness

**Definition 1.14.** Let  $X$  be a topological space. A separation of  $X$  is a pair  $U, V$  of non-empty disjoint open subsets such that  $X = U \cup V$ .

**Definition 1.15.** A connected topological space is one which cannot be separated.

**Lemma 1.16.** *A topological space  $X$  is connected if and only if the only sets which are both open and closed are  $\emptyset$  and  $X$ .*

*Example.* The intervals  $(a, b) \subset \mathbb{R}$  are connected. To see this, suppose that  $U, V$  is a separation of  $(a, b)$ . Pick  $x \in U, y \in V$ , and without loss of generality let  $x < y$ . Define  $S = [x, y] \cap U$ , and set  $c = \sup S$ . It can be argued that  $c \in (a, b)$ , but  $c \notin U, c \notin V$ , using the properties of the supremum.



**Theorem 1.17.** *The continuous image of a connected set is connected.*

*Proof.* Let  $f$  be a continuous map on the connected set  $X$ , and let  $Y$  be the image of  $X$ . If  $U, V$  is a separation of  $Y$ , then it can be shown that  $f^{-1}(U), f^{-1}(V)$  constitutes a separation of  $X$ , which is a contradiction.  $\square$

**Definition 1.16.** A path  $\gamma$  joining two points  $x, y \in X$  is a continuous map  $\gamma: [a, b] \rightarrow X$  such that  $\gamma(a) = x, \gamma(b) = y$ .

**Definition 1.17.** A set in  $X$  is path connected if given any two distinct points in  $X$ , there exists a path joining them.

**Lemma 1.18.** *Every path connected set is connected.*

*Proof.* Let  $X$  be path connected, and suppose that  $U, V$  is a separation of  $X$ . Then, pick  $x \in U, y \in V$ , and choose a path  $\gamma: [0, 1] \rightarrow X$  between  $x$  and  $y$ . The sets  $f^{-1}(U)$  and  $f^{-1}(V)$  separate the interval  $[0, 1]$ , which is a contradiction.  $\square$

*Example.* All connected sets are not path connected. Consider the topologist's sine curve,

$$\left\{ \left( x, \sin \frac{1}{x} \right) : 0 < x \leq 1 \right\} \cup \{(0, 0)\}.$$

**Definition 1.18.** The  $\epsilon$  neighbourhood of a set  $K$  in a metric space  $X$  is defined as

$$\bigcup_{a \in K} B_\epsilon(a) = \bigcup_{a \in K} \{x \in X : d(x, a) < \epsilon\}.$$

**Exercise 1.2.** Let  $K \subseteq \mathbb{R}^n$  be compact, and define  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$f(x) = \text{dist}(x, K) = \inf_{a \in K} d(x, a).$$

Show that  $f$  is continuous on  $\mathbb{R}^n$ , and  $f^{-1}(\{0\}) = K$ .

**Exercise 1.3.** If  $K \subseteq \mathbb{R}^n$  is compact and  $K \cap L = \emptyset$ , then

$$\text{dist}(K, L) = \inf_{a \in K} \text{dist}(a, L) > 0.$$

**Exercise 1.4.** If  $K \subseteq \mathbb{R}^n$  is compact and  $U$  is an open set containing  $K$ , then there exists  $\epsilon > 0$  such that  $U$  contains the  $\epsilon$  neighbourhood of  $K$ .

Is the compactness of  $K$  necessary?

## 2 Differential calculus

### 2.1 Differentiability

**Definition 2.1.** Let  $f: (a, b) \rightarrow \mathbb{R}^n$ , and let  $f_i = \pi_i \circ f$  be its components. Then,  $f$  is differentiable at  $t_0 \in (a, b)$  if the following limit exists.

$$f'(t_0) = \lim_{h \rightarrow 0} \frac{f(t_0 + h) - f(t_0)}{h}.$$

*Remark.* The vector  $f'(t_0)$  represents the tangent to the curve  $f$  at the point  $f(t_0)$ . The full tangent line is the parametric curve  $f(t) + f'(t_0)(t - t_0)$ .

**Definition 2.2.** Let  $U \subseteq \mathbb{R}^n$  be open, and let  $f: U \rightarrow \mathbb{R}^m$ . Then,  $f$  is differentiable at  $x \in U$  if there exists a linear transformation  $\lambda: \mathbb{R}^n \rightarrow \mathbb{R}^m$  such that

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

The derivative of  $f$  at  $x$  is denoted by  $\lambda = Df(x)$ .

*Remark.* In a neighbourhood of  $x$ , we may approximate

$$f(x + h) \approx f(x) + Df(x)(h).$$

*Remark.* The statement that this quantity goes to zero means that each of the  $m$  components must also go to zero. For each of these limits, there are  $n$  axes along which we can let  $h \rightarrow 0$ . As a result, we obtain  $m \times n$  limits, which allow us to identify the  $m \times n$  components of the matrix representing the linear transformation  $\lambda$  (in the standard basis). These are the partial derivatives of  $f$ , and the matrix of  $\lambda$  is the Jacobian matrix of  $f$  evaluated at  $x$ .

*Example.* Let  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear map. By choosing  $\lambda = T$ , we see that  $T$  is differentiable everywhere, with  $DT(x) = T$  for every choice of  $x \in \mathbb{R}^n$ . This is made obvious by the fact that the best linear approximation of a linear map at some point is the map itself; indeed, the ‘approximation’ is exact.

**Lemma 2.1.** If  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $x \in \mathbb{R}^n$ , with derivative  $Df(x)$ , then

1.  $f$  is continuous at  $x$ .
2. The linear transformation  $Df(x)$  is unique.

*Proof.* We prove the second part. Suppose that  $\lambda, \mu$  satisfy the requirements for  $Df(x)$ ; it can be shown that  $\lim_{h \rightarrow 0} (\lambda - \mu)h/\|h\| = 0$ . Now, if  $\lambda v \neq \mu v$  for some non-zero vector  $v \in \mathbb{R}^n$ , then

$$\lambda v - \mu v = \frac{\lambda(tv) - \mu(tv)}{\|tv\|} \cdot \|v\| \rightarrow 0,$$

a contradiction. □

## 2.2 Chain rule

**Exercise 2.1.** Let  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear transformation. Then, there exists  $M > 0$  such that for all  $\mathbf{x} \in \mathbb{R}^n$ , we have

$$\|T\mathbf{x}\| \leq M\|\mathbf{x}\|.$$

*Solution.* Set  $\mathbf{v}_i = T(\mathbf{e}_i)$  where  $\mathbf{e}_i$  are the standard unit basis vectors of  $\mathbb{R}^n$ . Then,

$$\|T\mathbf{x}\| = \left\| \sum_i x_i \mathbf{v}_i \right\| \leq \sum_i \|x_i \mathbf{v}_i\| \leq \max_i \|\mathbf{v}_i\| \sum_i |x_i|.$$

Since each  $|x_i| \leq \|\mathbf{x}\|$ , set  $M = n \max_i \|\mathbf{v}_i\|$  and write

$$\|T\mathbf{x}\| \leq \max_i \|\mathbf{v}_i\| \sum_i |x_i| \leq \max_i \|\mathbf{v}_i\| \cdot n \|\mathbf{x}\| = M\|\mathbf{x}\|.$$

**Theorem 2.2.** Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $g: \mathbb{R}^m \rightarrow \mathbb{R}^k$  where  $f$  is differentiable at  $a \in \mathbb{R}^n$  and  $g$  is differentiable at  $f(a) \in \mathbb{R}^m$ . Then,  $g \circ f$  is differentiable, with  $D(g \circ f)(a) = Dg(f(a)) \circ Df(a)$ . Note that this means that the Jacobian matrices simply multiply.

*Proof.* Set  $b = f(a) \in \mathbb{R}^m$ ,  $\lambda = Df(a)$ ,  $\mu = Dg(f(a))$ . Define

$$\begin{aligned} \varphi: \mathbb{R}^n &\rightarrow \mathbb{R}^m, & \varphi(x) &= f(x) - f(a) - \lambda(x - a), \\ \psi: \mathbb{R}^m &\rightarrow \mathbb{R}^k, & \psi(y) &= g(y) - g(b) - \mu(y - b). \end{aligned}$$

We claim that

$$\lim_{x \rightarrow a} \frac{g \circ f(x) - g \circ f(a) - \mu \circ \lambda(x - a)}{\|x - a\|} = 0.$$

Write the numerator as

$$g \circ f(x) - g \circ f(a) - \mu \circ \lambda(x - a) = \psi(f(x)) + \mu(\varphi(x)).$$

Note that

$$\lim_{x \rightarrow a} \frac{\varphi(x)}{\|x - a\|} = 0, \quad \lim_{y \rightarrow b} \frac{\psi(y)}{\|y - b\|} = 0.$$

Thus, find  $M > 0$  such that  $\|\mu(\varphi(x))\| \leq M\|\varphi(x)\|$  for all  $x \in \mathbb{R}^n$ , hence

$$\lim_{x \rightarrow a} \frac{\|\mu(\varphi(x))\|}{\|x - a\|} \leq \lim_{x \rightarrow a} \frac{M\|\varphi(x)\|}{\|x - a\|} = 0.$$

Now write

$$\lim_{f(x) \rightarrow b} \frac{\psi(f(x))}{\|f(x) - b\|} = 0,$$

hence for any  $\epsilon > 0$ , there is a neighbourhood of  $b$  on which

$$\|\psi(f(x))\| \leq \epsilon \|f(x) - b\| = \epsilon \|\varphi(x) + \lambda(x - a)\|.$$

Apply the triangle inequality and find  $M' > 0$  such that

$$\|\psi(f(x))\| \leq \epsilon \|\varphi(x)\| + \epsilon M' \|x - a\|.$$

Thus,

$$\lim_{x \rightarrow a} \frac{\|\psi(f(x))\|}{\|x - a\|} \leq \lim_{x \rightarrow a} \frac{\epsilon \|\varphi(x)\|}{\|x - a\|} + \epsilon M' = \epsilon M'.$$

Since  $\epsilon > 0$  was arbitrary, this limit is zero, completing the proof.  $\square$

## 2.3 Partial derivatives

**Definition 2.3.** Let  $U \subseteq \mathbb{R}^n$  be open, and let  $f: U \rightarrow \mathbb{R}$ . The partial derivative of  $f$  with respect to the coordinate  $x_j$  at some  $a \in U$  is defined by the following limit, if it exists.

$$\frac{\partial f}{\partial x_j}(a) = \lim_{h \rightarrow 0} \frac{f(a + h\mathbf{e}_j) - f(a)}{h}.$$

**Lemma 2.3.** If  $f: U \rightarrow \mathbb{R}$  is differentiable at a point  $a \in \mathbb{R}^n$ , then

$$Df(a)(x_1, \dots, x_n) = x_1 \frac{\partial f}{\partial x_1}(a) + \dots + x_n \frac{\partial f}{\partial x_n}(a).$$

*Example.* Consider

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto \begin{cases} xy/(x^2 + y^2), & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

Note that  $f$  is not differentiable at  $(0, 0)$ ; it is not even continuous there. However, both partial derivatives of  $f$  exist at  $(0, 0)$ .

**Lemma 2.4.** If  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $a \in \mathbb{R}^n$ , then the matrix representation of  $Df(a)$  in the standard basis is given by

$$[Df(a)] = \left[ \frac{\partial f_i}{\partial x_j}(a) \right]_{ij}.$$

**Lemma 2.5.** Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  be differentiable at  $a \in \mathbb{R}^n$ , and let  $g: \mathbb{R}^m \rightarrow \mathbb{R}^k$  be differentiable at  $f(a) \in \mathbb{R}^m$ . Then, the matrix representation of  $D(g \circ f)(a)$  in the standard basis is the product

$$[D(g \circ f)(a)] = [Dg(f(a))][Df(a)] = \left[ \sum_{\ell=1}^m \frac{\partial g_i}{\partial y_\ell} \frac{\partial f_\ell}{\partial x_j} \right]_{ij}.$$

In other words,

$$\frac{\partial}{\partial x_j}(g \circ f)_i(a) = \sum_{\ell=1}^m \frac{\partial g_i}{\partial y_\ell}(f(a)) \frac{\partial f_\ell}{\partial x_j}(a).$$

*Example.* Let  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  be differentiable, and let  $\Gamma(f) = \{(x, y, f(x, y)) : x, y \in \mathbb{R}\}$  be the graph of  $f$ . Now, let  $\gamma: [-1, 1] \rightarrow \Gamma(f)$  be a differentiable curve, represented by

$$\gamma(t) = (g(t), h(t), f(g(t), h(t))).$$

Then, we can compute the derivative

$$\gamma'(a) = \left( g'(a), h'(a), g'(a) \frac{\partial f}{\partial x} + h'(a) \frac{\partial f}{\partial y} \Big|_{(g(a), h(a))} \right)$$

**Exercise 2.2.** Consider the inner product map,  $\langle \cdot, \cdot \rangle: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ . What is its derivative?

*Solution.* We treat the inner product as a map  $g: \mathbb{R}^{2n} \rightarrow \mathbb{R}$ , which acts as

$$\langle \mathbf{x}, \mathbf{y} \rangle \cong g(x_1, \dots, x_n, y_1, \dots, y_n) = x_1 y_1 + \dots + x_n y_n.$$

Now, note that

$$\frac{\partial g}{\partial x_i} = y_i, \quad \frac{\partial g}{\partial y_i} = x_i.$$

Thus,

$$\begin{aligned} Dg(\mathbf{a}, \mathbf{b})(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^n x_i \frac{\partial g}{\partial x_i}(\mathbf{a}, \mathbf{b}) + \sum_{i=1}^n y_i \frac{\partial g}{\partial y_i}(\mathbf{a}, \mathbf{b}) \\ &= \sum_{i=1}^n x_i b_i + \sum_{i=1}^n y_i a_i \\ &= \langle \mathbf{x}, \mathbf{b} \rangle + \langle \mathbf{y}, \mathbf{a} \rangle. \end{aligned}$$

In other words, the matrix representation of the derivative of the inner product map at the point  $(\mathbf{a}, \mathbf{b})$  is given by  $[\mathbf{b}^\top \ \mathbf{a}^\top]$ .

**Exercise 2.3.** Let  $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$  be a differentiable curve. What is the derivative of the real map  $t \mapsto \|\gamma(t)\|^2$ ?

*Solution.* We write this map as  $t \mapsto \langle \gamma(t), \gamma(t) \rangle$ . Consider the scheme

$$\mathbb{R} \rightarrow \mathbb{R}^{2n} \rightarrow \mathbb{R}, \quad t \mapsto \begin{bmatrix} \gamma(t) \\ \gamma(t) \end{bmatrix} \mapsto \langle \gamma(t), \gamma(t) \rangle.$$

Pick a point  $t \in \mathbb{R}$ , whence the derivative of the map at  $t$  is

$$\begin{bmatrix} \gamma(t)^\top & \gamma(t)^\top \end{bmatrix} \begin{bmatrix} \gamma'(t) \\ \gamma'(t) \end{bmatrix} = 2\langle \gamma(t), \gamma'(t) \rangle.$$

*Remark.* Consider the surface  $S^{n-1} \subset \mathbb{R}^n$ , and pick an arbitrary differentiable curve  $\gamma: \mathbb{R} \rightarrow S^{n-1}$ . Now, the tangent vector  $\gamma'(t)$  is tangent to the sphere  $S^{n-1}$  at any point  $\gamma(t)$ . We claim that this tangent drawn at  $\gamma(t)$  is always perpendicular to the position vector  $\gamma(t)$ . This is made trivial by our exercise: the map  $t \mapsto \|\gamma(t)\|^2 = 1$  is a constant map since  $\gamma$  is a curve on the unit sphere. This means that it has zero derivative, forcing  $\langle \gamma(t), \gamma'(t) \rangle = 0$ .

### 2.3.1 Directional derivatives

**Definition 2.4.** Let  $U \subseteq \mathbb{R}^n$  be open, and let  $f: U \rightarrow \mathbb{R}$ . The directional derivative of  $f$  along a direction  $\mathbf{v} \in \mathbb{R}^n$  at a point  $a \in U$  is defined by the following limit, if it exists.

$$\nabla_{\mathbf{v}} f(a) = \lim_{h \rightarrow 0} \frac{f(a + h\mathbf{v}) - f(a)}{h}.$$

*Example.* Consider

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y) \mapsto \begin{cases} x^3/(x^2 + y^2), & \text{if } (x, y) \neq (0, 0), \\ 0, & \text{if } (x, y) = (0, 0). \end{cases}$$

Note that  $f$  is not differentiable at  $(0, 0)$ . However, all directional derivatives of  $f$  exist at  $(0, 0)$ . Indeed, consider a direction  $(\cos \theta, \sin \theta)$ , and examine the limit

$$\lim_{t \rightarrow 0} \frac{1}{t} [f(t \cos \theta, t \sin \theta) - f(0, 0)] = \cos^3 \theta.$$

**Definition 2.5.** Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable. The gradient of  $f$  is defined as the map

$$\nabla f: \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad x \mapsto \left[ \frac{\partial f}{\partial x_i}(x) \right]_i.$$

*Remark.* The gradient at a point  $x \in \mathbb{R}^n$  is thought of as a vector. In contrast, the derivative is thought of as a linear transformation. Otherwise, we see that  $\nabla f(x) = [Df(x)]$ .

**Definition 2.6.** Let  $C^1(\mathbb{R}^n)$  be the set of real-valued differentiable functions on  $\mathbb{R}^n$ . Fix a point  $a \in \mathbb{R}^n$ , then fix a tangent vector  $v \in \mathbb{R}^n$ . Then, the map

$$\nabla_v: C^1(\mathbb{R}^n) \rightarrow \mathbb{R}, \quad f \mapsto Df(a)(v)$$

is a linear functional. The quantity  $\nabla_v f$  is called the directional derivative of  $f$  in the direction  $v$  at the point  $a$ .

*Remark.* We can represent  $\nabla_v$  as the operator

$$\nabla_v(\cdot) = D(\cdot)(a)(v) = \sum_i v_i \frac{\partial}{\partial x_i} \Big|_a = v \cdot \nabla(\cdot).$$

**Lemma 2.6.** The directional derivatives  $\nabla_v$  form a vector space called the tangent space, attached to the point  $a \in \mathbb{R}^n$ . This can be identified with the vector space  $\mathbb{R}^n$  by the natural map  $\nabla_v \mapsto v$ . The standard basis can be informally denoted by the vectors

$$\nabla_{e_1} \equiv \frac{\partial}{\partial x_1}, \dots, \nabla_{e_n} \equiv \frac{\partial}{\partial x_n}.$$

### 2.3.2 Differentiation on manifolds \*

**Definition 2.7.** A homeomorphism is a continuous, bijective map whose inverse is also continuous.

**Lemma 2.7.** Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be continuous. Denote the graph of  $f$  as

$$\Gamma(f) = \{(x, f(x)) : x \in \mathbb{R}^n\}.$$

Then,  $\Gamma(f)$  is a smooth manifold.

*Proof.* Consider the homeomorphism

$$\varphi: \Gamma(f) \rightarrow \mathbb{R}^n, \quad (x, f(x)) \mapsto x.$$

This is clearly bijective, continuous (restriction of a projection map), with a continuous inverse (from the continuity of  $f$ ). Call this homeomorphism  $\varphi$  a coordinate map on  $\Gamma(f)$ .  $\square$

**Definition 2.8.** Let  $f: M \rightarrow \mathbb{R}$  where  $M$  is a smooth manifold, with a coordinate map  $\varphi: M \rightarrow \mathbb{R}^n$ . We say that  $f$  is differentiable at a point  $a \in M$  if  $f \circ \varphi^{-1}: \mathbb{R}^n \rightarrow \mathbb{R}$  is differentiable at  $\varphi(a)$ .

**Definition 2.9.** Let  $f: M \rightarrow \mathbb{R}$  where  $M$  is a smooth manifold, let  $\varphi: M \rightarrow \mathbb{R}^n$  be a coordinate map, and let  $a \in M$ . Let  $\gamma: \mathbb{R} \rightarrow M$  be a curve such that  $\gamma(0) = a$ , and further let  $\gamma$  be differentiable in the sense that  $\varphi \circ \gamma: \mathbb{R} \rightarrow \mathbb{R}^n$  is differentiable. The directional derivative of  $f$  at  $a$  along  $\gamma$  is defined as

$$\left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0} = \lim_{h \rightarrow 0} \left. \frac{f(\gamma(t+h)) - f(\gamma(t))}{h} \right|_{t=0}.$$

Note that we are taking the derivative of  $f \circ \gamma: \mathbb{R} \rightarrow \mathbb{R}$  in the conventional sense.

**Lemma 2.8.** Let  $\gamma_1$  and  $\gamma_2$  be two curves in  $M$  such that  $\gamma_1(0) = \gamma_2(0) = a$ , and

$$\left. \frac{d}{dt} \varphi \circ \gamma_1(t) \right|_{t=0} = \left. \frac{d}{dt} \varphi \circ \gamma_2(t) \right|_{t=0}.$$

In other words,  $\gamma_1$  and  $\gamma_2$  pass through the same point  $a$  at  $t = 0$ , and have the same velocities there. Then, the directional derivatives of  $f$  at  $a$  along  $\gamma_1$  and  $\gamma_2$  are the same.

**Definition 2.10.** Let  $M$  be a smooth manifold, and let  $a \in M$ . Consider the following equivalence relation on the set of all curves  $\gamma$  in  $M$  such that  $\gamma(0) = a$ .

$$\gamma_1 \sim \gamma_2 \iff \left. \frac{d}{dt} \varphi \circ \gamma_1(t) \right|_{t=0} = \left. \frac{d}{dt} \varphi \circ \gamma_2(t) \right|_{t=0}.$$

Each resultant equivalence class of curves is called a tangent vector at  $a \in M$ . Note that all these curves in a particular equivalence class pass through  $a$  with the same velocity vector.

The collection of all such tangent vectors, i.e. the space of all curves through  $a$  modulo the equivalence relation which identifies curves with the same velocity vector through  $a$ , is called the tangent space to  $M$  at  $a$ , denoted  $T_a M$ .

*Remark.* Each tangent vector  $v \in T_a M$  acts on a differentiable function  $f: M \rightarrow \mathbb{R}$  yielding a (well-defined) directional derivative at  $a$ .

$$v: C^1(M) \rightarrow \mathbb{R}, \quad f \mapsto \left. \frac{d}{dt} f(\gamma_v(t)) \right|_{t=0}.$$

Thus, the tangent space represents all the directions in which taking a derivative of  $f$  makes sense.

*Remark.* The tangent space  $T_a M$  is a vector space. Upon fixing  $f$ , the map  $Df(a): T_a M \rightarrow \mathbb{R}$ ,  $v \mapsto vf(a)$  is a linear functional on the tangent space.

*Remark.* Given a tangent vector  $v \in T_a M$ , it can be identified with its corresponding velocity vector in  $\mathbb{R}^n$ . Thus, the tangent space  $T_a M$  can be identified with the geometric tangent plane drawn to the manifold  $M$  at the point  $a$ .

## 2.4 Mean value theorem

Consider a differentiable function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , and fix  $a \in \mathbb{R}^n$ . Define the functions

$$g_i: \mathbb{R} \rightarrow \mathbb{R}, \quad g_i(x) = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n).$$



Then, each  $g_i$  is differentiable, with

$$g'_i(x) = \frac{\partial f}{\partial x_i}(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n).$$

By applying the Mean Value Theorem on some interval  $[c, d]$ , we can find  $\alpha \in (c, d)$  such that  $g_i(d) - g_i(c) = g'_i(\alpha)(d - c)$ . In other words,

$$f(\dots, d, \dots) - f(\dots, c, \dots) = \frac{\partial f}{\partial x_i}(\dots, \alpha, \dots)(d - c).$$

**Theorem 2.9.** *Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $a \in \mathbb{R}^n$ . Then,  $f$  is differentiable at  $a$  if all the partial derivatives  $\partial f / \partial x_j$  exist in a neighbourhood of  $a$  and are continuous at  $a$ .*

*Proof.* Without loss of generality, let  $m = 1$ . We claim that

$$\lim_{h \rightarrow 0} \frac{1}{\|h\|} \|f(a+h) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) h_i\| = 0.$$

Examine

$$\begin{aligned} f(a+h) - f(a) &= f(a_1 + h_1, \dots, a_n + h_n) - f(a_1, \dots, a_n) \\ &= f(a_1 + h_1, \dots, a_n + h_n) - f(a_1 + h_1, \dots, a_{n-1} + h_{n-1}, a_n) + \\ &\quad f(a_1 + h_1, \dots, a_{n-1} + h_{n-1}, a_n) - f(a_1 + h_1, \dots, a_{n-1}, a_n) + \\ &\quad \vdots \\ &\quad f(a_1 + h_1, a_2, \dots, a_n) - f(a_1, \dots, a_n) \\ &= \frac{\partial f}{\partial x_n}(c_n) h_n + \dots + \frac{\partial f}{\partial x_1}(c_1) h_1. \end{aligned}$$

The last step follows from the Mean Value Theorem. As  $h \rightarrow 0$ , each  $c_i \rightarrow a$ . Thus,

$$\begin{aligned} \frac{1}{\|h\|} \|f(a+h) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) h_i\| &= \frac{1}{\|h\|} \left\| \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right) h_i \right\| \\ &\leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right| \frac{|h_i|}{\|h\|} \\ &\leq \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i}(c_i) - \frac{\partial f}{\partial x_i}(a) \right|. \end{aligned}$$

Taking the limit  $h \rightarrow 0$ , observe that  $\partial f / \partial x_i(c_i) \rightarrow \partial f / \partial x_i(a)$  by the continuity of the partial derivatives, completing the proof.  $\square$

**Corollary 2.9.1.** *All polynomial functions on  $\mathbb{R}^n$  are differentiable.*

**Theorem 2.10.** *Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable with continuous partial derivatives, and let  $a \in \mathbb{R}^n$  be a point of local maximum. Then,  $Df(a) = 0$ .*

*Proof.* We need only show that each

$$\frac{\partial f}{\partial x_i}(a) = 0.$$

This must be true, since  $a$  is also a local maximum of each of the restrictions  $g_i$  as defined earlier.  $\square$

## 2.5 Inverse and implicit function theorems

**Theorem 2.11** (Inverse function theorem). *Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  be continuously differentiable on a neighbourhood of  $a \in \mathbb{R}^n$ , and let  $\det(Df(a)) \neq 0$ . Then, there exist neighbourhoods  $U$  of  $a$  and  $W$  of  $f(a)$  such that the restriction  $f: U \rightarrow W$  is invertible. Furthermore,  $f^{-1}$  is continuous on  $U$  and differentiable on  $U$ .*

**Lemma 2.12.** *Consider a continuously differentiable function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , and let  $M$  denote the surface defined by the zero set of  $f$ . Then,  $M$  can be represented as the graph of a differentiable function  $h: \mathbb{R}^{n-1} \rightarrow \mathbb{R}$  at those points where  $Df \neq 0$ .*

*Proof.* Without loss of generality, suppose that  $\partial f / \partial x_n \neq 0$  at some point  $a \in M$ . It can be shown that the map

$$F: \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad x \mapsto (x_1, x_2, \dots, x_{n-1}, f(x))$$

is invertible in a neighbourhood  $W$  of  $a$ , with a continuous and differentiable inverse of the form

$$G: \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad u \mapsto (u_1, u_2, \dots, u_{n-1}, g(u)).$$

Since  $F \circ G$  must be the identity map on  $W$ , we demand

$$(x_1, x_2, \dots, x_{n-1}, f(x_1, x_2, \dots, x_{n-1}, g(x))) = (x_1, x_2, \dots, x_{n-1}, x_n).$$

Thus, the zero set of  $f$  in this neighbourhood of  $a$  satisfies  $x_n = 0$ , hence

$$f(x_1, x_2, \dots, x_{n-1}, g(x_1, x_2, \dots, x_{n-1}, 0)) = 0.$$

In other words, the part of the surface  $M$  in the neighbourhood of  $a$  is precisely the set of points

$$(x_1, x_2, \dots, x_{n-1}, g(x_1, x_2, \dots, x_{n-1}, 0)).$$

Simply set

$$h: \mathbb{R}^{n-1} \rightarrow \mathbb{R}, \quad x \mapsto g(x_1, x_2, \dots, x_{n-1}, 0),$$

whence the surface  $M$  is locally represented by the graph of  $h$ . □

*Remark.* Note that by using

$$f(x_1, \dots, x_{n-1}, h(x_1, \dots, x_{n-1})) = 0$$

on the surface, we can use the chain rule to conclude that for all  $1 \leq i < n$ , we have

$$\frac{\partial f}{\partial x_i}(a) + \frac{\partial f}{\partial x_n}(a) \frac{\partial h}{\partial x_i}(a_1, \dots, a_{n-1}) = 0.$$

**Theorem 2.13** (Implicit function theorem). *Let  $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  be continuously differentiable in an open set containing  $(a, b)$ , with  $f(a, b) = 0$ . Let  $\det(\partial f^j / \partial x_{n+k}(a, b)) \neq 0$ . Then, there exists an open set  $U \subset \mathbb{R}^n$  containing  $a$ , an open set  $V \subset \mathbb{R}^m$  containing  $b$ , and a differentiable function  $g: U \rightarrow V$  such that  $f(x, g(x)) = 0$ .*

*Remark.* The condition on the determinant can be rephrased as  $\text{rank } Df(a, b) = m$ .

**Theorem 2.14.** *Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be continuously differentiable, and let  $M$  be the surface defined by its zero set. Furthermore, let  $\nabla f(a) \neq 0$  for some  $a \in M$ ; thus,  $M$  can be locally represented by a graph on  $\mathbb{R}^{n-1}$ . Then,  $\nabla f(a)$  is normal to the tangent vectors drawn at  $a$  to  $M$ ; in fact, the perpendicular space of  $\nabla f(a)$  is precisely the tangent space  $T_a M$ .*

*Proof.* Consider a tangent vector drawn at  $a$  to  $M$ , represented by the differentiable curve  $\gamma: \mathbb{R} \rightarrow M$ ,  $\gamma(0) = a$ ; note that we use the identification  $\gamma'(0) = v \in \mathbb{R}^n$ . Then, calculate

$$\left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0} = Df(\gamma(0))(\gamma'(0)) = Df(a)(v).$$

On the other hand, we have  $f(\gamma(t)) = 0$  identically. Thus,

$$v \cdot \nabla f(a) = Df(a)(v) = 0. \quad \square$$

## 2.6 Taylor's theorem

**Theorem 2.15.** *Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  have continuous second order partial derivatives. Then,*

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

**Theorem 2.16.** *Let  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  have continuous second order partial derivatives, and let  $(x_0, y_0) \in \mathbb{R}^2$ . Then, there exists  $\epsilon > 0$  such that for all  $\|(x - x_0, y - y_0)\| < \epsilon$ ,*

$$\begin{aligned} f(x, y) &= f(x_0, y_0) + \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0) \\ &\quad + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (x - x_0)^2 + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} (y - y_0)^2 \\ &\quad + \frac{\partial^2 f}{\partial x \partial y} (x - x_0)(y - y_0) + R(x, y), \end{aligned}$$

where as  $(x, y) \rightarrow (x_0, y_0) \rightarrow 0$ , the remainder term vanishes as

$$\frac{|R(x, y)|}{\|(x - x_0, y - y_0)\|^2} \rightarrow 0.$$

All partial derivatives here are evaluated at  $(x_0, y_0)$ .

*Proof.* This follows from applying the Taylor's Theorem in one variable to the real function  $g: \mathbb{R} \rightarrow \mathbb{R}$ ,  $t \mapsto f((1 - t)(x_0, y_0) + t(x, y))$ .  $\square$

## 2.7 Critical points and extrema

**Definition 2.11.** We say that  $a \in \mathbb{R}^n$  is a critical point of  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  if all  $\partial f / \partial x^j = 0$  there.

**Lemma 2.17.** *All points of extrema of a differentiable function are critical points.*

*Proof.* We already know that  $Df(a) = 0$  where  $a$  is either a point of maximum or minimum.  $\square$

*Example.* In order to find a point of extrema of a  $C^2$ -smooth function  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ , we first identify a critical point  $(x_0, y_0)$ . Next, we must find a neighbourhood of  $(x_0, y_0)$  which contains no other critical points – to do this, apply Taylor’s Theorem. Indeed, we see that

$$f(x, y) = f(x_0, y_0) + A(x - x_0)^2 + 2B(x - x_0)(y - y_0) + C(y - y_0)^2 + R_2.$$

For non-degeneracy of solutions, we demand  $AC - B^2 \neq 0$ , i.e. at  $(x_0, y_0)$ , we want

$$\left[ \frac{\partial^2 f}{\partial x \partial y} \right]^2 \neq \frac{\partial f}{\partial x} \frac{\partial f}{\partial y}.$$

If  $AC - B^2 > 0$  and  $\partial^2 f / \partial x^2 > 0$ , then we have found a point of minima; if  $\partial^2 f / \partial x^2 < 0$ , then we have found a point of maximum. If  $AC - B^2 < 0$ , then we have found a saddle point.

*Example.* Suppose that we wish to maximize the function  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ , given an equation of constraint  $g = 0$ , where  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ . Using the method of Lagrange multipliers, we look for solutions of the system

$$\begin{cases} \nabla f(x, y) + \lambda \nabla g(x, y) = 0, \\ g(x, y) = 0. \end{cases}$$

## 3 Integral calculus

### 3.1 Path integrals

**Definition 3.1.** A closed curve  $\gamma: [a, b] \rightarrow \mathbb{R}^n$  is closed if  $\gamma(a) = \gamma(b)$ . It is called simple if it has no self intersections.

**Definition 3.2.** Let  $p, q: U \rightarrow \mathbb{R}$  be continuous, where  $U \subseteq \mathbb{R}^2$  is an open set, and let  $\gamma: [a, b] \rightarrow U$  be piecewise smooth, i.e. smooth on  $(a, b)$  at all but finitely many points. Then, we define

$$\int_{\gamma} p \, dx + q \, dy = \int_a^b p(\gamma(t)) \gamma'_1(t) + q(\gamma(t)) \gamma'_2(t) \, dt.$$

**Lemma 3.1.** Let  $\gamma: [a, b] \rightarrow \mathbb{R}^2$  be a smooth curve, and let  $\varphi: [c, d] \rightarrow [a, b]$  be smooth, such that  $\varphi(c) = a$  and  $\varphi(d) = b$ . Then, the composition  $\gamma \circ \varphi: [c, d] \rightarrow \mathbb{R}^2$  is a smooth curve, and

$$\int_{\gamma \circ \varphi} p \, dx + q \, dy = \int_c^d [p(\gamma \circ \varphi(s)) \gamma'_1(\varphi(s)) + q(\gamma \circ \varphi(s)) \gamma'_2(\varphi(s))] \varphi'(s) \, ds.$$

By substituting the parameter  $\varphi(s) = t$ ,  $\varphi'(s) \, ds = dt$ , we retrieve

$$\int_{\gamma \circ \varphi} p \, dx + q \, dy = \int_a^b p(\gamma(t)) \gamma'_1(t) + q(\gamma(t)) \gamma'_2(t) \, dt = \int_{\gamma} p \, dx + q \, dy.$$

**Theorem 3.2.** Let  $p, q: U \rightarrow \mathbb{R}$  be continuous, and let  $\gamma: [a, b] \rightarrow U$  be a smooth curve. The integral

$$\int_{\gamma} p \, dx + q \, dy$$

depends only on the endpoints of  $\gamma$  if and only if there exists  $u: U \rightarrow \mathbb{R}$  such that

$$p = \frac{\partial u}{\partial x}, \quad q = \frac{\partial u}{\partial y}.$$

In other words, we demand that the vector field  $(p, q)$  be the gradient of  $u$ .

*Proof.* First suppose that there exists  $u$  such that  $\nabla u = (p, q)$ . Then,

$$\int_{\gamma} p \, dx + q \, dy = \int_a^b \frac{\partial u}{\partial x}(\gamma(t)) \gamma'_1(t) + \frac{\partial u}{\partial y}(\gamma(t)) \gamma'_2(t) \, dt.$$

The chain rule shows that this is simply

$$\int_a^b \frac{d}{dt} u(\gamma(t)) \, dt = u(\gamma(b)) - u(\gamma(a)).$$

Conversely, suppose that the given integral depends only on the endpoints of  $\gamma$ . Given two points  $\alpha, \beta \in U$ , we construct a path from  $\alpha$  to  $\beta$  by travelling only along the axes. Pick  $(x, y) \in U$ , and define  $u: U \rightarrow \mathbb{R}$ ,

$$u(x, y) = \int_{\gamma} p \, dx + q \, dy,$$

where  $\gamma$  is such a polygonal path from a fixed point  $\alpha$  to  $(x, y)$ . Note that  $u$  is well-defined by the independence of choice of path  $\gamma$ .  $\square$

*Example.* Let  $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be continuous, and let  $\gamma$  be a smooth curve in  $\mathbb{R}^2$ . We may denote

$$\int_{\gamma} \cdot ds = \int_{\gamma} f^1 \, dx + f^2 \, dy.$$

### 3.2 Multiple integrals

**Definition 3.3.** Let  $f: [a_1, b_1] \times [a_2, b_2] \rightarrow \mathbb{R}$  be continuous. Now, let  $P$  be a partition of the rectangular domain into  $n \times n$  sub-rectangles, and define

$$M_{ij} = \sup_{[x_i, x_{i+1}] \times [y_j, y_{j+1}]} f(x, y), \quad m_{ij} = \inf_{[x_i, x_{i+1}] \times [y_j, y_{j+1}]} f(x, y).$$

We also define,

$$U(f, P) = \sum_{i,j} M_{ij}(x_i - x_{i-1})(y_j - y_{j-1}), \quad L(f, P) = \sum_{i,j} m_{ij}(x_i - x_{i-1})(y_j - y_{j-1}).$$

Finally define the upper and lower sums

$$U(f) = \sup_P U(f, P), \quad L(f) = \sup_P L(f, P).$$

Then,  $f$  is Riemann integrable if  $U(f) = L(f)$ , and this integral is denoted by

$$\int_{[a_1, b_1] \times [a_2, b_2]} f$$

*Remark.* This definition naturally extends to integrals over any  $k$ -cell.

**Definition 3.4.** A measure zero set  $E \subset \mathbb{R}^n$  is such that given any  $\epsilon > 0$ , there exists a countable collection of rectangles  $\{A_j\}$  such that their union contains  $E$ , and the sum of their volumes is less than  $\epsilon$ .

*Example.* Any countable subset of  $\mathbb{R}^n$  has measure zero.

*Example.* Any line in  $\mathbb{R}^2$ , plane in  $\mathbb{R}^3$ , etc. has measure zero.

**Lemma 3.3.** The countable union of measure zero sets has measure zero.

**Theorem 3.4.** A bounded function  $f: A \rightarrow \mathbb{R}$  where  $A \subset \mathbb{R}^n$  is a rectangle is integrable if and only if for every  $\epsilon > 0$ , there exists a partition  $P$  of  $A$  such that  $U(f, P) - L(f, P) < \epsilon$ .

**Theorem 3.5.** Let  $f: A \rightarrow \mathbb{R}$  be bounded, where  $A \subset \mathbb{R}^n$  is a rectangle. Then,  $f$  is Riemann integrable if and only if its set of discontinuities has measure zero.

**Theorem 3.6.** Let  $f: A_1 \times A_2 \rightarrow \mathbb{R}$  be continuous, where  $A_1 \subset \mathbb{R}^n$  and  $A_2 \subset \mathbb{R}^m$  are closed rectangles. Then, we can write

$$\int_{A_1 \times A_2} f = \int_{A_2} \left( \int_{A_1} f(x, y) dx \right) dy = \int_{A_1} \left( \int_{A_2} f(x, y) dy \right) dx.$$

**Theorem 3.7** (Green's theorem). Let  $\gamma$  be a smooth simple closed curve in  $\mathbb{R}^2$  oriented counter-clockwise, and let  $\Omega$  be the region enclosed by  $\gamma$ . If  $p, q: \Omega \rightarrow \mathbb{R}$  have continuous partial derivatives, then

$$\int_{\gamma} p dx + q dy = \iint_{\Omega} \left( \frac{\partial q}{\partial x} - \frac{\partial p}{\partial y} \right) dx dy.$$

*Example.* Let  $\gamma: [0, 2\pi] \rightarrow \mathbb{R}^2$ ,  $\gamma(t) = (\cos t, \sin t)$ . Then,  $\gamma$  enclosed the unit disc in  $\mathbb{R}^2$ . To calculate its area, we can set  $p = 0$ ,  $q = x$ , giving

$$\iint_{\Omega} dx dy = \int_0^{2\pi} \cos^2 t dt = \pi.$$

Another option is to set  $p = -y/2$ ,  $q = x/2$ , giving

$$\iint_{\Omega} dx dy = \frac{1}{2} \int_0^{2\pi} \cos^2 t + \sin^2 t dt = \frac{1}{2} \int_0^{2\pi} dt = \pi.$$

**Theorem 3.8** (Change of variables). Let  $U \subset \mathbb{R}^n$  be an open set, and get  $g: U \rightarrow \mathbb{R}^n$  be a bijective, continuously differentiable map such that  $\det Dg(x) \neq 0$  on  $U$ . If  $f: g(U) \rightarrow \mathbb{R}$  is integrable, then

$$\int_{g(U)} f = \int_U f \circ g |\det Dg|$$