MULTIVARIABLE ANALYSIS

MARKUSS G. ĶĒNIŅŠ

Contents

University of Groningen
E-mail address: `m.g.kenins@student.rug.nl`.
Date: November 2025.

Dear reader,

In these notes we reproduce the theory of multivariable analysis based on the approach by Prof. Roland van der Veen (1) – the same approach as adopted in Spivak (2) –, however, in a more reader-friendly manner – the focus is on arriving at the definitions through simple examples from calculus 2, and to introduce all algebraic structures in the most natural way possible.

Tensors were reproduced with heavy modification from Postnikov, and, in a lesser sense, from Kiselev's notes, and serve as a first introduction.

NB! A core course consists of all §§ except §§ 10 to 13 (tensors) and appendices.

The 'goal' of multivariable analysis is to generalise the fundamental theorem of calculus to its multivariable version – Stokes' theorem and Poincaré lemma. Namely: (1) the integral of a derivative is the integral of the primitive over the boundary, (2) a primitive exists if and only if the derivative is zero. But (!) *what do we mean* by 'the' derivative...? We will see that it is dual to the boundary, making Stokes' theorem trivially true...

As an application of Stokes' powerful theorem, we give a very elegant proof of one of the core theorems of complex analysis (cf. Appendix A) – Cauchy's integral theorem: 'integral of a complex-differentiable function over a closed loop is zero'.

I hope you enjoy the notes! Please email all corrections to the email given above.

– MGĶ.

* * *

The course should prepare the student to begin reading differential geometry; a fantastic first reference is Postnikov's Lectures on Geometry, the second-half of Vol. II or the full Vol. III[1] as a first introduction. Parts of Vol. I introduce projective geometry (cf. contents of lectures). Further volumes (IV parts 1 and 2 and V) cover more advanced topics, including elements of algebraic topology. All volumes are available for free at `https://archive.org/search?query=creator%3A%22M.+Postnikov%22`.

A highly condensed text on differential geometry is (4).

* * *

Bibliography.

(1) R. van der Veen, Multivariable Analysis, 2021 [**NB!** *primary reference*];
(2) M. Spivak, Calculus on Manifolds [*reproduces a very similar theory to (1)*];
(3) M. Postnikov, Lectures on Geometry, Vol. II [*for tensors – §§ 4–9*];
(4) A.V. Kiselev, Tensor Calculus on Manifolds [*advanced course on differential geometry*].

[1] The beginning of Vol. III reproduces the end of Vol. II.

Part 1. Linear algebra

1. Definition and maps

We begin by recapitulating the basics of linear algebra. The reader is familiar with the canonical space $\mathbb{R}^n = \{(x^1, x^2, ..., x^n) : x^1, ..., x^n \in \mathbb{R}\}$ of n-tuples of real numbers (**NB!** we will use superscripts for the coefficients) with the usual addition and scalar multiplication (rescaling of vectors) rules. The typical generalisation of the definition is to list 8 important properties; we break them up via an algebraic approach. Now, for addition we merely need an Abelian group V under $+$, namely, there exists $0 \in V$ such that for all $u, v, w \in V$ we have

(1) $u + 0 = u$,
(2) $u + (v + w) = (u + v) + w$, and thus we write $u + v + w$,
(3) there exists $-u \in V$ such that $u + (-u) = 0$,
(4) $u + v = v + u$,

and scalar multiplication is an action from the real numbers into the group: for all $\lambda, \mu \in \mathbb{R}$ we have

(5) $1 \cdot u = u$,
(6) $(\lambda + \mu) \cdot v = \lambda v + \mu v$,
(7) $(\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v)$,
(8) $\lambda \cdot (u + v) = \lambda \cdot u + \lambda \cdot v$.

1.1. **Definition.** Let V be a set with operations $+ : V \times V \to V$ and $\cdot : \mathbb{R} \times V \to V$. If the operations satisfy the 8 properties above, we call V a real vector space.

We define maps to be functions that preserve both operations, similarly as in group theory.

1.2. **Definition.** Let V and W be real vector spaces. We call a map $L : V \to W$ linear if L is a group homomorphism and preserves scalar multiplication, namely, for all $u, v \in V$ and $\lambda \in \mathbb{R}$ we have

(1) $L(u + v) = L(u) + L(v)$,
(2) $L(\lambda v) = \lambda L(v)$.

We denote the set of all linear maps from V to W by $\mathrm{Hom}(V, W)$.

Note that the $\mathrm{Hom}(V, W)$ is itself a vector space under the operations $(L + L')(v) := L(v) + L'(v)$ and $(\lambda L)(v) = \lambda \cdot L(v)$.

The two conditions above may be combined into $L(u + \lambda v) = L(u) + \lambda L(v)$. We often omit brackets and write $L(u) = Lu$. Injectivity $L(u) = L(v)$ implies $u = v$ and subjectivity $L(V) = W$ may be defined as usual; just as in group theory, the maps define the kernel $\ker L = \{v \in V : Lv = 0\}$ and the image $L(V) = \{L(v) \in W : v \in V\}$ – subspaces! Indeed, injectivity is equivalent to $\ker L = \{0\}$ (two-line proof).

Wonderfully interesting constructions emerge after considering how subspaces fit into the whole space. Let U be a subspace of V; firstly, note that if we add two vectors from the same subspace U, the sum will remain in U (in principle, by definition); what happens if we add a vector not in U? Let $U \neq V$ be a proper subspace; if we add a vector $v \in V \backslash U$ to any vector in U, we will obtain a vector outside U; if we consider the set of vectors we obtain from this process, it turns out that this construction allows us to define a vector space.

1.3. **Definition.** Let U be a subspace of V. We define the quotient space $V/U = \{v + U : v \in V\}$, where $v + U = \{v + u : u \in U\}$, with addition $(v + U) + (w + U) = (v + w) + U$ and multiplication $\lambda(v + U) = (\lambda v) + U$.

The following fact follows immediately from rewriting the definition.

1.4. **Lemma.** *Let $V = U \oplus U'$ be a direct sum of two subspaces: $V = \{u + u' : u \in U, u' \in U'\}$. Then $U' \cong V/U$.*

This gives a way to *completely* characterise the way linear maps act. Let $\pi_U : V \to V/U$ be the map that sends $v \mapsto v + U$. This map is clearly surjective. We call a linear bijection an *isomorphism* of spaces.

1.5. **Theorem** (Canonical decomposition)**.** *Let $L : V \to W$ be a linear map. Then there exists a* unique *isomorphism (linear bijective map) $\tilde{L}$ such that the diagram*

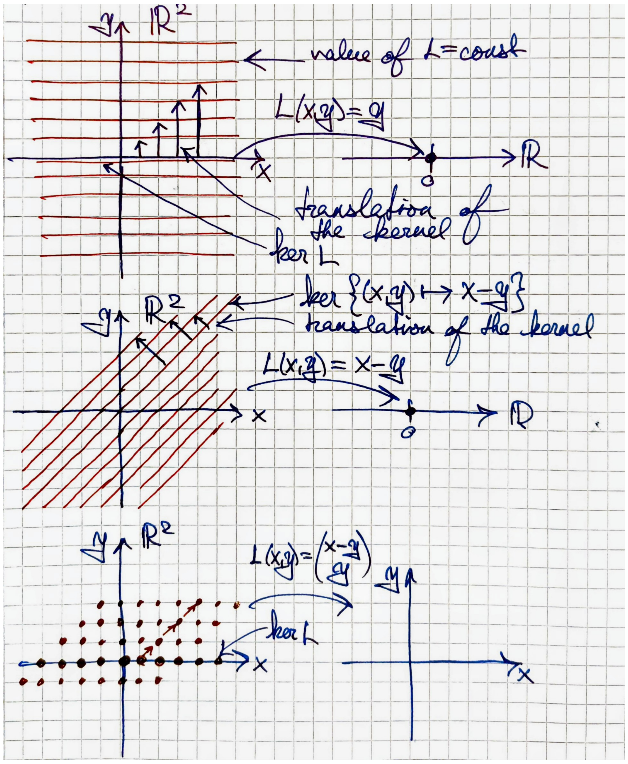
$$\begin{tikzcd} V \arrow[r, "\pi"] \arrow[rrr, bend left, "L"] & V/\ker L \arrow[r, "\tilde{L}"] & L(V) \arrow[r, hook] & W \end{tikzcd}$$

where $\hookrightarrow$ denotes set inclusion, commutes.

Proof. We require $L(v) = (\tilde{L} \circ \pi)(v) = \tilde{L}(v + \ker L)$, thus $\tilde{L}$ is uniquely determined. By defining $\tilde{L}(v + \ker L) = L(v)$, we see that $\tilde{L}$ is surjective; we compute $\ker \tilde{L} = \{v + \ker L : \tilde{L}(v + \ker L) = L(v) = 0\}$, meaning that $v \in \ker L$, hence $\ker \tilde{L} = \{\ker L\}$, which is the 'zero' (identity) in $V/\ker L$, and thus $\tilde{L}$ is injective. □

Namely, any linear map sends the kernel to zero, and all non-zero values depend on their relation to the kernel.

1.6. **Example.** The following examples illustrate this principle.



What makes linear maps so special is the fact that the map is completely determined by the way it acts on a specific set of vectors; indeed, this (ordered) set must generate the entire vector space – be a basis. **NB!** The basis is an *ordered* set!

1.7. **Definition.** Vectors $v_1, ..., v_n \in V$ are independent if no vector is a linear combination of the others; equivalently, if $a^1 v_1 + \cdots + a^n v_n = 0$ implies $a^1 = ... = a^n = 0$. The span of the vectors is the set of all linear combinations made by them

$\mathrm{Span}\{v_1, ..., v_n\} = \{a^1 v_1 + ... + a^n v_n \in V : a^1, ..., a^n \in \mathbb{R}\}$; the vectors span the space if $\mathrm{Span}\{v_1, ..., v_n\} = V$.

The *ordered set* $\mathbf{b} = (v_1, ..., v_n)$ is a basis for V if the vectors are independent and span V. The number of vectors for which this is true is unique (cf. LA2), and is called the dimension $\dim V = n$.

Intuitively, the independence requirement restricts the maximal number of vectors in a basis, whereas the span requirement guarantees that we have enough vectors. Note that n may not be finite, however, in this course (virtually) all spaces will be finitely-dimensional.

Indeed, the basis induces an isomorphism $\mathbf{b} : \mathbb{R}^n \to V$ (where $n = \dim V$) given by $\mathbf{b}(e_i) = v_i$, where $\mathbf{e} = (e_1, ..., e_n)$ is the standard basis $e_1 = (1, 0, ...), e_2 = (0, 1, 0, ...), ..., e_n = (0, ..., 0, 1)$. Since we know that in $\mathbb{R}^n$ any vector may be written uniquely in terms of the standard basis, the isomorphism guarantees that this is true in V.

Any linear map must respect the dimensions: intuitively, for every (independent) vector mapped to 0 we lose an independent image vector; this way, a linear map $L : V \to W$ splits V into the kernel and non-kernel subspaces, where the non-kernel is mapped injectively. This is formalised by the dimension theorem.

1.8. **Theorem** (Dimension). *For any linear $L : V \to W$, where $\dim V < \infty$ (but $\dim W$ may be ∞), we have*

$$\dim V = \dim \ker L + \dim L(V). \tag{1.1}$$

Proof. By Theorem 1.5, $V/\ker L \cong L(V)$, therefore $\dim L(V) = \dim V/\ker L = \dim V - \dim \ker L$. □

Note that without Theorem 1.5, the standard proof requires a lemma for basis extension, and is quite long and inelegant.

We provide the most general example of a vector space below, which we will often meet later in the course.

1.9. **Definition.** Let S be a finite set. We define the free vector space $\mathbb{R}^{(S)}$ by

$$\mathbb{R}^{(S)} = \left\{ \sum_{s \in S} c^s s : c^s \in \mathbb{R} \right\}, \tag{1.2}$$

where sums and scalar multiplication are defined term-wise for each distinct s.

2. Matrices and change of basis

It is well-known that multiplying a column vector by a matrix yields a linear map $L_A : \mathbb{R}^n \to \mathbb{R}^m$, namely, $L_A(v) = Av$, where $A \in \mathbb{R}^{m \times n}$, $A = (A^i_j)^{i=1,...,m}_{j=1,...,n}$ (notation: row index above, column index below). This correspondence goes both ways: every linear map is associated with a matrix given by $L(e^V_j) = A^1_j e_1 + A^2_j e_2 + \cdots + A^n_j e_n$, where $e^V_j \in V$ and $e_i \in W$; it is very important to notice that $e_1, ..., e_m$ are independent, thus the numbers A^i_j are *uniquely* determined. Rewriting in column form reveals that

$$L(e^V_j) = A^1_j e_1 + A^2_j e_2 + \cdots + A^m_j e_m = \begin{pmatrix} A^1_j \\ \vdots \\ A^m_j \end{pmatrix},$$

namely, *the columns of A represent the images of base vectors.*

We may extend the matrix description to arbitrary spaces by describing how linear maps act on basis vectors – *to use a matrix, the bases must be fixed!* Indeed, let $v_1, ..., v_n$ be a basis of V and $w_1, ..., w_m$ of W, inducing isomorphisms $\mathbf{b} : \mathbb{R}^n \to$

$V, \mathbf{b}(e_i) = v_i$ and $\mathbf{c} : \mathbb{R}^m \to W, \mathbf{c}(e_i) = w_i$, and let $L : V \to W$ be a linear map. The following commutative diagram

$$\begin{tikzcd} V \arrow[r, "L"] & W \\ \mathbb{R}^n \arrow[u, "\mathbf{b}"] \arrow[r, "\mathbf{c}^{-1}L\mathbf{b}"] & \mathbb{R}^m \arrow[u, "\mathbf{c}"'] \end{tikzcd}$$

yields the associated matrix $[L]^i_j = (\mathbf{c}^{-1} \circ L \circ \mathbf{b})^i_j$; explicitly, the matrix with columns given by

$$\begin{aligned} [L] &= [(\mathbf{c}^{-1}L\mathbf{b})(e_1), ..., (\mathbf{c}^{-1}L\mathbf{b})(e_n)] \\ &= [(\mathbf{c}^{-1}L)(v_1), ..., (\mathbf{c}^{-1}L(v_n)] \\ &= [\mathbf{c}^{-1}(L(v_1)), ..., \mathbf{c}^{-1}(L(v_n))], \end{aligned} \tag{2.1}$$

where $\mathbf{c}^{-1}(w) = c^1 w_1 + \cdots + c^m w_m$ gives the coefficients of a vector in the $w_1, ..., w_n$ basis of W. This proves the following.

2.1. **Lemma.** $\dim \mathrm{Hom}(V, W) = \dim V \dim W$.

3. Dual space

Up until now, the notes have focused primarily on recapitulation; already, or in a moment, this will no longer be the case.

The concept of a dual space is one of the most important concepts in linear algebra; indeed, the definition of tensors arises through functionals; let us jump right into the definition.

3.1. **Definition.** The dual[2] space $V^* = \mathrm{Hom}(V, \mathbb{R})$ is the space of linear maps from the vector space into the real numbers; the elements of V^* are called (linear) functionals or covectors.[3]

Let $b_1, ..., b_n$ be a basis for V. Indeed, for $v = a^1 b_1 + \cdots + a^n b_n$ the functional must act linearly on each basis vector, meaning that any functional $\varphi \in V^*$ is of the form $\varphi(v) = \varphi_1 a^1 + \cdots + \varphi_n a^n$, where $\varphi_i = \varphi(b_i)$. Since n arbitrary numbers specify φ, it follows that the dimension of V^* is n. Indeed, setting $\beta^i \in V^*$ given by $\beta^i(b_j) = \delta^i_j$, yields $\varphi(v) = \varphi_1 \beta^1 + \cdots + \varphi_n \beta^n$, as

$$(\varphi_1 \beta^1 + \cdots + \varphi_n \beta^n)(v) = \varphi_1 \beta^1(v) + \cdots + \varphi_n \beta^n(v) = \varphi_1 a^1 + \cdots + \varphi^n a^n = \varphi(v).$$

The basis $\beta^1, ..., \beta^n$ of V^* (dual basis of V) defines the corresponding isomorphism $\text{ß} : \mathbb{R}^n \to V^*$ given by $\text{ß}(e_i) = \beta^i$. This shows that if $\dim V < \infty$, then $\dim V = \dim V^*$.

3.2. **Theorem.** $(V^*)^* \cong V$ *with a natural identification.*

Proof. Let $v \in V$ be arbitrary. Denote $V^* = W$, and let $\hat{v} \in W^*$ be a functional on $W = V^*$ defined by $\hat{v}(\varphi) = \varphi(v) \in \mathbb{R}$. We claim that $\Phi : V \to (V^*)^*$ is an isomorphism (linear bijection). Indeed, $\Phi(v) = \Phi(w)$ implies that $\varphi(v) = \hat{v}(\varphi) = \hat{w}(\varphi) = \varphi(w)$ for all φ; thus it is true for all $\varphi = \beta^i$, which yields $v^1 = w^1, ..., v^n = w^n$, hence $x = y$, showing that Φ is injective.

We compute

$$\begin{aligned} \Phi(v + \lambda w)(\varphi) = (\widehat{v + \lambda w})(\varphi) = \varphi(v + \lambda w) = \varphi(v) + \lambda \varphi(w) \\ = \Phi(v)(\varphi) + \lambda \Phi(w)(\varphi); \end{aligned} \tag{3.1}$$

hence Φ is linear.

---

[2] In some older resources called the conjugate space.

[3] Addition and scalar multiplication are defined as usual: $(\varphi + \varphi')(v) = \varphi(v) + \varphi'(v)$ and $(\lambda\varphi)(v) = \lambda \cdot \varphi(v)$ under multiplication of real numbers.

By the dimension theorem, $\dim \Phi(V) = \dim V - \dim\ker\Phi = \dim V = \dim(V^*)^*$, hence Φ is surjective, proving that it is an isomorphism. $\square$

3.3. *Remark.* Annulets allow to characterise subspaces not as spans, but by dual vectors. We will often discuss level sets of maps; these further formalise that notion. We will not use these in the notes; cf. Postnikov, § L4, Def. 4.

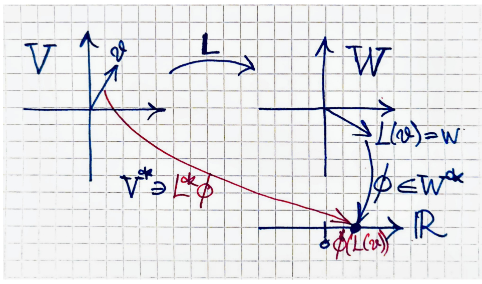
3.4. **Theorem.** *$V = P \oplus Q$ implies $V^* = \operatorname{Ann} P \oplus \operatorname{Ann} Q$ and $\operatorname{Ann} P \cong Q^*$, $\operatorname{Ann} Q \cong P^*$, where the annulet $\operatorname{Ann} S \subseteq V^*$ of $S \subseteq V$ is defined as*

$$\operatorname{Ann} S = \{\varphi \in V^* : \varphi(x) = 0 \ \forall x \in S\}. \tag{3.2}$$

4. Pull-backs of linear maps

A central theme throughout this topic will be that of pull-backs, which deserve its own section. A very common construction will be the following: given a structure on space W, we wish to define a similar structure at a point of V, such that the structures are related by some mapping $V \to W$.[4] We say that we *pull back the structure.*

The simplest example is the case, where the map $L : V \to W$ is linear, and the structure is a functional, as is illustrated below:



4.1. **Definition.** Let $L : V \to W$ be a linear map and $\phi \in W^*$ be any functional on W. Define the pull-back L^* of ϕ along L to be $L^*\phi = \phi \circ L$. Note that $L^* : W^* \to V^*$ is linear.

Indeed, the definition states that the following diagram *commutes*:

$$\begin{tikzcd} V \arrow[r, "L"] \arrow[dr, "L^*\phi"'] & W \arrow[d, "\phi"] \\ & \mathbb{R} \end{tikzcd}$$

where all arrows are linear maps. When we say that a diagram commutes we mean that traversing arrows from one set to another one way must yield the same result as by going any other way; in the diagram the sets V and $\mathbb{R}$ are the only ones connected by arrows in more than one way: either via $L^*\phi$ or via L followed by ϕ (i.e. composition $\phi \circ L$); commutativity means that these must be equal $L^*\phi = \phi \circ L$.

The reader will meet many other examples of pull-backs; these are in fact all the *same...* diagram. Namely, the sets will typically remain vector spaces ($\mathbb{R}$ is a vector space), but the nature of the arrows will be different – linear maps, differentiable functions, forms, etc.

One may attach more stuff around the 'triangle' in the diagram to indicate how some of the maps are constructed, however, they cannot influence the triangle.

4.2. *Remark* (What is Algebra, really?). During the second year of mathematics the reader will discover a new (correct? ☺) way to think about mathematics. So far, we have thought of linear maps as functions that satisfy this extra structure; of

[4] Cf. `https://en.wikipedia.org/wiki/Pullback_(differential_geometry)`.

homomorphisms as functions that respect group operations; etc. others – functions with fixed properties. And when we think of these maps, we think of sets and functions acting on sets; when these sets have structure, the functions may or may not satisfy the conditions.

Now, a better way is as follows: do not think of a structure (group, vector space, topological space, etc.) as 'sets' that have other additions, and functions simply being from the underlying sets to sets, but instead(!) fix structures and consider *only* the maps that preserve the structure.

For example, only group homomorphisms are called 'maps' between groups, only continuous functions between topological spaces, only linear maps between vector spaces. That is, we do not consider 'sets' and 'maps' separately – they are one; you cannot consider groups without group homorphisms, you cannot consider vector spaces without linear maps, and so on.

This idea is typically called *category theory* – a category consists of objects with maps. (Such that composition behaves well.) The maps are restricted – i.e. not just 'functions' –; indeed, the objects do not even have to be sets.

Examples: any sets and any functions – category of sets, **Set**; groups and group homomorphisms – category of groups **Grp**; category of vector spaces over a field[5] $\mathbb{K}$ with linear maps – $\mathbf{Vec}_{\mathbb{K}}$; smooth manifolds with p-times continuously differentiable maps – $\mathbf{Man}^p$; and so on.[6]

4.3. **Lemma.** *Let $U \xrightarrow{M} V \xrightarrow{L} W$ be linear maps. Then $(L \circ M)^* = M^* \circ L^*$ as a linear map $W^* \to U^*$.*

Proof. For any $\phi \in W^*$ we compute both sides separately
$$(L \circ M)^*\phi = \phi \circ L \circ M$$
and
$$(M^* \circ L^*)(\phi) = M^*(L^*(\phi)) = M^*(\phi \circ L) = \phi \circ L \circ M,$$
proving the statement. □

Therefore, pull-backs may be computed separately or from a single composition.

5. Metric and topological structure of vector spaces

We will wish to work with open neighbourhoods of vector spaces; as such we need a metric to define open balls. Recall the definition of a metric.

5.1. **Definition.** A metric $d : X \times X \to [0, \infty)$ on X is a function that satisfies for all $x, y, z \in X$:

(1) $d(x, y) = 0$ if and only if $x = y$ (positivity);
(2) $d(x, y) = d(y, x)$ (symmetry);
(3) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

We then call (X, d) a metric space.

A very special type of metric defined on vector spaces that preserves scalar multiplication is called a norm $d(x, y) = ||x - y||$, where $||\cdot|| : V \to V$, or $||x|| = d(x, 0)$, satisfies $||av|| = |a|\,||v||$, where $a \in [0, \infty)$ and $v \in V$. Note that

[5] A field (*Körper* – 'body') is an Abelian group with commutative multiplication, where every nonzero element has a multiplicative inverse (i.e. you can divide).

[6] Cf. probably one of the best algebra books 'Algebra: Chapter 0' by Paolo Aluffi – `https://agorism.dev/book/math/alg/algebra_chapter-0_paolo-aluffi.pdf` The book contains 9 chapters... Chapter 1 is a first (really, an introduction for absolute beginners, full of examples and motivation) introduction to category theory.

$||x+y|| = d(x,-y) \leq d(x,0)+d(0,-y) = ||x|| + ||-y|| = ||x|| + ||y||$. Similarly, the reverse triangle inequality holds:

$$| \, ||x|| - ||y|| \, | \leq ||x-y||. \tag{5.1}$$

We call norms $||\cdot||_1, ||\cdot||_2 : V \to [\infty)$ Lipshitz equivalent if there exist constants $c, C > 0$ such that $c||v||_2 \leq ||v||_1 \leq C||v||_2$ for all $v \in V$. Then the induced metrics $d_1(x,y) = ||x-y||_1$ and d_2 are Lipshitz equivalent:

$$d_1(x,y) = ||x-y||_1 \leq C||x-y||_2 = Cd_2(x,y),$$
$$cd_2(x,y) = c||x-y||_2 \leq ||x-y||_1 = d_1(x,y),$$

thus

$$cd_2(x,y) \leq d_1(x,y) \leq Cd_2(x,y). \tag{5.2}$$

Recall that Lipshitz equivalent metrics are topologically equivalent[7] – a subset $U \subseteq V$ is d_1-open if and only if it is d_2-open.

5.2. **Theorem.** *If* $\dim V < \infty$*, then all norms on* V *are Liphsitz equivalent.*

The proof is somewhat long and technical, and is skipped;[8] the crucial fact is to use the Heine–Borrel theorem – in $\mathbb{R}^n$ being closed and bounded is equivalent to being compact – for the unit $(n-1)$-sphere $\mathbb{S}^{n-1} \subsetneq \mathbb{R}^n$.

It can be easily shown that any norm and linear map are continuous. Similarly, that the image of the vectors with norm 1 is compact.

Since the set of linear maps $\mathrm{Hom}(V,W)$ from V to W is a vector space, we may define a norm on it as follows, called the *operator norm*

$$||L|| = \sup_{||v||=1} ||L(v)|| \tag{5.3}$$

for any linear $L : V \to W$. It is related to eigenvalues as is given by the lemma below. Recall that an eigenvalue and eigenvector pair (λ, v_λ) of $L : V \to V$ satisfies $L(v) = \lambda v$; the map L is diagonalisable if there exists an independent set of n eigenvectors $v_1, ..., v_n$ such that $L(v_i) = \lambda_i v_i$ (**NB!** no summation by Einstein notation.).

5.3. **Lemma.** *Let* $\dim V = n < \infty$ *and* $L : V \to V$ *be linear. Then the operator norm* $||L||^2 \leq \sum_{i,j=1}^n (L^i_j)^2$*, where* $L^i_j = \varepsilon^i L(e_j)$ *are basis elements with respect to basis* $e_1, ..., e_n$ *of* V*. If* L *is diagonalisable, then* $||L||$ *equals the largest eigenvalue of* L*.*

We recall continuous maps.

5.4. **Definition.** A map $f : V \to W$, with V and W considered as metric spaces with the metrics arising from norms, is *continuous* if the preimage $f^{-1}(A)$ of any open set $A \subset W$ is open in V.

We say f is continuous at $x \in V$ if for any $\epsilon > 0$ there exists a $\delta(x,\epsilon) > 0$ such that $f(B_\delta(x)) \subset B_\epsilon(f(x))$, where $B_r(x)$ is the open ball of radius $r > 0$ around x.[9]

Then continuity is equivalent to f being continuous at all $x \in V$. We say that $f : P \to W$ for some subset $P \subset V$ is continuous on P if f is continuous at all $x \in P$; equivalently, that the preimage of open is open in P.

[7] Cf. Sutherland, Prop. 6.34.

[8] It will be replicated in the functional analysis course.

[9] **NB!** Each of the two balls above is with respect to a different metric – d_V on V and d_W on W.

Part 2. Differentiation and integration

In this part of the notes we handle the definitions of the derivative (emphasis), and the integral, where we integrate multivariate functions; we will make no mention of 'integrating derivatives', as in the fundamental theorem of calculus – this belongs to the integration of forms.

6. The derivative

We now have all the tools to jump right into the definition of the derivative. Let V, W be real vector spaces with $\dim V, \dim W < \infty$, and let $f : P \to W$ be some continuous function defined on an open neighbourhood U of V. In essence, we will want to define the subspace of W that approximates $f(p)$ at some point $p \in P$ *to first order*; this subspace will be the image of some linear map. Equivalently, we look for the linear map $f'(p)$ that best approximates f at p to first order, namely:

$$f(p+v) = f(p) + f'(p)v \tag{6.1}$$

6.1. **Definition.** Let $||\cdot||_V$ and $||\cdot||_W$ be norms on V and W. Let $P \subset V$ be open. A function $f : P \to W$ is *differentiable* at $p \in P$ if there exists a *linear map* $L : V \to W$ that satisfies the following limit:

$$\lim_{v\to 0} \frac{f(p+v) - f(p) - L(v)}{||v||_V} = 0. \tag{6.2}$$

NB! Convergence is in W under the norm $||\cdot||_W$. We call $L = f'(p)$ the *derivative* $f'(p) \in \mathrm{Hom}(V, W)$ of f at p; the dependence upon the point p is crucial. If f is differentiable at all $p \in P$, we say f is differentiable on P; then $f' : P \to \mathrm{Hom}(V, W)$, $p \mapsto f'(p)$ is a well-defined function; if f' is continuous, we say that f is *continuously-differentiable.*

The emphasis is on the statement that $f'(p)$ *is a linear map.* Indeed, for $f : \mathbb{R} \to \mathbb{R}$ we have that $f'(p)$ as a number represents the linear map $x \mapsto f'(p)\cdot x$, namely, real number multiplication. We will see shortly that we can work with $f'(p)$ as a fully general linear map, or its matrix representation, whichever is more convenient (depends on the example!); this matrix representation will be of a form most familiar to the reader... can you guess what it is?

6.2. **Example.** Let $||\cdot||_2$ be the standard Euclidean norm on $\mathbb{R}^2$, namely, $||(x,y)||_2 = \sqrt{x^2+y^2}$. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x,y) = (x^2+y^2, xy)$. We check whether f is differentiable at p and, if so, find $f'(p)$ for any $p = (a,b) \in \mathbb{R}^2$. Recall that the derivative will be the linear map that *best-approximates the f to first order*; for that reason we may compute the difference $f(p+v) - f(p)$ and 'read off' the linear part (in v), if possible. Let $v = (h,k) \in \mathbb{R}^2$. Then

$$\begin{aligned} f(p+v) - f(p) &= f(a+h, b+k) - f(a,b) \\ &= \begin{pmatrix} (a+h)^2 + (b+k)^2 \\ (a+h)(b+k) \end{pmatrix} - \begin{pmatrix} a^2+b^2 \\ ab \end{pmatrix} \\ &= \begin{pmatrix} a^2 + 2ah + h^2 + b^2 + 2bk + k^2 - a^2 - b^2 \\ ab + ak + bh + hk - ab \end{pmatrix} \\ &= \begin{pmatrix} 2ah + 2bk \\ ak + bh \end{pmatrix} + \begin{pmatrix} h^2 + k^2 \\ hk \end{pmatrix} \\ &= \begin{pmatrix} 2a & 2b \\ b & a \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} + \begin{pmatrix} h^2+k^2 \\ hk \end{pmatrix}, \end{aligned}$$

which suggests that $f'(p)$ is represented by the 2×2 matrix above, which we call L. We indeed check this:

$$\begin{aligned}\lim_{v\to 0}\frac{f(p+v)-f(p)-L(v)}{||v||_2} &= \lim_{h,k\to 0}\frac{1}{\sqrt{h^2+k^2}}\begin{pmatrix}h^2+k^2\\ hk\end{pmatrix}\\ &= \lim_{h,k\to 0}\begin{pmatrix}\sqrt{h^2+k^2}\\ \frac{hk}{\sqrt{h^2+k^2}}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix},\end{aligned}$$

where the limit in the second argument $\lim_{h,k\to 0} hk/\sqrt{h^2+k^2}$ is zero because the numerator is of order 2, whereas the denominator of order 1. Therefore, $f'(a,b) : \mathbb{R}^2 \to \mathbb{R}^2$, with the standard bases e_1, e_2 of $\mathbb{R}^2$, is given by

$$f'(a,b)(v) = \begin{pmatrix}2a & 2b\\ b & a\end{pmatrix} v.$$

Notice the following:

$$\begin{pmatrix}\frac{\partial f}{\partial x}(p) & \frac{\partial f}{\partial y}(p)\end{pmatrix} = \begin{pmatrix}\left.\frac{\partial x^2+y^2}{\partial x}\right|_p & \left.\frac{\partial (x^2+y^2)}{\partial y}\right|_p\\ \left.\frac{\partial xy}{\partial x}\right|_p & \left.\frac{\partial xy}{\partial y}\right|_p\end{pmatrix} = \begin{pmatrix}2a & 2b\\ b & a\end{pmatrix}.$$

This is not a coincidence. We will later see that the linear map $f'(p)$ is represented by the matrix of partial derivatives in some basis (to be defined for cases other than $\mathbb{R}^n$ with the standard basis).

Let us first show that the definition of *the* derivative actually makes sense.

6.3. **Lemma.** *The differentiability condition does not depend upon the chosen norms.*

More precisely, if a function is differentiable with respect to norms $|\cdot|_V$ and $|\cdot|_W$, then it is differentiable with respect to any other norms $||\cdot||_V$ and $||\cdot||_W$. Fix $p \in P \subseteq V$. Let $f : V \to W$ be differentiable with respect to $|\cdot|_V$ and $|\cdot|_W$, namely, there exists a linear map $L : V \to W$ such that

$$\lim_{v\to 0}\frac{f(p+v)-f(p)-L(v)}{|v|_V} = 0,$$

which means that for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, p) > 0$ such that

$$\frac{|f(p+v)-f(p)-L(v)|_W}{|v|_V} < \epsilon \quad \text{whenever} \quad |v|_V < \delta.$$

We will prove a more general statement that implies it for the special case above.

6.4. **Lemma.** *The existence and value of a limit is independent upon the chosen norms.*

Proof. Suppose that $\lim_{v\to a} g(v) = w$ ($a \in V$, $w \in W$, $g : V \to W$), namely, for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, a) > 0$ such that $|g(v) - w|_W < \epsilon$ whenever $|v - a|_V < \delta$. We will show that the same is true with any other norms $||\cdot||_V$ and $||\cdot||_W$. Since all norms on a finite-dimensional space are equivalent, there exist $c, C, c', C' > 0$ such that $c||v||_V \le |v|_V \le C||v||_V$ and $c'||v||_V \le |v|_V \le C'||v||_V$.

Let $\epsilon > 0$ be arbitrary; then there exists a $\delta > 0$ such that $|g(v) - w|_W < c'\epsilon$ whenever $|v - a|_V < \delta$. Pick $\delta' = C\delta$. Whenever $||v - a||_V < \delta' = C\delta$, we have that $|v - a|_V \le ||v - a||_V /C < C\delta/C = \delta$, therefore

$$||g(v) - w||_W \le |g(v) - w|_W /c' < c'\epsilon/c' = \epsilon,$$

completing the proof. □

6.5. **Lemma.** *The derivative map is unique.*

Proof. Let $f : P \subseteq V \to W$ be differentiable at $p \in P$. Then there exists a linear map $L : V \to W$ such that

$$\lim_{v \to 0} \underbrace{\frac{f(p+v) - f(p) - L(v)}{||v||_V}}_{=: g(v)} = 0.$$

Suppose that some linear map $M : V \to W$ also satisfies

$$\lim_{v \to 0} \underbrace{\frac{f(p+v) - f(p) - M(v)}{||v||_V}}_{=: h(v)} = 0.$$

We will show that $M = L$. Let h, g be given as shown above; then

$$L(v) = f(p+v) - f(p) - ||v||_V g(v), \qquad M(v) = f(p+v) - f(p) - ||v||_V h(v).$$

We find for any $k \in \mathbb{N}$:

$$h(u/k) - g(u/k) = \frac{L(u/k) - M(u/k)}{||u/k||_V} = \frac{(1/k)(L(u) - M(u))}{(1/k)||u||_V} = h(u) - g(u).$$

Finally, for any $u \in V$ we have

$$(L - M)u = Lu - Mu = ||u||_V (h(u) - g(u)) = ||u||_V \lim_{k \to \infty} h(u/k) - g(u/k) = 0,$$

completing the proof. $\square$

6.6. **Example.** We will now illustrate that the ability to choose the norms freely allows for simpler computations. Note that $V \times V$ is a vector space with addition and scalar multiplication defined component-wise. We will find the derivative of the inner product map $g : V \times V \to \mathbb{R}$, where $\mathbb{R}$ is endowed with the standard absolute value norm, and the norm on V is derived from g, namely, $||v|| = \sqrt{g(v,v)}$; choose the norm on $V \times V$ to be $||(v,w)||_2 = \sqrt{||v||^2 + ||w||^2} = \sqrt{g(v,v) + g(w,w)}$. Let $(p,q) \in V \times V$. Recall the Cauchy–Schwarz inequality $|g(v,w)| \leq ||v||\,||w||$, and $(r-s)^2 \geq 0$ implies $r^2 + s^2 = 2rs + (r-s)^2 \geq 2rs \geq rs$ for any $r, s \in \mathbb{R}$.

As usual, we begin by meditating to find the linear term

$$\begin{aligned} g((p,q) + (v,w)) - g(p,q) &= g(p+v, q+w) - g(p,q) \\ &= g(p,q) + p(p,w) + g(v,q) + g(v,w) - g(p,q) \\ &= p(p,w) + g(v,q) + g(v,w), \end{aligned}$$

where $g(v,w)$ appears to be quadratic in v, w. Indeed, set $g'(p,q)(v,w) = g(p,w) + g(v,q)$ and compute

$$\begin{aligned} 0 &\leq \lim_{v,w \to 0} \frac{|g((p,q) + (v,w)) - g(p,q) - g'(p,q)(v,w)|}{||(v,w)||_2} \\ &= \lim_{v,w \to 0} \frac{|g(v,w)|}{||(v,w)||_2} \leq \lim_{v,w \to 0} \frac{||v||\,||w||}{\sqrt{||v||^2 + ||w||^2}} \leq \lim_{v,w \to 0} \frac{||v||\,||w||}{\sqrt{||v||\,||w||}} \\ &= 0. \end{aligned}$$

Therefore the limit on the first line is zero, showing that the proposed $g'(p,q)$ is indeed the derivative of g at $(p,q) \in V \times V$.

In fact, since $g' : V \times V \to \mathrm{Hom}(V \times V, \mathbb{R})$ is linear,

$$\begin{aligned} g'(a(p+q)+b(r,s))(v,w) &= g'(ap+br, aq+bs)(v,w) \\ &= g(ap+br, w) + g(v, aq+bs) \\ &= ag(p,w) + bg(r,w) + ag(v,q) + bg(v,s) \\ &= a(g(p,w)+g(v,q)) + b(g(r,w)+g(v,s)) \\ &= ag'(p,q)(v,w) + bg'(r,s)(v,w), \end{aligned}$$

it follows that g' is continuous; therefore, *any inner product is continuously differentiable.*

In the special case $V = \mathbb{R}$ we have $g(p,q) = pq$ and $g'(p,q)(x,y) = py + qx$; identifying $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ with basis e_1, e_2 and dual basis $\varepsilon^1, \varepsilon^2$ we have $g'(p,q) = p\varepsilon^2 + q\varepsilon^1$.

6.7. *Remark.* The special case is very much worth remembering: the derivative of $(x,y) \mapsto xy$ at (x,y) is $x\varepsilon^2 + y\varepsilon^1$, or in matrix form:

$$\begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix}.$$

Indeed, the matrix form is very often far easier to compute. If it contains many zeros, a form in terms of dual vectors is more convenient to work with. This begs the question: what is the relation between the matrix of partial derivatives (if they exist & conditions on components – ?) such that the function is differentiable & continuously differentiable? Part 5 of the following theorem yields the answer.

6.8. **Theorem** (Properties of the derivative)**.** [10] *Let V, W, X be finite-dimensional vector spaces. Let $P \subseteq V$, $Q \subseteq W$ be open and $P \xrightarrow{f} Q \xrightarrow{g} X$ be (any) maps. Then*

(1) if f is differentiable at $p \in P$, then f is continuous at p;
(2) if f is constant on P, then f is (continuously) differentiable on P, $f'(p) = 0$ map for all $p \in P$;
(3) if $P = V$ and f is linear,[11] *then f is differentiable on P with $f'(p) = f$ for all $p \in P$;*
(4) (chain rule) *if f is differentiable at $p \in P$ and g is differentiable at $f(p) \in Q$, then $g \circ f$ is differentiable at p, and*

$$(g \circ f)'(p) = g'(f(p)) \circ f'(p). \tag{6.3}$$

(5) (relation to Jacobian) *let $c_1, ..., c_m$ be a basis of W; then f may be written in basis components as $f = f^i c_i = f^1 c_1 + \cdots + f^m c_m$ for $f^i : P \to \mathbb{R}$; then f is differentiable at p if and only if each f^i is differentiable at p, and*

$$f'(p) = c_i (f^i)'(p). \tag{6.4}$$

Stated less formally, the theorem reads:

(1) discontinuous functions are not differentiable;
(2) constant functions have derivative zero (everywhere);
(3) a linear function is its own derivative;
(4) derivative of composition is composition of derivatives;
(5) a function is differentiable if and only if the components are differentiable.

[10] Reproduced almost verbatim from R. vd Veen, 2021. I make no claim of originality in reproducing this theorem.

[11] Or f may be linearly extended to V; this can be the case for piecewise functions where some piece is a ‘linear’ function.

Proof. Let $||\cdot||_V$ be a norm on V, $||\cdot||_W$ on W and $||\cdot||$ be the operator norm of linear maps $V \to W$, namely,

$$||L|| = \sup_{||v||_V=1} ||Lv||_W = \sup_{v\neq 0} \frac{||Lv||_W}{||v||_V}, \tag{6.5}$$

which gives $||Lv||_W \leq ||L||\,||v||_V$.

(1). Take $h(v) = f(p+v) - f(p) - f'(p)v$ to be a map $V \to W$, restricted to where h is defined.[12] The differentiability condition $\lim_{v\to 0} h(p)/||v||_V = 0$ yields that for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$||v||_V < \delta \qquad \Longrightarrow \qquad \left|\left|\frac{h(v)}{||v||_V}\right|\right|_W = \frac{||h(v)||_W}{||v||_V} < \epsilon.$$

Then

$$\begin{aligned} ||f(p+v) - f(p)||_W &= ||f'(p)v - h(v)||_W \leq ||f'(p)v||_W + ||h(v)||_W \\ &\leq ||f'(p)||\,||v||_V + ||h(v)||_W \frac{||v||_V}{||v||_V} \\ &= \left(||f'(p)|| + \frac{||h(v)||_W}{||v||_V}\right) ||v||_V \\ &< (||f'(p)|| + \epsilon)||v||_V \end{aligned}$$

implies that choosing $\delta' = \min\{\delta, \epsilon/(||f'(p)|| + \epsilon)\}$ yields for all $||v||_V < \delta'$

$$||f(p+v) - f(p)||_W < \epsilon,$$

proving continuity.

(2). We have $f(p+v) - f(p) = 0$ for any permissible choice of p and v; hence the linear map $v \mapsto 0 \in W$ satisfies the limit of the definition of derivative.

(3). We have $f(p+v) - f(p) = f(v)$, which is a linear map; setting $f'(p)(v) = f(v)$ yields

$$\lim_{v\to 0} \frac{f(p+v) - f(p) - f'(p)(v)}{||v||_V} = \lim_{v\to 0} \frac{0}{||v||_V} = 0.$$

(4). Let

$$\begin{aligned} h(v) &= f(p+v) - f(p) - f'(p)v, \\ k(w) &= g(f(p)+w) - g(f(p)) - g'(f(p))w. \end{aligned}$$

Then

$$\begin{aligned} \ell(v) &= (g\circ f)(p+v) - (g\circ f)(v) - g'(f(p))f'(p)v \\ &= g(f(p) + f'(p)v + h(v)) - g(f(p)) - g'(f(p))f'(p)v \qquad w := f'(p)v + h(v) \\ &= k(w) + g'(f(p))(f'(p) + h(v)) - g'(f(p))f'(p)v \\ &= k(f'(p)v + h(v)) + g'(f(p))h(v). \end{aligned}$$

Since $g'(f(p))$ is linear and $\lim_{v\to 0} h(v)/||v||_V = 0$ exists, $\lim_{v\to 0} g'(f(p))h(v)/||v||_V = g'(f(p)) \lim v \to 0 h(v)/||v||_V = 0$. Then $\lim \ell(v)/||v||_V = 0$ if and only if $\lim k(f(p+v) - f(p))/||v||_V = 0$.

Let $\epsilon > 0$. Differentiability of g gives the existence of $\delta' > 0$ such that $||k(w)||_X < \epsilon||w||_W/(||f'(p)||+1)$ for all $||w||_W < \delta'$; and differentiability of f gives $\delta < \delta'$ such that $||h(v)||_W/||v||_V < 1$ for all $||v||_V < \delta$.

---

[12]As P is open, there exits an open ball around p contained in P.

Then

$$\begin{aligned}\frac{||k(f'(p)v+h(v))||_X}{||v||_V} &= \epsilon\frac{1}{||f'(p)||+1}\frac{||f'(p)v+h(v)||_W}{||v||_V}\\ &\leq \epsilon\frac{1}{||f'(p)||+1}\frac{||f'(p)||\,||v||_V+||h(v)||_W}{||v||_V}\\ &= \epsilon\frac{1}{||f'(p)||+1}\left(||f'(p)||+\frac{||h(v)||_W}{||v||_V}\right)\\ &< \epsilon\frac{1}{||f'(p)||+1}(||f'(p)||+1)\\ &= \epsilon.\end{aligned}$$

(5). Let $\gamma^1,...,\gamma^m \in W^*$ be a dual basis for W. By part 3, these maps are linear, thus differentiable; by part 4, the composition $f^i = \gamma^i \circ f$ is differentiable if f is differentiable. We show the converse. Let f^i be differentiable at p for all i; hence

$$\lim_{v\to 0}\frac{|f^i(p+v)-f^i(p)-(f^i)'(p)v|}{||v||_V}=0.$$

Then (**NB!** Einstein notation: $(f^i)'(p)(v)c_i = (f^1)'(p)(v)c_1 + ... + (f^m)'(v)(p)c_m$)

$$\begin{aligned}\lim_{v\to 0}\frac{||f(p+v)-f(p)-(f^i)'(p)(v)c_i||_W}{||v||_V} &= \lim_{v\to 0}\frac{||f^i(p+v)c_i-f^i(p)c_i-(f^i)'(p)(v)c_i||_W}{||v||_V}\\ &\leq \lim_{v\to 0}\frac{|f^i(p+v)-f^i(p)-(f^i)'(p)(v)|\,||c_i||_W}{||v||_V}\\ &= 0,\end{aligned}$$

therefore, the limit exists and is zero, as required. Note that at the inequality summation goes from inside the norm to outside it. □

6.9. *Remark.* Indeed, part 5 of the preceding theorem gives the Jacobian matrix. Recall from early calculus that $(f^i)'(p)$ is represented by the gradient matrix (row! vector) $\nabla f^i(p) = (\frac{\partial f^i}{\partial x}(p), \frac{\partial f^i}{\partial y}(p), ...)$; indeed, the reader may easily show this from the definition. The gradient exists and *each component is continuous at* p if and only if f^i is differentiable; f is continuously differentiable if and only if each gradient component is continuous in p. Let $\mathbf{b} = b_1, ..., b_n$ be a basis for V and $\mathbf{c} = c_1, ..., c_m$ for W. This gives

$${}_{\mathbf{c}}[f'(p)]_{\mathbf{b}} = \begin{pmatrix} {}_{\mathbf{c}}[(f^1)'(p)]_{\mathbf{b}} \\ \vdots \\ {}_{\mathbf{c}}[(f^m)'(p)]_{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \nabla f^1(p) \\ \vdots \\ \nabla f^m(p) \end{pmatrix} = \begin{pmatrix} \frac{\partial f^1(p)}{\partial x_1} & \cdots & \frac{\partial f^1(p)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f^m(p)}{\partial x_1} & \cdots & \frac{\partial f^m(p)}{\partial x_n} \end{pmatrix}.$$

6.10. *Remark* (What criminal acts must you perform to compute a derivative?). If a basis representation is convenient, compute the matrix of partial derivatives; it should be fairly easy to check whether they exist; if not – f is not differentiable. If the components of the matrix are continuous functions at p, differentiability is implied. If no basis is convenient, repeatedly apply the chain rule.

This criminality is allowed precisely by part 5 of the preceding theorem.

But the best way: naïvely, 'professionally-stupidly' find the first-order term, and then check via the limit; namely, explicitly expand $F(p+v)$ and check, which terms depend 'in the norm' only on $||v||$. You should obtain that for some linear map L,

$$F(p+v) = F(p) + L(v) + O(||v||^2)$$

where the term L itself is your guess for the derivative. The map L could have a nice matrix form, or map matrices $H \mapsto AH + HA^\top$ or be an inner product

$v \mapsto \langle v, w\rangle$, or whatever... but contain v in the first order. Then plug L into the definition and compute the limit.

6.11. **Example.** I am tired, and getting old, thus I leave the list of examples as is; the reader should now compute $f'(p)$ for their favourite point for the following:

$$f(x,y) = \frac{(x,y)}{\sqrt{x^2+y^2}}, \quad f(x,y) = \frac{(x^2y, xy^2)}{(x^2+y^2)^{3/2}}, \quad f(x,y) = \left(\frac{\cos y}{x}, \frac{\cos x}{y}, xy\right).$$

Certain, more advanced, examples include: let $F : GL(n;\mathbb{R}) \to GL(n;\mathbb{R}) \subseteq \mathbb{R}^{n\times n}$ map $A \mapsto A^{-1}$; then

$$F'(A)(H) = \frac{-1}{1+\operatorname{Tr} HA^{-1}} A^{-1}HA^{-1}.$$

Let $s : \operatorname{Hom}(V,V) \to \operatorname{Hom}(V,V)$ be given by $s(L) = L \circ L$; then

$$s'(L)(K) = K \circ L + L \circ K.$$

The matrix of partial derivatives may be constructed in terms of columns instead.

6.12. **Definition.** The directional derivative $\partial_v f(p)$ of $f : P \to W$ at p in the direction of v is defined as

$$\partial_v f(p) = \lim_{h\to 0} \frac{f(p+vh) - f(p)}{h} \tag{6.6}$$

if the limit exists. Given a basis $b_1, ..., b_n$ of V, the partial derivatives are directional derivatives in the directions of the basis vectors $\partial_{b_i} f$.

6.13. **Lemma.** *Let $f : P \to W$ be differentiable at $p \in P$. Then $f'(p)(v) = \partial_v f(p)$. Let $\mathbf{b} = b_1, ..., b_n$ be a basis of V and $\mathbf{c} = c_1, ..., c_m$ be a basis of W; we write $f = f^i c_i$. Then*

$$_{\mathbf{c}}[f'(p)]_{\mathbf{b}} = \begin{pmatrix} \partial_{b_1} f^1(p) & \cdots & \partial_{b_n} f^1(p) \\ \vdots & & \vdots \\ \partial_{b_1} f^m(p) & \cdots & \partial_{b_n} f^m(p) \end{pmatrix}, \tag{6.7}$$

which is called the Jacobian.

7. Tangent plane

In some sense the motivation for the derivative comes from the wish to define tangent planes to surfaces. Indeed, the definition allows very clear definitions for regular, 2-dimensional surfaces, which the reader has encountered before; these generalise.

7.1. **Example.** Let $f(\theta) = (\cos\theta, \sin\theta)$ parametrise some open subset of the circle containing $f(p)$ for some fixed $0 < p < \pi/2$. Then to find the tangent line at p, we need the tangent vector at p, which, we recall from ordinary calculus, is given by the vector $\dot{f}(p) = (-\sin\theta, \cos\theta)$; then the tangent line is the span of this vector, $T_p f = \{\lambda(-\sin\theta, \cos\theta) : \lambda \in \mathbb{R}\}$.

Computing $f'(p) = (-\sin\theta\varepsilon^1, \cos\theta\varepsilon^1)$, notice that this tangent vector is in fact $\dot{f}(p) = f'(p)e_1$, where $\mathbb{R} = \operatorname{Span}(e_1)$ is the (only) basis vector of the domain of f and ε^1 is its dual. Then

$$\begin{aligned} T_{f(p)} f = \operatorname{Span} \dot{f}(p) &= \{\lambda \dot{f}(p) : \lambda \in \mathbb{R}\} = \{\lambda f'(p) e_1 : \lambda \in \mathbb{R}\} \\ &= \{f'(p)(\lambda e_1) : \lambda \in \mathbb{R}\} = f'(p)(\mathbb{R}) \end{aligned}$$

we see that the tangent line is the image of the derivative $T_{f(p)} f = f'(p)(V)$ for $V = \mathbb{R}$!

7.2. **Example.** Let $f(\theta, \phi) = (\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi)$ parametrise some open subset of the sphere containing $f(\theta_0, \phi_0)$ for some allowed $p = (\theta_0, \phi_0)$. We can fix θ_0 and only consider $f_{\theta_0}(\phi)$ as a function $\mathbb{R} \to \mathbb{R}^3$; the tangent vector is given by

$$\dot{f}_{\theta_0}(\phi_0) = (\cos\theta_0\cos\phi_0, \sin\theta_0\cos\phi_0, -\sin\phi_0).$$

Similarly,

$$\dot{f}^{\phi_0}(\theta_0) = (-\sin\theta_0\sin\phi_0, \cos\theta_0\sin\phi_0, 0).$$

Supposing that these two tangent vectors are independent, they span a 2-dimensional subspace of $\mathbb{R}^3$; this we call the tangent plane; note that it is by default a subspace, and thus passes through the origin of the vector space – if we want it to pass through a point, we must offset it by $f(p)$ – this is then called an 'affine subspace' (translated subspace).

Letting $\mathbb{R} = \operatorname{Span}(e_1) = \operatorname{Span}(e_2)$, we have $\dot{f}^{\phi_0}(\theta_0) = (f^{\phi_0})'(\theta_0)e_1$ and $\dot{f}_{\theta_0}(\phi_0) = (f_{\theta_0})'(\phi_0)e_2$. But notice that $\mathbb{R}^2 = \operatorname{Span}\{e_1, e_2\}$

We then write

$$\begin{aligned} T_{f(p)}f &= \operatorname{Span}\{\dot{f}^{\phi_0}(\theta_0), \dot{f}_{\theta_0}(\phi_0)\} \\ &= \operatorname{Span}\{(f^{\phi_0})'(\theta_0)e_1, (f_{\theta_0})'(\phi_0)e_2\} \\ &= \operatorname{Span}[(f^{\phi_0})'(\theta_0), (f_{\theta_0})'(\phi_0)] \\ &= \operatorname{Span}[\partial_{e_1}f(p), \partial_{e_2}f(p)] \\ &= f'(p)(\mathbb{R}^2) \end{aligned}$$

is once again the image – the span of the two partial derivative vectors, which is the span of the columns of the patrial derivative matrix.

7.3. **Example.** Alternatively, the sphere may be defined as the preimage $g^{-1}(\{0\})$ of $g(x, y, z) = x^2 + y^2 + z^2 - 1$. We recall that the normal to this surface is given by the gradient of g at p, namely, $\nabla g(p)$, which is the matrix (= row vector) representation of $g'(p)$. The tangent plane is given by all vectors of $\mathbb{R}^3$ that are perpendicular to the normal vector $n(p) = \nabla g(p)^\top$ (the transpose is there to get a column vector[13]).

Let v be in the tangent plane. But $v \perp n(p)$ means that $n(p)\dot{v} = 0$, or, equivalently, $[\nabla g(p)][v] = 0$ or $v \in \ker \nabla g(p)$. But as $\nabla g(p)$ is the matrix representation of $g'(p)$, the kernels coincide $\ker g'(p) = \ker \nabla g(p)$. Hence the tangent plane is given by

$$T_p g^{-1}(\{0\}) = \ker g'(p).$$

Note that for $f : V \to W$, $g : V \to \mathbb{R}$ we have $T_{f(p)}f \subseteq W$ and $T_p g^{-1}(\{0\}) \subseteq V$.

7.4. **Definition** (Tangent plane)**.** Let $f : P \subset V \to W$ and $g : P \subset V \to \mathbb{R}$ be differentiable at $p \in P$. Denote $S = g^{-1}(\{0\})$. Then define the *tangent plane*[14] at $q = f(p) \in W$ and $p \in S$, respectively, to be given by

$$T_q f = f'(p)(V) \subseteq W, \tag{7.1}$$

$$T_p S = \ker g'(p) \subseteq V. \tag{7.2}$$

8. Notation

In these notes we adopt $f'(p) : V \to W$ to denote a linear map, where upon evaluation the second argument $v \in V$ is written in second brackets or adjacently:

$$f'(p)(v) \in W \qquad\qquad f'(p)v \in W.$$

[13]But it is a pedantic addition, and is often omitted. The reader will quickly find that the really beautiful theories of mathematics make proper use of a criminal abuse of notation...

[14]Not necessarily 2-dimensional!

If f is differentiable over $P \subset V$ we denote $f' : P \to \mathrm{Hom}(V, W)$ to send each point p to it's associated linear map $f'(p)$ (that gives the first order approximation at p).

The following are various notations used for the derivative:

$$
\begin{array}{cccc}
Df(p) & Df|_p & D_p f & D|_p f \\
df(p) & df|_p & d_p f & d|_p f \\
\dfrac{df}{d\mathbf{x}}(p) & \left.\dfrac{df}{d\mathbf{x}}\right|_p & - & \left.\dfrac{d}{d\mathbf{x}}\right|_p f,
\end{array}
$$

where the bold $\mathbf{x}$ strictly emphasises that x is a *vector*!

The partial derivatives are denoted

$$\partial_{e_1} f(p) \qquad \partial_1 f(p) \qquad \partial_x f(p) \qquad \frac{\partial f}{\partial x}(p) \qquad f_x(p) \qquad f_1(p) \qquad (f'_x)(p),$$

with similarly wild opinions on where to stick that annoying p.

For $\mathbb{R}^3$ with standard basis e_1, e_2, e_3 the dual vectors are often denoted $\varepsilon^1 = dx$, $\varepsilon^2 = dy$, and $\varepsilon^3 = dz$. For $\mathbb{R}^n$ we denote $\varepsilon^i = dx^i$ for each $1 \le i \le n$. The reasoning behind this is the following: let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by $(x, y, z) \mapsto x$; then

$$dx := d((x, y, z) \mapsto x)(x, y, z) = \varepsilon^1.$$

Alternatively, to make nice formulae

$$df(p) = \frac{\partial f}{\partial x}(p)\,dx + \frac{\partial f}{\partial y}(p)\,dy + \frac{\partial f}{\partial z}(p)\,dz$$

instead of

$$f'(p) = \partial_{e_1} f(p)\varepsilon^1 + \partial_{e_2} f(p)\varepsilon^2 + \partial_{e_3} f(p)\varepsilon^3.$$

For integrals in the ordinary calculus sense the following notations are common:

$$\int_{x\in[0,1]} f \qquad\qquad \int_0^1 f\,|dx|$$

but in this course $\int_0^1 f\,dx$, namely, dx itself, will mean something different!

9. Definition of the Riemann integral

For now I will not write full formalism; it suffices to say that we can repeat the tedious real analysis arguments about partitions – now not of an interval $[a, b]$, but of a k-dimensional rectangle $[a_1, b_1] \times \cdots \times [a_k, b_k]$. One possibility is to define the Riemann integral as the value when upper and lower integrals (infima, suprema of sums over partitions) coincide;[15] another is to define it as the limit $n \to \infty$ upon equally dividing the rectangle in $n^k - 1$ pieces, each of 'side length' $(b - a)/n$.[16] Instead of dividing the original rectangle sides by n, one may also divide in half each time, creating these so-called 'dyadic cubes' of side length $(b - a)/2^n$.

Whatever the theory, the Riemann integral (if it exists) over a rectangle $B \subsetneq \mathbb{R}^k$ of a function $f : B \to \mathbb{R}$ is the best approximation to dividing up B into parts, evaluating f in each of these parts, and summing the values times volume of parts; in a way, measuring the $k + 1$ volume 'under f'.

Certainly, the definition is never used in practice, and is useful only for the following three statements:

(1) continuous functions are integrable;

[15] Cf. `https://math.okstate.edu/people/lebl/osu4153-s16/chapter10-ver1.pdf`.

[16] Cf. Roland, § 2.2.

(2) integration over $B = [a, b] \times [c, d]$ can be done by first dividing B into horizontal strips, or dividing into vertical strips; this generalises for $B = B' \times B''$, where B' is a p-rectangle and B'' is a q-rectangle – Fubini's theorem;
(3) characterising all Riemann-integrable functions – Riemann-Lebesgue theorem: a bounded $f : B \subsetneq \mathbb{R}^n \to \mathbb{R}$ is integrable if and only if the set of discontinuities of f has measure zero.

Indeed, point (4) shows that the theory of Riemann integration is weak in the analysis sense; indeed, the domain is even assumed to be a subset of $\mathbb{R}^n$ – what a terrible restriction! – to absolve these sins, one requires the Lebesgue integral:[17]

9.1. *Remark* (§ Lebesgue integral– *not part of this course*). Let X be any set, and $\mathfrak{M}$ be a collection of subsets containing X that is closed under complements and *countable* unions, called a σ-algebra (σ refers to *Summe* in German), and let $\mu : \mathfrak{M} \to [0, \infty]$ (yes, ∞ is included, with the convention $0 \cdot \infty = 0$) map countable unions to series $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$, called a *measure*; the *Lebesgue integral* of f over $E \in \mathfrak{M}$ with respect to μ is then the supremum

$$\int_E f \, d\mu := \sup_s \int_E s \, d\mu, \qquad \text{where} \qquad \int_E s \, d\mu := \sum_{i=1}^{n<\infty} \alpha_i \mu(A_i \cap E)$$

over all functions with finite image $s : X \to [0, \infty)$, $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$, where each $\alpha_i \in [0, \infty)$ and χ_{A_i} is 1 if $A_i \in \mathfrak{M}$ and 0 otherwise.

Ok.

Back to our theory. Since I do not possess the moral stamina to not skip the following proofs, I will also skip the formal definition of the Riemann integral – the statements hold for any acceptable definition.[18] Shortly: similarly as for $f : [a, b] \to \mathbb{R}$ but replacing the partition of $[a, b]$ by a partition of B using cubes.

9.2. **Lemma.** *Let $B = [a_1, b_1) \times \cdots \times [a_k, b_k)$ be a* half-open rectangle *($\bar{B}$ closed), $\alpha, \beta \in \mathbb{R}$ and $f, g : \bar{B} \to \mathbb{R}$ be continuous. Then*

(1) $\int_B f$ exists (any continuous function is integrable)*;*
(2) $\int_B (\alpha f + \beta g) = \alpha \int_B f + \beta \int_B g$ (the integral is linear)*;*

9.3. **Lemma.** *Let $f, g : U \subset \mathbb{R}^k \to \mathbb{R}$ be continuous over U open. If $\int_B f = \int_B g$ for all $B \subsetneq U \subseteq \mathbb{R}^k$ open subset, then $f = g$. Dropping the assumption of continuity, $f = g$ almost everywhere (i.e. except at most on a set of measure zero).*

9.4. **Theorem** (Fubini)**.** *Let $A \subsetneq \mathbb{R}^p$, $B \subsetneq \mathbb{R}^q$ be half-open rectangles, and $f : \bar{A} \times \bar{B} \to \mathbb{R}$ be continuous. Let $F : \bar{A} \to \mathbb{R}$, $a \to \int_B f(a, \cdot)$ and $G : \bar{B} \to \mathbb{R}$, $b \to \int_B f(\cdot, b)$. Then F and G are continuous and*

$$\int_{A \times B} f = \int_A F = \int_B G. \tag{9.1}$$

In essence, Fubini's theorem allows for computations of integrals in the 'usual' multivariable calculus way.

9.5. **Example.** Let $R = [0, 1]^n \subsetneq \mathbb{R}^n$ be the standard n-cube, and let $f : R \to \mathbb{R}$ be given by $f(x^1, x^2, ..., x^n) = x^1 \cdot x^2 \cdot \cdots \cdot x^n$ (recall: upper indices denote coefficients $x = x^i e_i$). We compute $I := \int_R f$.

[17] Cf. Rudin, Real and complex analysis, chapter 1 – a 30 page introduction to Lebesgue integration.

[18] And the proofs will be tedious for every one of them...

Let $g(x^n) = x^1 \cdots x^n$ and $h(x^n) = (1/2)(x^1 \cdots x^{n-1})(x^n)^2$, hence we have the relation $g(x^n) = h'(x^n)$. Note that $h(0) = 0$. Then

$$I = \int_R f = \int_{[0,1]^{n-1}} \int_{[0,1]} g = \int_{[0,1]^{n-1}} \int_{[0,1]} h' = \frac{1}{2} \int_{[0,1]^{n-1}} h(1) = \cdots = \frac{1}{2^n},$$

where we repeatedly apply a similar process of defining g and h for the last remaining variable (coordinate), obtaining $1/2^n$.

Fubini's theorem allows for a quick proof of the following.

9.6. **Proposition** (Mixed partials commute). *For any $f \in C^2$, $f : P \subseteq V \to W$ and any $u, v \in V$ the mixed directional derivatives are equal*

$$\partial_u \partial_v f = \partial_v \partial_u f. \tag{9.2}$$

Proof. Given f and any $p \in P$, let $\tilde{f} : U \subseteq \mathbb{R}^2 \to \mathbb{R}$ be given by $\tilde{f}(x, y) = f(p + xu + yv)$ with U open containing all allowed inputs; then $\partial_u \partial v f(p) = \partial_{e_1} \partial_{e_2} \tilde{f}(0, 0)$ and $\partial_v \partial_u f(p) = \partial_{e_2} \partial_{e_1} \tilde{f}(0, 0)$. Hence it suffices to prove the statement for $\tilde{f}$.

Let $[a, b) \times [c, d) \subsetneq U$ be arbitrary. Denote $g_x(q) = \partial_1 \tilde{f}(x, q)$ and $F(x) = \int_{[c,d]} g'_x$, which by the fundamental theorem of calculus[19] $F(x) = g(d) - g(c)$; letting $h(x) = \partial_{e_1} \tilde{f}(x, d) - \partial_{e_1} \tilde{f}(x, c)$, it follows that

$$\begin{aligned} I := \int_{[a,b]} F = \int_{[a,b]} \partial_{e_1} \tilde{f}(\cdot, d) - \partial_{e_1} \tilde{f}(\cdot, c) = \int_{[a,b]} h' = h(b) - h(a) \\ = f(b, d) - f(b, c) - f(a, d) + f(a, c). \end{aligned}$$

Similarly defining G, it can be shown that the same value is obtained for $J := \int_{[c,d]} G$, namely, $I = J$. By Fubini,

$$\int_{[a,b]\times[c,d]} \partial_{e_1} \partial_{e_2} \tilde{f} = I = J = \int_{[a,b]\times[c,d]} \partial_{e_2} \partial_{e_1} \tilde{f}.$$

Since this was true for any rectangles in U, by Lemma 9.3, it follows that over $U \ni (0, 0)$ the functions are equal $\partial_{e_1} \partial_{e_2} \tilde{f} = \partial_{e_2} \partial_{e_1} \tilde{f}$. As $\tilde{f}$ was defined for any $p \in P$, the statement holds for f, completing the proof. $\square$

[19]Cf. Abbott, Understanding analysis.

Part 3. Tensor algebra

The 'grand' theory that exists for integration (theorems of Stokes, Poincaré, and company), exists consistently only for these so-called forms. The reason behind this is perversely simple: we want a change of coordinates to work as usual.[20] We will explore this in Part 4.

The short story is that this forces 'linearly dependent' (whatever that means) forms to equal zero; similarly, swapping two elements gives a minus sign. What exactly are these 'elements'? For that we need the language of tensors – multilinear maps.

10. Tensors

Throughout this section let $e_1, ..., e_n$ be a fixed basis of V with dual basis $\varepsilon^1, ..., \varepsilon^n$. **NB!** Note that we use the symbols for the basis of V and not as the standard basis of $\mathbb{R}^n$.

From this point onwards (at least in the current section), we will adopt a way to significantly shorten notation; in many instances already, expressions of the form $a^1 x_1 + \cdots + a^n x_n$ have been enountered, where summation is done over the index $a^i x_i$. Probably the greatest invention of Einstein (I believe, in his own words), was the adoption of the notation

$$a^i x_i := a^1 x_1 + \cdots + a^n x_n, \tag{10.1}$$

where, upon encountering a repeated index (i above) *both above and below*, summation is done over all allowed values of i (typically from 1 to n). If there is any ambiguity about where to stop summation, it will be indicated in text.

Generalising, this works for any repeated indices, as illustrated below

$$b_{ij} x^i y^i := \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} x^i y^j. \tag{10.2}$$

This is called *Einstein's summation notation.* In case of division, $1/x^i$ counts as a lower index – *an inverse inverts the index.*

We will now develop the definition of tensors starting from bilinear functionals, ending with tensors in physics-compatible notation. The crucial idea is the following: *to define multiplication of linear maps, we borrow multiplication from the real numbers* – since functionals map into the real numbers, we can multiply values of functionals.

We will use the intuition from bilinear maps for defining tensors.

10.1. **Definition.** The map $B : V \times V \to \mathbb{R}$ is said to be a bilinear functional in V if it is linear in each argument, while keeping the other argument fixed:

$$\begin{aligned} B(x + \lambda x', y) &= B(x, y) + \lambda B(x', y), \\ B(x, y + \lambda y') &= B(x, y) + \lambda B(x, y') \end{aligned} \tag{10.3}$$

for all $x, x', y, y' \in V$ and $\lambda \in \mathbb{R}$.

[20] Change of volume is given by the determinant of the parametrisation; modulo details.

We shall now express B in a fixed basis $e_1, ..., e_n$ of V. Set the number $b_{ij} = B(e_i, e_j)$, which gives for $x = x^i e_i = x^1 e_1 + \cdots + x^n e_n$ and $y = y^j e_j$ the equality

$$
\begin{aligned}
B(x,y) &= B(x^i e_i, y^j e_j) = x^i B(e_i, y^j e_j) = x^i y^j B(e_i, e_j) = b_{ij} x^i y^j \\
&= \begin{Bmatrix} b_{11}x^1y^1 & + & \cdots & + & b_{1n}x^1y^n & + \\ b_{12}x^2y^1 & + & \cdots & + & b_{1n}x^2y^n & + \\ \vdots & & & & \vdots & \\ b_{n1}x^ny^1 & + & \cdots & + & b_{nn}x^ny^n & \end{Bmatrix}.
\end{aligned} \tag{10.4}
$$

It is evident that the coefficients $\{b_{ij} : i, j \in \{1, ..., n\}\}$ may be arranged in a matrix

$$
[B] = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix}, \tag{10.5}
$$

which is called the matrix of the bilinear functional B (in a given basis). This allows us to write $B(x, y) = [x]^\top B[y]$, where $[v] \in \mathbb{R}^n$ is the column vector of the coefficients of $v \in V$ in the $e_1, ..., e_n$ basis: $[v] = (v^1, ..., v^n)^\top$.

By adding two bilinear functionals $B + C$ term-wise, the resulting bilinear functional will be represented by the matrix $[B + C] = [B] + [C]$, and thus addition is linear; similarly, for scalar multiplication, $[\lambda B] = \lambda[B]$. Therefore bilinear functionals on V form a real vector space, denoted by $T_2(V)$. The correspondence between a bilinear functional and its matrix in some basis proves the following:

10.2. **Lemma.** $T_2(V) \cong \mathbb{R}^{n\times n} \cong \mathbb{R}^{nn}$.

10.3. **Definition.** Let $\varphi, \eta \in V^*$ be two linear functionals on V. Then the bilinear functional $\varphi \otimes \eta : V \times V \to \mathbb{R}$ defined by

$$
(\varphi \otimes \eta)(x, y) = \varphi(x)\eta(y) \tag{10.6}
$$

is called the tensor product of the functionals φ and η. The multiplication on the right is the standard multiplication of real numbers.

We now consider tensor products of dual basis vectors $\varepsilon^1, ..., \varepsilon^n$. Let $x, y \in V$. Since $\varepsilon^i(x) = x^i$ and $\varepsilon^j(y) = y^j$, we have

$$
(\varepsilon^i \otimes \varepsilon^j)(x, y) = x^i y^j. \tag{10.7}
$$

Therefore, for a bilinear functional $B = b_{ij}(\varepsilon^i \otimes \varepsilon^j)$ (note the sum over all i and j pairs) we have

$$
B(x, y) = b_{ij}(\varepsilon^i \otimes \varepsilon^j)(x, y) = b_{ij}\varepsilon^i(x)\varepsilon^j(y) = b_{ij}x^i y^j; \tag{10.8}
$$

indeed, $B(e_i, e_j) = b_{ij}$, which implies that B is represented by the matrix in Eq. (10.5).

It likewise shows that $\varepsilon^i \otimes \varepsilon^j$ are independent: if for all e_k, e_ℓ we have

$$
0 = 0(e_k, e_\ell) = B(e_k, e_\ell) = b_{ij}\varepsilon^i(e_k)\varepsilon^j(e_\ell) = b_{ij}\delta^i_k\delta^j_\ell = b_{k\ell},
$$

then the coefficients $b_{k\ell} = 0$ for all $k, \ell \in \{1, ..., n\}$. Since we have a total of n^2 independent bilinear functionals $\varepsilon^i \otimes \varepsilon^j$, and by Lemma 10.2 we have $\dim T_2(V) = \dim \mathbb{R}^{nn} = n^2$, it follows that $(\varepsilon^i \otimes \varepsilon^j : i, j = 1, ..., n)$ form a basis of $T_2(V)$.

A clever reader may already anticipate that every algebraic definition come with a dual definition. Indeed, the dual theory of bilinear functionals reverses vectors and covectors (functinals) in the definition.

10.4. **Definition.** The map $B : V^* \times V^* \to \mathbb{R}$ is said to be a bilinear functional of covectors of V if

$$
\begin{aligned}
B(\varphi + \lambda\varphi', \eta) &= B(\varphi, \eta) + \lambda B(\varphi', \eta), \\
B(\varphi, \eta + \lambda\eta') &= B(\varphi, \eta) + \lambda B(\varphi, \eta')
\end{aligned} \tag{10.9}
$$

for all $\varphi, \varphi', \eta, \eta' \in V$ and $\lambda \in \mathbb{R}$.

Indeed, $B(\varphi, \eta) = b^{ij}\varphi_i\eta_j$, where $b^{ij} = B(\varepsilon^i, \varepsilon^j)$, $\varphi_i = \varphi(e_i)$, and $\eta_j = \eta(e_j)$. Similarly as before, a basis is constructed via tensor products. The coefficients b^{ij} give rise to a matrix, as in Eq. (10.5).

We call the vector space of bilinear functionals on covectors $T^2(V)$. By comparing the two definitions, the following immediately follows.

10.5. **Lemma.** $T^2(V) = T_2(V^*)$.

10.6. **Definition.** The tensor product $x \otimes y \in T^2(V)$ as a bilinear functional of covectors of $x, y \in V$ is defined by

$$(x \otimes y)(\varphi, \eta) = \varphi(x)\eta(y) \tag{10.10}$$

for any $\varphi, \eta \in V^*$.

Notice how similar this definition is to the natural association $V \cong (V^*)^*$ given by $\Phi(x)(\varphi) = \hat{x}(\varphi) = \varphi(x)$, where $\Phi : V \to (V^*)^*$ is the natural isomorphism. Indeed, $T_2(V^*)$ generalises $(V^*)^*$ from linear to bilinear functionals, which could thus be called $T_1(V^*)$ by analogy (we will indeed provide a proper definition later); similarly, $T_2(V)$ generalises V^*.

We then have that $e_i \otimes e_j \in T^2(V)$ form a basis of $T^2(V)$, namely, any $B \in T^2(V)$ may be written as $B = b^{ij} e_i \otimes e_j$. It may be proven analogously as before, or, by exploiting $T^2(V) = T_2(V^*)$, which we present below:

$$T^2(V) \ni (x \otimes y)(\varphi, \eta) = \varphi(x)\eta(y) = \hat{x}(\varphi)\hat{y}(\varphi) = (\hat{x} \otimes \hat{y})(\varphi, \eta) \in T_2(V^*), \tag{10.11}$$

where the basis of $T_2(V^*)$ consists of tensor products of functionals $f_1, ..., f_n \in (V^*)^*$ acting on basis vectors $\varepsilon^1, ..., \varepsilon^n$ of V^* as $f_i(\varepsilon^j) = \delta^i_j$; this is satisfied by setting $f_i := \hat{e}_i$, thus completing the proof.

We now generalise the two definitions of $T_2(V)$ and $T^2(V)$ above.

10.7. **Definition.** The mixed bilinear functional $B : V \times V^* \to \mathbb{R}$ is defined by being linear: vectors in the first argument and covectors in the second. Namely,

$$\begin{aligned} B(x + \lambda x', \eta) &= B(x, \eta) + \lambda B(x', \eta), \\ B(x, \eta + \lambda \eta') &= B(x, \eta) + \lambda B(x, \eta') \end{aligned} \tag{10.12}$$

for all $x, x', \eta, \eta' \in V$ and $\lambda \in \mathbb{R}$. The space of mixed bilinear functionals is denoted by $T^1_1(V)$.

10.8. **Definition.** The tensor product $\eta \otimes y \in T^1_1(V)$ is defined by

$$(\eta \otimes y)(x, \varphi) = \eta(x)\varphi(y), \tag{10.13}$$

where $x \in V$ and $\varphi \in V^*$.

Similarly as before, $B(x, \varphi) = b^j_i x_i \varphi_j$, where $b^j_i = B(e_i, \varepsilon^j)$, and $\varepsilon^i \otimes e_j$ form a basis of $T^1_1(V)$: $B = b^j_i \varepsilon^i \otimes e_j$.

Indeed, a further generalisation is possible still! Bilinearity itself may easily be generalised to multilineality: linear in each argument. This may be done both for the number of vector components and covector components; this is the final generalisation, and leads to the general definition of a tensor.[21]

10.9. **Definition** (Tensor)**.** Let $p, q \geq 0$. A (p, q)-tensor in a space V is a multilinear functional $T : V^p \times (V^*)^q \to \mathbb{R}$ taking p vector and q covector arguments. The space of (p, q)-tensors is denoted by $T^q_p(V)$.

[21]**NB!** Kiselev uses the *opposite* convention for tensors: our (p, q)-tensors are his (q, p)-tensors.

We have already met the following examples: $(1,0)$-tensors are covectors, $(0,1)$-tensors are vectors (as $V = (V^*)^*$), bilinear functionals are $(2,0)$-tensors, whereas bilinear functionals on covectors – $(0,2)$-tensors. We let $T^0_0(V) = \mathbb{R}$.

We now construct a basis. We compute (writing index summation explicitly):

$$T(x_1, .., x_p, \varphi^1, ..., \varphi^q) = \sum_{i_1,...,i_p=1}^{n} \sum_{j_1,...,j_q=1}^{n} T^{j_1\cdots j_q}_{i_1\ldots i_p} x^{i_1}_1 \cdots x^{i_p}_p \varphi^1_{j_1} \cdots \varphi^q_{j_q}, \tag{10.14}$$

where $T^{j_1\cdots j_q}_{i_1\ldots i_p} = T(e_{i_i}, ..., e_{i_p}, \varepsilon^{j_1}, ..., \varepsilon^{j-q})$ is the value on the basis vectors and covectors, and is called a *tensor component.* Let us introduce particularly useful notation: let $I = (i_1, ..., i_p)$ and $J = (j_1, ..., j_q)$ denote ordered lists of indices; then

$$T(x_1, ..., x_p, \varphi^1, ..., \varphi^q) = T^J_I x^I \varphi_J, \tag{10.15}$$

where summation (Einstein notation still holds!) occurs over all combinations of indices under I and J.

With the basis $e_1, ..., e_n$ and dual basis $\varepsilon^1, ..., \varepsilon^n$ fixed on V, we then have a bijective correspondence between (p,q)-tensors and multilinear forms $T^J_I x_I \varphi^J$ determined by the coefficients T^J_I; for every index i_k and j_ℓ there are n choices in total, therefore $n^p n^q = n^{p+q}$ numbers specify the multilinear form. Therefore it follows that $\dim T^q_p(V) = n^{p+q}$.

11. Transformation of tensors

Let $b_1, ..., b_n$ be a basis of V transformed from $e_1, ..., e_n$ by $b_k = c^i_k e_i = c^1_k e_1 + \cdots + c^n_k e_n$; the change-of-basis matrix (of coefficients) ${}_{\mathbf{b}}[I]_{\mathbf{e}}$ is then given by

$${}_{\mathbf{b}}[I]^{-1}_{\mathbf{e}} = {}_{\mathbf{e}}[I]_{\mathbf{b}} = (c^i_j)^j_i =: C, \tag{11.1}$$

namely, the coefficient column vector is transformed as $[v]_{\mathbf{e}} = C[v]_{\mathbf{b}}$. The dual basis vectors $\beta^1, ..., \beta^n$ are then transformed as $[\eta]^{\boldsymbol{\varepsilon}} = [\eta]^{\boldsymbol{\beta}} C^{-1}$, where the brackets denote row vectors [ADD EXPLANATION]. Let $B : V \times V^* \to \mathbb{R}$ be a bilinear mixed functional; then $B(x, \varphi) = [\varphi]^{\boldsymbol{\varepsilon}}[B]_{\mathbf{e}}[x]_{\mathbf{e}}$, but also $B(x, \varphi) = [\varphi]^{\boldsymbol{\beta}}[B]_{\mathbf{b}}[x]_{\mathbf{b}}$. Therefore

$$[\varphi]^{\boldsymbol{\varepsilon}}[B]_{\mathbf{e}}[x]_{\mathbf{e}} = [\varphi]^{\boldsymbol{\beta}} C^{-1}[B]_{\mathbf{e}} C[x]_{\mathbf{b}} = [\varphi]^{\boldsymbol{\beta}}[B]_{\mathbf{b}}[x]_{\mathbf{b}}, \tag{11.2}$$

which shows that $[B]_{\mathbf{b}} = C^{-1}[B]_{\mathbf{e}} C$. In Einstein summation notation, the coefficients $\tilde{b}^{i'}_{i'}$ of $[B]_{\mathbf{b}}$ are given by $\tilde{b}^{j'}_{i'} = c^i_{i'} c^{j'}_j b^j_i$ – note that $c^i_{i'}$ arises from the vector argument of B, whereas $c^{j'}_j$ from the covector argument.

Indeed, generalising for an arbitrary (p,q)-tensor T, we obtain

$$\begin{aligned} \tilde{T}^{j'_1\cdots j'_q}_{i'_1\ldots i'_p} &= T(b_1, ..., b_p, \beta^1, ..., \beta^q) \\ &= c^{i_1}_{i'_1} \cdots c^{i_p}_{i'_p} \cdot c^{j'_1}_{j_1} \cdots c^{j'_q}_{j_q} \cdot T^{j_1\ldots j_q}_{i_1\ldots i_p}, \end{aligned} \tag{11.3}$$

which is better expressed in contracted notation by $\tilde{T}^{\beta'}_{\alpha'} = c^\alpha_{\alpha'} c^{\beta'}_\beta T^\beta_\alpha$. This transformation allows us to write the 'physics definition' of tensors as follows.

11.1. **Theorem.** *A tensor of type (p,q) may uniquely be given by n^{p+q} numbers $T^{j_1\ldots j_q}_{i_1\ldots i_p}$ called tensor components that transform under change-of-basis transformations as given by Eq.* (11.3)*.*

As $T^q_p(V)$ is a vector space, two (p,q)-tensors may be added, whereby

$$(T + S)^\beta_\alpha = T^\beta_\alpha + S^\beta_\alpha. \tag{11.4}$$

The power of Theorem 11.1 is that it allows us to describe multiplication of tensors – this operation takes us out of the space! Namely, let T be a (p,q)-tensor and S be an (r,s)-tensor; then $T\otimes S$ is defined as the $(p+r,q+s)$-tensor given by

$$\begin{aligned}(T\otimes S)(x_1,...,x_{p+r},\varphi^1,...,\varphi^{q+s}) =& \\ = T(x_1,...,x_p,\varphi^1,...,\varphi^q)&S(x_{p+1},...,x_{p+r},\varphi^{p+1},...,\varphi^{p+s}),\end{aligned}\tag{11.5}$$

with the coefficients $(T\otimes S)^{j_1\cdots j_{q+s}}_{i_1\ldots i_{p+r}}$ given by

$$(T\otimes S)^{j_1\cdots j_{q+s}}_{i_1\ldots i_{p+r}} = T^{j_1\cdots j_q}_{i_1\ldots i_p}S^{j_{q+1}+\cdots j_{q+s}}_{i_{p+1}\ldots i_{p+r}},\tag{11.6}$$

or

$$(T\otimes S)^{\beta\nu}_{\alpha\mu} = T^\beta_\alpha S^\nu_\mu\tag{11.7}$$

in contracted notation $\alpha=(i_1,...,i_p),\mu=(i_{p+1},...,i_{p+r})$ and $\beta=(j_1,...,j_q),\nu=(j_{q+1},...,j_{q+s})$. The following algebraic properties immediately follow (proof is tedious and left to the reader).

11.2. **Lemma.** *Let $T,S\in T^q_p(V)$ and $R\in T^r_s(V)$. Then*

$$(T+S)\otimes R = T\otimes R + S\otimes R \qquad \text{(distributivity I)},\tag{11.8}$$

$$R\otimes(T+S) = R\otimes T + R\otimes S \qquad \text{(distributivity II)},\tag{11.9}$$

$$(T\otimes S)\otimes R = T\otimes(S\otimes R) \qquad \text{(associativity)}.\tag{11.10}$$

We call the algebra of tensors created by $+$ and $\otimes$ from $T^q_p(V): p,q\geq 0$ is called the *tensor algebra* of a vector space V and denoted by $T(V)$. Indeed, $T^q_p(V)\subset T(V)$ for any $p,q\geq 0$ with $T^0_0(V)=\mathbb{R}$; $k\otimes T = T\otimes k = kT$.

11.3. **Theorem.** *Tensor products of the form $\varepsilon^\alpha\otimes e_\beta$ for all α,β form a basis for $T^q_p(V)$, allowing us to write $T=T^\beta_\alpha\varepsilon^\alpha\otimes e_\beta$ (in Einstein notation).*

Proof. Let $T(x,\varphi)=T^\beta_\alpha x^\alpha\varphi_\beta\in T^q_p(V)$. We compute

$$(T^\beta_\alpha\varepsilon^\alpha\otimes e_\beta)(x_1,...,x_p,\varphi^1,...,\varphi^q) = T^\beta_\alpha x^\alpha\varphi_\beta,$$

as $\varepsilon^\alpha(x_1,...,x_p)=x_1^{i_1}\cdots x_p^{i_p}=x^\alpha$ and $e_\beta(\varphi^1,...,\varphi^q)=\varphi^1_{j_1}\cdots\varphi^q_{j_q}=\varphi_\beta$, completing the proof. $\square$

11.4. **Definition.** A *free vector space* of any[22] set S over $\mathbb{R}$ is the set of *formal* (finite **NB!**) linear combinations

$$\mathbb{R}S := \{\lambda_1 s_1+\cdots\lambda_n s_{n<\infty} : s_1,...,s_n\in S, n<\infty\}.\tag{11.11}$$

Each element $s\in S$ gives a (canonical) basis vector $1s$.

11.5. **Example.** Let $S=\{a,b,c\}$. Then $2a+3b\in\mathbb{R}S$. The space is isomorphic to $\mathbb{R}^3$ in the following way: $a\mapsto e_1$, $b\mapsto e_2$, $c\mapsto e_3$ and extended linearly.

Indeed, $E=\{e_1,...,e_n\}$ gives $\mathbb{R}E\cong\mathbb{R}^n$ by $E\ni e_i\mapsto e_i\in\mathbb{R}^n$, where in E we have that e_i is a 'letter', but in $\mathbb{R}^n$ it denotes a vector.

11.6. *Remark.* For a finite set $S=\{s_1,...,s_n\}$ we often write

$$\mathbb{R}S = \mathbb{R}s_1\oplus\mathbb{R}s_2\oplus\cdots\oplus\mathbb{R}s_n,\tag{11.12}$$

where we call $\oplus$ the *formal sum.*

[22]Really, any; it can be of your favourite types of apples or contain each toe of your feet (preferably, intact).

11.7. **Theorem** ($T(V)$ is a twice-graded algebra)**.** *The tensor algebra* $T(V)$ *is a* vector space *given by a direct (formal) sum of spaces of tensors as follows:*

$$T(V) = \bigoplus_{k=0}^{\infty} \bigoplus_{p+q=k} T_p^q(V) \subsetneq \prod_{k=0}^{\infty} \prod_{p+q=k} T_p^q(V), \tag{11.13}$$

that is, any element T *of the tensor algebra* $T(V)$ *may be written as a* finite[23] *linear combination of tensors*

$$T = T_{k_1} + \cdots + T_{k_m},$$

for $0 \le k_1 < \cdots < k_m$*, where*

$$T_k \in \bigoplus_{p+q=k} T_p^q(V)$$

is a linear combination of tensors of order k*.*

12. Tensor construction of the exterior algebra

Recall the group of permutations on $\{1, ..., p\}$ denoted by S_p. Let $\sigma \in S_p$ be a permutation; we can count how many elements have their positions switched by σ – this is called the *sign* of σ, denoted by $(-1)^\sigma$.

From now on we will only work with tensors on vectors (i.e. no covectors), namely, of type $(p, 0)$; by generalisation, we call these *multilinear functionals.* We begin by defining σT to be T with the inputs permuted by σ, namely,

$$(\sigma T)(x_1, ..., x_p) := T(x_{\sigma(1)}, ..., x_{\sigma(p)}). \tag{12.1}$$

We call $T \in T_p(V)$ *symmetric* if T is unchanged by permuting (switching around) its inputs

$$\sigma T = T, \tag{12.2}$$

or, equivalently, the coefficients are invariant under any permutation of indices:

$$T_{\sigma(i_1,...,i_p)} = T_{i_1,...,i_p} \tag{12.3}$$

for all $\sigma \in S_p$.

Similarly, we call T *skew-symmetric* if T changes sign upon switching inputs $T(x, y) = -T(x, y)$; of course, switching again or switching another pair yields $+$, hence, we define it formally as follows:

$$\sigma T = (-1)^\sigma T, \tag{12.4}$$

or, equivalently,

$$T_{\sigma(i_1,...,i_p)} = (-1)^\sigma T_{i_1,...,i_p} \tag{12.5}$$

for all $\sigma \in S_p$.

There is a 'natural' way to turn any tensor into a symmetric or skew-symmetric tensor. The reader is already familiar with the fact that the cardinality of S_p is $\#S_p = p! = p \cdot (p-1) \cdots 2 \cdot 1$; in the definitions below, in some sense, we take the 'average' over all permutations, but in the skew-symmetric case we additionally 'force' skewness by introducing the sign!

12.1. **Definition.** Let $T \in T_p(V)$. Define the *symmetrisation* $\operatorname{Sym} T$ and *alternation* $\operatorname{Alt} T$ of T to be the $(p, 0)$ tensor given by

$$\operatorname{Sym} T = \frac{1}{p!} \sum_{\sigma \in S_p} \sigma T \qquad \text{and} \qquad \operatorname{Alt} T = \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^\sigma \sigma T, \tag{12.6}$$

respectively. We denote the coefficients of $\operatorname{Alt} T$ by $\operatorname{Alt} T_{i_1 \ldots i_p}$ or $T_{[i_1 \ldots i_p]}$.

[23]This is why $T(V)$ is a *proper* subset of the product space.

It is clear that these mappings $T \mapsto \operatorname{Sym} T$ and $T \mapsto \operatorname{Alt} T$ are linear.

12.2. **Lemma.** *For any $T \in T_p(V)$ we have that $\operatorname{Sym} T$ is symmetric and $\operatorname{Alt} T$ is skew-symmetric.*

12.3. **Lemma.** $\operatorname{Alt} \circ \operatorname{Alt} = \operatorname{Alt}$ *and* $\operatorname{Alt} \circ \sigma = (-1)^\sigma \operatorname{Alt}$, $\sigma \circ \operatorname{Alt} = (-1)^\sigma \operatorname{Alt}$ *for any $\sigma \in S_p$.*

12.4. **Lemma** (Characterisation on skew-symmetry)**.** *Let $T \in T_p(V)$ be a multilinear functional. The following are then equivalent:*

(1) T is skew-symmetric;
(2) $\operatorname{Alt} T = T$;
(3) $T_{i_{\sigma(1)} \dots i_{\sigma(p)}} = (-1)^\sigma T_{i_1 \dots i_p}$ for any $\sigma \in S_p$;
(4) $T_{[i_1 \dots i_p]} = T_{i_1 \dots i_p}$;
(5) There exists a $B \in T_p(V)$ such that $T = \operatorname{Alt} B$.

12.5. **Theorem.** *We recover the* sign definition[24] *of the determinant:*

$$\det \begin{pmatrix} x_1^1 & \cdots & x_1^p \\ \vdots & & \vdots \\ x_p^1 & \cdots & x_p^p \end{pmatrix} = p!\, x_1^{[1} \cdots x_p^{p]} = p!\, x_{[1}^1 \cdots x_{p]}^p. \tag{12.7}$$

Proof. Indeed, for any $T \in T_p(V)$ we have

$$\begin{aligned} T_{[i_1 \dots i_p]} &= \operatorname{Alt} T(e_{i_1}, ..., e_{i_p}) \\ &= \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^\sigma (\sigma T)(e_{i_1}, ..., e_{i_p}) \\ &= \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^\sigma T(e_{\sigma(i_1)}, ..., e_{\sigma(i_p)}) \\ &= \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^\sigma T_{\sigma(i_1) \dots \sigma(i_p)}. \end{aligned}$$

Then,

$$\begin{aligned} p!\, x_1^{[1} \cdots x_p^{p]} &= \sum_{\sigma \in S_p} (-1)^\sigma x_1^{\sigma(1)} \cdots x_p^{\sigma(p)}, \\ p!\, x_{[1}^1 \cdots x_{p]}^p &= \sum_{\sigma \in S_p} (-1)^\sigma x_{\sigma(1)}^1 \cdots x_{\sigma(p)}^p, \end{aligned}$$

which matches the sign definition of determinant. □

We now have the necessary tools to construct the exterior algebra. Recall that the cross product of two vectors is anti-symmetric: $v \times w = -w \times v$; considering tensors, this requires skew-symmetry.

12.6. **Definition.** The *external product* $T \wedge S \in T_{p+q}(V)$ of two skew-symmetric multilinear functionals $T \in T_p(V)$ and $S \in T_q(V)$ is defined as[25]

$$T \wedge S = \frac{(p+q)!}{p!\, q!} \operatorname{Alt}(T \otimes S). \tag{12.8}$$

[24]This is NOT the definition we will ultimately adopt.

[25]It must be remarked that from this point onwards *we will differ from Postnikov*; there are two conventions for the wedge: one with the factorial factor and one without. Ultimately, there is one extremely important and useful formula formula we wish to reach, namely, Theorem 12.11, stating how to evaluate a wedge product; adopting Postnikov's convention, this formula will have a factorial factor and ruin the geometric interpretation and its use in integrals. It will similarly ruin defining a natural identification between external product tensors acting on wedges of vectors. For a contemporary discussion (by proper mathematicians), see `https://mathoverflow.net/questions/54343/is-there-a-preferable-convention-for-defining-the-wedge-product`.

12.7. **Proposition** (Associativity). *The external product is associative:*

$$(A \wedge B) \wedge C = A \wedge (B \wedge C). \tag{12.9}$$

Proof. For any $\sigma \in S_p$ define $\sigma' \in S_{p+q}$ to be σ on the first p numbers and the identity on the others; then $\sigma \mapsto \sigma'$ is clearly an injective group homomorphism that preserves the sign $(-1)^{\sigma'} = (-1)^{\sigma}$. For any $T \in T_{p+q}(V)$ define the alternator on the first p numbers to be

$$\operatorname{alt}_p T = \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^{\sigma} \sigma' T. \tag{12.10}$$

We quickly show $\operatorname{Alt}(\operatorname{alt} T) = \operatorname{Alt} T$; indeed, by linearity,

$$\begin{aligned}
\operatorname{Alt}(\operatorname{alt} T) &= \operatorname{Alt}\left(\frac{1}{p!} \sum_{\sigma \in S_p} (-1)^{\sigma} \sigma' T\right) = \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^{\sigma} \operatorname{Alt}(\sigma' T) \\
&= \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^{\sigma} (-1)^{\sigma'} \operatorname{Alt} T = \operatorname{Alt} T \cdot \frac{1}{p!} \sum_{\sigma \in S_p} 1 \\
&= \operatorname{Alt} T.
\end{aligned}$$

Now, let $A \in T_p(V)$, $B \in T_q(V)$, and $C \in T_r(V)$. Then

$$\begin{aligned}
(A \wedge B) \wedge C) &= \frac{(p+q+r)!}{(p+q)!\, r!} \operatorname{Alt}((A \wedge B) \otimes C) \\
&= \frac{(p+q+r)!}{(p+q)!\, r!} \operatorname{Alt}(\frac{(p+q)!}{p!\, q!} \operatorname{alt}_{p+q}(A \otimes B \otimes C)) \\
&= \frac{(p+q+r)!}{p!\, q!\, r!} \operatorname{Alt}(A \otimes B \otimes C)
\end{aligned}$$

and, similarly (we redefine alt^p for the last p elements),

$$A \wedge (B \wedge C) = \frac{(p+q+r)!}{p!\, q!\, r!} \operatorname{Alt}(A \otimes B \otimes C),$$

completing the proof. □

12.8. **Proposition** (Skew-commutativity). *Let $T \in T_p(V)$ and $S \in T_q(V)$. Then*

$$S \wedge T = (-1)^{pq}\, T \wedge S. \tag{12.11}$$

Proof. We compute

$$\begin{aligned}
(S \otimes T)(x_1, ..., x_{p+q}) &= S(x_1, ..., x_q) T(x_{q+1}, ..., x_{q+p}) \\
&= T(x_{q+1}, ..., x_{q+p}) S(x_1, ..., x_q) \\
&= (T \otimes S)(x_{q+1}, ..., x_{q+p}, x_1, ..., x_q) \\
&= \sigma(T \otimes S)(x_1, ..., x_q),
\end{aligned}$$

which gives

$$S \wedge T = \frac{(p+q)!}{p!\, q!} \operatorname{Alt}(A \otimes T) = \frac{(q+p)!}{q!\, p!} (-1)^{\sigma} \operatorname{Alt}(T \otimes S) = (-1)^{\sigma} T \wedge S,$$

where $(-1)^{\sigma} = (-1)^{pq}$ as p exchanges are required to move x_1 to the front, and similarly for x_2 to the 2nd position, and so on; as there are a total of q such elements to be moved, and each requires p flips, $(-q)^{\sigma} = ((-1)^p)^q = (-1)^{pq}$, completing the proof. □

Finally, distributivity

$$(A+B)\wedge C = A\wedge C + B\wedge C \tag{12.12}$$

is obvious (or left to the reader ☺).

Now, $\Lambda_p(V)\subseteq T_p(V)$, the set of all skew-symmetric multilinear functionals in $T_p(V)$, is a subspace (under $+$). In the cases $p=0,1$ we have no restriction imposed by skew-symmetry, hence

$$\Lambda_0(V) = T_0(V) = \mathbb{R}, \tag{12.13}$$

$$\Lambda_1(V) = T_1(V) = V^*. \tag{12.14}$$

Similarly as for $T(V)$, now $+$ and $\wedge$ make an algebra from each $\Lambda_p(V)$, which we call the *exterior algebra of the dual of* V, denoted by $\Lambda_*(V)$. Indeed, the 'dual' in the name is evident from the fact that $\Lambda_1(V) = V^*$, and all others are multilinear *functionals*.

Whereas $T(V)$ was a twice graded algebra – once from order $r=p+q$ and twice from all $T_p^q(V)$ with $p+q=r$, or, equivalently, by $p\geq 0$ and $q\geq 0$ –, the exterior algebra of the dual $\Lambda_*(V)$ is graded only by one parameter $p\geq 0$ in $\Lambda_p(V)$. Even worse than that, whereas $\dim T(V)=\infty$ as a formal vector space (cf. [ADD DEFINITITION]), instead $\dim\Lambda_*(V)<\infty$. The value is, in fact, determined by skew-symmetry interacting with linearly dependent vectors.

12.9. **Lemma.** *Let $x_p\in\operatorname{Span}\{x_1,...,x_{p-1}\}$. Then for any $T\in\Lambda_p(V)$ we have*

$$T(x_1,...,x_p) = 0. \tag{12.15}$$

Proof. First note that if (any) one of the vectors repeats, namely, $T(x,x,...)$, then by skew-symmetry $T(x,x,...) = -T(x,x,...)$ after exchanging $x\leftrightarrow x$, but this is the same expression, therefore $T(x,x,...)=0$. We then have, by multilinearity,

$$\begin{aligned} T(x_1,...,x_p) &= T(x_1,...,x_{p-1}, a^1x_1+\cdots+a^{p-1}x_{p-1}) \\ &= a^1T(x_1,...,x_{p-1},x_1)+\cdots+a^{p-1}T(x_1,...,x_{p-1},x_{p-1}) \\ &= 0 \end{aligned}$$

as the vectors repeat, completing the proof. □

12.10. **Theorem.** *Let $\dim V<\infty$. Then $\Lambda_*(V)$ is a finitely-graded algebra, namely,*

$$\Lambda_*(V) = \bigoplus_{p=0}^{\dim V}\Lambda_p(V). \tag{12.16}$$

Remark: Since the direct product is finite, it is isomorphic to the product space.

Proof. Let $n=\dim V<\infty$. Written out, the statement reads

$$\Lambda_*(V) = \mathbb{R}\oplus V^*\oplus\Lambda_2(V)\oplus\cdots\oplus\Lambda_n(V). \tag{12.17}$$

Now, in a space of dimension n there can be no more than n independent vectors; then for any $p>n$ we will have that at least one vector in the arguments of $T\in\Lambda_p(V)$ is a linear combination of the others, thus, by Lemma 12.9, $T(x_1,...,x_p)=0$ for every choice of $x_1,...x_p$. Therefore $T=0$ and $\Lambda_p(V)=\{0\}$ is the trivial space; note that $\{0\}\oplus W\cong\{0\}\times W\cong W$ for any space W, proving the theorem. □

We have already met one instance of the determinant; indeed, the determinant belongs completely to the field of multivariable analysis! The reader is likely familiar only with the algebraic aspects; indeed, there is a far more geometric aspect. We shall now show an extremely important, yet so simple, identity involving the determinant (defined by the 'usual' sign definition) that will lead to our beautiful definition of the determinant. This identity is the one used in computations: how to actually compute the value of the tensor product?

12.11. **Theorem.** *For any $\varphi^1, ..., \varphi^p \in V^*$ and any $x_1, ..., x_p \in V$ we have*

$$(\varphi^1 \wedge \cdots \wedge \varphi^p)(x_1, ..., x_p) = \det \begin{pmatrix} \varphi^1(x_1) & \cdots & \varphi^1(x_p) \\ \vdots & & \vdots \\ \varphi^p(x_1) & \cdots & \varphi^p(x_p) \end{pmatrix}. \tag{12.18}$$

Proof. We compute

$$\begin{aligned}(\varphi^1 \wedge \cdots \wedge \varphi^p)(x_1, ..., x_p) &= \frac{p!}{1! \cdots 1!}(\mathrm{Alt}(\varphi^1 \wedge \cdots \wedge \varphi^p))(x_1, ..., x_p) \\ &= p! \frac{1}{p!} \sum_{\sigma \in S_p} (-1)^\sigma (\varphi^1 \otimes \cdots \otimes \varphi^p)(x_1, ..., x_p) \\ &= \sum_{\sigma \in S_p} (-1)^\sigma \varphi^1(x_{\sigma(1)}) \cdots \varphi^p(x_{\sigma(p)}) \\ &= \det((\varphi^i(x_j))^i_j),\end{aligned}$$

where the last equality follows by Theorem 12.5. □

12.12. **Lemma.** *Let $\varphi^1, ..., \varphi^p \in V^*$. Then $\varphi^1 \wedge \cdots \wedge \varphi^p = 0$ if and only if $\varphi^1, ..., \varphi^p$ are linearly dependent.*

Proof. If $\varphi^p = a_i \varphi^i$, then, by distributivity and skew-commutativity,

$$\begin{aligned}\varphi^1 \wedge \cdots \wedge \varphi^p &= \varphi^1 \wedge \cdots \wedge \varphi^{p-1} \wedge (a_i \varphi^i) = a_i(\varphi_1 \wedge \cdots \wedge \varphi^{p-1} \wedge \varphi^i) \\ &= a_i(\varphi_1 \wedge \cdots \wedge \varphi^i \wedge \cdots \wedge \varphi^{p-1} \wedge \varphi^i) \\ &= (\pm)\, a_i(\varphi_1 \wedge \cdots \wedge \varphi^i \wedge \varphi^i \wedge \cdots \wedge \varphi^{p-1}) \\ &= 0.\end{aligned}$$

We now show the converse; suppose φ^i are all independent. Then they may be extended to a basis $\beta^1 = \varphi^1, ..., \beta^p = \varphi^p, \beta^{p+1}, ..., \beta^n$ of V^*; let the conjugate basis be $b_1, ..., b_n \in V$. Then

$$(\varphi^1 \wedge \cdots \wedge \varphi^p)(b_1, ..., b_p) = \det \begin{pmatrix} \varphi^1(b_1) & \cdots & \varphi^1(b_p) \\ \vdots & & \vdots \\ \varphi^p(b_1) & \cdots & \varphi^p(b_p) \end{pmatrix} = \det \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix} = 1,$$

meaning that $\varphi^1 \wedge \cdots \wedge \varphi^p \neq 0$, showing independence. □

12.13. **Theorem.** *Let $\varepsilon^1, ..., \varepsilon^n \in V^*$ be a dual basis for V. A basis for $\Lambda_p(V)$ is given by $\varepsilon^I := \varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_p}$, where $1 \leq i_1 \leq \cdots \leq i_p \leq n$. We use ε^I to denote a wedge of an increasing sequence. Additionally, the dimension is*

$$\dim \Lambda_p(V) = \binom{n}{p}. \tag{12.19}$$

Proof. By Lemma 12.12, products ε^I are independent. Let $T \in \Lambda_p(V)$ be any skew-symmetric tensor $T = \mathrm{Alt}\, T$; any tensor is of the form $T = T_{i_1 \ldots i_p} \varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_p}$. Then $T = T_{i_1 \ldots i_p} \varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_p}$, which equals zero upon repeated indices; therefore, the sum consists only of different indices $i_1, ..., i_p$. Indeed, by skew-commutativity, upon exchanging the order by $\sigma \in S_p$ only the sign changes as follows:

$$\begin{aligned} T &= \sum_{i_1<\cdots<i_p} \sum_{\sigma \in S_p} T_{i_{\sigma(1)}\dots i_{\sigma(p)}} \varepsilon^{i_{\sigma(1)}} \wedge \cdots \wedge \varepsilon^{i_{\sigma(p)}} \\ &= \sum_{i_1<\cdots<i_p} \sum_{\sigma \in S_p} (-1)^\sigma T_{i_1\dots i_p} (-1)^\sigma \varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_p} \\ &= p! \sum_{i_1<\cdots<i_p} T_{i_1\dots i_p} \varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_p} \\ &= p!\, T_I \varepsilon^I \end{aligned}$$

This means that $\Lambda_p(V)$ is spanned by ε^I for increasing I, which completes the proof. □

13. Determinant

??

Given a linear map $L : V \to V$, and a basis $e_1, ..., e_n$ of V we can express

$$L(v) = \varphi^i(v) e_i = \varphi^1(v) e_1 + \cdots + \varphi^n(v) e_n \tag{13.1}$$

for some $\varphi^1, ..., \varphi^n \in V^*$. This immediately suggests a relation between the (unique) number $\det L \in \mathbb{R}$ and the wedge-property of the determinant – Theorem 12.11. Namely, let $v_1, ..., v_n$ be completely arbitrary vectors in V, and consider the matrix $A = {}_{\mathbf{e}}[L]_{\mathbf{e}}$ associated with L and bases $\mathbf{e}$ for both the domain and codomain; we arrange the vector coefficients in the $\mathbf{e}$ basis into a matrix $B = [[v_1]_{\mathbf{e}}, ..., [v_n]_{\mathbf{e}}]$.

The notation $(\Lambda^n L)(v_1, ..., v_n) := (\varphi^1 \wedge \cdots \wedge \varphi^n)(v_1, ..., v_n)$ is very well chosen – indeed, we say that $\Lambda^n V$ is the *n-multilinear map associated with L*; namely: how does L change the volume of the n-parallelogram spanned by any n vectors? Recall that $[L(v_i)]_{\mathbf{e}}$ denotes the coefficients of the vector $L(v_i)$ in the $\mathbf{e}$ basis. Then one final computation reveals the magic:

$$\begin{aligned} (\Lambda^n L)(v_1, ..., v_n) &:= (\varphi^1 \wedge \cdots \wedge \varphi^n)(v_1, ..., v_n) \\ &= \det \begin{pmatrix} \varphi^1(x_1) & \cdots & \varphi^1(x_p) \\ \vdots & & \vdots \\ \varphi^p(x_1) & \cdots & \varphi^p(x_p) \end{pmatrix} \\ &= \det[[L(v_1)]_{\mathbf{e}}, ..., [L(v_n)]_{\mathbf{e}}] \\ &= \det[A[v_1]_{\mathbf{e}}, ..., A[v_n]_{\mathbf{e}}] \\ &= \det(AB) \\ &= \det A \cdot \det B \\ &= \det L \cdot \det[[v_1]_{\mathbf{e}}, ..., [v_n]_{\mathbf{e}}] \\ &= \det L \cdot (\varepsilon^1 \wedge \cdots \wedge \varepsilon^n)(v_1, ..., v_n) \\ &= \det L \cdot (\Lambda^n \,\mathrm{id})(v_1, ..., v_n), \end{aligned} \tag{13.2}$$

namely:

13.1. **Theorem** (Geometrical definition of the determinant). *The determinant* $\det L$ *of any linear map* $L : V \to V$ *is the* unique *constant with which L linearly changes the n-volume of any n-parallelogram, where* $n = \dim V$.

This immediately shows that if $\ker L \neq \{0\}$, namely, L sends some line to zero, or L is not invertible, then $\det L = 0$.

Already, this suggests that we should formalise n-parallelograms and maps on n-parallelograms in such a way that the fact that L is associated with a *unique* such map is encompassed. We have defined the exterior algebra as consisting of skew-symmetric tensors; the tensor is zero if the arguments are linearly dependent: why not encompass this same idea by 'wedging' vectors?

14. Exterior algebra

By some cruel irony, what we constructed in the previous section was in fact the *dual* exterior algebra ☺ – we have, in essence, defined linear functions on vectors before defining vectors. This is not all that crazy if we recall how subspaces could be defined by zeros of linear functionals; indeed, in these notes the concept of duality is King. His Queen is a pull-back.[26]

The two are closely linked, as we shall see.

As per the previous §§, we wish to look at k-dimensional parallelograms (also called parallelipipeds, parallelotopes, or blades) spanned by $v_1, ..., v_k \in V$ on which we can evaluate multilinear maps by Theorem 12.11; the dimension will, of course, equal k if and only if the vectors are independent. Let $\dim V = n$. We first observe the properties of $\Lambda_k(V)$ and $\Lambda_*(V)$ constructed previously:

1. $\Lambda_k(V), \Lambda_*(V)$ are vector spaces under $+$;
2. $\Lambda_*(V)$ is a skew-commutative ring under $+, \wedge$, additionally requiring:
   (a) $\Lambda_0(V) \cong \mathbb{R}$ and $1 \wedge T = T \wedge 1 = T$ for any $T \in \Lambda_*(V)$,
   (b) $(A \wedge B) \wedge C = A \wedge (B \wedge C)$ for all $A, B, C \in \Lambda_*(V)$,
   (c) $(A + B) \wedge C = (A \wedge C) + (B \wedge C)$ for all $A, B, C \in \Lambda_*(V)$,
   (d) $A \wedge B = (-1)^{pq} B \wedge A$ for all $A \in \Lambda_p(V)$, $B \in \Lambda_q(V)$;
3. $\varphi^1 \wedge \cdots \wedge \varphi^k \neq 0$ if and only if $\varphi^1, ..., \varphi^k \in V^*$ are independent;
4. $\Lambda_*(V) = \bigoplus_{k=0}^n \Lambda_k(V)$ as a direct product;
5. the elements $\varepsilon^I = \varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_k}$ form a basis of $\Lambda_k(V)$, where $I = i_1 \dots i_k$ is an increasing sequence from $1, ..., n$.

We are now in a position to form the definition along the lines of $\Lambda(V) \cong (\Lambda_*(V))^*$ and $\Lambda^k V \cong (\Lambda_k V)^*$. The intuition behind what follows is less that of tensors, but more of parallelograms: for two vectors $v_1, v_2 \in V$ we think of $v_1 \wedge v_2$ as the oriented parallelogram spanned by v_1 and v_2; indeed, we can identify this wedge with the cross product. Then wedges of more vectors correspond to higher dimensional oriented parallelograms (in three dimensions – prisms), with the sums being disjoint unions.

Indeed, thinking of all parallelograms is too general; for the purposes of integration we wish to consider parallelograms with equal volume and the same orientation as being equivalent – the wedge $v_1 \wedge v_2$ is such a parallelogram up to equivalence; we will show in a moment that this follows from the definition.

14.1. **Definition** (Exterior algebra). Let $\dim V = n < \infty$. The *exterior algebra* ΛV, $+$, $\wedge$ is a real *vector space* under $+$ that contains as subspaces V and $\mathbb{R} \not\subseteq V$ (formally distinct[27] from V), and is the direct sum

$$\Lambda V = \bigoplus_{k=0}^{n} \Lambda^k V, \tag{14.1}$$

where $\Lambda^0 V = \mathbb{R}$, $\Lambda^1 V = V$, and $\Lambda^k V = \mathrm{Span}\{v_1 \wedge \cdots \wedge v_k : v_1, ..., v_k \in V\}$ for $k \geq 2$, such that:

1. $1 \wedge x = x \wedge 1 = x$ for all $x \in \Lambda V$ (*unit*),
2. $(x \wedge y) \wedge z = x \wedge (y \wedge z)$ for all $x, y, z \in \Lambda V$ (*associativity*),

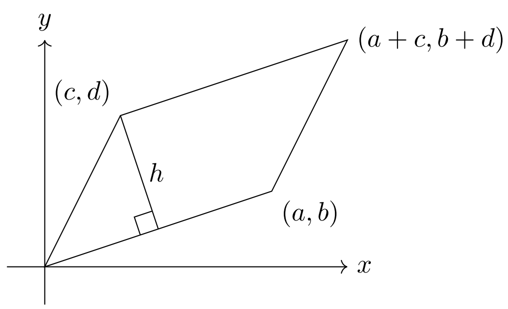
[26] If the reader objects to this order, I claim that it would have been foolish to begin the previous section with $(0, q)$ tensors instead – these tensors are in principle maps on functionals; we would be defining how to evaluate maps before defining the functions... Indeed, the essence is faulty.

[27] If needed, $\mathbb{R}\mathbf{1} = \mathrm{Span}\{\mathbf{1}\}$ with $\mathbf{1} \notin V$.

(3) $(x + ay) \wedge z = x \wedge z + a(y \wedge z)$ and $x \wedge (y + az) = x \wedge y + a(x \wedge z)$ for any $x, y, z \in \Lambda V$ and $a \in \mathbb{R}$ (*bilinearity/distributivity*),
(4) $v \wedge w = -w \wedge v$ for any vectors $v, w \in V$ (*antisymmetry*),
(5) $v_1 \wedge \cdots \wedge v_k \neq 0$ if $v_1, ..., v_k \in V$ are independent (*basis*).

14.2. *Remark.* The definition calls for formal remarks[28] and examples.

(1) For any $v \in V$ we have $v \wedge v = 0$: by antisymmetry $v \wedge v = -v \wedge v$, hence $2(v \wedge v) = 0$, which implies $v \wedge v = 0$.
(2) For dependent $v_1, ..., v_k \in V$ we have that some v_j may be expressed as a linear combination of the others; let $v_k = a^1 v_i := a^1 v_1 + \cdots + a^{k-1} v_{k_1}$. Then
$$v_1 \wedge \cdots \wedge v_k = v_1 \wedge \cdots \wedge v_{k-1} \wedge (a^i v_i) = a^1 (v_1 \wedge \cdots \wedge v_{k-1} \wedge v_i)$$
$$= \pm a^i (v_1 \wedge \cdots \wedge v_i \wedge v_i \wedge \cdots \wedge v_{k-1}) = \pm a^i \cdot 0 = 0.$$
Hence $v_1 \wedge \cdots \wedge v_k \neq 0$ if and only if $v_1, ..., v_k \in V$ are independent – the 'if' in the definition (5) implies 'if and only if'.
(3) Let us see an example for $V = \mathbb{R}^2$. We have that $\Lambda^0 V = \mathbb{R}\mathbf{1}$, where we indicate a span of $\mathbf{1}$ to emphasise that $\Lambda^0 V \not\subseteq V$; these are simply numbers, e.g. $1, \pi, 6$. Similarly, $\Lambda^1 V = \mathbb{R}^2$ consists of vectors in $\mathbb{R}^2$: $u = (a, b) = ae_1 + be_2$ (in the standard basis e_1, e_2). Now, the first interesting case is $\Lambda^2 V = \{u \wedge v : u, v \in \mathbb{R}^2\}$; we consider all possible wedges of basis vectors: $e_1 \wedge e_1$, $e_1 \wedge e_2$, $e_2 \wedge e_1$, $e_2 \wedge e_2$. Now, for equal vectors the wedges are zero, namely, $e_1 \wedge e_1 = e_2 \wedge e_2 = 0$, whereas for non-equal, however, linearly dependent: $e_2 \wedge e_1 = -e_1 \wedge e_2$. By bilinearity, any $u \wedge v$ will be a linear combination of wedges of basis vectors; we find only one independent wedge $e_1 \wedge e_2$, which forms the basis of $\Lambda^2 \mathbb{R}^2 = \operatorname{Span}\{e_1 \wedge e_2\}$, which implies that $\dim \Lambda^0 \mathbb{R}^2 = 1$, $\dim \Lambda^1 \mathbb{R}^2 = 2$, and $\dim \Lambda^2 \mathbb{R}^2 = 1$. These appear to be coefficients of Pascal's triangle...
An element in $\Lambda \mathbb{R}^2$ is of the form
$$a + (be_1 + ce_2) + d\, e_1 \wedge e_2,$$
where $a, b, c, d \in \mathbb{R}$ are constants.
(4) Indeed, for $V = \mathbb{R}^3$ we find the following independent wedges of basis vectors: $e_1 \wedge e_2, e_1 \wedge e_3, e_2 \wedge e_3 \in \Lambda^2 \mathbb{R}^3$ and $e_1 \wedge e_2 \wedge e_3 \in \Lambda^3 \mathbb{R}^3$. This gives the dimensions
$$\dim \Lambda^0 \mathbb{R}^3 = 1, \qquad \dim \Lambda^1 \mathbb{R}^3 = 3, \qquad \dim \Lambda^2 \mathbb{R}^3 = 3, \qquad \dim \Lambda^3 \mathbb{R}^3 = 1,$$
which again are coefficients in Pascal's triangle.
(5) We explore an example of the 'parallelogram up to equal area and orientation' argument: consider the space $V = \mathbb{R}^2$ and vectors $v_1 = ae_1 + be_2$ and $v_2 = ce_1 + de_2$ with e_1, e_2 being the standard basis vectors; the parallelogram spanned by these vectors is given below.



[28]As in, more explicit than endless yapping in the main text...

The area of this parallelogram may be computed geometrically as base times height: the base is given by $(0,0)-(a,b)$ with length $\sqrt{a^2+b^2}$; we find the height from its vector: h is perpendicular to $(0,0)-(a,b)$, thus is oriented by the unit vector $\hat{n} = n/||n||$, where n is given by

$$n = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} v_1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -b \\ a \end{pmatrix},$$

namely, $\hat{n} = n/\sqrt{a^2+b^2}$. Now, $v_2 + \alpha n$ must intersect the base, namely,

$$\begin{pmatrix} c \\ d \end{pmatrix} + \alpha \begin{pmatrix} -b \\ a \end{pmatrix} = \beta \begin{pmatrix} a \\ b \end{pmatrix}$$

or, rewriting to matrix form,

$$\begin{pmatrix} b & a \\ -a & b \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}$$

which has a solution by taking the inverse of the matrix

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{a^2+b^2} \begin{pmatrix} b & -a \\ a & b \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix},$$

namely, $\alpha = (bc-ad)/(a^2+b^2)$, hence $h = ||\alpha n|| = |\alpha|\sqrt{a^2+b^2}$. Therefore, the area is

$$\text{Area} = \sqrt{a^2+b^2} \cdot h = \sqrt{a^2+b^2}\frac{|bc-ad|}{\sqrt{a^2+b^2}} = |bc-ad| = |ad-bc|.$$

This should remind the reader of the determinant.

Now, instead, we compute the wedge

$$\begin{aligned} v_1 \wedge v_2 &= (ae_1 + be_2) \wedge (ce_2 + de_2) \\ &= ae_1 \wedge (ce_1 + de_2) + be_2 \wedge (ce_1 + de_2) \\ &= ac(e_1 \wedge e_1) + ad(e_1 \wedge e_2) + bc(e_2 \wedge e_1) + bd(e_2 \wedge e_2) \\ &= 0 + ad(e_1 \wedge e_2) - bd(e_1 \wedge e_2) + 0 \\ &= (ad - bd)(e_1 \wedge e_2). \end{aligned}$$

Therefore, we see that the coefficient in front of the wedge is the signed area! Indeed, we shall see that area evaluates as the determinant of the parametrisation (cf. Theorems 12.11 and 14.4) – in linear algebra it was the formula $A = |\det[v_1, v_2]| = |ad - bd|$, which we show in this course.

Needless to say, this linear algebra method allows for far easier computation of areas than via Euclidean geometry.

(6) We can identify the cross product in $V = \mathbb{R}^3$, namely, $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, with a wedge of $u, v \in \mathbb{R}^3$ by $u \times v = \Phi(u \wedge v)$, where $\Phi : \Lambda^2\mathbb{R}^3 \to \mathbb{R}^3$ is given by

$$\Phi(e_1 \wedge e_2) = e_1, \qquad \Phi(e_1 \wedge e_3) = -e_2, \qquad \Phi(e_2 \wedge e_3) = e_3,$$

and extended linearly (we will see below that the three wedge products constitute a basis for $\Lambda^2\mathbb{R}^3$). Let $u = (u^1, u^2, u^3)$ and $v = (v^1, v^2, v^3)$. Indeed, a computation shows

$$u \times v = \begin{vmatrix} e_1 & e_2 & e_3 \\ u^1 & u^2 & u^3 \\ v^1 & v^2 & v^3 \end{vmatrix} = (u^2v^3 - u^3v^2)e_1 - (u^1v^3 - u^3v^1)e_2 + (u^1v^2 - u^2v^1)e_3,$$

$$u \wedge v = (u^2v^3 - u^3v^2)e_1 \wedge e_2 - (u^1v^3 - u^3v^1)e_1 \wedge e_3 + (u^1v^2 - u^2v^1)e_2 \wedge e_3.$$

(7) The reader may have seen to $\text{Area}(v_1, v_2) = |\det[v_1, v_2]|$ a similar expression $\text{Vol}(v_1, v_2, v_3) = |\det[v_1, v_2, v_3]|$ for the volume of a prism spanned by 3 vectors in $\mathbb{R}^3$. Indeed, this is equal to the triple product $\text{Vol} = |v_1 \cdot (v_2 \times v_3)|$. Notably, however,

$$v_1 \wedge v_2 \wedge v_3 = \det[v_1, v_2, v_3] e_1 \wedge e_2 \wedge e_3,$$

and thus the constant in front of $e_1 \wedge e_2 \wedge e_3$ is the signed volume.

This immediately generalised to any dimension: the signed n-volume $\text{Vol}_n(v_1 \wedge \cdots \wedge v_n)$ of an n-parallelogram spanned by n vectors $v_1, ..., v_n$ in $\mathbb{R}^n$ is the unique constant that satisfies

$$v_1 \wedge \cdots \wedge v_n = \text{Vol}_n(v_1 \wedge \cdots \wedge v_n) e_1 \wedge \cdots \wedge e_n,$$

alternatively: $\text{Vol}_n(v_1 \wedge \cdots \wedge v_n) = \det[v_1, ..., v_n]$.

(8) A good shorthand notation is the use the multi-index $e_1 \wedge e_2 = e_{12}$, and, similarly, $e_1 \wedge \cdots \wedge e_k = e_{1...k}$ and $\varepsilon^1 \wedge \cdots \wedge \varepsilon^k = \varepsilon^{1...k}$, or, for some arbitrary increasing sequence $I = (i_1, ..., i_k) : 1 \leq i_1 \leq \cdots \leq i_k \leq n$ we write $\varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_k} = \varepsilon^I$.

14.3. **Lemma** (Basis). *Given a basis $e_1, ..., e_n$ of V, we have that $e^I = e^{i_1} \wedge \cdots \wedge e^{i_k}$ is a basis for $\Lambda^k V$, where I is an increasing sequence $1 \leq i_1 < \cdots < i_k \leq n$. Additionally,*

$$\dim \Lambda^k V = \binom{n}{k}. \tag{14.2}$$

Proof. Since any $x \in \Lambda^k V$ is a linear combination of terms $v_1 \wedge \cdots \wedge v_k$, we have by bilinearity,

$$\begin{aligned} v_1 \wedge \cdots \wedge v_k &= (v_1^{i_1} e_{i_1}) \wedge \cdots \wedge (v_k^{i_k} e_{i_k}) \\ &= v_1^{i_1} \cdots v_k^{i_k} e_{i_1} \wedge \cdots \wedge e_{i_k} \\ &= (-1)^\sigma v^I e_I, \end{aligned}$$

where $i_1, ..., i_k$ may be reordered by $\sigma \in S_k$ to an increasing I with sign $(-1)^\sigma$, completing the basis proof. Combinatorially, the number of ways to pick increasing k-tuples from $\{1, ..., n\}$ is the same as picking k distinct elements, which is $\binom{n}{k}$ by definition. $\square$

14.4. **Theorem** (Evaluation). *Canonically, $\Lambda^k(V^*) \cong \Lambda_k(V) \cong (\Lambda^k V)^*$ by*

$$(\varphi^1 \wedge \cdots \wedge \varphi^k)(v_1 \wedge \cdots \wedge v_k) = (\varphi^1 \wedge \cdots \wedge \varphi^k)(v_1, ..., v_k) = \det \begin{pmatrix} \varphi^1(v_1) & \cdots & \varphi^1(v_p) \\ \vdots & & \vdots \\ \varphi^p(v_1) & \cdots & \varphi^p(v_p) \end{pmatrix}$$

extended linearly is an isomorphism.

Proof. Let $e_1, ..., e_n$ be a basis of V with dual basis $\varepsilon^1, ..., \varepsilon^n \in V^*$. By Lemma 14.3, we have the basis e_J for $\Lambda^k V$ and ε^I for $\Lambda^k(V^*)$. Then let $\mathscr{I} : \Lambda^k(V^*) \to (\Lambda^k V)^*$ be given by

$$\mathscr{I}(\varphi^1 \wedge \cdots \wedge \varphi^k)(v_1 \wedge \cdots \wedge v_k) = \det \begin{pmatrix} \varphi^1(v_1) & \cdots & \varphi^1(v_p) \\ \vdots & & \vdots \\ \varphi^p(v_1) & \cdots & \varphi^p(v_p), \end{pmatrix}$$

hence

$$\mathscr{I}(\varepsilon^I)(e_J) = \begin{cases} 1 & \text{if} \quad I = J \\ 0 & \text{if} \quad I \neq J \end{cases} = \delta^I_J,$$

which shows that $\mathscr{I}(\varepsilon^I)$ is a dual basis to e_J, hence $(\Lambda^k V)^* \cong \Lambda^k(V)$. The middle isomorphism is revealed by Theorem 12.11. $\square$

Due to the natural isomorphism between $\Lambda^k(V^*)$ and $(\Lambda^k V)^*$, we define an evaluation of $\varphi^1 \wedge \cdots \wedge \varphi^k$ on $v_1 \wedge \cdots \wedge v_k$ as above.

We may define how a linear map acts on k-parallelograms.

14.5. **Definition.** Let $L : V \to W$ be linear; define the induced map $\Lambda^k L : \Lambda^k V \to \Lambda^k W$ to be the linear map given by

$$(\Lambda^k L)(v_1 \wedge \cdots \wedge v_k) = (Lv_1) \wedge \cdots \wedge (Lv_k) \tag{14.3}$$

and extended linearly

$$(\Lambda^k L)(ax + by) = a(\Lambda^k L)(x) + b(\Lambda^k L)(y). \tag{14.4}$$

We let $\Lambda^0 L(1) = 1$.

14.6. **Example.** Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $Le_1 = ae_1 + ce_2$ and $Le_2 = be_1 + de{+}2$, namely, represented by matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then $\Lambda^0 L = \mathrm{id}_{\mathbb{R}}$, $\Lambda^1 L = L$, and

$$\begin{aligned}(\Lambda^2 L)(e_1 \wedge e_2) &= (Le_1) \wedge (Le_2) = (ae_1 + ce_2) \wedge (be_1 + de_2) \\ &= (ab - cd)e_1 \wedge e_2,\end{aligned}$$

hence we see that $\Lambda^2 L = (\det L)\,\mathrm{id}_{\Lambda^2\mathbb{R}^2}$. This is not a coincidence; indeed, recall the construction in § 13 – any linear map may be decomposed into functionals, which makes $\Lambda^n L$ an n-wedge of functionals that evaluates as a determinant.

This shows that the sign definition of the determinant agrees with the geometrical definition below. Recall that for $\dim V = n$ we have $\dim \lambda^n V = \binom{n}{n} = 1$, and the only linear map from a 1-dimensional space into itself is scalar multiplication by a constant.

14.7. **Definition.** Let $L : V \to V$ be linear, and denote $\dim V = n$; the *determinant* of L is defined as the unique real number satisfying

$$\Lambda^n L = (\det L)\,\mathrm{id}_{\Lambda^n V}\,. \tag{14.5}$$

We define the determinant of the matrix $A \in \mathbb{R}^{n\times n}$ to be $\det A = \det L_A$, where $L_A(v) = Av$.

15. Pull backs of k-forms

Given some $\omega \in \Lambda^k(W^*)$, we want to define an analogous form on $\Lambda^k(V^*)$ that is related to ω through a linear map $L : V \to W$. Recall that L defines the pull-back map $L^* : W^* \to V^*$ given by $(L^*\varphi)(v) = (\varphi \circ L)(v) = \varphi(L(v))$, which induces $\Lambda^k(L^*) : \Lambda^k(W^*) \to \Lambda^k(V^*)$ given by $\Lambda^k(L^*)(\varphi^1 \wedge \cdots \wedge \varphi^k) = L^*\varphi^1 \wedge \cdots \wedge L^*\varphi^k$. The diagram below illustrates the pull-back definition for any $\omega \in \Lambda^k(W^*)$; all arrows represent linear maps.

$$
\begin{tikzcd}
(v_1, ..., v_k) \arrow[d, mapsto] \arrow[rrr, mapsto, bend left, "L\times\cdots\times L"] & V^k \arrow[r, "L\times\cdots\times L"] \arrow[d, two heads, "\pi"'] & W^k \arrow[d, two heads, "\pi'"] & (Lv_1, ..., Lv_k) \arrow[d, mapsto] \\
v_1 \wedge \cdots \wedge v_k \arrow[rrrd, mapsto, bend right, "L^*\omega = \Lambda^k(L^*)(\omega)"'] & \Lambda^k V \arrow[r, "\Lambda^k L"] \arrow[rd, dashed, "L^*\omega"'] & \Lambda^k W \arrow[d, "\omega"] & Lv_1 \wedge \cdots \wedge Lv_k \arrow[d, mapsto] \\
& & \mathbb{R} & \omega(Lv_1 \wedge \cdots \wedge Lv_k)
\end{tikzcd}
$$

Indeed, in principle, only the bottom triangle is needed to define the pull-back; all arrows are linear maps; the rest defines[29] the construction of $\Lambda^k L$.

15.1. **Definition.** Let $L : V \to W$ be a linear map. The pull-back of $\omega \in \Lambda^k(W^*)$ along L is the form in $\Lambda^k(V^*)$ given by $L^*\omega := \Lambda^k(L^*)(\omega) = \omega \circ \Lambda^k L$, namely, for any k-form $\omega = \varphi^1 \wedge \cdots \wedge \varphi^k$ we have

$$
\begin{aligned}
L^*(\varphi^1 \wedge \cdots \wedge \varphi^k) &= L^*\varphi^1 \wedge \cdots \wedge L^*\varphi^k \\
&= (\varphi^1 \circ L) \wedge \cdots \wedge (\varphi^k \circ L).
\end{aligned}
\tag{15.1}
$$

15.2. **Lemma.** *Let $L : V \to W$ be linear, $\varphi^1 \wedge \cdots \wedge \varphi^k \in \Lambda^k(W^*)$ and $v_1, ..., v_k \in V$. We have*

$$
L^*(\varphi^1 \wedge \cdots \wedge \varphi^k)(v_1 \wedge \cdots \wedge v_k) = \det \begin{pmatrix} \varphi^1(Lv_1) & \cdots & \varphi^1(Lv_k) \\ \vdots & & \vdots \\ \varphi^k(Lv_1) & \cdots & \varphi^k(Lv_k) \end{pmatrix}.
$$

Proof. The statement follows by the evaluation of k-forms (Theorem 14.4). □

15.3. **Lemma** (Properties of the pull-back of forms)**.** *Let $V \xrightarrow{L} W \xrightarrow{M} U$ be linear. Let ω, α be k-forms on W, let η be an ℓ-form on W; namely, $\omega, \alpha \in \Lambda^k(W^*)$, $\eta \in \Lambda^\ell(W^*)$. Let $r, s \in \mathbb{R}$. Then*

(1) $L^*(r\alpha + s\omega) = rL^*\alpha + sL^*\omega$ (pull-back is linear)*;*
(2) $L^*(\eta \wedge \omega) = (L^*\eta) \wedge (L^*\omega)$ (each component is pulled independently)*;*
(3) $L^*M^* = (M \circ L)^*$ (composition of pull-backs is a pull-back of the composition)*.*

Proof. (1). We have that $L^* = \Lambda^k(L^*)$ is linear.

(2). Let $\omega = \omega^1 \wedge \cdots \wedge \omega^k$ and $\eta = \eta^1 \wedge \cdots \wedge \eta^\ell$. We have

$$
\begin{aligned}
L^*(\eta \wedge \omega)(v_1 \wedge \cdots \wedge v_{k+\ell}) &= (\eta \wedge \omega)(Lv_1 \wedge \cdots \wedge Lv_{k+\ell}) \\
&= (\eta^1 \wedge \cdots \wedge \eta^\ell \wedge \omega^1 \wedge \cdots \wedge \omega^k)(Lv_1 \wedge \cdots \wedge Lv_{k+\ell}) \\
&= (L^*\eta^1 \wedge \cdots \wedge L^*\eta^\ell \wedge L^*\omega^1 \wedge \cdots \wedge L^*\omega^k)(v_1 \wedge \cdots \wedge v_{k+\ell}) \\
&= ((L^*\eta) \wedge (L^*\omega))(v_1, \wedge \cdots \wedge v_{k+\ell}),
\end{aligned}
$$

proving the statement.

[29] I decided to include it because we want to make everything dependent on the linear map L; that is, $\Lambda^k L$ is not some arbitrary linear map, but is uniquely determined by L!

(3). Let $\mathbf{b} = (b_J)_{\#J=k}$, for $J \subseteq \{1, ..., n\}$ increasing, be a basis of $\Lambda^k V$. Then

$$\begin{aligned}\Lambda^k M \circ \Lambda^k L(b_J) &= \Lambda^k M(Lb_{J_1} \wedge \cdots \wedge Lb_{J_k}) \\ &= (M \circ L)b_{J_1} \wedge \cdots \wedge (M \circ L)b_{J_k} \\ &= \Lambda^k (M \circ L)b_J.\end{aligned}$$

Then for any $\theta \in \Lambda^k(U)$ we have

$$L^* M^* \theta = L^*(\theta \circ \Lambda^k M) = \theta \circ \Lambda^k M \circ \Lambda^k L = \theta \circ \Lambda^k (M \circ L) = (M \circ L)^* \theta.$$

□

Part 4. Differentiation and integration of forms

Essentially in this section we genralise the fundamental theorem of calculus to higher dimensions; indeed, the reader has already met many of these in calculus: Green's, Stokes', and Gauss' divergence theorems are all examples of a more general statement.

Recall the fundamental theorem:

(1) firstly, for any f there exists a function F 'primitive' such that f may be written as the derivative of F, namely, $f = F'$;
(2) secondly, integrating f' from a to b yields $f(b) - f(a)$, namely, f evaluated on the boundary of $[a, b]$

The fundamental theorem of multivariable analysis will be:

(1) firstly, Poincar e lemma guarantees the existence of a primitive under a condition:
(2) secondly, Stokes' theorem states that evaluation of a 'derivative' is the same as evaluating the primitive along the boundary.

16. Forms

What is a 'derivative', really? We know of three examples already; we give them in a specific order – can you guess what it is? –:

(1) gradient,
(2) curl, and
(3) divergence.

Recall that for $f : \mathbb{R} \to \mathbb{R}$ we call f' the first derivative, and the function f itself – the zeroth derivative. We want to call k-th order derivatives 'k-forms'; as such, we can say that functions $f : P \subseteq V \to \mathbb{R}$ are 0-forms, and their derivatives $f' : P \to \mathrm{Hom}(V, \mathbb{R})$ are 1-forms.

(For completeness, in the preceding text, let V be a finite-dimensional real vector space, and $P \subseteq V$ be an open subset (with respect to a norm).)

For now consider $V = \mathbb{R}^3$. Then $f' = \nabla f$. We know that a curl takes a vector-valued function, namely, $F : P \to V$ and returns $\nabla \times F$, which we have seen for exterior algebra, is a wedge of two functionals: a '2-form'. Recall likewise the divergence $\nabla \cdot F$ that returns a scalar; but the dimension of $\Lambda^3(\mathbb{R}^3) = 1$, meaning that any 3-form is just a scalar multiple of $dx \wedge dy \wedge dz$ (or $\varepsilon^{123} = \varepsilon^1 \wedge \varepsilon^2 \wedge \varepsilon^3$ in different notation); the divergence at any point considers all three partials, thus can be considered to map into '3-forms'.

Let us adopt $\mathrm{Grad}\, f$, $\mathrm{Curl}\, f$, and $\mathrm{Div}\, f$ instead of the notation with nabla. The exact domains and codomains we do not yet formalise; this is still a motivating discussion. Gradient sends 0-forms into 1-forms, curl 1-forms into 2-forms, and, finally, divergence – 2-forms into 3-forms. We know that $\mathrm{Curl}\,\mathrm{Grad}\, f = 0$ and $\mathrm{Div}\,\mathrm{Curl}\, F = 0$; or, image of Grad lives in the kernel of Curl, i.e. $\operatorname{im}\mathrm{Grad} \subseteq \ker \mathrm{Curl}$ and $\operatorname{im}\mathrm{Curl} \subseteq \ker \mathrm{Div}$.

Whenever $\operatorname{im}\varphi \subseteq \ker\psi$, we denote this fact with the diagram

$$0 \to X \xrightarrow{\varphi} Y \xrightarrow{\psi} Z \to 0,$$

called a *chain (or co-chain) complex.* If $\operatorname{im}\varphi = \ker\psi$ we call it can exact.

Denoting[30] k-forms ($k = 0, 1, 2, 3$) by $\Omega^k(P)$, we thus obtain a chain complex[31]

$$0 \longrightarrow \Omega^0(P) \xrightarrow{\mathrm{Grad}} \Omega^1(P) \xrightarrow{\mathrm{Curl}} \Omega^2(P) \xrightarrow{\mathrm{Div}} \Omega^3(P) \longrightarrow 0. \tag{16.1}$$

[30]Cf. definition below.

[31]Note that $\Omega^{n\geq 4}(P) = 0$ (trivial vector space), hence we have the zero at the end.

Crucially, we see that we want to define a derivative on k-forms as something that takes k-forms and returns $(k+1)$-forms, whereby the usual differentiation operation is carried out on components. And we want this derivative to satisfy the diagram in (16.1); namely, denoting the derivative by d, we want $d \circ d = 0$.

We have just seen that 0-forms are functions from P to $\mathbb{R}$, 1-forms are maps from P to $\mathrm{Hom}(V, \mathbb{R}) = V^* = \Lambda^1(V^*)$, 2-forms are maps from P to 2-wedges of functionals $\Lambda^2(V^*)$; and so k-forms are $P \to \Lambda^k(V^*)$, as formalised below.

16.1. **Definition** (k-form)**.** Let V be a finite-dimensional vector space, and P be an open subset with respect to a norm. A (differential) k-form ω on $P \subseteq V$ is a map $\omega : P \to \Lambda^k(V^*)$; we denote the set of all k-forms on P by $\Omega^k(P)$.

Sometimes we wish to only consider forms that are continuous, denoted $\Omega^k_1(P)$; in case of continuous higher-order derivatives, j-times continuously differentiable – $\Omega^k_j(P)$.

We define multiplication by functions $f : P \to \mathbb{R}$ and wedges of forms pointwise and linearly:

$$(\omega + f\eta)(p) = \omega(p) + f(p)\eta(p), \qquad (\omega \wedge \eta)(p) = \omega(p) \wedge \eta(p). \tag{16.2}$$

16.2. **Example.** Some examples:

(1) Take $V = \mathbb{R}^2$; then $\omega(x, y) = x\varepsilon^1$ is a 1-form and $\omega(x, y) = xy\varepsilon^1 \wedge \varepsilon^2$ is a 2-form.
(2) Take $V = \mathbb{R}^3$; then $\omega(x, y, z) = xy(z + 1)\varepsilon^{12} + y\varepsilon^{13}$ is a 2-form, and $\omega(x, y, z) = f(x, y, z)\varepsilon^{123}$ characterises all 3-forms as $\dim \Lambda^3(\mathbb{R}^3) = 1$.
(3) [MORE INTERESTING EXAMPLES? SOMETHING FROM GEOMETRY?]

We will define the derivative of k-forms in terms of a basis; there is an alternative coordinate-free definition, however, it involves integration.

Recall that $\Lambda^k(V^*)$ has a basis $\mathfrak{B} = (\beta^I)$, for I increasing and of length $\#I = k$, and each β^i being the dual basis vector to b_i for basis $b_1, ..., b_n$ of V. Then a function $\omega : P \to \Lambda^k(V^*)$ can be decomposed into basis functions $\omega_I : P \to \mathbb{R}$, namely,

$$\omega(p) = \sum_I \omega_I(p)\beta^I. \tag{16.3}$$

In the Einstein convention, we will drop the sum symbol.

As per the usual spiel we will soon define pull-backs, show linearity and evaluating wedges, $d \circ d = 0$, et cetera.

16.3. **Definition** (Exterior derivative)**.** Let $\omega = f^I\beta^I$ be a k-form on $P \subseteq V$. We define the *exterior derivative* $d\omega : P \to \Lambda^{k+1}(V^*)$ to be the $(k + 1)$-form differentiated on basis vectors:

$$d(f_I\beta^I) := df_I \wedge \beta^I, \tag{16.4}$$

where $df_I : P \to V^*$ is the 1-form given by $df_I = f_I'$.

Note that for 0-forms, or, functions $f : P \to \mathbb{R}$, we have $df(p) = f'(p)$.

16.4. *Remark* (Forms and level sets.). [REWRITE BETTER LATER[32]]

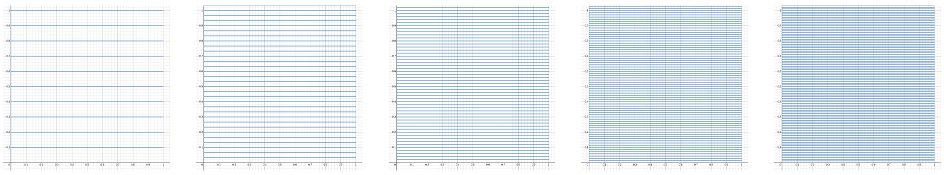
Wedges $\varphi^1 \wedge \cdots \wedge \varphi^k$ may be described by the intersection of level sets of $\varphi^1, ..., \varphi^k$; now, a k-form assigns a different such wedge at each point, meaning that these level sets change with it. For the form to not be identically zero, by adding another functional, the level sets decrease in dimension by 1: for a single dual vector the level set is $n - 1$ dimensional, for two – $n - 2$, and so on. Hence the level sets of $d\omega$ are one dimension lower than those of ω.

[32] `https://mathoverflow.net/questions/21024/what-is-the-exterior-derivative-intuitively`

Consider $V = \mathbb{R}^2$ and a 1-form $\omega(p) = p_1\varepsilon^2$, where $p_1 = \varepsilon^1(p)$; the form changes in the e_1 direction, but acts only on the e_2 component of $v = (x, y)$. The level set

$$\omega(p)^{-1}(\{c\}) = \{(x, y) \in V : p_1 y = c\} = \{(x, c/p_1) : x \in \mathbb{R}\}$$

is parallel to e_1 and as we move closer to $p_1 = 0$ we find fewer level lines, as illustrated below:



$p_1 = 0.1$ $\quad$ $p_1 = 0.3$ $\quad$ $p_1 = 0.5$ $\quad$ $p_1 = 0.7$ $\quad$ $p_1 = 0.9$

Namely, in each figure p_1 is fixed and c changes in equal increments. We observe that the number of lines increases linearly as we move away from $(0, 0)$.

We find $(d\omega)(p) = ((d\varepsilon^1) \wedge \varepsilon^2)(p) = \varepsilon^1 \wedge \varepsilon^2$ as ε^1 is linear, thus its own derivative.

We see that $d\omega$ is linear and has level sets

$$d\omega(p)^{-1}(\{c\}) = \{(x, y) \in V : xy = c\}$$

we obtain hyperbolae. These describe the intersection between a level set of ϵ^2 and $\epsilon^1(p)$ – how many level sets are created where; namely, as you move p, how many lines must *start* there. Indeed, this is true for any $\omega(p) = f(p)\eta$, where η does not depend on p.

Conversely, let $\omega(p) = \varepsilon^1(p)\varepsilon^1$. Then the level sets are vertical, and only depend on p_1 again. We have $d\omega(p) = \varepsilon^1 \wedge \varepsilon^1 = 0$; [...]

16.5. **Lemma.** *Let $\alpha, \omega : Q \to \Lambda^k(W^*)$ be k-forms and η be an ℓ-form on Q, $s, r \in \mathbb{R}$ be numbers, and $f : Q \to \mathbb{R}$ be differentiable. Then*

(1) $d(s\alpha + r\omega) = sd\alpha + rd\omega$ (linearity of d)*;*
(2) $d(f\omega) = (df) \wedge \omega + fd\omega$ (product rule 1)*;*
(3) $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k\omega \wedge d\eta$ (product rule 2).

Proof. (1). Writing out both forms in terms of the basis, we obtain linear combinations in each basis vector $(s\alpha + r\omega_I)\beta^I$, which becomes $(s\alpha'_I + r\omega'_I) \wedge \beta^I$ by linearity of the ordinary derivative.

(2). We have the product (also called Leibniz) rule for the ordinary derivative $(fg)'(p)(v) = f'(p)(v)g(p) + f(p)g'(p)(v)$, or, compactly, $d(fg) = (df)g + fdg$. Letting $\omega = g_I\beta^I$, in Einstein notation, we have

$$d(f\omega) = d(fg_I\beta^I) = d(fg_I) \wedge \beta^I = ((df)g_I + fdg_I) \wedge \beta^I = (df)\omega + fd\omega.$$

(3). For β^I we have $d\beta^I = 0$. Then for $\omega = f_I\beta^I$ and $\eta = g_J\beta^J$ we have that, by part 2,

$$\begin{aligned} d(f_I\beta^I \wedge g_J\beta^J) &= d(f_Ig_J\beta^{IJ}) = d(f_Ig_J) \wedge \beta^{IJ} + f_Ig_Jd\beta^{IJ} = d(f_Ig_J) \wedge \beta^{IJ} \\ &= (df_Ig_J + f_Idg_J) \wedge \beta^{IJ} = df_Ig_J \wedge \beta^I \wedge \beta^J + f_Idg_J \wedge \beta^I \wedge \beta^J \\ &= (df_I \wedge \beta^I) \wedge (g_J\beta^J) + f_Idg_J \wedge \beta^I \wedge \beta^J \\ &= d\omega \wedge \eta + f_Idg_J \wedge \beta^I \wedge \beta^J \end{aligned}$$

Now, in the second term we need to move the 1-form dg_J (which is a linear combination of β^i functionals) through β^I; by anti-symmetry, this gives a minus sign for every term of β_I we move through, thus the total number of minus signs will be the length of I, which is k, namely, $(-1)^k$, hence $f_Idg_J \wedge \beta^I \wedge \beta^J = (-1)^kf_I\beta^I \wedge (dg_J \wedge \beta^J) = (-1)^k\omega \wedge d\eta$, completing the proof. ☐

Recall that we want a double application of the derivative to cancel, namely, $\ker d \subseteq \operatorname{im} d$ (the two d's are between different spaces). This is true.

16.6. **Theorem.** *For all k-forms $\omega : P \to \Lambda^k(V^*)$ we have*

$$dd\omega = 0. \tag{16.5}$$

Proof. We can write any k-form as $\omega = f_I \beta^I$ for a basis β^i of V^* dual to b_i of V; then $d\omega = df_I \wedge \beta^I$, where we may decompose the (ordinary) derivative of f in terms of partial derivatives as (in Einstein notation!)

$$df = \partial_{b_i} f \beta^i,$$

and we may decompose the (ordinary) derivative of each $\partial_{b_i} f$ in terms of partials, giving

$$ddf = (\partial_{b_j}(\partial_{b_i} f)\beta^j) \wedge \beta^i = (\partial_{b_j}\partial_{b_i} f)\beta^j \wedge \beta^i.$$

Notice that the sum goes over all possible values of i and j, hence for every (i,j) pair there is a term (j,i). However, by commutativity of mixed partials (cf. Proposition 9.6), $\partial_{b_i}\partial_{b_j} f = \partial_{b_j}\partial_{b_i} f$, whereas $\beta_i \wedge \beta^j = -\beta^j \wedge \beta^i$, which means that every term in the sum cancels, giving $ddf = 0$. $\square$

17. Integrals & pull-backs

Now we look at how to move a form from one space to another along differentiable functions.

$$\begin{tikzcd} P \arrow[r, "f"] \arrow[d, "f^*\omega"'] & Q \arrow[d, "\omega"] \\ \Lambda^k(V^*) & \Lambda^k(W^*) \end{tikzcd}$$

$$\begin{tikzcd} \Lambda^k(V) \arrow[r, "\Lambda^k f'(p)"] \arrow[dr, "(f^*\omega)(p)"'] & \Lambda^k(W) \arrow[d, "\omega(f(p))"] \\ & \mathbb{R} \end{tikzcd}$$

17.1. **Definition.** Let $\omega : Q \to \Lambda^k(W^*)$ be a k-form on $Q \subseteq W$ and $f : P \to Q$ be a differentiable function with derivative $f'(p) : V \to W$. Then the pull-back of ω along f is the form $f^*\omega$ on $P \subseteq V$ given by

$$\begin{aligned} (f^*\omega)(p) &= (f'(p))^*(\omega(f(p))) \\ &= \omega(f(p)) \circ \Lambda^k f'(p). \end{aligned} \tag{17.1}$$

The pull-back of a function is just precomposition $F^* f = f \circ F$ for $f : Q \to \mathbb{R}$ and $F : P \to Q$.

17.2. **Lemma.** *Let $P \xrightarrow{f} Q \xrightarrow{g} R$ be differentiable; let ω, α be k-forms and η be an ℓ-form on Q, and $r : Q \to R$ be a function. Then*

(1) $f^*(r\alpha + \omega) = (r \circ f) f^*\alpha + f^*\omega$ (pull-back is linear)*;*
(2) $f^*(\eta \wedge \omega) = (f^*\eta) \wedge (f^*\omega)$ (components pulled independently)*;*
(3) $(g \circ f)^* = f^* g^*$ (chain rule)*.*

Proof. (1). We use the linearity property of pull-backs of constant k-forms by linear maps:

$$\begin{aligned} f^*(r\alpha + \omega)(p) &= (f'(p))^*(r\alpha + \omega)(f(p)) \\ &= r(f(p))\alpha(f(p)) \circ \Lambda^k f'(p) + \omega(f(p)) \circ \Lambda^k f'(p) \\ &= (r \circ f)(p)(f^*\alpha)(p) + (f^*\omega)(p). \end{aligned}$$

(2). Similarly, upon evaluation we recover pull-back of forms by linear maps

$$\begin{aligned}(f^*(\eta\wedge\omega))(p) &= (f'(p))^*((\eta\wedge\omega)(f(p)))\\ &= (f'(p))^*(\eta(f(p))\wedge\omega(f(p)))\\ &= (f'(p))^*\eta(f(p))\wedge(f'(p))^*\omega(f(p))\\ &= (f^*\eta)(p)\wedge(f^*\omega)(p).\end{aligned}$$

(3). Let ω be a k form on R. Recall that $(g\circ f)'(p) = g'(f(p))\circ f'(p)$. Denote $y = (g\circ f)(p)$. Then

$$\begin{aligned}((g\circ f)^*\omega)(p) &= ((g\circ f)'(p))^*\omega(y)\\ &= (g'(f(p))\circ f'(p))^*\omega(y)\\ &= (f'(p))^*(g'(f(p)))^*\omega(y)\\ &= (f'(p))^*(g^*\omega)(f(p))\\ &= f^*(g^*\omega)(p).\end{aligned}$$

□

17.3. **Lemma.** *Let $\omega : Q\to\Lambda^k(W^*)$ be a k-form and $F : P\to Q$ be a differentiable function. Then the pull-back of the derivative is the derivative of the pulled-back form:*

$$F^*d\omega = d(F^*\omega). \tag{17.2}$$

Proof. Induction on k. For $k = 0$, by definition of the pull-back of the 1-form $df = f'$ along a differentiable map,

$$(F^*f')(p) = (F'(p))^*(f'(F(p))) = f'(F(p))\circ\Lambda^1F'(p) = f'(F(p))\circ F'(p),$$

which equals $(f\circ F)'(p)$ by the chain rule. Then $F^*f' = (f\circ F)' =: (F^*f)'$.

Suppose the statement is true for all ℓ-forms η for all $\ell\le k$, $k\ge 1$. We can write any $(k+1)$-form as $\omega = \eta_i\wedge\beta^i$ (namely, in each $\omega_I\beta^I$ term write out the last functional β^i of β^I out separately, leaving a linear combination of k-forms wedged with β^i). Then, by assumption, $F^*(d\eta_i) = d(F^*\eta_i)$, and, using $d\beta^i = 0$ and the pull-back of a wedge, we have

$$\begin{aligned}F^*(d\omega) &= F^*d(\eta_i\wedge\beta^i) = F^*(d\eta_i\wedge\beta^i + (-1)^k\eta_i\wedge d\beta^i)\\ &= F^*(d\eta_i\wedge\beta^i) = F^*(d\eta_i)\wedge F^*\beta_i = d(F^*\eta_i)\wedge F^*\beta^i,\end{aligned}$$

but

$$\begin{aligned}d(F^*\omega) &= d(F^*(\eta_i\wedge\beta^i)) = d(F^*\eta_i\wedge F^*\beta^i)\\ &= d(F^*\eta_i)\wedge F^*\beta^i + (-1)^kF^*\eta_i\wedge d(F^*\beta_i)\\ &= d(F^*\eta_i)\wedge F^*\beta^i + (-1)^kF^*\eta_i\wedge F^*(d\beta_i)\\ &= d(F^*\eta_i)\wedge F^*\beta^i,\end{aligned}$$

where $d(F^*\beta_i) = F^*(d\beta_i)$ follows by hypothesis, proving the statement. □

We will now see how the integral is essentially the pull-back along the parametrisation of the integration domain – 'k-cube'.

We recall line and surface integrals from calculus; we will use these to generalise a 'k-surface' integral to any k-dimensional parametrised surface $\gamma : [0,1]^k\to V$ (with the condition that γ' exists).[33]

For lines $\gamma : [0,1]\to V$, and an open set $P\subseteq V$ that contains the image $\gamma[0,1]\subsetneq P$, we have that the integral of a 1-form $\omega = f_i\beta^i : P\to V^*$ is the sum

[33]The 'dimension' of the surface is k only if $\gamma'(t)$ is injective at all $t\in[0,1]^k$, namely, of maximal rank.

over $\omega(\gamma(t_i))$ evaluated on tangent vectors $\dot\gamma(t_i) = \partial_{e_1}\gamma(t_i)$ separated by $t_{i+1} - t_i$ intervals between them, namely,

$$\begin{aligned}\int_\gamma \omega &= \lim_{r\to\infty} \sum_{i=1}^{r} \omega(\gamma(t_i))(\partial_{e_1}\gamma(t_i)) \cdot (t_{i+1} - t_i) \\ &= \int_{t\in[0,1]} \omega(\gamma(t))(\partial_{e_1}\gamma(t)).\end{aligned} \tag{17.3}$$

Now, for a surface $\gamma : [0,1]^2 \to V$ and $\operatorname{im}\gamma \subsetneq P$, the integral of the 2-form $\omega : P \to \Lambda^2(V^*)$ is the sum of the form evaluated at the surface on tangent planes: $\omega(\gamma(t))(\partial_{e_1}\gamma(t) \wedge \partial_{e_2}\gamma(t))$, namely,

$$\int_\gamma \omega = \int_{t\in[0,1]^2} \omega(\gamma(t))(\partial_{e_1}\gamma(t) \wedge \partial_{e_2}\gamma(t)). \tag{17.4}$$

This gives a clear generalisation.

17.4. **Definition.** Let $\gamma : [0,1]^k \to P \subseteq V$ be a continuously differentiable function, called a k-cube or k-surface. Let $\omega : P \to \Lambda^k(V^*)$ be a k-form on P. Then the integral of ω over the k-surface γ is defined as

$$\int_\gamma \omega = \int_{t\in[0,1]^k} \omega(\gamma(t))(\partial_{e_1}\gamma(t) \wedge \cdots \wedge \partial_{e_k}\gamma(t)). \tag{17.5}$$

The special case $k = 0$ is defined a evaluation at the only point of $\gamma : \{0\} \to P$, namely, $\int_\gamma \omega = \omega(\gamma(0))$.

Notice that

$$\begin{aligned}(\gamma^*\omega)(t)(e_1 \wedge \cdots \wedge e_k) &= (\omega(\gamma(t)) \circ \Lambda^k\gamma'(t))(e_1 \wedge \cdots \wedge e_k) \\ &= \omega(\gamma(t))(\Lambda^k\gamma'(t)(e_1 \wedge \cdots \wedge e_k)) \\ &= \omega(\gamma(t))(\gamma'(t)(e_1) \wedge \cdots \wedge \gamma'(t)(e_k)) \\ &= \omega(\gamma(t))(\partial_{e_1}\gamma(t) \wedge \cdots \wedge \partial_{e_k}\gamma(t)),\end{aligned}$$

which is the term under the integral in the definition above. We may thus write each integral over the image of γ as the integral over the standard-cube of the pull-back:

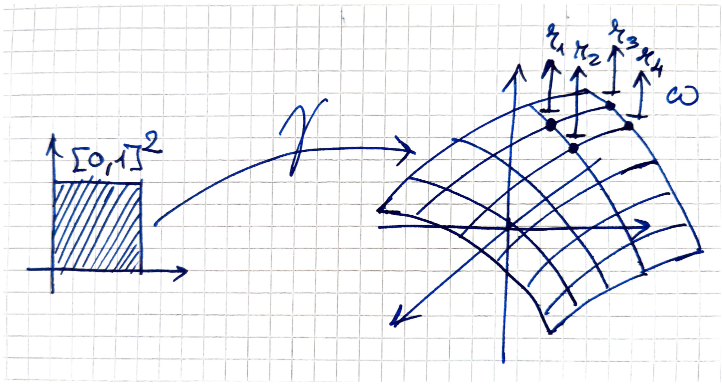
17.5. **Theorem** (Integral of k-form)**.** *Definition 17.4 may be rewritten as*

$$\int_\gamma \omega = \int_{[0,1]^k} \gamma^*\omega, \tag{17.6}$$

where we evaluate $\gamma^*\omega$ *on* $e_1 \wedge \cdots \wedge e_k$*.*[34]

[MAKE THIS THE DEFINITION INSTEAD?] This is better illustrated geometrically:

[34] This corresponds to the integral over the standard cube $I^k : [0,1]^k \to [0,1]^k$ – see §§ below.



In essence, $\gamma^*\omega$ assigns a linear map at each point of the rectangle $[0,1]^k$ with the standard directions $e_1 \wedge \cdots \wedge e_k$; evaluating $\gamma^*\omega$ in the direction $e_1 \wedge \cdots \wedge e_k$ we obtain a number, hence a function $[0,1]^k \to \mathbb{R}$. We know how to integrate functions over rectangles in $[0,1]^k \subsetneq \mathbb{R}^k$.

17.6. **Lemma** (Substitution)**.** *Let $F : P \to Q$ be a continuously differentiable function, $\gamma : [0,1]^k \to P$, and $\omega : Q \to \Lambda^k(W^*)$ a k-form on Q; then*

$$\int_{F\circ\gamma} \omega = \int_\gamma F^*\omega. \tag{17.7}$$

Proof. We employ the property of pull-back (Theorem 17.5):

$$\int_{F\circ\gamma} \omega := \int_{[0,1]^k} (F\circ\gamma)^*\omega = \int_{[0,1]^k} \gamma^* F^*\omega = \int_\gamma F^*\omega,$$

completing the proof. $\square$

We give some examples of how to compute the integrals of forms below.

17.7. **Example.**

17.8. **Example.**

17.9. **Example.**

18. Chains

Recall that we defined the integral over a point $b = \gamma(0)$ given by $\gamma : \{0\} \to P$ to be evaluation at this point $\int_\gamma f = f(\gamma(0)) = f(b)$. We allow ourselves to write $\int_{\{b\}} = \int_\gamma$ for shorthand. Likewise the statement we wish to generalise is of the form $\int_{[a,b]} f' = f(b) - f(a)$, which we see corresponds to $\int_{[a,b]} f' = \int_{\{b\}} f - \int_{\{a\}} f$; it is immediately clear that the 'order' of the derivative decreases together with the dimension of the 'surface' over which we are integrating. Now, the detail of having to evaluate two integrals on the left is bothersome – we wish to combine $\{a\}$ and $\{b\}$ into a single integral. The reader should now recall free vector spaces.

We will proceed with the definition shortly, where we will want to write $\{b\} - \{a\}$ as a meaningful statement: integrate over $\{b\}$ with a constant $+1$ in front, whereas a -1 before $\{a\}$. We see that the basis vectors will be cubes. Let us begin with the standard cube.

The simplest case is $[0,1]$, which has endpoints $\{0\}$ and $\{1\}$. A dimension higher gives $[0,1]^2$, which has four boundary faces: the two horizontal edges $\{(x,0) : x \in [0,1]\}$, $\{(x,1) : x \in [0,1]\}$ and the two verticals $\{(0,y) : y \in [0,1]\}$, $\{(1,y) : y \in [0,1]\}$. There are both one-dimension lower and can be described as maps from the interval $[0,1]$. Generalising further is clear: for every dimension $1 \geq i \leq k$ in $[0,1]^k$

we will have two faces: one on the left with the i-th coordinate being 0 and one on the right with 1. The set notation is bothersome, hence we will adopt functions.

18.1. **Definition.** We define the standard k-cube to be the identity map $I^k : [0,1]^k \to [0,1]^k$, $x \mapsto x$, where $x = (x^1, ..., x^k)$. For each coordinate of x we define the left ($\sigma = 0$) and right ($\sigma = 1$) faces to be the $(k-1)$-cubes mapping into the k-cube, namely, $I_{i,\sigma} : [0,1]^{k-1} \to [0,1]^k$, with a fixed i-th coordinate:

$$I^k_{i,\sigma}(t^1, ..., t^{k-1}) = (t^1, ..., t^{i-1}, \sigma, t^i, ..., t^{k-1}). \tag{18.1}$$

Given any[35] (i.e. curved; non-standard) k-cube $\gamma : [0,1]^k \to P \subseteq V$ we use the faces of the standard cube to define the faces of γ as $\gamma_{i,\sigma} = \gamma \circ I^k_{i,\sigma}$.

18.2. *Remark.* The faces of γ do not have to be separated by an edge – a line of singular points. The simplest example is the circle & sphere: the edges of $\gamma : [0,1] \to \mathbb{R}^2$, $t \mapsto (\cos 2\pi t, \sin 2\pi t)$ will just be some point(-s) $\gamma(\{0\}) = \{(1,0)\} = \gamma(\{1\})$ on the circle, but the image γ is smooth through them. For the 2-sphere $\gamma : [0,1]^2 \to \mathbb{R}^3$, $(t,s) \mapsto (\cos 2\pi t \sin 2\pi s, \sin 2\pi t \sin 2\pi s, \cos 2\pi s)$ we have the 1-faces

$$\gamma_{1,0}(s) = \gamma \circ \{s \mapsto (0,s)\}(s) = \gamma(0,s) = \begin{pmatrix} \sin 2\pi s \\ 0 \\ \cos 2\pi s \end{pmatrix},$$

which gives the unit circle on the xz-plane.

We can now formalise the notion of linear combinations of faces.

18.3. **Definition.** We call the free vector space over all k-cubes $\gamma : [0,1]^k \to P$ in $P \subseteq V$ the space of k-chains in P, denoted by

$$C_k(P) = \mathbb{R}^{\{k-\text{cubes in } P\}} = \mathbb{R}^{\{\gamma:[0,1]^k \to P \ : \ \gamma \text{ differentiable}\}}. \tag{18.2}$$

Indeed, it is immediately clear that $C_k(P)$ is infinite-dimensional.[36] The point is that we may now extend our definition of an integral to k-cubes in the most natural way.

18.4. **Definition.** The integral over any linear combination $r\mu + s\gamma$ $(r, s \in \mathbb{R})$, where γ is a k-cube and $\mu \in C_k(P)$, of k-chains in $C_k(P)$ is defined as

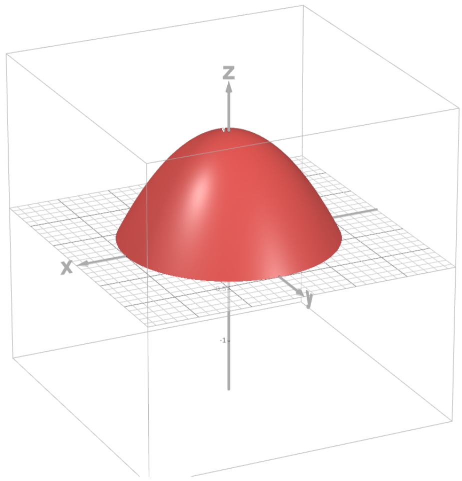
$$\int_{r\mu + s\gamma} \omega = r \int_\mu \omega + s \int_\gamma \omega. \tag{18.3}$$

If we wish to integrate over the (surface of a) paraboloid (unit disk in xy-plane and paraboloid $z = x^2 + y^2$ over it) as shown below:

---

[35]For regularity, we may restrict γ to be differentiable.

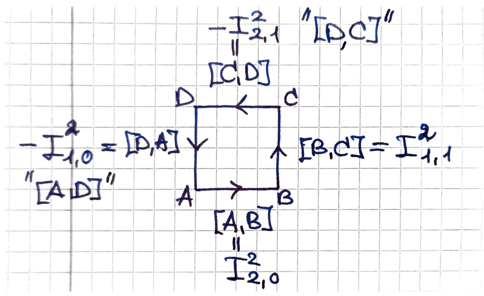
[36]We will not encounter a basis for it here, thus we are not interested whether the dimension is countable or uncountable.



Then we can break up this paraboloid into a sum of two 2-cubes: one over the curbed part, and the other over the disk in the plane – both of these are 2-chains. Note that we CANNOT take the sum 'pointwise' – the sum of k-cubes is not a k-cube.

Indeed, this allows us to write $\int_{[a,b]} f' = \int_{\{b\}-\{a\}} f$. We already know that in Stoke's theorem, we will need to relate the integral over γ to the integral over the boundary of γ; but this boundary has to be oriented in a way consistent with the surface.

As a preliminary, the boundary of $[0,1]$ is the endpoint minus the beginning point $\{1\} - \{0\}$. What we want is for the boundary of a k-cube to detect whether whether the k-cube is closed; we know that all standard cubes are closed. For the 2-cube $[0,1]^2$ we can write vertices A, B, C, D and traverse the edges clockwise, namely, denoting (*for the present moment only*) the edge from A to B by $[A, B]$:



We see that the boundary of I^2 becomes

$$[A, B] + [B, C] + [C, D] + [D, A].$$

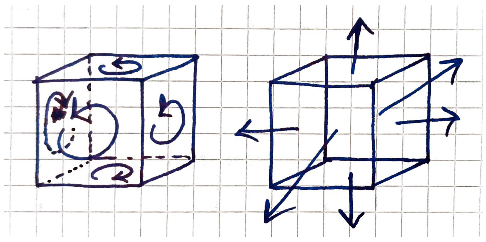
But(!), now observe, how these edges relate to the *functions* $I^2_{i,\sigma}$ – these functions are always oriented towards the positive direction. We see that the edge $[A, B]$ is pointed in the positive e_1 direction, and thus we can assign $+I^2_{2,0}$; similarly for $I^2_{1,1}$. But for the edge $[C, D]$ we see that it points towards D, which is in the $-e_1$ direction, but $I^2_{2,1}$ points in the $+e_1$ direction, therefore, we have to assign $-I^2_{2,1}$; similarly for $[D, A]$ we assign $-I^2_{1,0}$. The boundary is then

$$I^2_{2,0} + I^2_{1,1} - I^2_{2,1} - I^2_{1,0}.$$

And if we take the boundary again, we will obtain

$$B - A + C - B + D - C + A - D = 0.$$

For I^3 we again simply proceed with counterclockwise-oriented faces (corresponds to the right-hand rule around the normal vectors):



And the reason why: the boundary of the cube is, of course, all the faces; the normal vector orients the faces such that for an edge the two neighbouring faces orient it oppositely (and thus will cancel). The question becomes which sign to assign to which face. We will succeed by setting the sign to be $(-1)^{i+\sigma}$, meaning that for odd-numbered $e_1, e_3, \ldots$ the sign will be positive when the normal points along $+e_i$, whereas for $e_2, e_4, \ldots$ it will be opposite.

18.5. **Definition.** We set the boundary $\partial\gamma$ of a k-cube $\gamma : [0,1] \to P$ to be the $(k-1)$-chain

$$\partial\gamma = \sum_{i=1}^{k} \sum_{\sigma\in\{0,1\}} (-1)^{i+\sigma}\gamma_{i,\sigma}. \tag{18.4}$$

We set the boundary of a k-chain $\partial : C_k(P) \to C_{k-1}(P)$ to be the boundary of k-cubes extended linearly – for any k-cubes μ, γ and $r, s \in \mathbb{R}$:

$$\partial(r\mu + s\gamma) = r\partial\mu + s\partial\gamma. \tag{18.5}$$

18.6. **Lemma** (Boundary of boundary is zero)**.** *$\partial\partial\gamma = 0$ for all chains γ.*

Proof. Observe that for the standard k-cube $I^k_{i,\sigma}$ is a $(k-1)$-cube (taking the place of γ before). We will show that the boundaries coming from adjacent faces will cancel;[37] indeed, we find

$$\partial\partial I^k = \sum_{i=1}^{k}\sum_{\sigma}(-1)^{i+\sigma}\partial I^k_{i,\sigma} = \sum_{i=1}^{k}\sum_{\sigma}\sum_{j=1}^{k-1}\sum_{\tau}(-1)^{i+\sigma}(-1)^{j+\tau} I^k_{i,\sigma}\circ I^{k-1}_{j,\tau}.$$

For $i < j$ (note: $(j+1)-1 = j > i$) we compute[38]

$$\begin{aligned}
(I^k_{i,\sigma}\circ I^{k-1}_{j,\tau})(t^1,\ldots,t^{k-2}) &= I^k_{i,\sigma}(t^1,\ldots,t^{j-1},\tau,t^j,\ldots,t^{k-2})\\
&= (t^1,\ldots,t^{i-1},\sigma,t^i,\ldots,t^{j-1},\tau,t^j,\ldots,t^{k-2}),\\
(I^k_{j+1,\tau}\circ I^{k-1}_{i,\sigma})(t^1,\ldots,t^{k-2}) &= I^k_{j+1,\tau}(t^1,\ldots,t^{i-1},\sigma,t^i,\ldots,t^{k-2})\\
&= (t^1,\ldots,t^{i-1},\sigma,t^i,\ldots,t^{j-1},\tau,t^j,\ldots,t^{k-2})\\
&= I^k_{i,\sigma}\circ I^{k-1}_{j,\tau}.
\end{aligned}$$

Note that the signs will be opposite: $(-1)^{i+\sigma+j+\tau} = -(-1)^{i+\sigma+j+1+\tau}$. When we consider any $(k-2)$-cube that appears in the $\partial\partial I^k$ sum, namely, with σ in the m-th component and τ in the ℓ-th component, it can be described both by $I^k_{m,\sigma}\circ I^{k-1}_{\ell-1,\tau}$ and by $I^k_{\ell,\tau}\circ I^{k-1}_{m,\sigma}$, hence the terms describing this $k-2$ cube cancel, leaving $\partial\partial I^k = 0$.

Lastly, for $\partial\partial\gamma$ we will have the same cancellation as above after composition with γ, hence $\partial\partial\gamma = 0$. $\square$

[37] The proof in Roland's notes contains a major typo; we fix it and complete the steps of the proof.

[38] The reader can explicitly check the case $i = 2$ and $j = 3$.

18.7. *Remark.* Notice the similarity between $\partial^2 = 0$ and $d^2 = 0$. Indeed, these concepts are dual; the names are homology and cohomology. We will return to these in **????**. Notice that $d^2 = 0$ means that $\operatorname{im} d \subseteq \ker d$ and $\operatorname{im} \partial \subseteq \ker \partial$. **NB!** the d's and ∂'s on each side map between different spaces:

$$\mathscr{X} \xrightarrow[\text{im}]{d \text{ or } \partial} \mathscr{Y} \xrightarrow[\text{ker}]{d \text{ or } \partial} \mathscr{Z},$$

where under the arrows we indicate which is which, and $\mathscr{Y} = \Omega^k(P)$ or $C_k(P)$, the space of forms or chains. Indeed, this allows us to write the quotient vector spaces of the form $\ker \partial / \operatorname{im} \partial$ (dually, $\ker d / \operatorname{im} d$), namely,

$$\underset{\text{‘}k\text{-th homology’}}{H_k(P) = \frac{\ker \partial}{\partial(C_k(P))}} \qquad \text{and} \qquad \underset{\text{‘}k\text{-th cohomology’}}{H^k(P) = \frac{\ker d}{d(\Omega^{k-1}(P))}},$$

where we used $\operatorname{im} \partial = \partial(C_k(P))$ and $\operatorname{im} d = d(\Omega^{k-1}(P))$ to explicitly indicate which order (k) we are talking about.[39] The space is trivial if there is no k-dimensional hole in P; if there is, the vector space structure allows for characterisation of it.

19. Stokes' theorem & Poincaré lemma

We are now ready to explore the interplay between $\partial^2 = 0$ and $d^2 = 0$ when integrating. Consider a blob γ with boundary $\partial\gamma$; when we integrate a form ω over $\partial\gamma$, we count how many level sets of ω are created or destroyed inside γ – indeed, if a level set both enters and leaves the blob, its contribution will cancel, adding 0 to the integral. And $d\omega$ is precisely what allows us to count the number of created level sets inside γ. The integral of the form over the bounary is thus the integral of the (exterior) derivative over the interior.

19.1. **Theorem** (Stokes')**.** *For any continuously-differentiable $(k-1)$-form on P $\omega \in \Omega_1^{k-1}(P)$ and any k-chain $\gamma \in C_k(P)$ we have*

$$\int_\gamma d\omega = \int_{\partial\gamma} \omega. \tag{19.1}$$

Proof. Let us first consider an integral of $d\omega$ over the standard-cube I^k; the standard cube lives in $\mathbb{R}^k$ and thus corresponds to the volume element $e_1 \wedge \cdots \wedge e_k$ (cf. Theorem 17.5). We consider a $(k-1)$-form ω in $\mathbb{R}^k$; by the basis of $\Lambda^{k-1}(\mathbb{R}^k)$, we can write $\omega = f_J \varepsilon^J$, where each J consists of an increasing sequence from $1, ..., k$ with some j removed; indeed, let $J = (1, ..., j-1, j+1, ..., k)$. For this fixed J, we have $\omega = f\varepsilon^J$, and $d\omega = df \wedge \varepsilon^J = (\partial_{e_i} f \varepsilon^i) \wedge \varepsilon^J = \partial_{e_j} f \varepsilon^j \varepsilon^J = (-1)^{j-1} \partial_{e_j} f \varepsilon^{1\dots k}$, namely, a multiple of the volume element in $(\mathbb{R}^k)^*$, whereby $\varepsilon^{1\dots k} e_{1\dots k} = 1$.

We now have

$$\int_{I^k} d\omega = \int_{I^k} df \wedge \varepsilon^J = (-1)^{j-1} \int_{[0,1]^k} \partial_{e_j} f \varepsilon^{1\dots k} e_{1\dots k} = (-1)^{j-1} \int_{[0,1]^k} \partial_{e_j} f,$$

where we may evaluate the final integral first in the j-th direction (by Fubini's theorem) using the fundamental theorem of calculus

$$\int_{t^j \in [0,1]} \partial_{e_j} f(t^1, ..., t^j, ..., t^k) = f(t^1, ..., 1, ..., t^k) - f(t^1, ..., 0, ..., t^k) = f \circ I^k_{j,1} - f \circ I^k_{j,0},$$

obtaining

$$\int_{I^k} d\omega = (-1)^{j-1} \int_{[0,1]^{k-1}} (f \circ I^k_{j,1} - f \circ I^k_{j,0}).$$

We will now show that this equals $\int_{\partial I^k} \omega$.

[39] Alternatively, one gives different names to each map in the quotient $\partial_k : C_k(P) \to C_{k-1}(P)$ and $d_{k-1} : \Omega^{k-1}(P) \to \Omega^k(P)$ to distinguish the ∂ or d above and below.

Evaluating the integral over a chain yields the linear combination of integrals evaluated over its $(k-1)$-cubes $\partial I^k = \sum_{i,\sigma}(-1)^{i+\sigma} I^k_{i,\sigma}$:

$$\int_{\partial I^k} \omega = \sum_{i,\sigma}(-1)^{i+\sigma}\int_{I^k_{i,\sigma}} \omega = \sum_{i,\sigma}(-1)^{i+\sigma}\int_{[0,1]^{k-1}} (I^k_{i,\sigma})^*\omega,$$

where we evaluate the pulled-back form on the volume element of $\mathbb{R}^{k-1}$, namely, $\tilde{e}_1 \wedge \cdots \wedge \tilde{e}_{k-1}$, where we distinguish $\tilde{e}_i \in \mathbb{R}^{k-1}$ from $e_i \in \mathbb{R}^k$. Now, $\partial_{\tilde{e}_1} I^k_{i,\sigma}(t) = e_1 \in \mathbb{R}^k$; similarly, $\partial_{\tilde{e}_{m<i}} I^k_{i,\sigma}(t) = e_m$, whereas $\partial_{\tilde{e}_{m>i}} I^k_{i,\sigma}(t) = e_{m+1}$. This yields

$$\begin{aligned}
((I^k_{i,\sigma})^*\omega)(t)(\tilde{e}_{1\ldots k-1}) &= f(I^k_{i,\sigma}(t))\varepsilon^J \circ \Lambda^{k-1}(I^k_{i,\sigma})'(\tilde{e}_{1\ldots k-1}) \\
&= f(I^k_{i,\sigma}(t))\varepsilon^J(\partial_{\tilde{e}_1} I^k_{i,\sigma}(t) \wedge \cdots \wedge \partial_{\tilde{e}_{k-1}} I^k_{i,\sigma}(t)) \\
&= f(I^k_{i,\sigma}(t))\varepsilon^J(e_1 \wedge \cdots \wedge e_{i-1} \wedge e_{i+1} \wedge \cdots \wedge e_k) \\
&= f(I^k_{i,\sigma}(t))\delta^j_i.
\end{aligned}$$

Therefore the sum over i reduces to just j, giving

$$\begin{aligned}
\int_{\partial I^k} \omega &= \sum_{\sigma}(-1)^{j+\sigma}\int_{[0,1]^{k-1}} f \circ I^k_{j,\sigma} \\
&= (-1)^{j+1}\int_{[0,1]^{k-1}} f \circ I^k_{j,1} + (-1)^j \int_{[0,1]^{k-1}} f \circ I^k_{j,0} \\
&= (-1)^{j-1}\left(\int_{[0,1]^{k-1}} f \circ I^k_{j,1} - \int_{[0,1]^{k-1}} f \circ I^k_{j,0}\right) \\
&= (-1)^{j-1}\int_{[0,1]^{k-1}} (f \circ I^k_{j,1} - f \circ I^k_{j,0}) \\
&= \int_{I^k} d\omega,
\end{aligned}$$

where we used $(-1)^{j+1} = (-1)^{j-1}$. This proves the theorem for the special case of integrating over the standard k-cube $\gamma = I^k$.

For the general case $\gamma : [0,1]^k \to P$ we employ the pull-back of a derivative (cf. Lemma 17.3) and use the special case ($\omega = f_J \varepsilon^J$ over ∂I^k):

$$\int_\gamma d\omega = \int_{I^k} \gamma^* d\omega = \int_{I^k} d(\gamma^*\omega) = \int_{\partial I^k} \gamma^*\omega = \int_{\partial\gamma} \omega,$$

where we recall that $\gamma \circ \partial I^k = \partial\gamma$. This completes the proof. □

Note that the proof is (perhaps) surprisingly short.[40]

There is a joke that Stoke's theorem is trivial; this is because: (1) all terms have been thoroughly defined, (2) it has significant consequences. One of the consequences is Cauchy's integral theorem from complex analysis; we first need to define complex-differentiation and see that it restricts $f'(p)$ to be of a certain form. We are able to give a 1-paragraph proof of the theorem,[41] which is truly incredible. This is developed in Appendix A.

We now turn to the second part of the Fundamental theorem – Poincaré lemma. We would ideally like to apply Stokes' theorem on a form, which turns out to be a derivative of another form; namely, we do not explicitly begin with the derivative – this is indeed the form most useful in practice. Firstly, a necessary condition.

19.2. **Lemma.** *Let $\omega \in \Omega^k_1(P)$ satisfy $d\omega \neq 0$. Then there exists no $\eta \in \Omega^{k-1}_2(P)$ such that $d\eta = \omega$. Equivalently, η can exist only if $d\omega = 0$.*

---

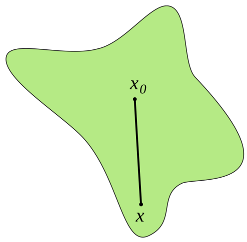
[40] Omitting full explanations, it can be written as just 3 lines of computation.

[41] And without skipping all the steps!

Proof. We compute $0 = dd\eta = d\omega \neq 0$ – a contradiction. ☐

The converse is true as long as the domain does not contain holes. We shall prove it for a less general case – a star-shaped domain.

19.3. **Definition.** A open subset $0 \in P \subseteq V$ is called *star-shaped domain* if it contains the line segment $[0,1]x = \{tx : t \in [0,1]\} \subseteq P$ between 0 and any $x \in P$. An example is given below ($x_0 = 0$).



Note that this means that P contains an open ball around 0.

19.4. **Lemma** (Poincaré). *Let $P \subseteq V$ be a star-shaped domain. If any k-form $\omega \in \Omega_1^k(P)$ satisfies $d\omega = 0$ on P, then there exists a primitive $\alpha \in \Omega_2^{k-1}(P)$ such that $d\alpha = \omega$.*

Proof. The idea is to integrate over $[0,1]x$. Cf. `https://www.math.brown.edu/reschwar/M114/notes7.pdf`. ☐

19.5. *Remark.* Some terminology: ω is called closed if $d\omega = 0$; exact if there exists a primitive $d\alpha = \omega$. Poincaré lemma states that in Euclidean spaces, being closed and exact are equivalent; this is not true in general spaces – manifolds.

19.6. **Theorem** (Reparametrisation). *Let $F : [0,1]^k \to [0,1]^k$ reparametrise the cube, namely, F is a k-cube with the boundary $\partial F = \partial I^k$. Then*

$$\int_{\gamma \circ F} \omega = \int_{\gamma} \omega \tag{19.2}$$

for any $\gamma : [0,1]^k \to P$ k-cube and any $\omega \in \Omega_1^k(P)$.

Proof. Note that $[0,1]^k \subseteq \mathbb{R}^k$ and $\dim \mathbb{R}^k = k$, hence $\gamma^*\omega \in \Omega_1^k([0,1]^k)$ is a form of maximal degree $d(\gamma^*\omega) = 0$. Let $\alpha \in \Omega_2^{k-1}([0,1]^k)$ be a primitive $d\alpha = \gamma^*\omega$. Then

$$\int_{\gamma\circ F} \omega = \int_F \gamma^*\omega = \int_F d\alpha = \int_{\partial F} \alpha = \int_{\partial I^k} \alpha = \int_{I^k} d\alpha = \int_{I^k} \gamma^*\omega = \int_{\gamma} \omega$$

by Stokes' theorem. ☐

20. Duals of forms – Hodge star

Recall the similarity between k forms and $n-k$ forms; indeed, $\dim \Lambda^k(V) = \binom{n}{k} = \dim \Lambda^{n-k}(V)$. It turns out that these spaces are indeed naturally isomorphic, where the isomorphism is called a Hodge star.

There exist an infinite family of isomorphisms, and the 'natural' one is determined by the geometry of the space; the simplest geometry is the standard dot product – indeed, we will begin with this simpler case.[42] Let $\dim V = n$ and $\mathbf{e} = e_1, ..., e_n \in V$ be a basis for V with dual basis $\boldsymbol{\varepsilon} = \varepsilon^1, ..., \varepsilon^n \in V^*$.

[42] No mention will be made of the inner product just yet; after we generalise, it will become clear that it is hidden in the definition.

20.1. **Definition.** The map $\star : \Lambda^k(V) \to \Lambda^{n-k}(V)$ defined by

$$\star e_I = e_J \qquad \text{such that} \qquad e_I \wedge e_J = e_1 \wedge \cdots \wedge e_n \tag{20.1}$$

and extended linearly is called the *Hodge star*.

The same definition holds when V is replaced by V^* (i.e. forms).

20.2. *Remark.* Note the order in which $e_I \wedge e_J$ form the full volume element – reversing two elements under the wedge gives a minus.

Not that for any V the map $\star : \Lambda^0(V) \to \Lambda^n(V)$ must be $\star 1 = e_{1\ldots n}$. Similarly, $\star : \Lambda^n(V) \to \Lambda^0(V)$ is given by $\star e_{1\ldots n} = 1$.

20.3. **Example.** (a) Let $\dim V = 2$. Then $\star : \Lambda^1(V) \to \Lambda^1(V)$ maps 1-forms into themselves according to

$$\star e_1 = e_2 \qquad \text{and} \qquad \star e_2 = -e_1,$$

where the minus appears as we have $e_2 \wedge e_1 = -(e_1 \wedge e_2)$.

(b) Let $\dim V = 3$. We compute $\star : \Lambda^1(V) \to \Lambda^2(V)$ by

$$\star e_1 = e_2 \wedge e_3, \qquad \star e_2 = -e_1 \wedge e_3, \qquad \star e_3 = (-1)^2 e_1 \wedge e_2 = e_1 \wedge e_2$$

as $e_2 \wedge e_1 \wedge e_3 = -e_1 \wedge e_2 \wedge e_3$ must be interchanged with one vector e_1, giving a minus sign.

20.4. **Lemma.** *The Hodge star is an isomorphism and is its own inverse up to sign.*

Proof. Let $\star : \Lambda^k(V) \to \Lambda^{n-k}(V)$ and $* : \Lambda^{n-k}(V) \to \Lambda^k(V)$ be Hodge stars. Then, letting $\star e_I = e_J$ (hence $e_I \wedge e_J = e_{1\ldots n}$) we have

$$(* \circ \star)(e_I) = *(\star e_I) = *(e_J) = e_I$$

as $e_J \wedge e_K = e_{1\ldots n}$ implies that e_K contains all e_i not in e_J, hence $e_J \wedge e_I = (-1)^{k(n-k)} e_I \wedge e_J$. Defining $\star^{-1} = (-1)^{k(n-k)} *$ gives $\star^{-1} \circ \star = \mathrm{id}_{\Lambda^k(V)}$. Now,

$$(\star \circ \star^{-1})(e_J) = \star(e_I) = e_J,$$

where we notice that every e_J is reached by $\star$, implying $\star \circ \star^{-1} = \mathrm{id}_{\Lambda^{n-k}(V)}$. Thus $\star^{-1}$ is the inverse of $\star$, completing the proof. □

The power of the star comes in showing the calculus identities Curl Grad = 0 and Div Curl = 0 by exploiting the exterior derivative d and $d^2 = 0$. For a differentiable function $f : V \to \mathbb{R}$ we have

$$\operatorname{Grad} f = \sum_j \frac{\partial f}{\partial x^j} e_j : V \to V \qquad \text{and} \qquad df = \sum_j \frac{\partial f}{\partial x^j} \varepsilon^j : V \to \mathbb{R},$$

which we see are eerily similar; indeed, if we make the identification $\mathscr{I}(e_j) = \varepsilon^j$ and extended linearly we obtain $\mathscr{I}(\operatorname{Grad} f) = df$ (pointwise), where $\mathscr{I} : V \to V^*$. For simpler notation we will say $e_j \simeq \varepsilon^j$ to mean that there is an $\mathscr{I}$ between them. We obtain Grad $\simeq d$ – a fact that was perhaps clear to the reader, but not fully formalised to this point.

For a differentiable vector field $F : \mathbb{R}^3 \to \mathbb{R}^3$ we identify F with $\star d\omega_F$, where $\omega_F \in \Lambda^1((\mathbb{R}^3)^*)$ is the 1-form identifying $\omega_F \simeq \mathscr{I}(F) = \mathscr{I}(f^j e_j) = f^j \mathscr{I}(e_j) = \sum_j f^j \varepsilon^j$. Indeed, a computation shows

$$\begin{aligned}
\star d\omega_F &= \star\Big(\sum_j df^j \wedge \varepsilon^j\Big) = \sum_j \partial_i f^j \star (\varepsilon^i \wedge \varepsilon^j) \\
&= (\partial_1 f^2 - \partial_2 f^1) \star \varepsilon^{12} + (\partial_2 f^3 - \partial_3 f^2) \star \varepsilon^{23} + (\partial_3 f^1 - \partial_1 f^3) \star \varepsilon^{13}) \\
&= \mathscr{I}(\operatorname{Curl} F),
\end{aligned}$$

confirming $\mathscr{I}(\operatorname{Curl} F) = \star d\omega_F$, or, Curl $\simeq \star d$.

A similar computation yields $\mathrm{Div} \simeq \star d \star$ identifying F with ω_F as before.

20.5. **Proposition.** *We have* $\mathrm{Curl}\,\mathrm{Grad} = 0$ *and* $\mathrm{Div}\,\mathrm{Curl} = 0$ *for* $V = \mathbb{R}^3$.

Proof. We compute

$$\mathrm{Curl} \circ \mathrm{Grad} \simeq \star d \circ d = \star d^2 = \star 0 = 0, \qquad \text{and} \tag{20.2}$$

$$\mathrm{Div} \circ \mathrm{Curl} \simeq \star d \star \circ \star d = \star d(\pm \mathrm{id})d = \pm \star d^2 = 0, \tag{20.3}$$

completing the proof. □

20.6. *Remark.* Notice the similarity between d and $\star d \star$. Indeed, we may define a 'dual' exterior derivative $d^* = \star d \star : \Lambda^k(V^*) \to \Lambda^{k-1}(V^*)$, which lowers the degree of a form. We can then define the Laplacian operator Δ

$$\Delta = dd^* + d^*d = d \star d \star + \star d \star d, \tag{20.4}$$

which $\Delta : \Lambda^k(V^*) \to \Lambda^k(V^*)$ does not change the order of the forms.

For $V = \mathbb{R}^3$ we recover $\Delta \simeq \mathrm{Div}\,\mathrm{Grad}$.

The Laplacian allows to connect differential geometry with partial differential equations – the topic is called Hodge theory.

The Hodge star may be defined in a coordinate-free form if we endow V with an inner product.

20.7. *Remark.* Finally, the Hodge star allows to rewrite the 4 Maxwell's equations for electromagnetism into two very compact equations for a two-form $F = - \star \hat{F}$ with $\hat{F} = \hat{F}_{\mu\nu} dx^\mu \wedge dx^\nu$ (combining the electric and magnetic field components), namely:

$$dF = 0 \qquad \text{and} \qquad d^\star F = J, \tag{20.5}$$

where J^μ combines current and charge. For further reading and full definitions see `https://staff.fnwi.uva.nl/h.b.posthuma/TIP2019/Lecture%202-2019.pdf`.

Appendix A. Proof of Cauchy's integral theorem (complex analysis)

A.1. **Definition.** We consider $\mathbb{C} = \mathbb{R}^2$ with basis $1, i$ (any $z = x + iy$). We call $f : P \subseteq \mathbb{C} \to \mathbb{C}$ (P open) *complex-differentiable on* P if the derivative of f at any point is a complex-linear[43] function: for every $p \in P$ there exists a number $w(p) \in \mathbb{C}$ such that

$$f'(p) = m_{w(p)}, \qquad \text{where} \qquad m_{w(p)} : \mathbb{C} \to \mathbb{C}, \qquad z \mapsto w(p)z. \tag{A.1}$$

A.2. **Lemma** (Cauchy–Riemann equations)**.** *For any complex-differentiable function $f : P \subseteq \mathbb{C} \to \mathbb{C}$ (P open) we have $\partial_i f = i\partial_1 f$. Writing $f(z) = u(z) + iv(z)$ this is equivalent to $u_x = v_y$ and $u_y = -v_x$ (notation: $u_x := \partial_1 u$ and $u_y := \partial_i u$).*

Proof. Note that multiplication by i is a linear map represented by the matrix

$$[i] = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

with respect to the standard basis $1, i$ of $\mathbb{C}$. As a linear map, $m_{w=w_1+iw_2}(x + iy) = (w_1x - w_2y) + i(w_2x + w_xy)$, and thus has matrix representation

$$[m_{w(p)}] = \begin{pmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{pmatrix} = [f'(p)] = [\partial_1 f(p) \;\; \partial_i f(p)] = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix},$$

from which we see that $[\partial_i f(p)] = [i][\partial_1 f(p)]$, or $\partial_i f(p) = i\partial_1 f(p)$, and the relations $u_x = v_y$ and $u_y = -v_x$. □

A.3. **Corollary** (Cauchy's integral)**.** *For any continuously-complex-differentiable function $f : P \subseteq \mathbb{C} \to \mathbb{C}$ (P open) and any 2-cube $\gamma : [0, 1]^2 \to P$ we have*

$$\int_{\partial\gamma} f(z)dz = 0, \tag{A.2}$$

where dz is the 1-form $dz = dx + idy$.

Proof. We show $d(f(z)dz) = 0$; by Stokes' theorem then $\int_{\partial\gamma} f(z)dz = \int_\gamma 0 = 0$.

Note that $dx(a + ib) = a$ and $dy(a + ib) = b$, which gives a natural meaning to the functional $dz = dx + idy$. Write $f(z) = u(z) + iv(z)$. Now,

$$\begin{aligned} fdz &= (u + iv(dx + idy) = (udx - vdy) + i(vdx + udy), && \text{therefore} \\ d(fdz) &= (du \wedge dx - dv \wedge dy) + i(dv \wedge dx + du \wedge dy), && du(p) := u'(p) \\ &= -(u_y + v_x)\, dx \wedge dy + i(u_x - v_y)\, dx \wedge dy = 0 \end{aligned}$$

by using the Cauchy–Riemann equations. The proof is complete. □

A.4. *Remark.* We need continuous (complex-)differentiability because this appears in the statement of Stokes' theorem. The theorem holds more generally without requiring continuous partial derivatives. One then first needs to show that a complex-differentiable function has a convergent power series (in terminology: holomorphic implies analytic); this is difficult.

[43] Note that $\dim_{\mathbb{C}} \mathbb{C} = 1$ as a complex-valued vector space.