MAS212 Lecture Notes

Dima Pasechnik

Draft April 24, 2006

Contents

Chapter 1

Derivatives

Notation and linear algebra, and other, preliminaries

Linear algebra. We work with vectors $v := (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$, where $\mathbb{R}$ denotes the real numbers. We use the (Euclidean) *norm* (sometimes denoted by $|v|$)

$$\|v\| := \sqrt{\sum_{i=1}^{n} v_i^2}. \tag{1.1}$$

It satisfies the following inequalities: "triangle"

$$\|v\| + \|w\| \geq \|v + w\|, \qquad \text{for all } u, v \in \mathbb{R}^n, \tag{1.2}$$

and Cauchy-Schwarz:

$$v \cdot w \leq \|v\| \|w\|, \qquad \text{for all } u, v \in \mathbb{R}^n. \tag{1.3}$$

A *linear mapping* $L : \mathbb{R}^n \to \mathbb{R}^m$ is a function satisfying the following properties:

$$L(\alpha v) = \alpha L(v) \qquad \forall v \in \mathbb{R}^n, \quad \forall \alpha \in \mathbb{R} \tag{1.4}$$

$$L(v + w) = L(v) + L(w) \qquad \forall v, w \in \mathbb{R}^n. \tag{1.5}$$

In turn, the linear mappings $\mathbb{R}^n \to \mathbb{R}^m$ form a vector space of dimension nm, usually denoted by $\mathbb{R}^{m \times n}$; when one fixes bases in $\mathbb{R}^n$ and $\mathbb{R}^m$, these mappings are represented by $m \times n$ *matrices* $A := A(L)$, and the image of $v \in \mathbb{R}^n$ under L is just $Av \in \mathbb{R}^m$, the matrix-vector product. Thus for the image of v under L we often write Lv instead of $L(v)$. (Note that linear mappings are also called linear transformations.)

For $L : \mathbb{R}^n \to \mathbb{R}^m$ define its *norm* (the matrix norm) as follows:

$$\|L\| := \sup_{\|v\| \leq 1} \|Lv\|, \qquad \text{(here of course } v \in R^n\text{)}. \tag{1.6}$$

This norm satisfies the following important properties:[1]

$$\|L\| < \infty, \qquad \forall L \in \mathbb{R}^{m\times n}, \tag{1.7}$$

$$\|BA\| \leq \|A\|\|B\|, \qquad \forall A \in \mathbb{R}^{m\times n}, \quad \forall B \qquad \in \mathbb{R}^{k\times m}. \tag{1.8}$$

Certainly (1.3) is a particular case of (1.8). Note that even when $m = n = k$, one can have $AB \neq BA$.

A linear mapping $L : \mathbb{R}^n \to \mathbb{R}^n$ is called *invertible* when $Lv = 0$ implies $v = 0$. In this case L defines a bijection on $\mathbb{R}^n$. It is equivalent to the existense of a linear mapping $B : \mathbb{R}^n \to \mathbb{R}^n$ satisfying $LB = I$, the identity transformation. Such B is denoted by L^{-1}, and it is in fact unique. The set of invertible L's is denoted by $GL_n(\mathbb{R})$. They form a group under composition (i.e. matrix multiplication), called *general linear* (GL is the abbreviation here).

More generally, a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ is called *nonsingular* when $Lv = 0$ implies $v = 0$. In this case is it necessary that $n \leq m$. That is, L is always singular when $n > m$.

Slighly more general than linear are *affine linear* (or just *affine*) mappings: given a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ and $w \in R^m$ we define

$$\begin{aligned} L_w : \mathbb{R}^n &\to \mathbb{R}^m \\ v &\mapsto Lv + w. \end{aligned} \tag{1.9}$$

The image of $\mathbb{R}^n$ under an affine mapping L_w is called *affine subspace*[2]. It obviously has the form $w + \{Lv \mid v \in \mathbb{R}^n\}$. A very familiar affine subspace of $\mathbb{R}^2$, the graph of a (affine) linear function is specified by the equation $y = \alpha x + \beta$.

Open and closed, etc., sets, limits. A subset $E \subseteq \mathbb{R}^n$ is called *open* when for any $x \in E$ there exists $r > 0$ so that the ball of radius r centered at x,

$$B_r(x) := \{y \in R^n \mid \|y - x\|^2 \leq r\},$$

satisfies $B_r(x) \subset E$.

A subset $F \subset \mathbb{R}^n$ is called *closed* when its complement $\mathbb{R}^n - F$ is open.

A subset $F \subseteq \mathbb{R}^n$ is called *convex* if for any $a, b \in F$ and any $0 \leq \tau \leq 1$ one has $(1 - \tau)a + \tau b \in F$.

A sequence of vectors $S = \{v_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ *converges* to $v \in \mathbb{R}^n$ if for any $r > 0$ the set $B_r(v) - \{v_k\}_{k=1}^{\infty}$ is finite. An equivalent definition (prove the equivalence) is that for any $r > 0$ there exists $N \in \mathbb{N}$ so that $i > N$ implies $\|v - v_i\| < r$. If such v exists it is called the *limit* of S and denoted by

$$v := \lim_{k\to\infty} v_k.$$

---

[1] Excercise: prove them!

[2] When $n = m$, the set of L_w, with L invertible, forms a group, called *affine general linear*, and denoted by $AGL_n(\mathbb{R})$.

Such an S is then called *convergent*.

The sequence S is called *Cauchy sequence* if for any $r > 0$ there exists $N \in \mathbb{N}$ such that $\|v_i - v_j\| < r$ whenever $i > N$ and $j > N$.

A subset $E \in \mathbb{R}^n$ is *bounded* if it is contained in some ball $B_r(x)$. A closed and bounded E is called *compact*.

A $v \in E$ is called a *limit point* of E if there exists a sequence $\{v_k\}_{k=1}^{\infty} \subset E$ so that $v = \lim_{k\to\infty} v_k$.

1.1 Vector-valued functions of several variables

The central place in our course is occupied by "nicely behaved" *functions*[3]

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

Sometimes one has to restrict f so that $f : E \to \mathbb{R}^m$ with $E \subset \mathbb{R}^n$. Usually E must be *open*.

Synonyms to *function* are *mapping*, *map*, *vector field*, etc. However, it is most common to talk about a mapping when f is (affine) linear.

Often one thinks of such an f as of a vector of functions $f_i : \mathbb{R}^n \to \mathbb{R}$, for $1 \leq i \leq m$.

It is useful to consider the *graph* of f, that is, a subset of $\mathbb{R}^m \times \mathbb{R}^n$ (or of $\mathbb{R}^{m+n}$) of the form $\{(f(x), x) \mid x \in \mathbb{R}^n\}$. When bases are fixed in $\mathbb{R}^n$ and $\mathbb{R}^m$, one can say that the graph is given by the *system of equations*

$$\begin{cases} y_1 = f_1(x_1, \ldots, x_n) \\ y_2 = f_2(x_1, \ldots, x_n) \\ \quad \cdots \\ y_m = f_m(x_1, \ldots, x_n). \end{cases} \tag{1.10}$$

Note that the graph of an (affine) linear mapping is an (affine) subspace of $\mathbb{R}^{m+n}$.

1.1.1 Multilinear and alternating multilinear functions.

Let $V = \mathbb{R}^n$. A function $f : V^k \to \mathbb{R}^m$ is called *multilinear* if it is linear with respect to each agrument. In other words,

$$\begin{aligned} f(x_1, \ldots, cx_k, \ldots, x_n) &= cf(x_1, \ldots, x_k, \ldots, x_n) && \text{for all } c \in \mathbb{R}, \\ f(x_1, \ldots, a+b, \ldots, x_n) &= f(x_1, \ldots, a, \ldots, x_n) + f(x_1, \ldots, b, \ldots, x_n) && \forall a, b \in V. \end{aligned}$$

An important distinction is that here we regard f as a function on k arguments, although each of them is an n-vector! Upon fixing a base $\{e_{ij} \mid 1 \leq j \leq n\}$ in

---

[3] Recall that a function $g : A \to B$ is a subset S of $A \times B := \{(a, b) \mid a \in A,\ b \in B\}$ satisfying $b = b'$ whenever $(a, b), (a, b') \in S$.

the i-th copy of V, one can write the i-th vector argument x_i as $x_i = \sum_j x_{ij} e_{ij}$. When all the other arguments fixed, the ℓ-th component f_ℓ of f takes the form $a \cdot x_i = \sum_j a_j x_{ij}$. Hence in general f_ℓ is a homogeneous degree k polynomial in variables x_{ij}, so that two variables x_{ij} and x_{pq} do not occur in the same monomoial when $i = p$, and for each i a variable x_{ij} does occur in the monomial, for some j. That is

$$f_\ell(x_1, \dots, x_k) = \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} a_{j_1,\dots,j_k} x_{1,j_1} x_{2,j_2} \dots x_{k,j_k}.$$

A function $f : V^k \to \mathbb{R}^m$ is called *alternating* if

$$f(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = 0 \qquad \text{for all } i \neq j, \text{ whenever } x_i = x_j.$$

Each multilinear alternating function is *antisymmetric*, that is

$$f(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = -f(x_1, \dots, x_j, \dots, x_i, \dots, x_n) \qquad \text{for all } i \neq j.$$

Indeed, linearity implies

$$f(x_1, \dots, a+b, \dots, a+b, \dots, x_n) = f(x_1, \dots, b, \dots, a, \dots, x_n) +$$
$$+ f(x_1, \dots, a, \dots, b, \dots, x_n) + f(x_1, \dots, a, \dots, a, \dots, x_n) + f(x_1, \dots, b, \dots, b, \dots, x_n),$$

and applying alternacy to both sides of the latter implies

$$0 = f(x_1, \dots, b, \dots, a, \dots, x_n) + f(x_1, \dots, a, \dots, b, \dots, x_n),$$

as claimed.

An important example of antisymmetric multilinear functions is the following bilinear (the prefix "bi-" points out that $k = 2$ here). $f : V \times V \to \mathbb{R}$ is $f(x, y) = x \cdot Ay$, with $A = -A^T \in \mathbb{R}n \times n$. Another (bilinear) example is the cross product $v \times u$ of two vectors u, v in $\mathbb{R}^3$.

$$v \times u = \det \begin{pmatrix} e_1 & e_2 & e_3 \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{pmatrix} = \det \begin{pmatrix} v_2 & v_3 \\ u_2 & u_3 \end{pmatrix} e_1 - \det \begin{pmatrix} v_1 & v_3 \\ u_1 & u_3 \end{pmatrix} e_2 + \det \begin{pmatrix} v_1 & v_2 \\ u_1 & u_2 \end{pmatrix} e_3.$$

The first determinant is actually a 1-form (a linear function) $\phi(e)$ in e_i's, and the 2×2-determinants (with signs) are coefficients of the vector z that represents $\phi(e) = z \cdot e$.

Obviously, k-alternating functions form a vector space. We will denote it by $\wedge^k(V)$.

In algebra one introduces an important operation of *external product* (also called *wedge product*) $\wedge$ that makes a $k + m$-alternating function $f \wedge g$ from a k-alternating form f and an m-alternating form g. This operation is

1. associative: $(f \wedge g) \wedge h = f \wedge (g \wedge h)$,
2. distributive: $(f_1 + f_2) \wedge h = f_1 \wedge h + f_2 \wedge h$, with $f_i \in \wedge^k(V)$

3. skew-commutative: $f \wedge g = (-1)^{km} g \wedge f$.

Note that in the case of the wedge product of 1-forms f_i (each 1-form is trivially alternating) one can define it as

$$f_1 \wedge \cdots \wedge f_k(v^1, \dots, v^k) = \det \begin{pmatrix} f_1(v^1) & \dots & f_k(v^1) \\ \dots & \dots & \dots \\ f_k(v^k) & \dots & f_k(v^k) \end{pmatrix}, \quad v^1, \dots, v^k \in V. \quad (1.11)$$

For instance, let π^i denote the linear function $V \ni x \mapsto x_i$, i.e. it is the i-th coordinate projection. Then by (1.11) one has

$$\pi^1 \wedge \cdots \wedge \pi^k(v^1, \dots, v^k) = \det \begin{pmatrix} v_1^1 & \dots & v_k^1 \\ \dots & \dots & \dots \\ v_1^k & \dots & v_k^k \end{pmatrix}, \quad v^1, \dots, v^k \in V.$$

E.g. when $n = k$ this is the usual determinant, vanishing iff v^j's are linearly independent. Note that the determinant is the *volume* of the parallelepiped spanned by the v^j's.

Generally speaking, k vectors $v^1, \dots, v^k \in V$ are linearly independent iff at least for one k-subset $\{\ell_1, \dots, \ell_k\} \subseteq \{1, \dots, n\}$ the k-alternating form $\pi^{\ell_1} \wedge \cdots \wedge \pi^{\ell_k}(v^1, \dots, v^k) \neq 0$.

Theorem 1.1.1. *The set of all* $\pi^{\ell_1} \wedge \cdots \wedge \pi^{\ell_k}$*, as* $\{\ell_1, \dots, \ell_k\}$ *ranges through* k*-subsets of* $\{1, \dots, n\}$*, forms a basis for* $\wedge^k(V)$*. In particular* $\dim \wedge^k(V) = \binom{n}{k}$*. Let* $L \in \wedge^k(V)$ *and* $e_1, \dots, e_k$ *a basis for* V*. Then*

$$L(v^1, \dots, v^k) = \sum_{1 \le i_1 < \cdots < i_k \le n} L(e_{i_1}, \dots, e_{i_k}) \pi^{i_1} \wedge \cdots \wedge \pi^{i_k}(v^1, \dots, v^k).$$

□

Using the aforementioned properties of $\wedge$ and (1.11), one can form external products of arbitrary (homogeneous) linear combinations of alternating forms.

Limits of functions. Let $f : E \to \mathbb{R}^m$, and $v \in E$ is a limit point of E. We write

$$\lim_{x \to v} f(x) = q$$

and say that the limit of f at v equals $q \in \mathbb{R}^m$ when for any $\epsilon > 0$ there exists $\delta > 0$ satisfyng $\|q - f(x)\| < \epsilon$ for all $x \in E$ with $0 < \|x - v\| < \delta$.

Note that v does not need to be in E.

An equivalent definition is that the preimage $f^{-1}(V)$ of any open subset of $V \subset f(E)$ is open.

Continuity. Our functions will usually be *continuous*: f is continuous at $p \in E$ when for any $\epsilon > 0$ there exists $\delta > 0$ such that $\|f(x) - f(p)\| < \epsilon$ for any $x \in B_\delta(p)$. When f is continuous at any $p \in E$ we say that f is continuous on E.

When $p \in E$ is a limit point of E then f is continuous in p iff $\lim_{x \to p} f(x) = f(p)$.

1.1.2 Differentiability

Continuous functions can still be quite strange, cf. *Cantor function*[4]. In order for functions to be "tame", it is important that locally (in a sufficiently small ball, that is) they behave like linear mappings. Recall the case of $n = m = 1$, when the tangent (the graph of an affine mapping) to a the graph of a differentiable function approximates the function near the point of tangency. One has the *derivative* of f at x given (provided it exists, certainly) by

$$f'(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{1.12}$$

The affine mapping (that defines the tangent at x to the graph of $f(x)$) is given by $y \mapsto f'(x)(y - x) + f(x)$, so the constant term $f(x) - f'(x)x$ is obtained "automatically". It will be convenient to rearrange (1.12) as follows:

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = 0. \tag{1.13}$$

Similarly, in general we try to approximate the (graph of) function $f : \mathbb{R}^n \to \mathbb{R}^m$ by (the graph of) an affine mapping. Note that (1.13) will make sense in this case if we take the norms in the numerator and denominator:

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - f'(x)h\|}{\|h\|} = 0. \tag{1.14}$$

Then we must view $f'(x)$ as a *linear mapping* $\mathbb{R}^n \to \mathbb{R}^m$. It is called the *derivative*[5] of f at x (certainly, it exists iff the limit (1.14) exists). To create the affine mapping that approximates $f(x)$ at x is now easy: it must have the form $y \mapsto f'(x)(y - x) + f(x)$, just as in the univariate case!

Looking at the definition of $f'(x)$ one sees that in some sense it is "defined" on vectors h. To emphasize this, with $x \in E$ one associates a copy of $\mathbb{R}^n$, denoted by TE_x, and called the *tangent space*[6] to E at x. So one can view $f'(x)$ as a linear mapping $f'(x) : TE_x \to T\mathbb{R}^m_{f(x)}$. When $n < m$ this has a natural geometric interpretation, as one can "visualise" $f'(T\mathbb{R}^n_x)$ as the tangent subspace approximating $f(E)$.

Remark 1.1.2. The derivative of an affine mapping L_w is L.

---

[4] Recall that the Cantor function maps the unit interval onto the *Cantor set.*

[5] Also called *total derivative*, or *Fréchet derivative*, or *differential.*

[6] Spivak [2] does not write TE_x, he just writes E_x.

Note that when we talk about the derivate of $f : E \to \mathbb{R}^m$ at $x \in E$, E an open set, we do not have to worry about the choice of h in (1.14), as there is a ball $B_r(x)$ around x that is contained in E, for all sufficiently small $r > 0$.

Lemma 1.1.3. *The derivative $f'(x)$ in (1.14) is unique (provided it exists).*

Proof. Suppose A_1 and A_2 are two derivatives of f at x. Then

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - A_i h\|}{\|h\|} = 0$$

for $i = 1, 2$. On the other hand

$$\|f(x+h)-f(x)-A_1h\| + \|f(x+h)-f(x)-A_2h\| = \|f(x+h)-f(x)-A_1h\| + \\ + \|-f(x+h)+f(x)+A_2h\| \geq \|(A_2 - A_1)h\|$$

by (1.2). Hence $\lim_{h\to 0} \frac{\|(A_2-A_1)h\|}{\|h\|} = 0$, by the usual "squeeze theorem" for limits. In particular we can fix $h \in \mathbb{R}^n$ and write

$$\lim_{\tau \to 0} \frac{\|(A_2 - A_1)\tau h\|}{\|\tau h\|} = 0 = \lim_{\tau \to 0} \frac{\|(A_2 - A_1)h\|}{\|h\|},$$

but by choice of h the latter can only hold when $A_1 = A_2$. □

It is easy to see that for any two diffirentiable at x functions $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ the following familiar in the case $n = m = 1$ formula holds.

$$(f+g)'(x) = f'(x) + g'(x). \tag{1.15}$$

The next important topic is the behaviour of the derivative when one composes two functions. Recall the univariate case (the chain rule):

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

Similarly one has for the composition of $f : \mathbb{R}^n \to \mathbb{R}^m$ and of $g : \mathbb{R}^m \to \mathbb{R}^k$, that is best written as

$$\mathbb{R}^n \xrightarrow{f} \mathbb{R}^m \xrightarrow{g} \mathbb{R}^k,$$

the following

Theorem 1.1.4. *Let $F = g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ (that is $F(x) = g(f(x))$) and let $g'(f(x))$ and $f'(x)$ exist for $x \in \mathbb{R}^n$. Then $F'(x) = g'(f(x)) \cdot f'(x)$.*

`proof will come later`

1.1.3 Partial derivatives

Consider the setting of (1.10), that is, when the bases $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$ and $\{u_1, \dots, u_m\}$ of $\mathbb{R}^m$ are fixed for $f : \mathbb{R}^n \to \mathbb{R}^m$. Another way of writing f_i (f_i's are called *components* of f) is

$$f_i(x) = f(x) \cdot u_i, \qquad \text{where } a \cdot b = \sum_j a_j b_j.$$

Define (provided, this limit exists, certainly) the *i-partial derivative* of f with respect to x_j as follows:

$$\left(\frac{\partial f_i}{\partial x_j}\right)(x) := \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t}.$$

This is of course nothing but the derivative of the univariate function $g(x_j) := f_i(x_1, \dots, x_j, \dots, x_n)$ with respect to x_j, provided that in f each x_k for $k \neq j$, are considered constant.

Remark 1.1.5. The existence of the $\frac{\partial f_i}{\partial x_j}(x)$'s does *not* imply the existence of $f'(x)$. For instance $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\begin{aligned} &f(x, y) = x^3/(x^2 + y^2), \quad (x, y) \neq (0, 0), \\ &f(0, 0) = 0 \end{aligned}$$

does not have $f'(0, 0)$, while $\frac{\partial f}{\partial x}(0, 0)$ and $\frac{\partial f}{\partial y}(0, 0)$ exist.

On the other hand,

Theorem 1.1.6. *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$ *be differentiable at* x*, and* e_j*'s,* u_i*'s as above. Then* $\frac{\partial f_i}{\partial x_j}(x)$ *exists, for* $1 \leq i \leq m$, $1 \leq j \leq n$, *and*

$$f'(x) \cdot e_j = \sum_{i=1}^{m} \frac{\partial f_i}{\partial x_j}(x) u_i.$$

Proof. (Sketch.) Choose $h = \tau e_j$ in (1.14). Then the differentiability of f at x implies

$$\lim_{\tau \to 0} \frac{\|f(x + \tau e_j) - f(x) - \tau f'(x)e_j\|}{\|\tau e_j\|} = 0.$$

By (1.2), $\|f(x + \tau e_j) - f(x) - \tau f'(x)e_j\| = (\|f(x + \tau e_j) - f(x) - \tau f'(x)e_j\| + \|\tau f'(x)e_j\|) - \|\tau f'(x)e_j\| \geq \|f(x + \tau e_j) - f(x)\| - \|\tau f'(x)e_j\|$, implying

$$\lim_{\tau \to 0} \frac{f(x + \tau e_j) - f(x)}{\tau} = f'(x)e_j.$$

Bul the LHS of the latter equals $\lim_{\tau \to 0} \sum_{i=1}^{m} \frac{f_i(x + \tau e_j) - f_i(x)}{t} u_i$, as claimed. □

Notation. We say that f' is *continuous* on an open set $E \subseteq \mathbb{R}^n$ if it is a continuous function $E \to \mathbb{R}^{m\times n}$, that is, for any $x \in E$ and any $r > 0$ there is $\delta > 0$ such that $\|f'(y) - f'(x)\| < r$ if $y \in E \cap B_\delta(x)$.

The set of *continuously differentiable* (i.e., having continuous f') functions $f : \mathbb{R}^n \to \mathbb{R}^m$ (more generally, $f : E \to \mathbb{R}^m$, with $E \subseteq \mathbb{R}^n$) will be denoted by $\mathcal{C}'$ (respectively, $\mathcal{C}'(E)$).

It turns out that $f \in \mathcal{C}'(E)$ iff all the partial derivatives of f on E exist and are continuous. To be able to prove this (as well as some other theorems below), we will need suitable generalisations of the mean value theorem.

Mean value theorem for curves

Recall the classical

Theorem 1.1.7. *(Mean Value theorem.) For a differentiable on $[a, b]$ function $f : \mathbb{R} \to \mathbb{R}$ there exists $a \le x \le b$ satisfying $f'(x) = (f(b) - f(a))/(b - a)$.* □

Here we generalise this to *curves.*

Proposition 1.1.8. *Let $f : \mathbb{R} \to \mathbb{R}^n$ be differentiable on (a, b) and continuous on $[a, b]$. Then there exists $a < x < b$ satisfying*

$$\|f(b) - f(a)\| \le (b - a)\|f'(x)\|.$$

Proof. For $a \le t \le b$, define

$$\phi(t) = (f(b) - f(a)) \cdot f(t).$$

Applying Theorem 1.1.7 to ϕ, one establishes the existenece of $a < u < b$ satisfying

$$\phi(b) - \phi(a) = (b - a)\phi'(t) = (b - a)(f(b) - f(a)) \cdot f'(u).$$

On the other hand

$$\phi(b) - \phi(a) = (f(b) - f(a)) \cdot (f(b) - f(a)) = \|f(b) - f(a)\|^2.$$

By (1.3) and the latter we obtain

$$\|f(b) - f(a)\|^2 = (b - a)(f(b) - f(a)) \cdot f'(u) \le (b - a)\|f(b) - f(a)\|\|f'(u)\|.$$

Cancelling $\|f(b) - f(a)\|$, we obtain the claim. □

Remark 1.1.9. For $n = 1$ Proposition 1.1.8 is implied by Theorem 1.1.7. Indeed,

$$|f(b) - f(a)| \le (b - a) \sup_{a<x<b} |f'(x)|.$$

$\mathcal{C}'$-functions and partial derivatives

We show here the following.

Theorem 1.1.10. *Let* $f : E \to \mathbb{R}^m$*, and* $E \subseteq \mathbb{R}^n$ *is open. Then* $f \in \mathcal{C}'$ *iff* $\frac{\partial f_i}{\partial x_j}$ *exist and are continuous on* E *for* $1 \leq i \leq m$*,* $1 \leq j \leq n$*.*

`proof to be added`

1.2 The inverse function theorem

We know well that a $\mathcal{C}'$ function $f : \mathbb{R} \to \mathbb{R}$ is “locally” 1-to-1, provided we are not close to a where $f'(a) = 0$. (Draw a picture to illustrate this)

More precisely, let $f'(a) \neq 0$ be continuous at a, and $f(a) = b$. Then there are open sets (i.e. open intervals) $U \ni a$, $V \ni b$ so that f is 1-to-1 on U, and $f(U) = V$. And thus there exists

$$g \in \mathcal{C}'(V) : \quad g(f(x)) = x \quad \text{for all} \quad x \in U.$$

Note that $f'(x)^{-1} = g'(f(x))$.

We would like to generalise this result to the case $n = m > 1$. (Certainly, $n = m$ is necessary, otherwise differentiable functions will not provide even local bijections.)

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and $f \in \mathcal{C}'$. The n-dimensional analogy of $f'(a) \neq 0$ above is the *invertibility* of $A = f'(a)$ at $a \in \mathbb{R}^n$.

Theorem 1.2.1. *Let* f *be as above, and* $A = f'(a)$ *invertible. Then*

1. *there exist open sets* $\mathbb{R}^n \supset U \ni a$*,* $\mathbb{R}^n \supset V \ni b = f(a)$*,* f *is 1-to-1 on* U*, and* $f(U) = V$*.*
2. *The inverse of* f *on* U *is given by* $g(f(x)) = x$*, for* $x \in U$*, and* $g \in \mathcal{C}'(V)$*.*

Remark 1.2.2. In the setting of (1.10), with $n = m$, this means that the system (1.10) can be solved for x in terms of y, and this solution is unique and $\mathcal{C}'$.

Our proof will use the following generalisation on the mean value Theorems 1.1.7 and 1.1.8.

Theorem 1.2.3. *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$ *be differentiable on a convex open set* $E \subseteq \mathbb{R}^n$*, so that* $\|f'(x)\| \leq M$ *for all* $x \in E$*. Then*

$$\|f(b) - f(a)\| \leq M\|b - a\| \qquad \textit{for any } a, b, \in E.$$

Proof. Let $\gamma(\tau) = (1 - \tau)a + \tau b$. By convexity of E, for any $0 \leq \tau \leq 1$ one has $\gamma(\tau) \in E$. By the chain rule, Theorem 1.1.4, one has

$$f(\gamma(\tau))' = f'(\gamma(\tau)) \cdot \gamma'(\tau) = f'(\gamma(\tau)) \cdot (b - a),$$

by (1.3) one has

$$\|f(\gamma(\tau))'\| \leq \|f'(\gamma(\tau))\|\|(b-a)\| \leq M\|b-a\|. \tag{1.16}$$

By Theorem 1.1.8,

$$\|f(\gamma(1)) - f(\gamma(0))\| \leq (1-0)\|f(\gamma(t))'\| \tag{1.17}$$

for some $0 < t < 1$. But $f(\gamma(1)) = f(b)$ and $f(\gamma(0)) = f(a)$, so combining (1.16) and (1.17) yields the claim of the statement. ☐

`proof of 1.2.1 to be added`

1.3 The implicit function theorem

Notation. For $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, write $(x,y) = (x_1, \dots, x_n, y_1, \dots, y_m) \in \mathbb{R}^{n+m}$.

Any linear transformation $A : \mathbb{R}^{n+m} \to \mathbb{R}^n$ can be "split" into A_x and A_y, defined by

$$A_x h = A(h,0), \quad h \in \mathbb{R}^n; \qquad A_y k = A(0,k), \quad y \in \mathbb{R}^m.$$

Then $A(h,k) = A_x h + A_y k$. The following a version of implicit function theorem for *linear mappings*.

Theorem 1.3.1. *Let $A : \mathbb{R}^{n+m} \to \mathbb{R}^n$ and A_x invertible. Then for each $k \in \mathbb{R}^m$ there exists unique $h \in \mathbb{R}^m$ satisfying $A(h,k) = 0$. Namely, $h = -A_x^{-1} A_y k$.*

Proof. Let $z = A(h,k) = A_x h + A_y k$. Then $A_x^{-1} z = h + A_x^{-1} A_y k$, implying the claim, by setting $z = 0$. ☐

As we know that the behavour of linear mappings approximates the behavour of "nice" functions, it should not be a big surprise to see the following.

Theorem 1.3.2. *Let $\mathcal{C}'(E) \ni f : E \to \mathbb{R}^n$, with $E \subseteq \mathbb{R}^{n+m}$, such that $f(a,b) = 0$ for some $(a,b) \in E$. Let $A = f'(a,b)$ and assume A_x invertible. Then there exist open sets $(a,b) \in U \subseteq \mathbb{R}^{n+m}$ and $b \in W \subseteq \mathbb{R}^m$, such that for every $y \in W$ there is a unique $x = g(y)$ so that $(x,y) \in U$ and $f(x,y) = 0$.*
Moreover, $g \in \mathcal{C}'(W)$, $g(b) = a$,

$$f(g(y), y) = 0 \qquad \text{for all } y \in W, \tag{1.18}$$

and $g'(b) = -A_x^{-1} A_y$.

We say that g is implicitly defined by (1.18), that's why Theorem 1.3.2 is called *Implicit function Theorem.* One can view $f(x,y) = 0$ as a system of equations

$$\begin{cases} f_1(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \\ \qquad\qquad\qquad\qquad \cdots \\ f_n(x_1, \dots, x_n, y_1, \dots, y_m) = 0 \end{cases} \tag{1.19}$$

and the invertibility of A_x as the invertibility of the $n \times n$ matrix $(\frac{\partial f_i}{x_j})_{1 \leq i,j \leq n}$. When (1.19) holds with $(x, y) = (a, b)$ then Theorem 1.3.2 says that (1.19) can be solved for x in terms of y for any y sufficiently close to b, and these solutions are $\mathcal{C}'$.

One can apply the chain rule to (1.18) to obtain relations, involving $g'(y)$ for $y \in W$, that sometimes can help determining g explicitly.

Example 1.3.3. Let $f : \mathbb{R}^{2+3} \to \mathbb{R}^2$ be given by

$$f_1(x_1, x_2, y_1, y_2, y_3) = 2e^{x_1} + x_2 y_1 - 4y_2 + 3$$
$$f_2(x_1, x_2, y_1, y_2, y_3) = x_2 \cos x_1 - 6x_1 + 2y_1 - y3.$$

Then $f(a, b) = 0$ for $a = (0, 1)$ and $b = (3, 2, 7)$. Then (w.r.t. the standart bases) $A = f'(a, b)$ is

$$\begin{pmatrix} 2 & 3 & 1 & -4 & 0 \\ -6 & 1 & 2 & 0 & -1 \end{pmatrix},$$

where the first 2 columns give A_x and the remaining 3 give A_y. Note that A_x is invertible and Theorem 1.3.2 provides $g \in \mathcal{C}'$ with $g(b) = a$ and $f(g(y), y) = 0$. One can compute, using this theorem, that

$$g'(b) = \begin{pmatrix} \frac{1}{4} & \frac{1}{3} & -\frac{3}{20} \\ -\frac{1}{2} & \frac{6}{5} & \frac{1}{10} \end{pmatrix}.$$

`proof of Theorem 1.3.2 to be added`

1.4 Directional derivative and the gradient

Let $f : E \to \mathbb{R}$ and $\gamma : (a, b) \to E$ be differentiable functions, with $a < b \in \mathbb{R}$ and $E \in \mathbb{R}^n$. We can define its composition and apply the chain rule to it:

$$g(\tau) = f(\gamma(\tau)), \quad g'(\tau) = f'(\gamma(\tau))\gamma'(\tau), \qquad a < \tau < b.$$

So $g'(\tau) : \mathbb{R} \to \mathbb{R}$ is a linear mapping (for fixed τ), that can be regarded as a number in $\mathbb{R}$, too. In terms of partial derivatives (they exist by Theorem 1.1.6) one computes

$$g'(\tau) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\gamma(\tau))\gamma_i'(\tau) = (\nabla f)(\gamma(\tau)) \cdot \gamma'(\tau), \tag{1.20}$$

where we denoted by $(\nabla f)(x)$ the "gradient" of f at x, namely, the vector

$$(\nabla f)(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) e_i,$$

where $\{e_i\}$ is, as usual, a base for $\mathbb{R}^n$.

Next, we specialise $\gamma((a,b))$ to be a *line*, that is, we fix $x \in E$ and $u \in \mathbb{R}^n$ with $\|u\| = 1$, so that

$$\gamma(t) = x + tu, \qquad -\infty < t < \infty, \qquad \gamma'(t) = u.$$

Thus (1.20) gives $g'(0) = (\nabla f)(x) \cdot u$.

On the other hand $g(t) - g(0) = f(x+tu) - f(x)$, so

$$\lim_{t \to 0} \frac{f(x+tu) - f(x)}{t} = (\nabla f)(x) \cdot u. \tag{1.21}$$

The limit in (1.21) is called the *directional derivative* of f at x in the direction u, and is denoted by $(D_u f)(x)$.

Remark 1.4.1. Similar notation is also sometimes used for $\frac{\partial f}{\partial x_i}$, namely one writes

$$(D_i f)(x) = \frac{\partial f}{\partial x_i}(x).$$

A justification for this is the fact that $(D_i f)(x)$ is nothing but $(D_{u_i} f)(x)$, the directional derivative in the direction u_i.

One can write $(D_u f)(x) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) u_i$.

If you fix x, but vary u, then (1.21) shows that $(D_u f)(x)$ attains its maximum when $u = \beta (\nabla f)(x)$, with $\beta > 0$ (provided $(\nabla f)(x) \neq 0$).

1.5 Maxima and minima

On the other hand, you can fix u and vary x. The values of x where $u = \beta(\nabla f)(x)$ holds has an interesting geometric meaning: namely, the points where the linear mapping $x \mapsto u{\cdot}x$ reaches its *local extrema* on the set $\{y \in \mathbb{R}^n \mid f(y) = 0\}$ are like this.

More concretely perhaps, one might like to know maxima and minima of $f : E \to \mathbb{R}$, with $E \subseteq \mathbb{R}^n$.

Theorem 1.5.1. *If a maximum (or a minimum) of f occurs in an interior point $a \in E$, and if $(\nabla f)(a)$ exists, then $(\nabla f)(a) = 0$.*

Proof. Consider the function $f_i : \mathbb{R} \to \mathbb{R}$ given by

$$f_i(t) = f(a_1, \dots, a_{i-1}, t, a_{i+1}, \dots, a_n).$$

Obviously f_i has a maximum (or minimum) at a_i. As a is interior, f_i is defined in an open interval containing a_i. Thus $g_i'(a_i) = \frac{\partial f}{\partial x_i}(a) = 0$, as claimed. ☐

Remark 1.5.2. The converse of Theorem 1.5.1 is false even for $n = 1$ (consider $f(x) = x^3$ at $x = 0$). More intrestingly, consider $f(x,y) = x^2 - y^2$ at $(x,y) = (0,0)$. This is so-called *saddle point.*

Chapter 2

Integration

2.1 Integrable functions

We begin by recalling the Fundamental Theorem of univariate calculus.

Theorem 2.1.1. *Let $f(x) = F'(x)$ be continuous on $[a, b]$, then*

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

Multivariate integration is a very natural generalisation of the univariate definite integral. In the simplest case, given a function $f : E \to \mathbb{R}$ and

$$A = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq E \subset \mathbb{R}^n,$$

we consider the restricted to A graph of f, that is, the set

$$G_A(f) = \{(x, f(x)) \in \mathbb{R}^{n+1} \mid x \in A\}.$$

It cuts out subsets

$$\begin{aligned} V_+ &= \{(x, y) \in \mathbb{R}^{n+1} \mid x \in A,\ 0 \le y \le f(x)\}, \\ V_- &= \{(x, y) \in \mathbb{R}^{n+1} \mid x \in A,\ 0 \ge y \ge f(x)\}. \end{aligned}$$

Then

$$\int_A f(x)dx = \mathrm{vol}(V_+) - \mathrm{vol}(V_-),$$

provided the *volume* $\mathrm{vol}(V)$ is defined for $V \subset \mathbb{R}^m$. We define

$$\mathrm{vol}(A) = \Pi_{i=1}^n \|a_i - b_i\|.$$

Certainly, the volume of the interior of A will be equal to that of A itself. Define

$$\mathcal{L}(V) = \sup_{\mathcal{F}} \sum_{\pi \in \mathcal{F}} \mathrm{vol}(\pi),$$

where $\mathcal{F}$ is a family of disjoint open parallelepipeds π contained in V. Similarly, define

$$\mathcal{U}(V) = \inf_{\mathcal{F}^*} \sum_{\pi\in\mathcal{F}^*} \mathrm{vol}(\pi),$$

where $\mathcal{F}^*$ is a family of disjoint open parallelepipeds π satisfying $V \subseteq \cup_{\pi\in\mathcal{F}^*}\overline{\pi}$. Then

$$\mathrm{vol}(V) := \mathcal{L}(V) = \mathcal{U}(V),$$

provided both quatities exist (and are equal).

Let P be a partition of A. That is, each interval $[a_i, b_i]$ is partitioned as $a_i = c_{i,0} < c_{i,1} < \cdots < c_{i,k_i} = b_i$, so we obtain parallelepipeds (or rectangles)

$$p = [c_{1,j_1}, c_{1,j_1+1}] \times [c_{2,j_2}, c_{2,j_2+1}] \times \cdots \times [c_{n,j_n}, c_{n,j_n+1}]$$

whose union is A. For each $p \in P$ define

$$m_p(f) = \inf_{x\in p} f(x); \qquad M_p(f) = \sup_{x\in p} f(x).$$

Then the *lower* and *upper sums* of f w.r.t. P are

$$L(f,P) = \sum_{p\in P} m_p(f)\,\mathrm{vol}(p); \qquad U(f,P) = \sum_{p\in P} M_p(f)\,\mathrm{vol}(p).$$

Partitions *refine* each other in the natural way, when each subrectangle of one is a union of subrectangles of another.

Lemma 2.1.2. *Let a partition P' refine P. Then*

$$L(f,P) \le L(f,P'), \qquad U(f,P) \ge U(f,P').$$

If Q is another partition (not necessarily a refinement) of A then

$$L(f,Q) \le U(f,P).$$

Definition. A function f is called *integrable* on A if f is bounded on A and

$$I := \sup_P L(f,P) = \inf_Q U(f,Q); \qquad \text{so } \int_A f(x)dx = I.$$

There are bounded, but not integrable functions, e.g. $f : [0,1]^2 \to \mathbb{R}$ so that $f(x,y) = 1$ if $x \in \mathbb{Q}$ and $f(x,y) = 0$ otherwise.

Computing $\int_A f$.

It turns out that in the most reasonable situations $\int_A f$ equals the iterated integral, so that the Fubini's theorem holds. Let $A = B \times C$, and let us write $x = (x_B, x_C)$, for projections x_B of x on B (resp. x_C on C). Assume for simplicity that $I_B = \int_B f(x_B, x_C)dx_B$ exists [1] for all $x_C \in C$.

---

[1] It's actually not needed, but then I_B needs to be replaced by two different "upper" and "lower" integrals.

Theorem 2.1.3. (Fubini) *Let* $B \subset \mathbb{R}^n$ *and* $C \subset \mathbb{R}^m$, *let* $f : B \times C \to \mathbb{R}$ *be integrable and* $f(x_B, x_C)$ *be integrable for any fixed* $x_C \in C$. *Then*

$$\int_{B\times C} f = \int_C \left(\int_B f(x_B, x_C) dx_B \right) dx_C.$$

When the area of integration A is not a rectangle, we still have do define $\int_A f$. The easiest way is to introduce the *characteristic function* χ_A of $A \subset B \subset \mathbb{R}^n$, where B is a rectangle: so that $\chi_A(x) = 1$ for any $x \in A$ and 0 for all $x \in B - A$. Then $\int_A f = \int_B f\chi_A$.

Analytically, one does the following. Let $A = g(B)$ be a compact, for $\mathcal{C}'(B) \ni g : \mathbb{R}^n \to \mathbb{R}^n$. Then for a continuous on A function f

$$\int_A f = \int_{g(B)} f = \int_B f(g(t)) \| \det g'(t) \| dt.$$

For instance, let

$$A = \{(x, y) \mid x^2 + y^2 \leq 1\} \subset \mathbb{R}^2.$$

Then

$$A = \{(r\cos\phi, r\sin\phi) \mid 0 \leq r \leq 1,\ 0 \leq \phi \leq 2\pi\}$$

and $g(r, \phi) = (r\cos\phi, r\sin\phi)$. It is easy to compute $\det(g'(r, \phi)) = r$. Thus

$$\iint_A \exp(x^2 + y^2) dx\, dy = \int_0^{2\pi} d\phi \int_0^1 re^{r^2} dr = \pi \int_0^1 \exp(t)\, dt = \frac{\pi(e-1)}{2}.$$

`something to be proved here, still.`

2.2 Introduction to differential forms via integration

The major players in this quite involved definition are the following [1].

- A compact set $D \subset \mathbb{R}^k$, that for the sake of clarity can be thought to be the unit k-cube.
- An open set $E \subseteq \mathbb{R}^n$; can often be taken to be the whole $\mathbb{R}^n$.
- Functions $\mathcal{C}' \ni \Phi : D \to E$, that are called *k-surfaces*; when $k = 1$ they are usually called just *curves*, and when $k = 2$ – just *surfaces*.
- Continuous functions $a_I : E \to \mathbb{R}$, indexed by $I = (i_1, \ldots, i_k)$, with $1 \leq i_j \leq n$ for $1 \leq j \leq k$. They need not be smooth, in general. It will become clear after a while that I can be taken to be subsets of $\{1, \ldots, n\}$.

Given all this, we introduce formal linear combinations, called *differential forms of order k*, or just *k-forms*:

$$\omega = \sum_{|I|=k} a_I(x),\ dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k},$$

where we often will write $\bigwedge_{i\in I} dx_i$ instead of $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$.

The sum above is taken over all k-tuples of indices $i_j \in \{1, 2, \ldots, n\}$. Thus $\omega = \omega(a_{I_1}, \ldots a_{I_N})$ is determined by the functions a_I.

A k-form ω is a function from the set of k-surfaces to $\mathbb{R}$, essentially "something that one can integrate over" a k-surface, as follows:

$$\Phi \mapsto \int_\Phi \omega := \int_D \sum_{|I|=k} a_I(\Phi(u)) \det \Phi_I'(u)\, du. \tag{2.1}$$

Here we write Ψ_I for the restriction of a k-surface Ψ to the k-tuple of indices $I = (i_1, \ldots, i_k)$, in the order prescribed by I. That is $\Phi_I = (\Phi_{i_1}, \Phi_{i_2}, \ldots, \Phi_{i_k})$. It can be immediately seen from (2.1) that I has all the indices different, without loss of generality. Indeed, if I has a repeated pair of indices then $\det \Phi_I'$, and thus the summand corresponding to a_I, vanishes identically. Later on we will see that I can be chosen to be unordered.

Example 2.2.1. Let $\omega = x\,dy + y\,dx$ be an 1-form, and $\mathcal{C}' \ni \gamma : [0, 1] \to \mathbb{R}^2$ a curve. So we have $a_1(x, y) = y$ and $a_2(x, y) = x$. Then using (2.1) we get

$$\int_\gamma \omega = \int_0^1 (\gamma_1(t)\gamma_2'(t) + \gamma_2(t)\gamma_1'(t))\, dt = \gamma_1(t)\gamma_2(t)|_0^1.$$

It is remarkable that the value of the integral depends ony upon the values of γ in the endpoints of the interval, but not in the rest of them.

Example 2.2.2. let $\omega = dx \wedge dy \wedge dz$ be a 3-form, and Φ a 3-surface. Then $\int_\Phi \omega$ is related to the volume of the body bounded by $\Phi(D)$. For instance, let $D = [0, R] \times [0, \pi] \times [0, 2\pi]$ and

$$\Phi(r, \theta, \phi) = r \cdot (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta).$$

Then using (2.1) one gets $\int_\Phi \omega = 4\pi R^3/3$, the volume of the 3-ball of radius R.

For uniformity it will help to consider also 0-forms, that are just functions $E \to \mathbb{R}$. The k-forms $\wedge_{i\in I} dx_i$ are called *basic*. They are an important object in algbera, namely in tensor algebra, where they are called *antisymmetric k-tensors.*

2.2.1 Some properties of k-forms: antisymmetricity, etc

Given two k-forms ω_1 and ω_2, we call then *equal* if $\int_\Phi \omega_1 = \int_\Phi \omega_2$ for any k-surface Φ. The following equalities are immediate from (2.1).

$$\int\limits_\Phi c\omega = c\int\limits_\Phi \omega, \qquad \text{for any } c \in \mathbb{R}, \tag{2.2}$$

$$\int\limits_\Phi (\omega_1+\omega_2) = \int\limits_\Phi \omega_1 + \int\limits_\Phi \omega_2. \tag{2.3}$$

Consider two almost identical k-forms

$$\omega = a(x)\bigwedge_{i\in I} dx_i = a(x)\,dx_{i_1}\bigwedge\cdots\wedge\ dx_{i_j}\wedge\cdots\wedge dx_{i\ell}\wedge\cdots\wedge dx_{i_k},$$

$$\omega' = a(x)\bigwedge_{i\in I'} dx_i = a(x)\,dx_{i_1}\wedge\cdots\wedge\ dx_{i_\ell}\wedge\cdots\wedge dx_{i_j}\wedge\cdots\wedge dx_{i_k}.$$

That is, I' is obtained from I by swapping i_j and i_ℓ, and this is the only difference. Then, as the determinant changes its sign when two of its rows are swapped, one observes from (2.1) that

$$dx_1\wedge dx_2 = -dx_2\wedge dx_1, \quad \text{and in general } \omega = -\omega'. \tag{2.4}$$

This is called *anticommutativity* of the operation $\wedge$, to emphasise the difference with operations like $+$.

For instance, (2.4) allows one to rewrite

$$x_1dx_2\wedge dx_1 - x_2\,dx_3\wedge dx_2 + x_3\,dx_2\wedge dx_3 + dx_1\wedge dx_2 = (1-x_1)dx_1\wedge dx_2 + (x_1+x_3)dx_2\wedge dx_3.$$

A representation of a k-form ω with all the I's being ordered, i.e. $i_1 < i_2 < \cdots < i_k$, is called *standard.*

Proposition 2.2.3. *Let a k-form ω be in a standard representation*

$$\omega = \sum_{I\in\binom{\{1,\dots,n\}}{k}} a_I(x)\bigwedge_{i\in I} dx_i.$$

Then $\omega = 0$ if and only if each $a_I = 0$.

Proof. Let $a_I(v) > 0$ for some $v \in E$. By continuity of a_I, there exists $h > 0$ so that $a_I(w) > 0$ for any w in the n-cube with centre at v and side size h. We can and will assume without loss in generality that D is the k-cube with side size h. Observe that the k-surface

$$\Phi(u) = v + \sum_{j=1}^k u_j e_{i_j}, \quad \text{for } I = (i_1,\dots,i_k),$$

satisfies $a_I(\Phi(u)) > 0$ for any $u \in D$. On the other hand, the term corresponding in (2.1) to any $I \neq I' \in \binom{\{1,\dots,n\}}{k}$ is identically 0, as the corresponding Jacobian vanishes. Indeed, $\frac{\partial\Phi}{\partial u_\ell} = 0$ for any $\ell \notin I$.

The Jacobian of Φ_I is 1. Hence $\int_\Phi \omega = \int_D a_I(\Phi(u))\,du$, a constradiction, as $a_I(\Phi(u)) > 0$. ☐

Note in particular that there is essentially unique nonzero basic n-form, see Example 2.2.2 for the case $n = 3$.

2.2.2 Products of k-forms

Given a basic p-form $\wedge_{i\in I} dx_i$ and a basic q-form $\wedge_{j\in J} dx_j$, one defines a

$$\omega = \bigwedge_{i\in I} dx_i \wedge \bigwedge_{j\in J} dx_j,$$

that is a $p+q$-form if $I \cap J = \emptyset$, and identically 0 otherwise (and so it can be viewed as $p+q$-form anyway, the zero $p+q$-form).

Lemma 2.2.4. *Let α denote the number of pairs (s,t), with $s \in I$, $t \in J$, and $s > t$. Then*

$$\omega = \bigwedge_{i\in I} dx_i \wedge \bigwedge_{j\in J} dx_j = (-1)^\alpha \bigwedge_{\ell\in I\cup J} dx_\ell,$$

where $I \cup J$ is ordered. ☐

Using this statement, we define the wedge product of a p-form $\omega = \sum_I a_I(x) \wedge_{i\in I} dx_i$ and a q-form $\lambda = \sum_J b_J(x) \wedge_{j\in J} dx_j$.

$$\omega \wedge \lambda = \sum_{I,J} a_I(x) b_J(x) \bigwedge_{i\in I} dx_i \wedge \bigwedge_{j\in J} dx_j. \tag{2.5}$$

Obviously, $p + q > n$ automatically implies $\omega \wedge \lambda = 0$. One can further see that $\wedge$ behaves like the "usual multiplication" with respect to addition of k-forms, and repeated products. That is, the following holds.

$$(\omega \wedge \lambda) \wedge \sigma = \omega \wedge (\lambda \wedge \sigma), \tag{2.6}$$

$$(\omega_1 + \omega_2) \wedge \sigma = \omega_1 \wedge \sigma + \omega_2 \wedge \sigma, \tag{2.7}$$

$$\sigma \wedge (\omega_1 + \omega_2) = \sigma \wedge \omega_1 + \sigma \wedge \omega_2. \tag{2.8}$$

2.2.3 Differential operator

There is also an additive, with respect to addition of k-forms, differential operator d that makes $(k+1)$-forms from k-forms, as follows. We need to assume as well that the finctions a_I are differentaible (sufficiently many times). For 0-form f, define

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\, dx_i. \tag{2.9}$$

Note that the familiar $d(fg) = f\,dg + g\,df$ will hold, then. Certainly, df is nothing but $f'(x)$, written in the fancy basis $\{dx_i\}$.

For a k-form ω define

$$d\omega = d(\sum_I a_I(x) \bigwedge_{i\in I} dx_i) = \sum_I (da_I(x)) \wedge \bigwedge_{i\in I} dx_i). \tag{2.10}$$

So "only functions are touched" by d. For instance $d(x\,dy) = dx \wedge dy$.

Proposition 2.2.5. *Let ω be a k-form and λ - a m-form. Then*

$$d(\omega \wedge \lambda) = (d\omega) \wedge \lambda + (-1)^k \omega \wedge (d\lambda), \tag{2.11}$$

$$d(d(\omega)) = 0. \tag{2.12}$$

Proof. To prove (2.11), it suffices to prove it for a k-form $\omega = f \wedge_{i\in I} dx_i$ and an m-form $\lambda = g \wedge_{j\in J} dx_j$. Then w.l.o.g. $I \cap J = \emptyset$, otherwise $\omega \wedge \lambda$ is identically 0 and also the right hand side of (2.11), by definition of d. By Lemma 2.2.4

$$\omega \wedge \lambda = fg \bigwedge_{i\in I} dx_i \wedge \bigwedge_{j\in J} dx_j = (-1)^\alpha fg \bigwedge_{\ell\in I\cup J} dx_\ell.$$

As $d(fg) = f\,dg + g\,df$, we obtain

$$d(\omega \wedge \lambda) = (-1)^\alpha (f\,dg + g\,df) \bigwedge_{\ell\in I\cup J} dx_\ell = (f\,dg + g\,df) \bigwedge_{i\in I} dx_i \wedge \bigwedge_{j\in J} dx_j.$$

Finally we derive, using anticommutativity k times, that

$$dg \wedge \bigwedge_{i\in I} dx_i = (-1)^k \bigwedge_{i\in I} dx_i \wedge dg,$$

implying

$$d(\omega \wedge \lambda) = (df \wedge \bigwedge_{i\in I} dx_i) + (-1)^k (f \bigwedge_{i\in I} dx_i) \wedge (dg \wedge \bigwedge_{j\in J} dx_j).$$

This proves (2.11).

To prove the remaining part of the Proposition, observe that for a 0-form f, the rule (2.12) is a simple consequence of $\frac{\partial^2 f}{\partial x_i}\partial x_j = \frac{\partial^2 f}{\partial x_j}\partial x_i$ and anticommutativity of $\wedge$. Proceeding to $k > 0$, a k-form $\omega = f \wedge_{i\in I} dx_i$, we know that $d\omega = df \wedge \bigwedge_{i\in I} dx_i$. Thus by (2.11) we have

$$d(d\omega) = d(df) \wedge \bigwedge_{i\in I} dx_i + (-1)^\ell df \wedge d(\bigwedge_{i\in I} dx_i) = 0.$$

Indeed, the first summand vanishes as, by above $d(df) = 0$, and the second summand vanishes as $d(\bigwedge_{i\in I} dx_i) = 0$ by definition of d in (2.10). ☐

2.2.4 Change of variables in k-forms

Next, we investigate the behavour of k-forms under the change of variables. Let $E \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^m$ be open and $\mathcal{C}' \ni T : E \to V$ provides such a change. In V let us have a k-form ω in standard representation $\omega = \sum_I b_I(y) \bigwedge\limits_{i\in I} dy_i$.

We make 1-forms from the components T_i of T (so that $y_i = T_i(x)$) as defined in (2.9).

$$dT_i = \sum_{j=1}^{n} \frac{\partial T_i}{\partial x_j}(x)dx_j, \quad 1 \le i \le m.$$

Using this, we define a k-form ω_T in E, as follows.

$$\omega_T = \sum_I b_I(T(x)) \bigwedge_{i\in I} dT_i.$$

It turns out that this operation is well-behaved with respect to the operations on differential forms we introduced before.

Proposition 2.2.6. *Let T be suffciently many times differentiable, ω a k-form on V, and λ an ℓ-form on V. Then*[2]

1. *for $k = \ell$, one has $(\omega + \lambda)_T = \omega_T + \lambda_T$.*
2. *$(\omega \wedge \lambda)_T = \omega_T \wedge \lambda_T$.*
3. *$d(\omega_T) = (d\omega)_T$.* □

The most important property is

Proposition 2.2.7. *With T and ω as above, Φ a k-surface in E, one has*

$$\int_{T(\Phi)} \omega = \int_{\Phi} \omega_T.$$

2.3 Differential forms via tangent spaces

Here, unlike above, we define differential forms as multilinear alternating functions on tangent spaces.

We can view a k-form ω on E (notation: $\omega \in \Omega^k(E)$) with a_I having derivatives of all orders as alternating multilinear k-forms on the tangent spaces, as in e.g. [3].

$$\omega(x) : (TE_x)^k \to \mathbb{R} \qquad \text{for each } x \in E.$$

In other words, $\omega(x) \in \wedge^k(TE_x)$. In general, $\omega(x)$ does not stay the same as x changes. One should not confuse ω and $\omega(x)$ – they are different objects, that is, $\omega(x)$ is the "value" of ω at x.

---

[2] In [1, Thm 10.22], one should replace "m-form" by "ℓ-form", to avoid notation clash.

For instance, for a differentiable $f : E \to \mathbb{R}$ one has the 1-form $df := f'$ defined by $f'(x) : TE_x \to \mathbb{R}$. Namely,

$$f'(x) \cdot h = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} h_i.$$

For the projection π^i one has $\pi^i(x) \cdot h = h_i$. For k differentiable functions f_1, …, f_k, one can use (1.11) to get

$$df_1 \wedge \cdots \wedge df_k(h^1, \ldots, h^k) = \det \begin{pmatrix} df_1(h^1) & \ldots & df_k(h^1) \\ \cdots & & \\ df_1(h^k) & \ldots & df_k(h^k) \end{pmatrix}, \quad h^1, h^2, \ldots, h^k \in TE_x.$$

Thus, from k one-forms df_i we obtained a k-form.

In particular, for $f_i = \pi^{\ell_i}$ one gets

$$d\pi^{\ell_1} \wedge \cdots \wedge \pi^{\ell_k}(h^1, \ldots, h^k) = \det \begin{pmatrix} h^1_{\ell_1} & \ldots & h^1_{\ell_k} \\ \cdots & & \\ h^k_{\ell_1} & \ldots & h^k_{\ell_k} \end{pmatrix}, \quad h^1, h^2, \ldots, h^k \in TE_x. \tag{2.13}$$

Often we simply write dx_{ℓ_j} instead of $d\pi^{\ell_j}$.

Generally speaking, we get for $\omega(x) = a(x)\, d\pi^{\ell_1} \wedge \cdots \wedge d\pi^{\ell_k}$, using Theorem 1.1.1, and denoting the basis for TD_x as $\{e_1(x), \ldots, e_n(x)\}$, we can write

$$\omega(x)(h^1, h^2, \ldots, h^k) = \sum_{1 \le \ell_1 < \cdots < \ell_k \le n} \omega(e_{\ell_1}(x), \ldots, e_{\ell_k}(x)) \\ \pi^{\ell_1} \wedge \cdots \wedge d\pi^{\ell_k}(h^1, h^2, \ldots, h^k), \ h^j \in V.$$

As $\omega(e_{\ell_1}(x), \ldots, e_{\ell_k}(x))$ depends upon x alone, we see that indeed it is $a(x)$.

2.3.1 Tangent spaces, k-surfaces, and coordinate changes.

Now, when we have a smooth k-surface $\Phi : D \to E$, where we denote $S = \Phi(D)$, we can consider the tangent space $TS_x \subseteq TE_x$ for any $x \in S$, and the natural restriction of a p-form $\omega(x)$ (w.l.o.g. $p \le k$) onto TS_x. Our next task is to see what happens with ω under the coordinate change defined by Φ. This will allow us here (and in the next subsection for the integral) to see the equivalence of two definitions of k-forms. Note that in [1]the notation ω_Φ is used for what we here (and [3]) denote by $\Phi^*\omega$.

Let $f : E \to \mathbb{R}$. Then $\Phi^* f : D \to \mathbb{R}$, by setting

$$(\Phi^* f)(t) := f(\Phi(t)), \qquad t \in D.$$

Thus Φ maps the functions (i.e., 0-forms) on E into functions (i.e., 0-forms) on D via $f \mapsto \Phi^* f$, i.e. $\Phi^* : \Omega^0(E) \to \Omega^0(D)$.

Now for p-forms we see that $\Phi^* : \Omega^p(E) \to \Omega^p(D)$. Let $t \in D$. Then we have the map of tangent spaces $\Phi'(t) : TD_t \to TE_{\Phi(t)}$. Define

$$\Phi^*\omega(t)(\tau^1, \ldots, \tau^p) = \omega(\Phi(t))(\Phi'(t)\tau^1, \ldots, \Phi'(t)\tau^p), \ \tau^j \in TD_t.$$

Remark 2.3.1. Moreover it follows from the chain rule, Theorem 1.1.4, that for

$$D \xrightarrow{\Phi} E \xrightarrow{\Psi} F$$

one has

$$(\Psi \circ \Phi)^* = \Phi^* \circ \Psi^* \tag{2.14}$$

and thus

$$\Omega^p(D) \xleftarrow{\Phi^*} \Omega^p(E) \xleftarrow{\Psi^*} \Omega^p(F).$$

Let us see how the p-forms change under coordinates changes. For instance, take the 2-form $\omega = dx_1 \wedge dx_2$, and a coordinate change Φ given by $x_i = x_i(t^1, \ldots, t^k)$, $1 \leq i \leq n$. We would like to find $\Phi^*\omega$. Take $t \in D$ and $\tau_1, \tau_2 \in TD_t$. Then in $TE_{\Phi(t)}$ their images are $\rho_i = \Phi^*(t)\tau_i$. Thus

$$\begin{aligned}
\Phi^*\omega(\tau_1, \tau_2) = \omega(\Phi(t))(\rho_1, \rho_2) = dx_1 \wedge dx_2(\rho_1, \rho_2) = \\
\det\begin{pmatrix} \rho_1^1 & \rho_1^2 \\ \rho_2^1 & \rho_2^2 \end{pmatrix} = \det\begin{pmatrix} \sum_{j=1}^k \frac{\partial x_1}{\partial t^j}\tau_1^j & \sum_{j=1}^k \frac{\partial x_2}{\partial t^j}\tau_1^j \\ \sum_{j=1}^k \frac{\partial x_1}{\partial t^j}\tau_2^j & \sum_{j=1}^k \frac{\partial x_2}{\partial t^j}\tau_2^j \end{pmatrix} = \\
= \sum_{j,j'=1}^k \frac{\partial x_1}{\partial t^j}\frac{\partial x_2}{\partial t^{j'}} \det\begin{pmatrix} \tau_1^j & \tau_1^{j'} \\ \tau_2^j & \tau_2^{j'} \end{pmatrix} = \\
= \sum_{j,j'=1}^k \frac{\partial x_1}{\partial t^j}\frac{\partial x_2}{\partial t^{j'}} dt^j \wedge dt^{j'}(\tau_1, \tau_2) = \\
= \sum_{1 \leq j < j' \leq k} \left(\frac{\partial x_1}{\partial t^j}\frac{\partial x_2}{\partial t^{j'}} - \frac{\partial x_1}{\partial t^{j'}}\frac{\partial x_2}{\partial t^j}\right) dt^j \wedge dt^{j'}(\tau_1, \tau_2) = \\
= \sum_{1 \leq j < j' \leq k} \phi'(t)_{j,j'}\, dt^j \wedge dt^{j'}(\tau_1, \tau_2).
\end{aligned}$$

Thus

$$\Phi^*\omega = \sum_{1 \leq j < j' \leq k} \phi'(t)_{j,j'}\, dt^j \wedge dt^{j'}.$$

Using $\Phi^*(a(x)\omega) = a(\Phi(t))\Phi^*\omega$, one derives for a k-form $\omega = a(x)dx_1 \wedge \cdots \wedge dx_k$:

$$\Phi^*\omega = \sum_{1 \leq j_1 < \cdots < j_k \leq n} a(\Phi(t))\Phi'(t)_{j_1,\ldots,j_k}\, dt^{j_1} \wedge \cdots \wedge dt^{j_k}.$$

2.3.2 Alternative definition of $\int_\Phi \omega$

Here we show that the definition of the integral $\int_\Phi \omega$ can also be given via tangent spaces.

Let $D = [0,1]^k \subset \mathbb{R}^k$ and $\Phi : D \to E$ smooth, ω a k-form on $\Phi(D)$. Consider a partition P of D induced by partitions of each $[0,1]$, and in each sub-k-cube I_i

take the vertex t_i with the minimal indexes of the coordinates. Associate to t_i vectors τ_1,...,τ_k joining t_i and its k adjacent vertices of I_i.

In the tangent space $T\Phi(D)_{x_i}$, for $x_i = \Phi(t_i)$, take the images $\rho_j = \Phi'(t_i)\tau_j$ of τ_j's under the mapping Φ' of tangent spaces. Compute

$$(\Phi^*\omega)(t_i)(\tau_1, \dots, \tau_k) := \omega(x_i)(\rho_1, \dots, \rho_k).$$

Let $\lambda(P) = \max_i \operatorname{vol} I_i$. Then we *define*

$$\int\limits_{\Phi} \omega \quad = \quad \lim_{\lambda(P)\to 0} \sum_i \omega(x_i)(\rho_1, \dots, \rho_k) \quad = \quad \lim_{\lambda(P)\to 0} \sum_i (\Phi^*\omega)(t_i)(\tau_1, \dots, \tau_k).$$

In particular when $\Phi = id$ and $\omega = f(t)\, dt^1 \wedge \cdots \wedge dt^k$ we have by the definition of the usual multiple integral that

$$\int_D f(t)\, dt^1 \wedge \cdots \wedge dt^k = \int_D f(t)\, dt^1 dt^k. \tag{2.15}$$

Thus we have $\int_\Phi \omega = \int_D \Phi^*\omega$, and the latter is just the usual multiple integral.

Now, to consider the case of general compact $D \subset \mathbb{R}^k$, we observe that here also (2.15) makes sense. Then we can define

$$\int_{\Phi(D)} \omega := \int_D \Phi^*\omega.$$

When $S = \cup_j S_j$ is a piecewise smooth union of smooth pieces S_j, each of them is k-surface $S_j = \Phi_j(D)$, define

$$\int_S \omega = \sum_j \int_{S_j} \omega.$$

Here we certainly need that the pieces S_j intersect in subsets of dimension $< k$.

This raises a number of questions, in particular, whether the integral is dependent upon a particular parametrisation. To begin with, let $\omega = f(x)\, dx^1 \wedge \cdots \wedge dx_k$, and $\Phi : G \to D$ – a diffeomorphism (i.e. $\det \Phi' \neq 0$). On one hand $\int_D \omega$ is given by (2.15). On the other hand

$$\int_D \omega = \int_G \Phi^*\omega = \int_G f(\Phi(t)) \det \Phi'(t) dt^1 \dots dt^k.$$

As w.l.o.g. $\det \Phi'(t) > 0$, by the theorem on variable change in multiple integral we have that the latter equal $\int_D f(x)\, dx^1 \dots dx^k$, as required. (we cheated with the sign a bit, we could have gotten the result with the minus sign instead. Be aware of the orientation!)

Now let $\Phi : D \to S$ and $\Psi : D \to S$ be two different parametrisations of a k-surface S. Then $\Gamma := \Phi^{-1} \circ \Psi : D \to D$ is a diffeomorphism. Then $\Phi^*\omega$ can be obtained from $\Psi^*\omega$ using the change of variables Γ, as $\Gamma^*\Phi^*\omega = (\Phi \circ \Gamma)^*\omega = \Psi^*\omega$ by (2.14).

2.4 Low-dimensional case

Here we concentrate on k-forms in $\mathbb{R}^2$ and $\mathbb{R}^3$.

2.4.1 1-forms and 2-forms in $\mathbb{R}^2$. Green's theorem.

Here we elaborate upon Example 2.2.1. One important question that we like to address and illustrate in this simple setting is what caused the integral in the example to be independent upon the intermediate values along the path. We will see that this is connected to the question whether the form ω is representable as df for some 0-form f.

A general 1-form in an open set $U \subseteq \mathbb{R}^2$ is

$$\omega = p_1(x)dx_1 + p_2(x)dx_2, \qquad \text{for } x = (x_1, x_2).$$

The function $(p_1, p_2) : U \to \mathbb{R}^2$ is often called a *vector field.*

Curves $\gamma : [a, b] \to U$ will be assumed to be continuos on $[a, b]$ and differentiable on (a, b). We shall also be assuming that γ can be "extended" a bit outside $[a, b]$, i.e. that γ is a restriction to $[a, b]$ of some continuous on $(a - \epsilon, b + \epsilon)$ function, with $\epsilon > 0$. Such curves will be called *smooth paths* with initial point $\gamma(a)$ and final point $\gamma(b)$ (that will also be called *endpoints*). We have

$$\int\limits_\gamma \omega = \int\limits_a^b \left(p_1(\gamma(t))\frac{d\gamma_1}{dt} + p_2(\gamma(t))\frac{d\gamma_2}{dt} \right) dt.$$

Lemma 2.4.1. *Let* $\omega = df$ *for some* $\mathcal{C}' \ni f : U \to \mathbb{R}$*. Then* $\int_\gamma \omega = f(\gamma(t)) \mid_a^b$.

Proof. By the chain rule

$$\frac{d}{dt}f(\gamma(t)) = \frac{\partial f}{\partial x_1}(\gamma(t))\frac{d\gamma_1}{dt} + \frac{\partial f}{\partial x_2}(\gamma(t))\frac{d\gamma_2}{dt}.$$

Thus by Theorem 2.1.1,

$$\int_\gamma \omega = \int_a^b \frac{d}{dt}f(\gamma(t))dt = f(\gamma(t)) \mid_a^b,$$

as required. □

For instance in Example 2.2.1 we had $\omega = d(x_1 x_2)$.

Excercise. Show that $df = dg$ on U if and only if $f - g$ is locally constant on U (locally constant means each $u \in U$ has a neighbourhood where $f - g$ is constant).

How does U matter? Let

$$\theta = (x_1^2 + x_2^2)^{-1}(-x_2 dx_1 + x_1 dx_2) \qquad \text{in } \mathbb{R}^2 - \{(0)\}. \tag{2.16}$$

Then for the (closed) path $\gamma(t) = (\sin t, \cos t)$, $\gamma : [0, 2\pi] \to \mathbb{R}^2 - \{(0)\}$ one computes $\int_\gamma \theta = 2\pi$, so by Lemma 2.4.1 $\theta \neq df$, but on the other hand

$$d\left(\tan \frac{x_2}{x_1}\right)^{-1} = \theta \qquad \text{for } x_1 > 0.$$

So Lemma 2.4.1 will work for θ if we restrict U to be the open right halfplane.

2.4.2 Segmented paths

Let γ_ℓ, $1 \leq \ell \leq n$ be n smooth paths in U so that the final point of each γ_ℓ is the inital point of $\gamma_{\ell+1}$ for $\ell < n$. We concatenate them into a *segmented path* γ with initial point $P = \gamma_1(a_1)$ and final point $Q = \gamma_n(b_n)$. Then it is natural to write $\gamma = \sum_\ell \gamma_\ell$ and set

$$\int_\gamma \omega = \sum_{\ell=1}^{n} \int_{\gamma_\ell} \omega.$$

In particular it follows from Lemma 2.4.1 that $\int_\gamma df = f(Q) - f(P)$. The converse is also true. Namely the following holds.

Proposition 2.4.2. *Let ω be a 1-form on U. Then the following are equivalent.*

- *(i)* $\int_\gamma \omega = \int_\delta \omega$ *for all segmented paths γ, δ in U with the same initial and final points;*
- *(ii)* $\int_\gamma \omega = 0$ *for all segmented closed paths γ in U;*
- *(iii)* $\omega = df$ *for some smooth $f : U \to \mathbb{R}$.*

Proof. We just observed that (*iii*) implies (*i*). To show that (*ii*) implies (*i*), we need the notion of *inverse* σ^{-1} of the path $\sigma : [a, b] \to U$ defined by $\sigma^{-1}(t) = \sigma(a + b - t)$. Obviously $\int_\sigma \omega = -\int_{\sigma^{-1}} \omega$. As $\omega + \delta^{-1}$ is closed, (*ii*) implies (*i*). The opposite direction is obvious.

It remains to show that (*i*) implies (*iii*). It suffices to consider the case of connected U, for we can take different f's on different components. Let $P_0 \in U$, and define $f : U \to \mathbb{R}$ by $f(P) = \int_\gamma \omega$, for $\gamma : [0, 1] \to U$ a segmented path with initial point P_0 and the final point P. By assumption f is well-defined. We claim that $\omega := \sum_i p_i(x)dx_i = \sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(x)dx_i$. We will show that

$$\lim_{w \to 0} \frac{f(x_1 + w, x_2) - f(x_1, x_2)}{w} = \frac{\partial f}{\partial x_1}(P), \qquad P = (x_1, x_2).$$

(the proof for the $\frac{\partial f}{\partial x_2}$ is basically the same, and is omitted.)

Let $w > 0$ and σ be the path from P to (x_1+w, x_2) given by $\sigma(t) = (x_1+t, x_2)$, and γ any segmented path from P_0 to P. Integrating along the segmented path $\gamma + \sigma$ we get

$$\begin{aligned}\frac{f(x_1+w,x_2)-f(x_1,x_2)}{w} &= \frac{1}{w}\left(\int\limits_{\gamma+\sigma}\omega - \int\limits_{\gamma}\omega\right) = \frac{1}{w}\int\limits_{\sigma}\omega = \\ &= \frac{1}{w}\int_0^w p_1(x_1+t,x_2)dt = \\ &= p_1(x_1+w^*,x_2),\ 0 \le w^* \le w,\end{aligned}$$

where the last equality is obtained by the mean value theorem. Thus

$$\lim_{w\to 0}\frac{1}{w}\int_0^w p_1(x_1+t,x_2)dt = p_1(x_1,x_2),$$

as required. The case $w < 0$ is treated very similarly. □

2.4.3 When does $\omega = df$, and 2-forms.

As for $f : U \to \mathbb{R}$ we have $\frac{\partial^2 f}{\partial x_1\partial x_2} = \frac{\partial^2 f}{\partial x_2\partial x_1}$ we obtain the following necessary condition on $\omega = df$.

Lemma 2.4.3. *Let $\omega = p_1(x)dx_1 + p_2(x)dx_2$. For $\omega = df$ it is necessary that $\frac{\partial p_2}{\partial x_1} = \frac{\partial p_1}{\partial x_2}$.* □

This somehow complicated condition has a better expression in terms of 2-forms. A 2-form ζ in U written in standard form is just $\zeta = h\, dx_1 \wedge dx_2$, for $\mathcal{C}' \ni h : U \to \mathbb{R}$. When ζ is obtained by differentiating a 1-form, we get

$$\begin{aligned}\zeta = d\omega &= dp_1(x)\wedge dx_1 + dp_2(x)\wedge dx_2 = \\ &= \left(\frac{\partial p_1}{\partial x_1}dx_1 + \frac{\partial p_1}{\partial x_2}dx_2\right)\wedge dx_1 + \left(\frac{\partial p_2}{\partial x_1}dx_1 + \frac{\partial p_2}{\partial x_2}dx_2\right)\wedge dx_2 = \\ &= \left(\frac{\partial p_2}{\partial x_1} - \frac{\partial p_1}{\partial x_2}\right)dx_1\wedge dx_2.\end{aligned} \tag{2.17}$$

So the the condition of Lemma 2.4.3 can be rewritten as $d\omega = 0$. This is a particular case of (2.11).

The example (2.16) shows that this condition is not sufficient. It turns out that U really matters here, too. For “easier” than $\mathbb{R}^2 - \{0\}$ open sets (in fact, for these ones that do not have closed paths that are not 0-homotopic) the criterion of Lemma 2.4.3 is sufficient. Namely, the following holds.

Theorem 2.4.4. *Let $U = (a_1, b_1) \times (a_2, b_2)$, that is*

$$U = \{(x_1,x_2) \mid a_1 < x_1 < b_1,\ a_2 < x_2 < b_2\},\ -\infty \le a_i < b_i \le \infty.$$

Then the following holds.

- (Green's theorem for rectangles). *Denoting by ∂U the closed rectangular path with corners $(a_1, b_1), (a_2, b_1), (a_2, b_2), (a_2, b_1)$, and assuming U finite,*

$$\int_{\partial U} \omega = \iint_{\overline{U}} d\omega.$$

- *If an 1-form $\omega = p_1(x)\, dx_1 + p_2(x)\, dx_2$ on U satisfies $d\omega = 0$ then there exists $f : U \to \mathbb{R}$ with $\omega = df$.*

Proof. To prove the first item, write out $d\omega$ as in (2.17), and integrate using Fubini Theorem 2.1.3 and Theorem 2.1.1:

$$\iint_{\overline{U}} \frac{\partial p_2}{\partial x_1} dx_1\, dx_2 = \int_{a_2}^{b_2} (p_2(b_1, t) - p_2(a_1, t))dt \tag{2.18}$$

$$\iint_{\overline{U}} \frac{\partial p_1}{\partial x_2} dx_1\, dx_2 = \int_{a_1}^{b_1} (p_1(t, b_2) - p_1(t, a_2))dt. \tag{2.19}$$

Subtracting (2.19) from (2.18) gives the claimed formula.

To prove the last item, fix $P_0 = (p_1, p_2) \in U$. Assume $x_i \geq p_i$, ($i = 1, 2$; the remaining cases are treated similarly). For any $P = (x_1, x_2) \in U$, set $f(P) = \int_\gamma \omega$ (respectively, $g(P) = \int_{\gamma^*} \omega$), where γ (respectively, γ^*) is the two-segment path from P_0 to (p_1, x_2) (respectively, to (x_1, p_2)) and then to P. So we change the 2nd (resp. 1st) coordinate first, and then the 1st (resp. 2nd) one. Proceeding as in the proof of Prop. 2.4.2 (*iii*), we derive that $\frac{\partial f}{\partial x_1} = p_1$ and $\frac{\partial g}{\partial x_2} = p_2$. We now apply Green's theorem for the rectangle formed by γ and γ^* with $d\omega = 0$ to obtain $f(P) = g(P)$. Thus $\frac{\partial f}{\partial x_2} = p_2$, that is $\omega = df$. ☐

Green's theorem for arbitrary compact region A can be derived from the one for the rectangle by considering partitions of increasingly tight approximations of A into subrectangles and taking appropriate sup and inf. The orientations of subrectangles must be taken into account, certainly. A better proof will follow from Stokes Theorem 2.5.1 below.

2.4.4 1-, 2-, and 3-forms in $\mathbb{R}^3$.

We continue to assume functions sufficiently many times differentiable. Given an open set $U \subset \mathbb{R}^3$, we have

- 0-forms – they are just functions $f : U \to \mathbb{R}$;
- 1-forms – $\phi = p_1\, dx_1 + p_2\, dx_2 + p_3\, dx_3$: they can be integrated over a differentiable path $\gamma : [a, b] \to U$, just as in $\mathbb{R}^2$-case:

$$\int_\gamma p_1\, dx_1 + p_2\, dx_2 + p_3\, dx_3 = \int_a^b \left(\sum_{i=1}^{3} p_i(\gamma(t)) \frac{d\gamma_i}{dt}(t) \right) dt.$$

- 2-forms – $\omega = \sum_{1\le i<j\le 3} p_{ij}\, dx_i \wedge dx_j$: they can be integrated over a differentiable surface $\Gamma : [a,b]\times[c,d] \to U$:

$$\int_\Gamma \sum_{1\le i<j\le 3} p_{ij}\, dx_i \wedge dx_j = \int_a^b \int_c^d \left(\sum_{1\le i<j\le 3} p_{ij}(\Gamma(s,t))\Gamma'_{ij}(s,t) \right) ds\, dt.$$

- 3-forms – $h\, dx_1 \wedge dx_2 \wedge dx_3$: they can be integrated over a differentiable 3-surface $\Pi : [a,b]\times[c,d]\times[e,f] \to U$:

$$\int_\Pi h\, dx_1 \wedge dx_2 \wedge dx_3 = \int_a^b \int_c^d \int_e^f h(\Pi(s,t,u))\Pi'(s,t,u) ds\, dt\, du.$$

By the fundamental theorem of calculus

$$\int_\gamma df = f(\gamma(b)) - f(\gamma(a)).$$

By the Green's theorem for the rectangle (Theorem 2.4.4),

$$\iint_\Gamma d\phi = \int_{\partial\Gamma} \phi.$$

The Stokes theorem for the box says

$$\iiint_\Pi \omega = \iint_{\partial\Pi} \omega.$$

Here $\partial\Pi$ (more precisely, the integral $\iint_{\partial\Pi}$) must be defined as a sum/difference of the restiction of Π to the 6 faces of the box $[a,b]\times[c,d]\times[e,f]$.

Recall that a k-form ω is closed when $d\omega = 0$, and is exact when $\omega = d\mu$, for a $(k-1)$-form μ. So all exact forms are closed. But as in the $\mathbb{R}^2$-case, the opposite is not the case. Similarly, a closed 1-form on U is exact iff all the integrals of it over closed paths are 0. For instance, when U is the complement of the x_3-axis, the 1-form $\omega = \frac{x_1 dx_2 - x_2 dx_1}{x^2+y^2}$ is closed but not exact.

Although if only 1 point (or a closed ball) is removed from $\mathbb{R}^3$ to get U, all the closed 1-forms on U are exact. However, for 2-forms this is no longer the case. E.g. let $U = \mathbb{R}^3 - \{0\}$, and

$$\omega = \frac{x_1 dx_2 \wedge dx_3 - x_2 dx_1 \wedge dx_3 + x_3 dx_1 \wedge dx_2}{(x_1^2 + x_2^2 + x_3^2)^{3/2}}.$$

Then ω is closed, but not exact, as can be seen integrating over the surface of the spherical coordinate mapping.

2.4.5 An application of Green's theorem

Here we give an application of Green's theorem to a property of smooth bijections on the closed unit disk $B = \{(x, y) \mid x^2 + y^2 \leq 1\}$.

Namely, we show the following.

Proposition 2.4.5. *Let $f : B \to B$ be a smooth bijection. Then there exists $p \in B$ so that $f(p) = p$.*

Proof. Assume the contrary. Then for *any* $p \in B$ define $\phi(p)$ to be the intersection of the ray starting at $f(p)$ and the boundary ∂B of B. It turns out that $\phi : B \to \partial B$ is smooth. Obviously, ϕ is identity on ∂B. Consider the 1-form θ given in (2.16). Then θ is closed, but not exact. (i.e. $d\theta = 0$, but $\theta \neq dg$ for any 0-form g.) Let $\omega = \phi^*\theta$. Then $d\omega = \phi^*(d\theta) = \phi^*0 = 0$. By Green's theorem

$$\int_{\partial B} \omega = \int_B d\omega = \int_B 0 = 0.$$

But, as the restriction of ϕ to ∂B is the identity mapping, we have

$$\int_{\partial B} \omega = \int_{\partial B} \theta \neq 0,$$

where the latter unequality was verified after (2.16), a contradiction. □

2.4.6 Volumes and areas

Recall that the volume of a parallelepiped in $\mathbb{R}^n$ spanned by the rows J_i of an $n \times n$ square matrix J is

$$V(J_1, \ldots, J_n) = \pm \det J.$$

The sign above depends upon *orientation* of J, i.e. the ordering of the rows in J, but we will not dwell upon this here. On the other hand

$$(\det J)^2 = \det J \det J = detJ \det J^T = \det JJ^T = \det G, \quad \text{where } G_{ij} = J_i \cdot J_j.$$

The matrix G is called the *Gram matrix* of the vectors $J_1, \ldots, J_n$. In particular $V(J_1, \ldots, J_n) = \sqrt{\det G}$.

More generally, for k vectors $J_1, \ldots, J_k$ we *define* the *k-volume* (usually called *area* when $k = 2$, and *volume* when $k = 3$) as follows.

$$V(J_1, \ldots, J_k) = \sqrt{\det G}, \quad \text{where } G_{ij} = J_i \cdot J_j, \quad 1 \leq i, j \leq k. \tag{2.20}$$

This allows one to compute the area of a smooth *parametrically defined k-surface* $S \subset \mathbb{R}^n$ given by $r : D \to S$, for D a k-cube in $\mathbb{R}^k$ (smooth here just means that r is differentiable). Let $t_0 \in D$, and $e_1, \ldots, e_k$ span the tangent space $T\mathbb{R}^k_{t_0}$, and $h_1 > 0$, …, $h_k > 0$ be small enough so that the parallelepiped I spanned by the $h_i e_i$ is contained in D. Then

$$r(t_0 + h_i e_i) - r(t_0) = \frac{\partial r}{\partial t_i}(t_0) h_i + o(h_i),$$

as follows by considering the image of $T\mathbb{R}^k_{t_0}$ under the linear map $r'(t_0)$. Thus as $h_j \to 0$, the parallelepiped spanned by the vectors $h_i r'(t_0) e_i$ approximate $r(I)$ increasingly well. We obtain, using (2.20), that

$$V(h_1 r'(t_0) e_1, \ldots, h_k r'(t_0) e_k) = \sqrt{\det G(r'(t_0))} h_1 \ldots h_k,$$

where $G(r'(t_0))$ is the Gram matrix of columns of $r'(t_0)$. Passing to the limit, arrive at the following definition of the k-volume of S:

$$V(S) = \int\limits_D \sqrt{\det G(r'(t))}\ dt_1 \ldots dt_k.$$

The function $\sqrt{\det G(r'(t)}$ is called the k-volume element, or the *area element.* When $k = n$, we get familiar

$$V(S) = \int\limits_D \sqrt{\det G(r'(t))}\ dt_1 \ldots dt_n = \int\limits_D |\det r'(t)| dt_1 \ldots dt_n = \int\limits_S dt_1 \ldots dt_n.$$

When $k = 1$, we get an expression for the length of the parametric curve, with $D = [a, b]$:

$$V(S) = \int_a^b \|r'(t)\| dt = \int_a^b \sqrt{\sum_{i=1}^n (D_i r(t))^2}\, dt.$$

In the case $k = 2$, $n = 3$ we get

$$V(S) = \int\limits_D \sqrt{EG - F^2}\ dt_1 dt_2,$$

where traditionally one denotes

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} = r'(t)^T r'(t).$$

Another nice way to write $\sqrt{\det G(r'(t)}$, actually valid whenever we deal with a *hypersurface*, i.e. when $k = n - 1$, is as follows:

$$\sqrt{\det G(r'(t)} = \|r'(t)e_1 \times r'(t)e_2 \times \cdots \times r'(t)e_k, \|$$

where $r'(t)e_j$ is just the j-th column of $r'(t)$. The vector

$$n(t) := r'(t)e_1 \times \cdots \times r'(t)e_k$$

has a geometric meaning: it is perpendicular to S at $r(t)$. It is called the *normal* vector to S at $r(t)$.

The *unit normal* to S at $r(t)$ is just $\frac{1}{\|n(t)\|} n(t)$.

2.5 Chains of k-surfaces

Let D be the k-cube $[0,1]^k$ in $\mathbb{R}^k$. Recall that a k-form ω in $E \subseteq \mathbb{R}^n$ is a function from the set of k-surfaces $\Phi : D \to E$ to $\mathbb{R}$, essentially "something that one can integrate over" a k-surface Φ, as in (2.1). On the other hand, we can view k-surfaces Φ, Ψ as functions from the set of k-forms to $\mathbb{R}$. As we can add functions, we can write

$$\int_\Phi \omega + \int_\Psi \omega = \int_{\Phi+\Psi} \omega.$$

This is how the formal linear combinations of k-surfaces, called k-chains, or *singular k-cubes*, arise. More generally, with $a = (a_1, \ldots, a_m) \in \mathbb{Z}^m$, we can define

$$\int\limits_{\sum_{i=1}^m a_i \Phi_i} \omega = \sum_{i=1}^m a_i \int\limits_{\Phi_i} \omega,$$

and k-chain $\Gamma = \sum_{i=1}^m a_i \Phi$. We have already seen something like this in the case of 2-forms in $\mathbb{R}^2$.

However, we now need to define *boundary* $\partial\Gamma$ of Γ. An important example of k-surface is the *standard k-cube* $I^k : D \to D$, so that $I^k(x) = x$. First, we define ∂I^k. For $1 \le i \le k$, define the $(k-1)$-cube

$$I^k_{(i,\alpha)}(x) = I^k(x_1, \ldots, x_{i-1}, \alpha, x_i, \ldots, x_{k-1}).$$

Then $I^k_{(i,0)}$ and $I^k_{(i,1)}$ correspond to the 2 faces of $I^k(D)$. Define

$$\partial I^k = \sum_{i=1}^k \sum_{\alpha=0}^1 (-1)^{i+\alpha} I^k_{(i,\alpha)}.$$

For a general k-surface Φ, define

$$\Phi_{(i,\alpha)} = \Phi(I^k_{(i,\alpha)})$$

and

$$\partial\Phi = \sum_{i=1}^k \sum_{\alpha=0}^1 (-1)^{i+\alpha} \Phi_{(i,\alpha)}.$$

Finally,

$$\partial\Gamma = \sum_j a_j \partial\Phi_j.$$

Remark. Although we will not need this, note that $\partial(\partial\Gamma) = 0$.

Now we can state the general Stokes theorem.

Theorem 2.5.1. *For a k-chain Γ in an open set $E \subseteq \mathbb{R}^n$ and ω a $(k-1)$-form on E,*

$$\int_\Gamma d\omega = \int_{\partial\Gamma} \omega. \tag{2.21}$$

Bibliography

[1] W. Rudin. *Principles of mathematical analysis.* McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.

[2] M. Spivak. *Calculus on manifolds. A modern approach to classical theorems of advanced calculus.* W. A. Benjamin, Inc., New York-Amsterdam, 1965.

[3] V. A. Zorich. *Mathematical analysis. II.* Universitext. Springer-Verlag, Berlin, 2004. Translated from the 2002 fourth Russian edition by Roger Cooke.

Index