Fields, Forms and Flows 3/34

Lecture Notes 2018

School of Mathematics, University of Bristol

Contents

1 Vector Fields, Flows and Diffeomorphisms

1.1 Notation

We denote points in $\mathbb{R}^m$ as follows:

$$x = (x^1, \ldots, x^m) \in \mathbb{R}^m. \tag{1}$$

Note that points are denoted in plain text, not boldface (we write x rather than $\mathbf{x}$). Note too that indices are written as superscripts, not subscripts. This will take some getting used to, but it turns out to be a very useful convention (writing indices as superscripts is standard in discussions of differentiable manifolds).

The inner product is denoted as follows:

$$x \cdot y = x^1 y^1 + \cdots + x^m y^m. \tag{2}$$

The norm is denoted as follows:

$$||x|| = (x \cdot x)^{1/2} = \left((x^1)^2 + \cdots + (x^m)^2\right)^{1/2}. \tag{3}$$

1.2 Open and closed sets in $\mathbb{R}^m$

The notion of open and closed sets is basic in calculus and analysis, in particular to the notions of continuous and differentiable functions and maps. If you have taken the unit Metric Spaces, the definitions will be familiar to you. Otherwise, open and closed sets in $\mathbb{R}^m$ generalise the familiar notions of open and closed intervals in $\mathbb{R}$. Besides their definitions, we will state some of the basic properties of open and closed sets. I'll just remark that the notion of open and closed sets may be extended from $\mathbb{R}^m$ to more general spaces. In the most general setting, this subject is called point-set topology, or general topology. But we will stick to $\mathbb{R}^m$, and our discussion will be brief and elementary. A number of the results appear as exercises in Problem Sheet 1, and proofs can be found in the solutions.

1.2.1 Open sets

Definition 1.2.1 (Open ball)**.** Given $x \in \mathbb{R}^m$, the **open ϵ-ball about x**, also called the **ϵ-neighbourhood of x**, denoted $B_\epsilon(x)$, is given by

$$B_\epsilon(x) = \{y \in \mathbb{R}^m \,|\, ||x - y|| < \epsilon\}.$$

See Figure 1.

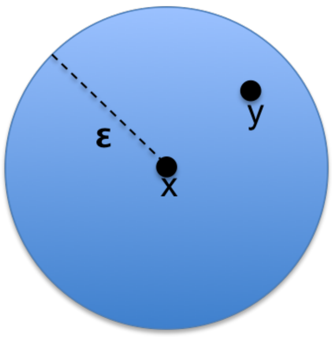


Figure 1: An open ϵ-ball about x.

For example, for $m = 1$, $B_\epsilon(x)$ is just the open interval $(x - \epsilon, x + \epsilon)$.

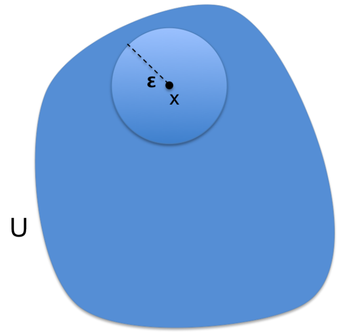


Figure 2: An open set U.

Definition 1.2.2 (Open set)**.** Let $U \subset \mathbb{R}^m$. U is **open** if $\forall x \in U$, $\exists \epsilon > 0$ such that $B_\epsilon(x) \subset U$. See Figure 2.

Proposition 1.2.3 (Properties of open sets)**.**

(i) The union of two (in fact, arbitrarily many) open sets is open.

(ii) The intersection of two (or any finite collection of) open sets is open.

Proof. See Problem Sheet 1. □

Note: The intersection of infinitely many open sets may not be open. For example, the intervals $(-1/N, 1/N)$, where N is a positive integer, are open. 0 is the only point which belongs to every interval. Therefore,

$$\bigcap_{N=1}^{\infty} (-1/N, 1/N) = \{0\}.$$

But $\{0\}$ is not open; it contains no ϵ-neighbourhood of 0.

Example 1.2.4 (Examples and non-examples of open sets)**.**

a) $B_\epsilon(x)$ is open (See Problem Sheet 1).

b) $[0,1] \subset \mathbb{R}$ is not open, because it contains no ϵ-neighbourhood of 0 or 1.

c) The set of points $x \in \mathbb{R}^m$ with x^1 rational is not open.

Remark 1.2.5. Open sets will be important for us because we will be studying maps between $\mathbb{R}^m$ and $\mathbb{R}^n$, as in Multivariable Calculus, which are defined only on a subset of $\mathbb{R}^m$. On an open subset, it makes sense to ask whether the map is differentiable at any point, because the difference $f(x+\epsilon) - f(x)$ is defined for ϵ sufficiently small.

1.2.2 Closed sets

Definition 1.2.6 (Limit point)**.** Let $X \subset \mathbb{R}^m$. A point $p \in \mathbb{R}^m$ is a **limit point** of X if for all ϵ, every ϵ-neighbourhood of p contains at least one point of X, i.e.

$$\forall \epsilon > 0, B_\epsilon(p) \cap X \neq \emptyset,$$

where $\emptyset$ denotes the empty set. See Figure 3.

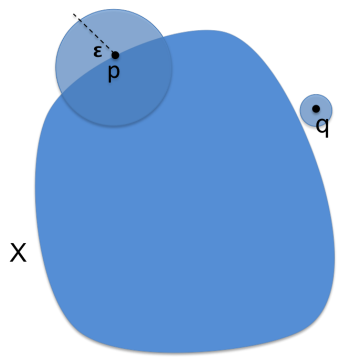


Figure 3: p is a limit point of X, since every neighbourhood of p has a nonempty intersection with X. q is not a limit point of X, as it has a neighbourhood which does not intersect with X.

Limit points are also called **boundary points**. Note that every $x \in X$ is a limit point of X (why?).

Definition 1.2.7 (Closed set)**.** $X \subset \mathbb{R}^m$ is **closed** if X contains all of its limit points.

Example 1.2.8 (Examples and non-examples of closed sets)**.**

a) $[0,1] \subset \mathbb{R}$ is closed

b) $(0,1) \subset \mathbb{R}$ is not closed, since 0 and 1 are limit points (why?) but are not contained in $(0,1)$.

Remark 1.2.9. A set can be both open and closed. For example, $\mathbb{R}^m$ is both open and closed. The empty set is also regarded as open and closed.

Definition 1.2.10 (Complement)**.** Given $X \subset \mathbb{R}^m$, its **complement**, denoted $\tilde{X}$, is given by

$$\tilde{X} = \{y \in \mathbb{R}^m \,|\, y \notin X\}.$$

Proposition 1.2.11 (Relation between open and closed sets)**.** Let $U \subset \mathbb{R}^m$. Then U is open if and only if $\tilde{U}$ is closed.

Proof. See Problem Sheet 1 and solutions. □

1.3 Continuous and differentiable maps

1.3.1 Continuous maps

Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$, and let

$$F : U \to V; x \mapsto F(x)$$

denote a map from U to V. We write

$$F(x) = \left(F^1(x), \ldots, F^n(x)\right),$$

where $F^j : U \to \mathbb{R}$ denotes the jth component of F.

Definition 1.3.1 (Continuous maps)**.** Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$ be open sets, and let $F : U \to V$ be a map from U to V. F is **continuous** if for all $x \in U$ and for all $\epsilon > 0$, there exists $\delta > 0$, which may depend on ϵ and x, such that if $x' \in U$ and $||x' - x|| < \delta$, then $||F(x') - F(x)|| < \epsilon$. Equivalently, $x' \in B_\delta(x)$ implies that $F(x') \in B_\epsilon(F(x))$.

Notation. We denote the set of continuous maps from U to V by $C^0(U, V)$.

There is a nice characterisation of continuous maps in terms of open sets. Before giving this characterisation, we need to define the inverse image of a set under a map.

Definition 1.3.2 (Inverse image)**.** Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$, and let $F : U \to V$ be a map from U to V. Let $Y \subset V$. The **inverse image of Y under F**, denoted $F^{-1}(Y)$, is the subset of U given by

$$F^{-1}(Y) = \{x \in U \,|\, F(x) \in Y\}.$$

See Figure 4.

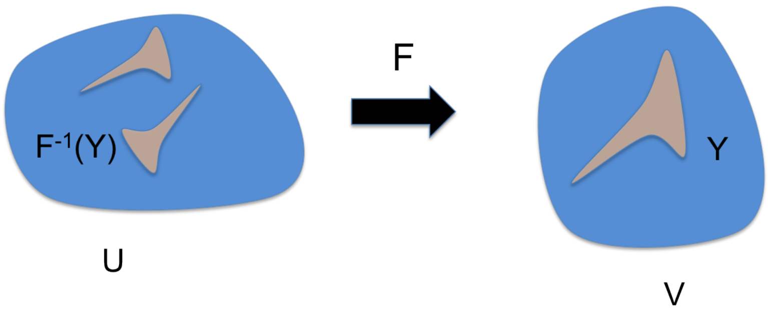


Figure 4: The inverse image $F^{-1}(Y)$ of the set Y.

Example 1.3.3 (Example of inverse image)**.** Let $U = V = \mathbb{R}$, and $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Let $Y = (1, 4) \subset \mathbb{R}$. Then

$$f^{-1}(Y) = (-2, -1) \cup (1, 2).$$

As this example shows, f needn't be invertible in order to define the inverse image. Note that when U and V are subsets of $\mathbb{R}$, the map is a function, and we will often denote it by a small letter instead of a capital letter, e.g. f instead of F.

We shall also have occasion to refer to the image of a set under a map.

Definition 1.3.4 (Image)**.** Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$, and let $F : U \to V$ be a map from U to V. Let $X \subset U$. The **image of X under F**, denoted $F(X)$, is the subset of V given by

$$F(X) = \{y \in V \,|\, y = F(x) \text{ for some } x \in X\}.$$

See Figure 5.

Example 1.3.5 (Example of image)**.** Let $U = V = \mathbb{R}$, and $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. Let $X = (-1, 1) \subset \mathbb{R}$. Then

$$f(X) = [0, 1).$$

Proposition 1.3.6 (Continuity and open sets)**.** Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$ be open sets, and let $F : U \to V$ be a map from U to V. Then F is continuous if and only if for all open sets $Y \subset V$, $F^{-1}(Y)$ is open. That is, F is continuous if and only if the inverse image of every open set is open.

Proof. See Problem Sheet 1. $\square$

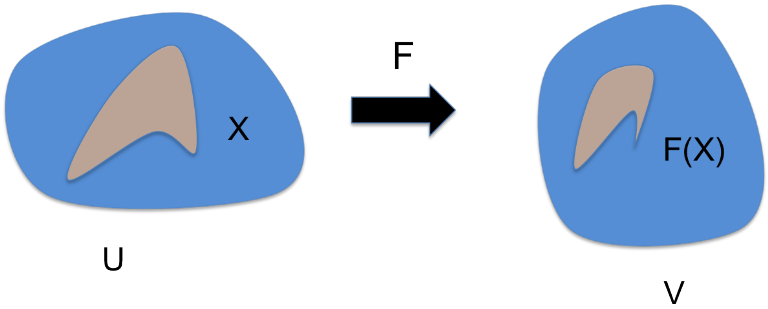


Figure 5: The image $F(X)$ of the set X

Note: The *image* of an open set under a continuous map is *not* necessarily open, as Example 1.3.5 shows.

One advantage of this characterisation of continuity, in terms of open sets, is that it generalises to cases where the "ϵ/δ" description is either artificial or indeed not available (there may be no natural notion of distance, which is required for the "ϵ/δ" definition). Another advantage is that it simplifies certain arguments, like the fact that the composition of two continuous maps is continuous.

Definition 1.3.7 (Composition of maps)**.** Let $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^p$, and let $F : U \to V$ and $G : V \to W$. The **composition** of F and G, denoted $G \circ F$, is the map $G \circ F : U \to W$ defined by

$$(G \circ F)(x) = G(F(x)).$$

Proposition 1.3.8 (Composition of continuous maps)**.** Let U, V, W and F, G be as above. Suppose that U, V, W are open sets and that F and G are continuous. Then $G \circ F$ is continuous. In other words, if $F \in C^0(U, V)$ and $G \in C^0(V, W)$, then $G \circ F \in C^0(U, W)$.

Proof. Let $Z \subset W$. Then $(G \circ F)^{-1}(Z) = F^{-1}(G^{-1}(Z))$, since $x \in (G \circ F)^{-1}(Z)$ if and only if $G(F(x)) \in Z$, which holds if and only if $F(x) \in G^{-1}(Z)$, which holds if and only if $x \in F^{-1}(G^{-1}(x))$.

Now suppose that Z is open. Since G is continuous, $G^{-1}(Z)$ is open. Since F is continuous, $F^{-1}(G^{-1}(Z))$ is open, which from the preceding is equivalent to saying that $(G \circ F)^{-1}(Z)$ is open. This implies that $G \circ F$ is continuous. □

1.3.2 Differentiable maps

This is a brief review of some material from Multivariable Calculus.

Let $e_{(k)}$ denote the kth unit vector in $\mathbb{R}^m$, i.e.

$$e_{(k)} = (0, \ldots, 0, 1, 0, \ldots, 0),$$

where the '1' occurs in the kth component. Note that the subscript (k) is a label, not an index.

Definition 1.3.9 (Partial derivative)**.** Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$ be open, and let $F : U \to V$ be a map. The **kth partial derivative of F^j**, denoted $\partial F^j / \partial x^k\,(x)$ is defined by

$$\frac{\partial F^j}{\partial x^k}(x) = \lim_{t \to 0} \frac{F^j\left(x + te_{(k)}\right) - F^j(x)}{t}.$$

Note that $F^j(x + te_{(k)})$ is defined for t sufficiently small, since U is open. Note too, however, that the limit might not exist!

Definition 1.3.10 (Continuously differentiable maps)**.** The map $F : U \to V$ is **continuously differentiable** if the partial derivatives $\partial F^j/\partial x^k\,(x)$ exist for all $1 \le k \le m$ and for all $1 \le j \le n$, and moreover are continuous functions on U.

Notation. Let $C^1(U, V)$ denote the set of continuously differentiable maps $F : U \to V$. Let $F'(x)$ denote the $n \times m$ matrix of partial derivatives of F, i.e.

$$[F'(x)]_{jk} = \frac{\partial F^j}{\partial x^k}(x). \tag{4}$$

Proposition 1.3.11 (Linear approximation)**.** Let $F \in C^1(U, V)$. Then $\forall x \in U$, $\forall v \in \mathbb{R}^m$ and sufficiently small $\epsilon > 0$,

$$F(x + \epsilon v) = F(x) + \epsilon F'(x) \cdot v + r(\epsilon, x, v),$$

where the remainder term $r(\epsilon, x, v)$ goes to zero faster than ϵ, i.e.

$$\lim_{\epsilon \to 0} \frac{||r(\epsilon, x, v)||}{\epsilon} = 0.$$

See Figure 6.

Proof. See Multivariable Calculus notes. ◻

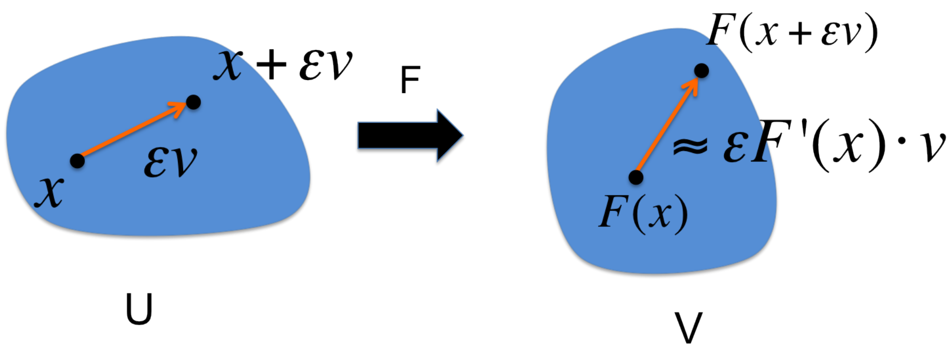


Figure 6: F maps x to $F(x)$, and F maps a nearby point, $x + \epsilon v$, to $F(x + \epsilon v)$. $F(x + \epsilon v) - F(x)$ is given approximately by $F'(x) \cdot \epsilon v$. Note: this figure is NOT to scale. You should think of the vector displacements as being small, of order ϵ, even though they look big in the picture.

The Chain Rule establishes that the composition of differentiable maps is differentiable and gives a formula for the derivative of the composition.

Proposition 1.3.12 (Chain rule)**.** Let U, V and W be open sets in $\mathbb{R}^m$, $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively. Let $F \in C^1(U, V)$ and $G \in C^1(V, W)$. Then $G \circ F \in C^1(U, W)$, and

$$(G \circ F)'(x) = G'(F(x))F'(x).$$

Proof. See Multivariable Calculus. ◻

Definition 1.3.13 (Second-order partial derivatives)**.** Let $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ be open sets, and let $F : U \to V$ be a map from U to V. The **second-order partial derivatives of F**, denoted $\partial^2 F^j/\partial x^k \partial x^l\,(x)$, where $1 \le j \le n$, $1 \le k, l \le m$, are given by

$$\frac{\partial^2 F^j}{\partial x^k \partial x^l}(x) = \frac{\partial}{\partial x^k}\left(\frac{\partial F^j}{\partial x^l}\right)(x).$$

Note that the second-order partial derivatives might not all exist.

Higher-order partial derivatives are defined similarly, by induction. Suppose that we have defined partial derivatives up to order r, for $r \geq 1$.

Definition 1.3.14 (($r+1$)st-order partial derivatives)**.** Let $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ be open sets, and let $F : U \to V$ be a map from U to V. The **($r+1$)st -order partial derivatives of F**, denoted $\partial^{r+1}F^j/\partial x^{k_{r+1}} \cdots \partial x^{k_1}(x)$, are given by

$$\frac{\partial^{r+1}F^j}{\partial x^{k_{r+1}}\partial x^{k_r} \cdots \partial x^{k_1}}(x) = \frac{\partial}{\partial x^{k_{r+1}}}\left(\frac{\partial^r F^j}{\partial x^{k_r} \cdots \partial x^{k_1}}\right)(x),$$

where $1 \leq j \leq n$ and $1 \leq k_1, \ldots, k_{r+1} \leq m$.

Notation. The set of maps $F : U \to V$ for which all rth-order partial derivatives exist and are continuous functions on U is denoted by $C^r(U, V)$. The set of maps $F : U \to V$ for which partial derivatives of all orders exist and are continuous is denoted $C^\infty(U, V)$.

Definition 1.3.15 (Smooth maps)**.** A map $F \in C^\infty(U, V)$ is called a **smooth map**.

In this course, we will be primarily concerned with smooth maps.

Proposition 1.3.16 (Equality of mixed partials)**.** Let $F \in C^r(U, V)$, $r \geq 2$. Then

$$\frac{\partial^2 F^j}{\partial x^k \partial x^l}(x) = \frac{\partial^2 F^j}{\partial x^l \partial x^k}(x).$$

Proof. See Multivariable Calculus notes. (In fact, the result holds under much weaker assumptions. For simplicity, we state the assumptions for the case where $U = \mathbb{R}^2$ and $V = \mathbb{R}$, i.e. for a function $f(x, y)$. Suppose $f \in C^1(\mathbb{R}^2, \mathbb{R})$ and that $\partial(\partial f/\partial x)/\partial y$ exists and is continuous at some point (x_*, y_*). Then $\partial(\partial f/\partial y)/\partial x$ also exists at (x_*, y_*) and moreover is equal to $(\partial(\partial f/\partial x)/\partial y)(x_*, y_*)$. That is, for the equality of mixed partials to hold at a point, it is enough that both first derivatives are continuous and that one of the mixed partials exists and is continuous at that point.) $\square$

The equality of mixed partials will be very important in this course. A number of results make essential use of this fact. Note that the equality of second-order mixed partials implies that for smooth functions, the ordering of any number of partial derivatives does not matter. For example, if $f : \mathbb{R}^2 \to \mathbb{R}; (x, y) \mapsto f(x, y)$ is a smooth function of two variables, then

$$\frac{\partial^3 f}{\partial y \partial x \partial x}(x, y) = \frac{\partial^3 f}{\partial x \partial y \partial x}(x, y) = \frac{\partial^3 f}{\partial x \partial x \partial y}(x, y).$$

Example 1.3.17 (Examples and non-examples of differentiable and smooth maps)**.**

a) Let $m = n = 1$, $U = V = \mathbb{R}$. Then

$$f(x) = \frac{1}{1 + x^2}$$

is smooth.

$$g(x) = \frac{1}{1 - x^2}$$

is not even continuous (it has a singularity at $x = \pm 1$). But if we take $U = (1, \infty)$ and $V = \mathbb{R}$, then $g : U \to V; x \mapsto 1/(1 - x^2)$ is smooth. (We've excluded the singularities at $x = \pm 1$ from the domain of definition.)

b) $m = n = 2$. Then
$$F(x, y) = (\sin y, x^2) \in C^\infty(\mathbb{R}^2, \mathbb{R}^2),$$
i.e. F is smooth. Let
$$G(x, y) = \begin{cases} \left(\frac{x^4}{x^2+y^2}, y\right), & (x, y) \neq 0, \\ 0, & (x, y) = 0. \end{cases}$$
Then
$$\begin{aligned} &G \in C^1(\mathbb{R}^2, \mathbb{R}^2), \\ &G \notin C^2(\mathbb{R}^2, \mathbb{R}^2), \\ &G \in C^\infty(\mathbb{R}^2 - (0,0), \mathbb{R}^2). \end{aligned}$$
That is, on $\mathbb{R}^2$, G has everywhere continuous first partial derivatives but does not have everywhere continuous second partial derivatives; the second partial derivatives of F^1 are singular at the origin. If, however, the origin is excluded from the domain, then G is smooth. Note that $\mathbb{R}^2 - (0,0)$ denotes the plane with the origin removed, and this is an open set. (Why?)

(Here are a few more details. The partial derivative of G^1 with respect to x is given by
$$\frac{-2x^5}{(x^2+y^2)^2} + \frac{4x^3}{x^2+y^2},$$
for $(x, y) \neq 0$. To compute $\partial G^1/\partial x$ at the origin, however, we can't use the preceding expression (because the definition of G^1 at the origin is special). However, we can resort to the definition of the partial derivative:
$$\frac{\partial G^1}{\partial x}(0,0) = \lim_{h \to 0} \frac{G^1(h,0) - G^1(0,0)}{h} = \lim_{h \to 0} \frac{h^2 - 0}{h} = 0.$$
We obtain
$$\frac{\partial G^1}{\partial x}(x, y) = \begin{cases} \frac{-2x^5}{(x^2+y^2)^2} + \frac{4x^3}{x^2+y^2}, & (x, y) \neq 0, \\ 0, & (x, y) = 0. \end{cases}$$
You can check that $\partial G^1/\partial x$ is continuous at the origin – for (x, y) small, $\partial G^1/\partial x$ is small.

Next, let's compute the second partial of G^1 with respect to x. We obtain
$$\frac{\partial^2 G^1}{\partial x^2}(x, y) = \frac{8x^6}{(x^2+y^2)^3} - \frac{18x^4}{(x^2+y^2)^2} + \frac{12x^2}{x^2+y^2}, \quad (x, y) \neq 0.$$
$\frac{\partial^2 G^1}{\partial x^2}(x, y)$ is *not* continuous at the origin. Indeed, we have for all $\epsilon \neq 0$, that
$$\frac{\partial^2 G^1}{\partial x^2}(\epsilon, 0) = 2, \quad \frac{\partial^2 G^1}{\partial x^2}(0, \epsilon) = 0.$$
Thus, $\frac{\partial^2 G^1}{\partial x^2}$ is constant and equal to 2 along the x-axis, while it is constant but equal to 0 along the y-axis.)

c) Linear maps. Let $A \in \mathbb{R}^{n \times m}$; that is, A is a real $n \times m$ matrix. Let $F(x) = A \cdot x$. Then $F'(x) = A$, so that the first-partial derivatives of F are constants. All second- and higher-order partial derivatives of F are identically 0, so $F \in C^\infty(\mathbb{R}^m, \mathbb{R}^n)$, i.e. F is smooth.

1.4 The Inverse Function Theorem

Let $U, V \subset \mathbb{R}^n$ be open sets (so both U and V belong to $\mathbb{R}^n$ in this case). Let $F \in C^r(U, V)$.

Here are some basic questions we might ask. Given $y \in V$, can we find $x \in U$ such that $F(x) = y$? That is, can we solve the equation $F(x) = y$ to obtain x in terms of y; in other words, is F onto? If so, is x unique? In other words, is F one-to-one, or 1-1; that is, if $F(x) = F(x')$, does it follow that $x = x'$? If F is both 1-1 and onto, then we can define the inverse of F, which we denote by $F^{-1} : V \to U$. We say that F is invertible. With x and y as above, we have that $F^{-1}(y) = x$. Equivalently,

$$F^{-1} \circ F = Id_U, \quad F \circ F^{-1} = Id_V, \tag{5}$$

where Id_U and Id_V denote the identity maps on U and V, i.e.

$$Id_U(x) = x, \forall x \in U, \quad Id_V(y) = y, \forall y \in V. \tag{6}$$

A final question: if F^{-1} exists, does it follow that $F^{-1} \in C^s(V, U)$ for some s?

Example 1.4.1 (Examples and non-examples of maps with smooth inverses)**.**

a) Linear maps. Let $F(x) = A \cdot x$, $A \in \mathbb{R}^{n \times n}$. That is, A is a real $n \times n$ matrix. Then F is invertible if and only if $\det A \neq 0$, i.e. if and only if A is invertible. Note that $F'(x) = A$, so we can write that F is invertible if and only if $F'(x)$ is invertible. In this case, $F^{-1}(y) = A^{-1} \cdot y$. Thus F^{-1} is also a linear map. It follows that F^{-1} is smooth, since linear maps are smooth.

b) $n = 1$, $U = V = \mathbb{R}$, $f : U \to V; x \mapsto f(x)$.

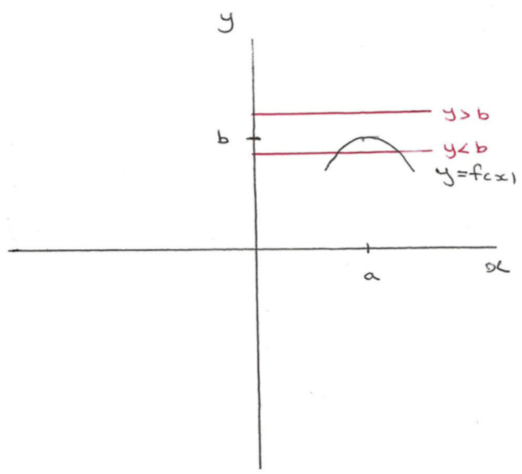


Figure 7: f is not invertible. For $y > b$, $f(x) = y$ has no solutions x near a; for $y < b$, $f(x) = y$ has two solutions near a.

See Figure 7. Suppose $f'(a) = 0$, $f''(a) < 0$. Let $f(a) = b$ (an example would be $f(x) = b - (x - a)^2$). Given y near b, can we solve $f(x) = y$ for x? For $y > b$, there are no solutions, at least near $x = a$ (so f may not be onto). For $y < b$, there are two solutions near $x = a$ (so f is not 1-1).

c) $f(x) = x^3$. Does f^{-1} exist? Let $x^3 = y$. Then $x = y^{1/3}$. So $f^{-1}(y) = y^{1/3}$, and the inverse exists. Let's consider whether the inverse is smooth. f itself is smooth, i.e. $f \in C^\infty(\mathbb{R}, \mathbb{R})$. But f^{-1} is not smooth, since

$$(f^{-1})'(y) = \frac{1}{3} y^{-2/3},$$

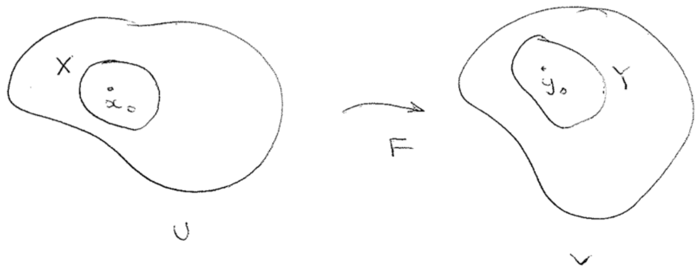


Figure 8: The Inverse Function Theorem. If $F'(x_0)$ is nonsingular, then $F : X \to Y$ is invertible.

which blows up at $y = 0$. You can show that f^{-1} is continuous, i.e. $f^{-1} \in C^0(\mathbb{R}, \mathbb{R})$. But f^{-1} does not have a continuous derivative, so $f^{-1} \notin C^1(\mathbb{R}, \mathbb{R})$.

What these examples suggest is that difficulties in finding the inverse can arise when $f' = 0$, in case $n = 1$, or when $\det F' = 0$, in case $n > 1$.

Theorem 1.4.2 (Inverse Function Theorem)**.** Let $U, V \subset \mathbb{R}^n$ be open, and let $F \in C^r(U, V)$ (where $1 \leq r \leq \infty$). Take $x_0 \in U$ and let $F(x_0) = y_0$. If $\det F'(x_0) \neq 0$, then F has a locally defined inverse near $y = y_0$. That is, there exist open sets $X \subset U$ and $Y \subset V$ with $x_0 \in X$ and $y_0 \in Y$ such that $F : X \to Y$ is invertible, with inverse $F^{-1} : Y \to X$. Moreover, $F^{-1} \in C^r(Y, X)$, and the derivative of F^{-1} is related to the derivative of F by the formula

$$\left(F^{-1}\right)'(F(x)) = (F'(x))^{-1}.$$

See Figure 8.

Note: the formula for $\left(F^{-1}\right)'$ follows from the Chain Rule, Proposition 1.3.12. Indeed, since $F^{-1} \circ F = Id_X$ and $Id_X' = I$ (the identity map is a linear map whose derivative is the identity matrix), it follows that

$$(F^{-1} \circ F)'(x) = \left(F^{-1}\right)'(F(x))F'(x) = I,$$

which yields the formula for the derivative of F^{-1}.

Proof. Proofs may be found in the course references, including Spivak and Hubbard and Hubbard. $\square$

Here is an idea of the proof. Given $F(x_0) = y_0$, we want to solve

$$F(x_0 + h) = y_0 + k$$

to get h as a function of k. We know that for $k = 0$, one solution is given by $h = 0$. We try an approximation, based on k and h being small, which follows from Proposition 1.3.11:

$$F(x_0 + h) \approx F(x_0) + A \cdot h = y_0 + k,$$

where $A = F'(x_0)$. Eliminating y_0 from both sides, we get that $A \cdot h \approx k$, or $k \approx A^{-1} \cdot h$. Note that this approximation makes sense only if A is invertible. Of course, this is not a proof, and showing that the inverse actually exists takes more work. In the end, one cannot hope to get an explicit general formula for the inverse.

Example 1.4.3 (Polar coordinates)**.** The transformation between cartesian and polar coordinates provides a familiar example where the Inverse Function Theorem applies, and where the role of the condition on F' is apparent. In this case, $n = 2$, and we let $U = V = \mathbb{R}^2$ to start with. We'll denote coordinates on V by (x, y), and coordinates on U by (r, θ). This might look funny at first, since normally we don't think of r as being allowed to be negative. But for now, r and θ are just names of coordinates.

We define the map

$$F : U \to V; (r, \theta) \mapsto F(r, \theta) = (x(r, \theta), y(r, \theta)),$$

where

$$x(r, \theta) = r \cos \theta, \quad y(r, \theta) = r \sin \theta.$$

Clearly $F \in C^\infty(U, V)$. Let us compute the derivative $F'(r, \theta)$.

$$F'(r, \theta) = \begin{pmatrix} \partial x/\partial r & \partial x/\partial \theta \\ \partial y/\partial r & \partial y/\partial \theta \end{pmatrix} (r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}.$$

Then

$$\det F'(r, \theta) = r.$$

Let $(r_0, \theta_0) \in U$, where $r_0 > 0$. Let $x_0 = r_0 \cos \theta_0$, $y_0 = r_0 \sin \theta_0$. According to the Inverse Function Theorem, since $r_0 \neq 0$, there exist open sets $X, Y \subset \mathbb{R}^2$ with $(r_0, \theta_0) \in X$, $(x_0, y_0) \in Y$, such that $F : X \to Y$ is invertible and $F^{-1} \in C^\infty(Y, X)$.

In fact, we can see this directly. As in Figure 9, we can choose an interval around r_0 of half-width a and an interval around θ_0 of half-width b. We take

$$X = \{(r, \theta) \,|\, r_0 - a < r < r_0 + a, \ \theta_0 - b < \theta < \theta_0 + b\}.$$

We take $Y = F(X)$. Then we can write down the inverse map, as follows:

$$F^{-1}(x, y) = (r(x, y), \theta(x, y)),$$

where

$$r(x, y) = (x^2 + y^2)^{1/2}, \quad \theta(x, y) = \tan^{-1}(y/x).$$

The only ambiguity is which branch to take for $\tan^{-1}(y/x)$ (there are many angles whose tangent is y/x, which differ by integer multiples of π). This ambiguity is resolved by taking $\tan^{-1}(y_0/x_0) = \theta_0$, and defining the branch elsewhere to make θ continuous in Y.

Note that $F : X \to Y$ does not have a smooth inverse if X contains a point with $r_0 = 0$ (in this case, Y would contain the point $(0, 0)$, and $(x^2 + y^2)^{1/2}$ is not smooth in a neighbourhood of $(0, 0)$, nor can θ be continuously defined).

Note also that in order for an inverse to exist, X can't be "too big". In particular, if $b > \pi$, then $F : X \to Y$ is not 1-1.

1.5 Diffeomorphisms

Definition 1.5.1 (Diffeomorphism)**.** Let $U, V \subset \mathbb{R}^n$ be open. A **diffeomorphism from U to V** is a map $F : U \to V$ such that

(i) $F \in C^\infty(U, V)$

(ii) F is 1-1 and onto (so that $F^{-1} : V \to U$ exists)

(iii) $F^{-1} \in C^\infty(V, U)$

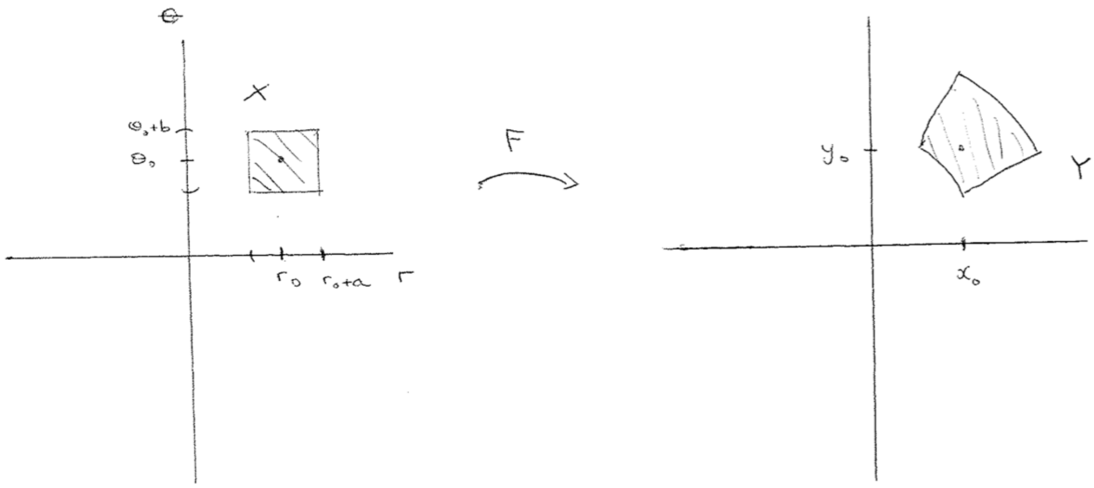


Figure 9: Polar to cartesian coordinates – see Example 1.4.3.

The Inverse Function Theorem (Theorem 1.4.2) generates examples of diffeomorphisms. Let $\mathrm{Diff}(U, V)$ denote the set of diffeomorphisms from U to V.

Example 1.5.2 (Examples and non examples of diffeomorphisms)**.** $U = V = \mathbb{R}$ throughout. See Figure 10.

a) $f(x) = \begin{cases} x, & x < 0 \\ 2x, & x \geq 0 \end{cases}$. Then $f \notin \mathrm{Diff}(\mathbb{R}, \mathbb{R})$, since f is not smooth. However, f is 1-1 and onto.

b) $f(x) = x^2$. $f \notin \mathrm{Diff}(\mathbb{R}, \mathbb{R})$, since f is not 1-1 nor onto. However, f is smooth.

c) $f = x^3$. $f \notin \mathrm{Diff}(\mathbb{R}, \mathbb{R})$, even though f is smooth, 1-1 and onto, since $f^{-1}(y) = y^{1/3}$ is not smooth.

d) $f = x^3 + x$. $f \in \mathrm{Diff}(\mathbb{R}, \mathbb{R})$, since f is smooth, $1 - 1$ and onto, and f^{-1} is smooth.

   Let us verify these assertions for this last example.

   It's clear that $f(x) = x^3 + x$ is smooth.

   The fact that f is 1-1 follows from the Mean Value Theorem, which says that

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$$

   for some $c \in [x_1, x_2]$, and the fact that, in this case, $f'(x) = 3x^2 + 1$, so that $f'(c)$ is strictly positive, and therefore nonzero. It follows that $f(x_2) = f(x_1)$ implies that $x_2 = x_1$. (Note: this argument really shows that, in general, if f has nonvanishing derivative, then f is 1-1.)

   The fact that f is onto can be established by appealing to the Intermediate Value Theorem. Note that as $x \to \pm\infty$, $f(x) \to \pm\infty$. Since f is continuous, the Intermediate Value Theorem applies, which in this case says that f must assume every value between $-\infty$ and ∞.

   Since f is 1-1 and onto, it follows that its inverse exists. For convenience, let's denote f^{-1} by g. It remains to show that g is smooth. First, we'll show that g is differentiable, and obtain an explicit formula for the derivative. Recalling the Mean Value Theorem above and letting $f(x_2) = y_2$, $f(x_1) = y_1$, $x_2 = g(y_2)$, and $x_1 = g(y_1)$, we obtain

$$y_2 - y_1 = f'(c)(g(y_2) - g(y_1)),$$

or

$$\frac{g(y_2) - g(y_1)}{y_2 - y_1} = \frac{1}{f'(c)},$$

where c lies between $g(y_1)$ and $g(y_2)$. Taking the limit as $y_1 \to y_2$, we see that g is differentiable with derivative given by

$$g'(y) = \frac{1}{f'(g(y))} = \frac{1}{3g^2(y) + 1}.$$

It is clear that repeated differentiation will produce continuous functions (you can show, say by induction, that the nth derivative of g will be of the form $P(g)/(3g^2 + 1)^N$, where $P(g)$ is a polynomial in g and N is a positive integer), so that g is smooth.

You might wonder why we didn't just invoke the Inverse Function Theorem to establish that f is a diffeomorphism. After all, since $f' \neq 0$, the conditions of the theorem are satisfied. The point is that the Inverse Function Theorem is a local result; it would establish that f is a diffeomorphism which maps some open interval U to another open interval V. However, it would not automatically imply that U and V could both be taken to be all of $\mathbb{R}$. In Problem Sheet 2.3 b), there is a two-dimensional example where $\det F' \neq 0$, so that the Inverse Function Theorem applies, but nevertheless F is not invertible on all of $\mathbb{R}^2$.

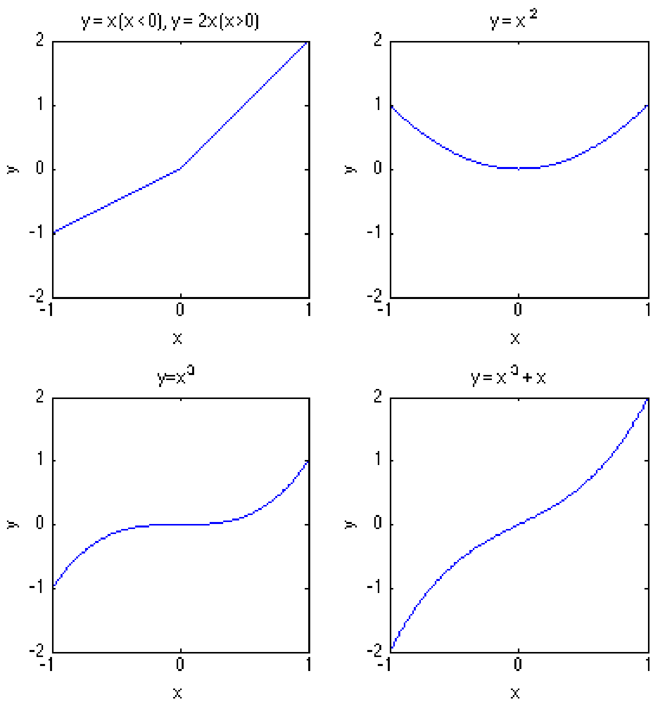


Figure 10: Example 1.5.2

Let $U \subset \mathbb{R}^n$ be open. We write $\mathrm{Diff}(U) := \mathrm{Diff}(U, U)$ to denote the set of diffeomorphisms from U to itself. Here are some observations about $\mathrm{Diff}(U)$:

i) $Id_U \in \text{Diff}(u)$

ii) If $F, G \in \text{Diff}(U)$, then $F \circ G \in \text{Diff}(U)$. This follows from the Chain Rule (Proposition 1.3.12), which implies that $F \circ G$ is smooth, and the fact that $(F \circ G)^{-1} = G^{-1} \circ F^{-1}$, which again by the Chain Rule is smooth (since F^{-1} and G^{-1} are smooth).

iii) If $F, G, H \in \text{Diff}(U)$, then $(F \circ G) \circ H = F \circ (G \circ H)$. In fact, this is true for all maps from U to itself, not just diffeomorphisms; the point is that composition of maps is associative.

iv) If $F \in \text{Diff}(U)$, then $F^{-1} \in \text{Diff}(U)$.

These observations can be summarised by the following:

Proposition 1.5.3 (Diffeomorphism group)**.** Let $U \subset \mathbb{R}^n$ be open. Then $\text{Diff}(U)$ is a group, with product given by composition and identity given by Id_U.

1.6 ODE's and vector fields

As we shall see, solutions to first-order systems of ordinary differential equations, or ODE's, are another source of diffeomorphisms.

Definition 1.6.1 (Vector field)**.** Let $U \subset \mathbb{R}^n$ be open. A **vector field** on U is a map

$$\mathbb{X} : U \to \mathbb{R}^n; x \mapsto \mathbb{X}(x) = \left(\mathbb{X}^1(x), \dots, \mathbb{X}^n(x)\right).$$

Definition 1.6.2 (First-order system of ODE's)**.** Let $U \subset \mathbb{R}^n$ be open, and let $\mathbb{X}$ be a vector field on U. A **first-order system of autonomous ODE's** on U for a curve $x(t) \in U$ parameterised by t is given by

$$\dot{x} = \mathbb{X}(x), \quad x(0) = x_0, \tag{$*$}$$

where $x_0 \in U$. The equation $x(0) = x_0$ is the **initial condition**. In terms of components, the system is given by

$$\dot{x}^i = \mathbb{X}^i(x), \quad 1 \le i \le n, \quad x(0) = x_0. \tag{$*$}$$

"First-order" means that the first derivative of $x(t)$, $\dot{x}$, is expressed as a function of x. (A second-order equation would be one where the second derivative $\ddot{x}$ is expressed in terms of x and $\dot{x}$). "Autonomous" means that $\mathbb{X}$ does not depend explicitly on t (towards the end of the course, we shall have occasion to consider nonautonomous systems, or equivalently vector fields $\hat{\mathbb{X}}$ that depend explicitly on t).

From the ODE2 course, you will know that an mth-order nonautonomous system is equivalent to a first-order autonomous system in a higher dimensional space. Specifically, if the mth-order nonautonomous system is defined on $U \subset \mathbb{R}^n$, then the equivalent first-order autonomous system is defined on $V \subset \mathbb{R}^{nm+1}$. See Problem Sheet 2.7 and 2.8 for examples. Thus, first-order systems are quite general.

A fundamental result in the theory of ODE's is the following:

Theorem 1.6.3 (Existence and uniqueness of solutions of ODE's)**.** Suppose that $\mathbb{X} : U \to \mathbb{R}^n$ is smooth. Then $\forall x_0 \in U$, $\exists T > 0$ (which may depend on x_0) such that (*) has a unique solution $x(t)$ defined for $-T < t < T$.

Proof. See ODE2. A classic reference is VI Arnold, Ordinary Differential Equations. □

In fact, $\mathbb{X}$ need not be smooth. It is sufficient that $\mathbb{X}$ satisfies a Lipschitz condition. (While we won't have much occasion to refer to the Lipschitz condition, this means that there exists $C > 0$ such that for all $x, x' \in U$, $||\mathbb{X}(x') - \mathbb{X}(x)|| < C||x - x'||$. C is called the Lipschitz constant.) If $\mathbb{X}$ is merely continuous, but does not satisfy a Lipschitz condition, then a solution exists but in general is not unique. See Problem Sheet 2.10 for an example.)

Note that even if $\mathbb{X}$ is smooth, the solution $x(t)$ may not be defined for all t. See Problem Sheet 2.9.

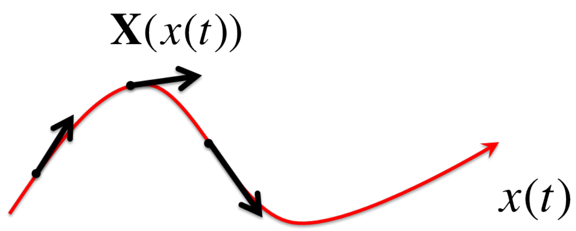


Figure 11: Geometrical description of the solution to an autonomous first-order system. The derivative $\dot{x}(t)$ of the solution curve $x(t)$ is everywhere given by the vector field $\mathbb{X}$.

Definition 1.6.4 (Complete vector field)**.** If solutions $x(t)$ to (*) exist for all initial conditions $x_0 \in U$ and for all t, and if $x(t) \in U$ for all t, then $\mathbb{X}$ is said to be a **complete** vector field on U.

Usually in this course we will restrict our attention to complete vector fields.

Definition 1.6.5 (Linear vector field)**.** Let A be a real $n \times n$ matrix. Let $U = \mathbb{R}^n$. The vector field $\mathbb{X} : U \to \mathbb{R}^n$ given by

$$\mathbb{X}(x) = A \cdot x$$

is a **linear vector field** on $\mathbb{R}^n$.

As is shown in Example 1.6.10 below, linear vector fields are complete.

Remark 1.6.6 (Vector fields vs maps)**.** You might be wondering, What is the difference between a vector field $\mathbb{X} : U \to \mathbb{R}^n$ and a map $F : U \to \mathbb{R}^n$? As we have defined them, there is in fact no difference. However, in more advanced treatments, and in particular in the context of differentiable manifolds, they are defined differently; indeed, vector fields are defined as maps from U to $U \times \mathbb{R}^n$, under which $x \mapsto (x, \mathbb{X}(x))$.

For our purposes, it will suffice to have in mind *different interpretations* of maps and vector fields, as shown in Figure 12. We shall think of a map $F : U \to \mathbb{R}^n$ in the usual way, as taking points in U to points in $\mathbb{R}^n$. Vector fields, on the other hand, we shall think of as assigning a vector to each point $x \in U$, with the base of the vector sitting at x. In physical terms, we think of $\mathbb{X}$ as assigning a velocity at each point of U (for example, the velocity of a fluid or gas moving in U).

It is useful to indicate explicitly the dependence of the solution $x(t)$ on the initial condition x_0. Thus we will write $x(t, x_0)$ instead of $x(t)$. Then (*) looks like

$$\frac{\partial x}{\partial t}(t, x_0) = \mathbb{X}(x(t, x_0)), \quad x(0, x_0) = x_0. \tag{$*$}$$

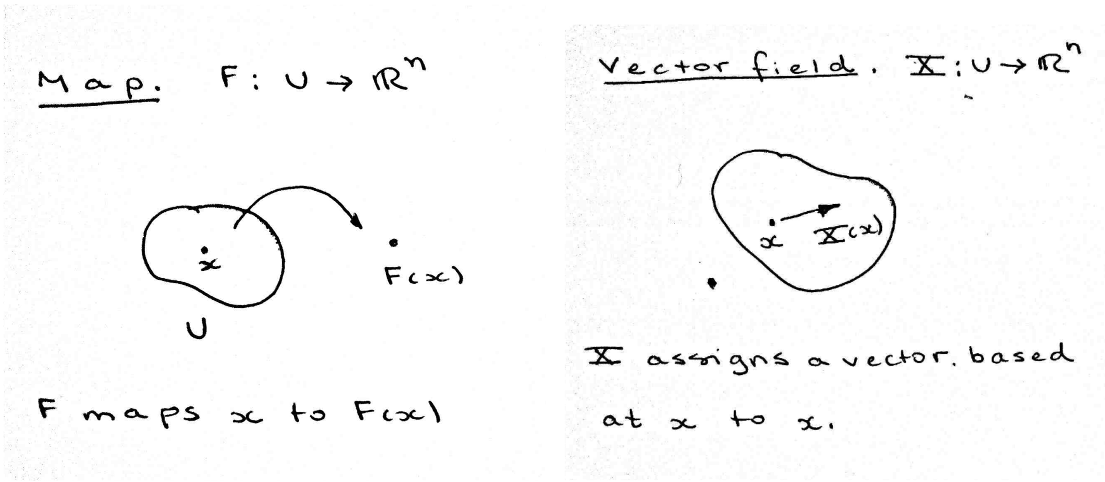
The nature of the dependence on initial conditions is given by the following:

Theorem 1.6.7 (Smooth dependence on initial conditions)**.** Suppose $\mathbb{X} : U \to \mathbb{R}^n$ is smooth, and that $x(t, x_0) \in U$ is defined for all $x_0 \in U$ and $-T < t < T$ for some $T > 0$. Then for all $t \in (-T, T)$, the map which takes initial conditions to solutions at time t,

$$x(t, \cdot) : U \to U; x_0 \mapsto x(t, x_0),$$

is smooth.

Proof. See ODE2 references; VI Arnold, Ordinary Differential Equations. $\square$



(a) Map from U to $\mathbb{R}^n$ (b) Vector field on U

Figure 12: Distinction between maps and vector fields

Definition 1.6.8 (Flow)**.** Let $\mathbb{X} : U \to \mathbb{R}^n$ be a smooth, complete vector field on U, and let $x(t, x_0)$ denote the solution of (*). The **flow** of $\mathbb{X}$, denoted Φ, is the map on $\mathbb{R} \times U$ defined by

$$\Phi : \mathbb{R} \times U \to U; (t, x_0) \mapsto \Phi_t(x_0) = x(t, x_0).$$

Note that for all t, Φ_t is a map from U to U. Often we will refer to Φ_t (as well as Φ) as the flow of $\mathbb{X}$. Φ_t maps initial conditions at $t = 0$ to solutions at t. See Figure 13. In terms of the flow, the system (*) can be written as

$$\frac{\partial \Phi_t}{\partial t}(x_0) = \mathbb{X}(\Phi_t(x_0)), \quad \Phi_0(x_0) = x_0. \tag{**}$$

If we omit the argument x_0, this can be written more concisely as

$$\frac{\partial \Phi_t}{\partial t} = \mathbb{X} \circ \Phi_t, \quad \Phi_0 = Id_U. \tag{***}$$

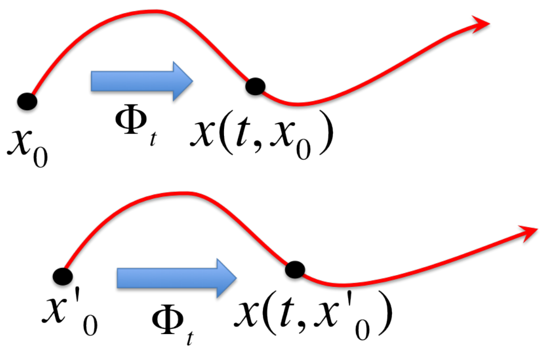


Figure 13: The flow map Φ_t

The flow map Φ_t is really just another notation for the solutions to a system of first-order ODE's. However, it brings to the fore certain properties of these solutions, which will be important to us. The following result follows straightforwardly from Theorems 1.6.3 and 1.6.7 (indeed, you might regard it as a restatement of these theorems), but it will be basic to much of what follows:

Proposition 1.6.9 (Properties of the flow)**.** Let Φ be the flow of a smooth, complete vector field $\mathbb{X} : U \to \mathbb{R}^n$ on U. Then

i) $\Phi_0 = Id_U$

ii) $\Phi_t \circ \Phi_s = \Phi_{t+s}$

iii) $\Phi_t \in \mathrm{Diff}(U)$ (hence, flows provide examples of diffeomorphisms).

iv) $\Phi \in C^\infty(\mathbb{R} \times U, U)$,

Proof.

i) This is clear from (**) above.

ii) See Figure 14. This follows from the uniqueness of solutions to (*). Fix $x_0 \in U$. Let

$$x_1(t) = \Phi_t \circ \Phi_s(x_0) = \Phi_t(\Phi_s(x_0)),$$
$$x_2(t) = \Phi_{t+s}(x_0).$$

We will show that $x_1(t)$ and $x_2(t)$ satisfy the same system of ODE's and the same initial condition.

First, we compute $\dot{x}_1(t)$. We may write that $x_1(t) = \Phi_t(\Phi_s(x_0)) = \Phi_t(y)$, where we have introduced $y = \Phi_s(x_0)$ for convenience. Then using (**),

$$\dot{x}_1(t) = \frac{\partial \Phi_t}{\partial t}(y) = \mathbb{X}(\Phi_t(y)).$$

But $\mathbb{X}(\Phi_t(y))$ is just $\mathbb{X}(x_1(t))$, so we get that $x_1(t)$ satisfies the system

$$\dot{x}_1 = \mathbb{X}(x_1).$$

Next, we compute $\dot{x}_2(t)$. From (**), replacing Φ_t by Φ_{t+s}, we have that

$$\dot{x}_2(t) = \frac{\partial \Phi_{t+s}}{\partial t}(x_0) = \mathbb{X}(\Phi_{t+s}(x_0)).$$

But this is equivalent to

$$\dot{x}_2 = \mathbb{X}(x_2).$$

Thus, $x_1(t)$ and $x_2(t)$ satisfy the same first-order system, namely (*). At $t = 0$, we have that $x_1(0) = \Phi_0(\Phi_s(x_0)) = \Phi_s(x_0)$, while $x_2(0) = \Phi_s(x_0)$. Thus, $x_1(t)$ and $x_2(t)$ satisfy the same initial condition, and therefore must be the same, by Theorem 1.6.3.

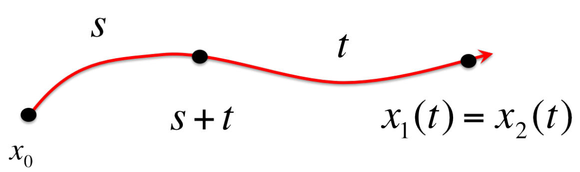


Figure 14: $\Phi_t \circ \Phi_s = \Phi_{t+s}$. That is, you can evolve from initial conditions to $t+s$ in a single step of duration $t+s$, or in two successive steps of durations s and t.

iii) The fact that $\Phi_t : U \to U$ is smooth follows from Theorem 1.6.7. It remains to show that Φ_t has a smooth inverse. In fact, it is easily seen that $(\Phi_t)^{-1} = \Phi_{-t}$, since

$$\Phi_t \circ \Phi_{-t} = \Phi_0 \text{ (by (ii))} = Id_U \text{ (by (i))}.$$

Also, $\Phi_{-t} \in C^\infty(U, U)$, by Theorem 1.6.7. Therefore, Φ_t has a smooth inverse, so that $\Phi_t \in \mathrm{Diff}(U)$.

iv) *Theorem 1.6.7 tells us that for fixed t, $\Phi_t : U \to U$ is smooth, so that all partial derivatives of Φ_t with respect to components of x exist. We need to show that all partial derivatives with respect to both components of x *as well as* t exist. For this calculation, it is probably safest to use components (although it makes for longer and more cumbersome formulas). From (**),

$$\frac{\partial \Phi_t^i}{\partial t}(x) = \mathbb{X}^i(\Phi_t(x))$$

(note that we are using x rather than x_0 for the argument of Φ_t – this makes no real difference, and is only for the sake of clarity). The expression obtained on the right-hand side is evidently smooth in x (it is the composition of functions which are smooth in x), so it follows that we can apply a partial derivative with respect to t to Φ_t followed by any number of partial derivatives with respect to x and obtain something smooth. By the equality of mixed partials (the stronger version given in the note in the proof of Proposition 1.3.16), it follows that all mixed partials of Φ involving at most one partial derivative with respect to t are smooth. Returning to Φ_t, taking a second derivative with respect to t and using the Chain Rule, we get that

$$\frac{\partial^2 \Phi_t^i}{\partial t^2}(x) = \sum_{j=1}^{n} \frac{\partial \mathbb{X}^i}{\partial x^j}(\Phi_t(x))\mathbb{X}^j(\Phi_t(x)).$$

The resulting expression is evidently smooth in x, so by repeating the preceding argument we may conclude that all mixed partials of Φ involving at most two partial derivative with respect to t are smooth. We can continue in this way, showing that any number of partial derivatives of Φ_t with respect to t followed by any number of partial derivatives with respect to components of x is smooth. It then follows by the strong version of the equality of mixed partials that all partial derivatives of Φ_t exist (and are therefore smooth). Thus, $\Phi \in \mathbb{C}^\infty(\mathbb{R} \times U, U)$.

$\square$

Example 1.6.10 (Linear vector fields)**.** Let A be an $n \times n$ real matrix. Let $U = \mathbb{R}^n$, and let $\mathbb{X}(x) = A \cdot x$, so that $\mathbb{X}$ is a linear vector field. Consider the system

$$\dot{x} = \mathbb{X}(x) = A \cdot x, \quad x(0) = x_0.$$

As was shown in ODE2, the solution is given by

$$x(t) = e^{At} x_0,$$

where the matrix exponential may be defined by the power series

$$e^{At} := \sum_{j=0}^{\infty} \frac{A^j t^j}{j!}.$$

Thus, the flow is given by

$$\Phi_t(x_0) = e^{At} \cdot x_0.$$

Let us verify Proposition 1.6.9 (ii). We have that

$$
\begin{aligned}
(\Phi_t \circ \Phi_s)(x_0) &= \Phi_t(\Phi_s(x_0)) = \Phi_t\left(e^{As} \cdot x_0\right) = e^{At}\ e^{As} \cdot x_0, \\
\Phi_{t+s}(x_0) &= e^{A(t+s)} \cdot x_0.
\end{aligned}
$$

The fact that these are equal is equivalent to the statement

$$
e^{At}e^{As} = e^{A(t+s)}. \tag{7}
$$

This is the familiar formula for the product of exponentials, here applied to the matrix exponential. Note that it is *not true* that $e^A e^B = e^{A+B}$ for general matrices A and B; this holds only if the matrices A and B commute, i.e. $AB = BA$.

Let us verify (7) directly, using the power series definition of the exponential. We have that

$$
e^{At}e^{As} = \left(\sum_{j=0}^{\infty} \frac{A^j t^j}{j!}\right)\left(\sum_{k=0}^{\infty} \frac{A^k s^k}{k!}\right) = \sum_{j,k=0}^{\infty} \frac{A^{j+k} t^j s^k}{j!\,k!}.
$$

Rearrange the sum as follows: Let $m = j + k$, so that m takes values between 0 and ∞. For given m, the index j can take values between 0 and m. For given m and j, we have that $k = m - j$. Thus, if $F(j,k)$ denotes an arbitrary summand depending on j and k, we have that

$$
\sum_{j,k=0}^{\infty} F(j,k) = \sum_{m=0}^{\infty}\sum_{j=0}^{m} F(j, m-j).
$$

Applying this to the preceding, we get that

$$
e^{At}e^{As} = \sum_{m=0}^{\infty}\sum_{j=0}^{m} \frac{A^m t^j s^{m-j}}{j!\,(m-j)!} = \sum_{m=0}^{\infty} \frac{A^m}{m!} \sum_{j=0}^{m} \frac{m!}{j!\,(m-j)!} t^j s^{m-j},
$$

where in the last equality we have multiplied and divided by $m!$. We can evaluate the sum over j using the binomial theorem,

$$
\sum_{j=0}^{m} \frac{m!}{j!\,(m-j)!} t^j s^{m-j} = \sum_{j=0}^{m} \binom{m}{j} t^j s^{m-j} = (s+t)^m.
$$

Substituting into the preceding, we get that

$$
e^{At}e^{As} = \sum_{m=0}^{\infty} \frac{A^m (s+t)^m}{m!} = e^{A(t+s)},
$$

as required.

Definition 1.6.11 (One-parameter subgroup of diffeomorphisms)**.** Let $U \subset \mathbb{R}^n$ be open. A map $\Phi : \mathbb{R} \times U \to U; (t,x) \mapsto \Phi_t(x)$ such that $\Phi \in C^\infty(\mathbb{R} \times U, U)$, $\Phi_t \in \mathrm{Diff}(U)$ and

i) $\Phi_0 = Id_U$,

ii) $\Phi_t \circ \Phi_s = \Phi_{t+s}$

is called a **one-parameter subgroup of diffeomorphisms**.

Thus, Proposition 1.6.9 says that the flow of a smooth, complete vector field is a one-parameter subgroup of diffeomorphisms. Example 1.6.10 says that $\Phi_t(x) = e^{At} \cdot x$ is a one-parameter subgroup of diffeomorphisms.

Proposition 1.6.9 shows how vector fields give rise to flows. The following is a sort of converse; it shows that a one-parameter subgroup of diffeomorphisms give rise to a vector field.

Proposition 1.6.12 (Flows and vector fields)**.** Let $U \subset \mathbb{R}^n$ be open, and let $\Phi : \mathbb{R} \times U \to U$ be a one-parameter subgroup of diffeomorphisms on U. Let

$$\mathbb{X}(x) = \frac{\partial \Phi_t}{\partial t}(x)\bigg|_{t=0}.$$

Then Φ is the flow of $\mathbb{X}$.

Proof. Let $x(t) = \Phi_t(x_0)$. We need to show that

$$\dot{x} = \mathbb{X}(x), \quad x(0) = x_0.$$

The initial condition follows from the fact that $\Phi_0 = Id_U$. As for the differential equation, we have that

$$\dot{x}(t) = \lim_{h\to 0} \frac{x(t+h) - x(t)}{h} = \lim_{h\to 0} \frac{\Phi_{t+h}(x_0) - \Phi_t(x_0)}{h}.$$

Since $\Phi_{t+h} = \Phi_h \circ \Phi_t$, we have that

$$\Phi_{t+h}(x_0) - \Phi_t(x_0) = \Phi_h(\Phi_t(x_0)) - \Phi_t(x_0) = \Phi_h(y) - y,$$

where for convenience we have introduced $y = \Phi_t(x_0)$ (note that y is just $x(t)$, but for now it is easier to write y). Then

$$\dot{x}(t) = \lim_{h\to 0} \frac{\Phi_h(y) - y}{h} = \frac{\partial \Phi_h}{\partial h}(y)\bigg|_{h=0} = \mathbb{X}(y) = \mathbb{X}(x(t)),$$

as required. □

Terminology. We will say that the one-parameter subgroup of diffeomorphisms Φ_t is **generated** by $\mathbb{X}$.

Example 1.6.13 (Linear maps)**.** Let A be an $n\times n$ real matrix. From Example 1.6.10, we know that $\Phi_t(x) = \exp(tA)\cdot x$ is a one-parameter subgroup of diffeomorphisms on $\mathbb{R}^n$. Let us compute the associated vector field directly, using Proposition 1.6.12. Differentiating term-by-term, we have that

$$\mathbb{X}(x) = \frac{\partial \Phi_t}{\partial t}(x)\bigg|_{t=0} = \frac{\partial}{\partial t}(e^{tA}\cdot x)\bigg|_{t=0} = \frac{\partial}{\partial t}\left(x + tA\cdot x + \tfrac{1}{2}t^2A^2\cdot x + \cdots\right)\bigg|_{t=0} = A\cdot x.$$

Not surprisingly, we recover the linear vector field $\mathbb{X}(x) = A\cdot x$.

1.7 Pushforward map

Notation: Let $\mathcal{X}(U)$ denote the set of smooth vector fields on an open set $U \subset \mathbb{R}^n$. Usually we will assume that vector fields in $\mathcal{X}(U)$ are complete, but this isn't automatically the case. We note that if $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$, then $\mathbb{X} + \mathbb{Y} \in \mathcal{X}(U)$. Also, if $f \in C^\infty(U)$, then $f\mathbb{X} \in \mathcal{X}(U)$.

We can motivate the pushforward map by considering a change of variables in a first-order system of ODE's. Let $\mathbb{X} \in \mathcal{X}(U)$, and consider the system

$$\dot{x} = \mathbb{X}(x), \quad x(0) = x_0. \tag{$*$}$$

Let $U, V \subset \mathbb{R}^n$ be open, and let $F \in \mathrm{Diff}(U, V)$. Given $x(t)$ satisfying (*), let

$$y(t) = F(x(t)).$$

Question: What system of ODE's does $y(t)$ satisfy?

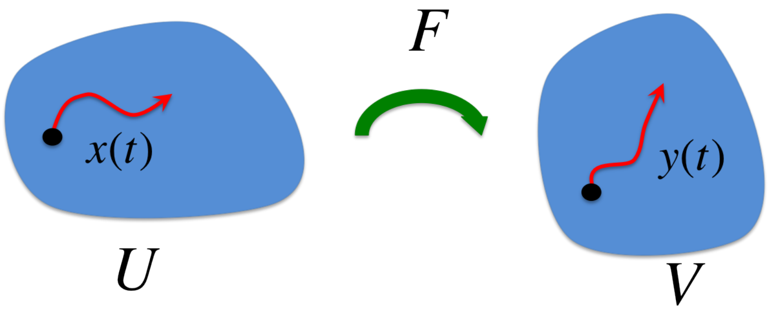


Figure 15: Change of variables in system of ODE's

This is a computation. It will be clearer if we consider components.

$$
\begin{aligned}
\dot{y}^i(t) &= \frac{d}{dt}F^i(x(t)) = \sum_{j=1}^{n} \frac{\partial F^i}{\partial x^j}(x(t))\frac{dx^j}{dt} \qquad \text{(by the Chain Rule)} \\
&= \sum_{j=1}^{n} \frac{\partial F^i}{\partial x^j}(x(t))\mathbb{X}^j(x(t)) \qquad \text{(using (*))}.
\end{aligned}
$$

The rhs of the preceding is a function of $x(t)$. We would like to express it instead as a function of $y(t)$. Since F is a diffeomorphism, we may write

$$x(t) = F^{-1}(y(t)).$$

Therefore,

$$\dot{y}^i(t) = \sum_{j=1}^{n} \frac{\partial F^i}{\partial x^j}(F^{-1}(y(t))\,\mathbb{X}^j(F^{-1}(y(t))) = \mathbb{Y}^i(y(t)),$$

where

$$\mathbb{Y}^i(y) = \sum_{j=1}^{n} \frac{\partial F^i}{\partial x^j}(F^{-1}(y))\,\mathbb{X}^j(F^{-1}(y)).$$

From now on, to simplify notation, we will use the *summation convention*. This saves having to write the summation symbol, $\sum$, in situations where its presence can be inferred. According to the summation convention, we agree that if an index j appears twice on one side of an equation, once as an upper index and once as a lower index, then we sum over that index. Note that an upper index in the "denominator" of a derivative, e.g. the index j in $\partial F^i/\partial x^j$, counts as a lower index. Therefore, according to the summation convention, we may write

$$\mathbb{Y}^i(y) = \frac{\partial F^i}{\partial x^j}(F^{-1}(y))\mathbb{X}^j(F^{-1}(y)),$$

since the index j is to be summed over.

The preceding change-of-variables calculation motivates the following definition:

Definition 1.7.1. Let $U, V \subset \mathbb{R}^n$ be open, $F \in \mathrm{Diff}(U, V)$ and $\mathbb{X} \in \mathcal{X}(U)$. The **pushforward of $\mathbb{X}$ by F**, denoted $F_*\mathbb{X}$, is the vector field in $\mathcal{X}(V)$ defined by either of the following equivalent formulas:

$$
\begin{aligned}
\mathbb{Y}^i(y) &= \frac{\partial F^i}{\partial x^j}(F^{-1}(y))\,\mathbb{X}^j(F^{-1}(y)), \\
\mathbb{Y}^i(F(x)) &= \frac{\partial F^i}{\partial x^j}(x)\,\mathbb{X}^j(x).
\end{aligned}
$$

More compactly, we may omit arguments and component indices and write these formulas as

$$\mathbb{Y} = \left(F' \cdot \mathbb{X}\right) \circ F^{-1},$$
$$\mathbb{Y} \circ F = F' \cdot \mathbb{X}.$$

Note that the pushforward by the identity map Id_U is just the identity map on $\mathcal{X}(U)$. That is,

$$Id_{U*}\mathbb{X} = \mathbb{X}$$

for all $\mathbb{X} \in \mathcal{X}(U)$. A pictorial description follows below.

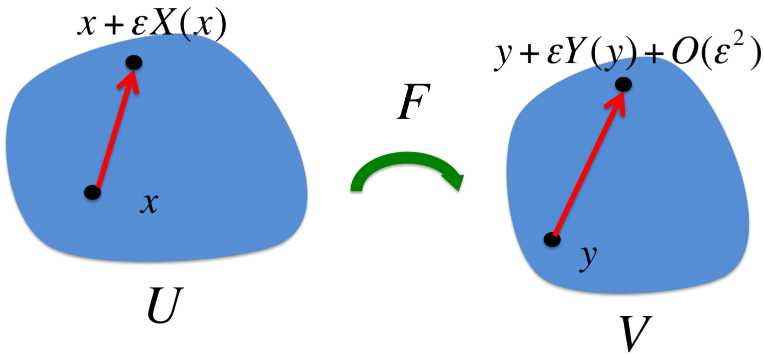


Figure 16: The pushforward of a vector field. Take $x \in U$, and consider the nearby point $x + \epsilon\mathbb{X}(x)$ (which can be thought of as the point reached from x after moving with velocity $\mathbb{X}(x)$ for a short time ϵ). F maps x into y, and it maps $x + \epsilon\mathbb{X}(x)$ into a point near y. Up to $O(\epsilon^2)$ corrections, $F(x + \epsilon\mathbb{X}(x))$ is $y + \epsilon\mathbb{Y}(y)$ (which may be thought of as the point reached from y after moving with velocity $\mathbb{Y}(y)$ for a short time ϵ). Here, of course, $\mathbb{Y}$ is the pushforward $F_*\mathbb{X}$.

Note that the pushforward can be regarded as a map $F_* : \mathcal{X}(U) \to \mathcal{X}(V)$.

Example 1.7.2 (Linear vector fields)**.** We take $U = V = \mathbb{R}^n$. Let A be an $n \times n$ real matrix, and $\mathbb{X}(x) = A \cdot x$ the associated linear vector field. Let S be an invertible $n \times n$ real matrix, and $F(x) = S \cdot x$ a linear diffeomorphism (c.f. Example 1.4.1 a), with inverse $F^{-1}(y) = S^{-1} \cdot y$. Then $F'(x) = S$, from Example 1.3.17 c).

We compute the pushforward $F_*\mathbb{X}$ as follows. We have that $\mathbb{Y}(F(x)) = F'(x) \cdot \mathbb{X}(x)$, which gives $\mathbb{Y}(S \cdot x) = SA \cdot x$. Letting $y = S \cdot x$, we get that

$$\mathbb{Y}(y) = SAS^{-1} \cdot y.$$

Thus, in the linear case, the pushforward corresponds to matrix conjugation.

Further examples of calculations of the pushforward can be found in Problem Sheet 3.2 and 3.3, as well as previous exam papers (see Question 1 or 2 of the recent exams).

Our original motivation for introducing the pushforward – changing coordinates in systems of ODE's - is formalised by the following:

Proposition 1.7.3 (Pushforward and flow)**.** Let $\mathbb{X} \in \mathcal{X}(U)$, and let Φ be the flow of $\mathbb{X}$. Let $F \in \mathrm{Diff}(U, V)$. For $t \in \mathbb{R}$, let us define

$$\Psi_t := F \circ \Phi_t \circ F^{-1}.$$

Then Ψ is the flow of $F_*\mathbb{X}$.

Proof. The plan is as follows: First, we'll show that Ψ_t is a one-parameter subgroup of diffeomorphisms. Then we'll use Proposition 1.6.12 to establish that Ψ_t is generated by $F_*\mathbb{X}$.

Clearly $\Psi \in C^\infty(\mathbb{R} \times V, V)$, since Ψ is a composition of maps which are smooth (from the Chain Rule, again). Also, $\Psi_t \in \mathrm{Diff}(V)$, since Ψ_t is smooth, and the inverse of Ψ_t is given by

$$(\Psi_t)^{-1} = (F \circ \Phi_t \circ F^{-1})^{-1} = F \circ (\Phi_t)^{-1} \circ F^{-1} = F \circ \Phi_{-t} \circ F^{-1} = \Psi_{-t},$$

and Ψ_{-t} (the inverse of Ψ_t) is smooth. Finally, $\Psi_0 = F \circ \Phi_0 \circ F^{-1} = F \circ F^{-1} = Id_V$, and

$$\Psi_t \circ \Psi_s = \left(F \circ \Phi_t \circ F^{-1}\right) \circ \left(F \circ \Phi_s \circ F^{-1}\right) = F \circ \Phi_t \circ \Phi_s \circ F^{-1} = F \circ \Phi_{t+s} \circ F^{-1} = \Psi_{t+s}.$$

Thus, from Definition 1.6.11, Ψ is a one-parameter subgroup of diffeomorphisms. .

By Proposition 1.6.12, the vector field $\mathbb{Y}$ that generates Ψ is given by

$$\mathbb{Y}(y) = \frac{\partial \Psi_t}{\partial t}(y)\bigg|_{t=0}.$$

Taking one component at a time for clarity, and using the Chain Rule, we get

$$\mathbb{Y}^i(y) = \frac{\partial}{\partial t} F^i(\Phi_t(F^{-1}(y)))\bigg|_{t=0} = \frac{\partial F^i}{\partial x^j}(\Phi_t(F^{-1}(y)))\frac{\partial \Phi_t^j}{\partial t}(F^{-1}(y))\bigg|_{t=0}.$$

Evaluating at $t = 0$ and using (**) (see page 17), we get

$$\mathbb{Y}^i(y) = \frac{\partial F^i}{\partial x^j}(F^{-1}(y))\, \mathbb{X}^j(F^{-1}(y)).$$

Comparing the preceding expression to Definition 1.7.1 of the pushforward, we see that $\mathbb{Y} = F_*\mathbb{X}$. □

Example 1.7.4 (Linear case)**.** Let A be a real $n \times n$ matrix, and let

$$\mathbb{X}(x) = A \cdot x.$$

The flow of $\mathbb{X}$ is given by

$$\Phi_t(x) = e^{tA} \cdot x.$$

Set S be an invertible real $n \times n$ matrix, and let

$$F(x) = S \cdot x.$$

From Example 1.7.2,

$$\mathbb{Y}(y) = F_*\mathbb{X}(y) = SAS^{-1} \cdot y.$$

As $\mathbb{Y}$ is a linear vector field, its flow, Ψ_t, is also given by a matrix exponential,

$$\Psi_t(y) = e^{tSAS^{-1}} \cdot y.$$

Let's verify that this expression for Ψ_t is consistent with Proposition 1.7.3. According to Proposition 1.7.3, we have that

$$\Psi_t(y) = F(\Phi_t(F^{-1}(y))) = Se^{tA}S^{-1} \cdot y.$$

Thus, we need to show that

$$e^{tSAS^{-1}} = Se^{tA}S^{-1}.$$

This can be seen from the power series for the matrix exponential, as follows:

$$e^{tSAS^{-1}} = \sum_{j=0}^{\infty} \frac{t^j}{j!}(SAS^{-1})^j = \sum_{j=0}^{\infty} \frac{t^j}{j!} SA^jS^{-1} = S\left(\sum_{j=0}^{\infty} \frac{t^jA^j}{j!}\right)S^{-1} = Se^{tA}S^{-1}.$$

Proposition 1.7.5 (Pushforward by composition)**.** Let $U, V, W \in \mathbb{R}^n$ be open. Let $F \in \mathrm{Diff}(U, V)$, $G \in \mathrm{Diff}(V, W)$. Then

$$(G \circ F)_* = G_* F_*$$

Note that F_* is a map from $\mathcal{X}(U)$ to $\mathcal{X}(V)$, G_* is a map from $\mathcal{X}(V)$ to $\mathcal{X}(W)$, and $(G \circ F)_*$ is a map from $\mathcal{X}(U)$ to $\mathcal{X}(W)$. Thus, the assertion makes sense.

Proof.
We will apply both maps to a vector field $\mathbb{X} \in \mathcal{X}(U)$. The result will follow from the Chain Rule.

For convenience, let $H = G \circ F$, $y = F(x)$, and $z = G(y) = H(x)$. Also, let $\mathbb{Y} = F_*\mathbb{X}$.

Consider the left-hand side first. We have that

$$H_*\mathbb{X}(z) = (G \circ F)'(x) \cdot \mathbb{X}(x) = G'(y)F'(x) \cdot \mathbb{X}(x),$$

where we have used the Chain Rule (Proposition 1.3.12) and the last formula for the pushforward given in Definition 1.7.1.

On the right-hand side, we have that

$$(G_*F_*\mathbb{X})(z) = (G_*\mathbb{Y})(z) = G'(y) \cdot \mathbb{Y}(y) = G'(y)F'(x) \cdot \mathbb{X}(x).$$

As the expressions obtained on both sides are the same, the result follows. □

Exercise. Show that in the linear case, Proposition 1.7.5 is equivalent to

$$T\left(Se^{tA}S^{-1}\right)T^{-1} = (TS)e^{tA}(TS)^{-1}.$$

1.8 Jacobi bracket

Next, we consider the pushforward of one vector field by the flow of another.

Let $U \subset \mathbb{R}^n$ be open, and let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$. Let Ψ_s denote the flow of $\mathbb{Y}$. As we know, $\Psi_s \in \mathrm{Diff}(U)$, so we may define

$$\mathbb{X}_s = \Psi_{s*}\mathbb{X}.$$

$\mathbb{X}_s$ is a family of vector fields in $\mathcal{X}(U)$ parameterised by s. The derivative of $\mathbb{X}_s$ with respect to s is also a vector field in $\mathcal{X}(U)$. We want to evaluate this derivative at $s = 0$. We proceed as follows: From the last formula in Definition 1.7.1, we have that

$$\mathbb{X}_s(\Psi_s(x)) = \Psi'_s(x) \cdot \mathbb{X}(x).$$

In terms of components (the second formula in Definition1.7.1),

$$\mathbb{X}^i_s(\Psi_s(x)) = \frac{\partial \Psi^i_s}{\partial x^j}(x)\mathbb{X}^j(x).$$

Next, differentiate both sides with respect to s and set $s = 0$. Note that, since $\Psi_0 = Id_U$, we have that $\mathbb{X}_0 = \mathbb{X}$. On the right-hand side, we get

$$\left.\frac{\partial^2 \Psi^i_s}{\partial s \partial x^j}(x)\right|_{s=0} \mathbb{X}^j(x) = \frac{\partial}{\partial x^j}\left(\left.\frac{\partial \Psi^i_s}{\partial s}\right|_{s=0}\right)\mathbb{X}^j(x) = \frac{\partial \mathbb{Y}^i}{\partial x^j}(x)\mathbb{X}^j(x), \tag{RHS}$$

where we have used the relation (**) between the flow Ψ_s and its generating vector field $\mathbb{Y}$. On the left-hand side, taking into account both instances of s, we get

$$\begin{aligned}
\left.\frac{\partial}{\partial s}\mathbb{X}^i_s(\Psi_s(x))\right|_{s=0} &= \left.\left(\frac{\partial \mathbb{X}^i_s}{\partial s}(\Psi_s(x))\right)\right|_{s=0} + \left.\frac{\partial \mathbb{X}^i_s}{\partial x^j}(\Psi_s(x))\frac{\partial \Psi^j_s}{\partial s}(x)\right|_{s=0} \\
&= \left.\frac{\partial \mathbb{X}^i_s}{\partial s}(x)\right|_{s=0} + \frac{\partial \mathbb{X}^i}{\partial x^j}(x)\mathbb{Y}^j(x), \qquad (LHS)
\end{aligned}$$

where once again we have used the relation (**) between the flow Ψ_s and its generating vector field $\mathbb{Y}$. Equating (RHS) and (LHS), we get

$$\left.\frac{\partial \mathbb{X}_s^i}{\partial s}(x)\right|_{s=0} = \frac{\partial \mathbb{Y}^i}{\partial x^j}(x)\mathbb{X}^j(x) - \frac{\partial \mathbb{X}^i}{\partial x^j}(x)\mathbb{Y}^j(x).$$

Notation. Let

$$\nabla = \left(\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\right)$$

We write

$$(\mathbb{X}\cdot\nabla)\,\mathbb{Y}^i = \left(\mathbb{X}^j\frac{\partial}{\partial x^j}\right)\mathbb{Y}^i = \mathbb{X}^j\frac{\partial \mathbb{Y}^i}{\partial x^j}.$$

Then, dropping the component index i, we may write

$$\left.\frac{\partial \mathbb{X}_s}{\partial s}\right|_{s=0} = (\mathbb{X}\cdot\nabla)\,\mathbb{Y} - (\mathbb{Y}\cdot\nabla)\,\mathbb{X}.$$

Definition 1.8.1 (Jacobi bracket)**.** Let $U \subset \mathbb{R}^n$ be open, and let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$. The **Jacobi bracket** of $\mathbb{X}$ and $\mathbb{Y}$, denoted $[\mathbb{X}, \mathbb{Y}]$, is the vector field in $\mathcal{X}(U)$ given by

$$[\mathbb{X}, \mathbb{Y}] := (\mathbb{X}\cdot\nabla)\,\mathbb{Y} - (\mathbb{Y}\cdot\nabla)\,\mathbb{X}.$$

Proposition 1.8.2 (Jacob bracket and flows)**.** Let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$, and let Ψ_s be the flow of $\mathbb{Y}$. Then

$$\left.\frac{\partial}{\partial s}\right|_{s=0} \Psi_{s*}\mathbb{X} = [\mathbb{X}, \mathbb{Y}].$$

Proof. See the preceding calculation. □

Example 1.8.3 (Simple Jacobi bracket calculation)**.** The following is a typical short part-question from previous exams: Let $\mathbb{X} = (y, x)$ and $\mathbb{Y} = (y^2, x^2)$ be vector fields on $\mathbb{R}^2$. Compute $[\mathbb{X}, \mathbb{Y}]$.

We have that

$$(\mathbb{X}\cdot\nabla)\,\mathbb{Y} = \left(y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}\right)(y^2, x^2) = (2xy, 2xy).$$

Also,

$$(\mathbb{Y}\cdot\nabla)\,\mathbb{X} = \left(y^2\frac{\partial}{\partial x} + x^2\frac{\partial}{\partial y}\right)(y, x) = (x^2, y^2).$$

Then

$$[\mathbb{X}, \mathbb{Y}] = (2xy - x^2, 2xy - y^2).$$

Example 1.8.4 (Jacobi bracket of linear vector fields)**.** See Problem Sheet 3.6(a).

Proposition 1.8.5 (Simple properties of the Jacobi bracket)**.** Let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$. Then

$$[\mathbb{X}, \mathbb{Y}] = -[\mathbb{Y}, \mathbb{X}] \qquad \text{(antisymmetry).} \tag{a}$$

If $\mathbb{Z} \in \mathcal{X}(U)$, then

$$[\mathbb{X}, \mathbb{Y} + \mathbb{Z}] = [\mathbb{X}, \mathbb{Y}] + [\mathbb{X}, \mathbb{Z}]. \qquad \text{(linearity).} \tag{b}$$

Also, if $f \in C^\infty(U)$, then

$$[\mathbb{X}, f\mathbb{Y}] = f[\mathbb{X}, \mathbb{Y}] + (\mathbb{X}\cdot\nabla f)\mathbb{Y} \qquad \text{(product rule).} \tag{c}$$

Proof. The first and second properties follow immediately from the formula for the Jacobi bracket in Definition 1.8.1. The third follows from noting that

$$(\mathbb{X} \cdot \nabla)(f\mathbb{Y}) = f(\mathbb{X} \cdot \nabla)\mathbb{Y} + (\mathbb{X} \cdot \nabla f)\mathbb{Y}.$$

□

The following shows that the pushforward preserves the Jacobi bracket.

Proposition 1.8.6 (Pushforward of Jacobi bracket)**.** Let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$, and let $F \in \mathrm{Diff}(U, V)$. Then

$$[F_*\mathbb{X}, F_*\mathbb{Y}] = F_*[\mathbb{X}, \mathbb{Y}].$$

Proof. It is possible to prove this result directly from the formulas for the pushforward and the Jacobi bracket, but the calculations are surprisingly long and rather tricky. The following argument uses the relations between brackets and flows.

From Proposition 1.8.2,

$$[\mathbb{X}, \mathbb{Y}] = \left.\frac{\partial}{\partial s}\right|_{s=0} \Psi_{s*}\mathbb{X}.$$

Then

$$F_*[\mathbb{X}, \mathbb{Y}] = F_* \left.\frac{\partial}{\partial s}\right|_{s=0} \Psi_{s*}\mathbb{X} = \left.\frac{\partial}{\partial s}\right|_{s=0} F_*\Psi_{s*}\mathbb{X}.$$

$\Big($Note that we have interchanged the partial derivative with respect to s and the pushforward F_*. This can be justified by expressing the partial derivative as a limit of quotients, as follows:

$$F_* \left.\frac{\partial}{\partial s}\right|_{s=0} \Psi_{s*}\mathbb{X} = \lim_{s\to 0} F_* \left(\frac{\Psi_{s*}\mathbb{X} - \mathbb{X}}{s}\right) = \lim_{s\to 0} \left(\frac{F_*\Psi_{s*}\mathbb{X} - F_*\mathbb{X}}{s}\right) = \left.\frac{\partial}{\partial s}\right|_{s=0} F_*\Psi_{s*}\mathbb{X}.\Big)$$

From Proposition 1.7.5,

$$F_*\Psi_{s*}\mathbb{X} = (F \circ \Psi_s)_*\mathbb{X} = (F \circ \Psi_s \circ F^{-1} \circ F)_*\mathbb{X} = (F \circ \Psi_s \circ F^{-1})_* F_*\mathbb{X}.$$

Let $\Gamma_s = F \circ \Psi_s \circ F^{-1}$. We have shown that

$$F_*[\mathbb{X}, \mathbb{Y}] = \left.\frac{\partial}{\partial s}\right|_{s=0} \Gamma_{s*}F_*\mathbb{X}.$$

By Proposition 1.7.3, Γ_s is the flow of the vector field $F_*\mathbb{Y}$. It follows from Proposition 1.8.2 that

$$\left.\frac{\partial}{\partial s}\right|_{s=0} \Gamma_{s*}F_*\mathbb{X} = [F_*\mathbb{X}, F_*\mathbb{Y}].$$

Therefore,

$$F_*[\mathbb{X}, \mathbb{Y}] = [F_*\mathbb{X}, F_*\mathbb{Y}].$$

□

Example 1.8.7 (Pushforward of the Jacobi bracket – Linear case)**.** See Problem Sheet 3.6(b).

Next, we consider the pushforward of the Jacobi bracket of two vector fields by the flow of a third. Differentiating with respect to the flow parameter yields an important result, called the Jacobi identity. Let $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \in \mathcal{X}(U)$, and let Γ_s be the flow of $\mathbb{Z}$. We consider $\Gamma_{s*}[\mathbb{X}, \mathbb{Y}]$. From Proposition 1.8.6,

$$\Gamma_{s*}[\mathbb{X}, \mathbb{Y}] = [\Gamma_{s*}\mathbb{X}, \Gamma_{s*}\mathbb{Y}].$$

Differentiating the left-hand side and setting $s = 0$, we obtain from Proposition 1.8.2 that

$$\left.\frac{\partial}{\partial s}\right|_{s=0} \Gamma_{s*}[\mathbb{X}, \mathbb{Y}] = [[\mathbb{X}, \mathbb{Y}], \mathbb{Z}] = -[\mathbb{Z}, [\mathbb{X}, \mathbb{Y}]]. \tag{LHS}$$

Differentiating the right-hand side and setting $s = 0$, we get that

$$\left.\frac{\partial}{\partial s}\right|_{s=0} [\Gamma_{s*}\mathbb{X}, \Gamma_{s*}\mathbb{Y}] = \left[\left.\frac{\partial}{\partial s}\right|_{s=0} \Gamma_{s*}\mathbb{X}, \mathbb{Y}\right] + \left[\mathbb{X}, \left.\frac{\partial}{\partial s}\right|_{s=0} \Gamma_{s*}\mathbb{Y}\right] = -[[\mathbb{Z}, \mathbb{X}], \mathbb{Y}] + -[\mathbb{X}, [\mathbb{Z}, \mathbb{Y}]], \tag{RHS}$$

where we have used Proposition 1.8.2 again. $\Bigg($In fact, some argument is required to justify the first equality in the preceding. This can be done by expressing the partial derivative as the limit of a quotient and using the linearity of the Jacobi bracket (Proposition 1.8.5(b)).

$$\begin{aligned}
\left.\frac{\partial}{\partial s}\right|_{s=0} [\Gamma_{s*}\mathbb{X}, \Gamma_{s*}\mathbb{Y}] &= \lim_{s\to 0} \frac{1}{s}\left([\Gamma_{s*}\mathbb{X}, \Gamma_{s*}\mathbb{Y}] - [\mathbb{X}, \mathbb{Y}]\right) \\
&= \lim_{s\to 0} \frac{1}{s}\left([\Gamma_{s*}\mathbb{X}, \Gamma_{s*}\mathbb{Y}] - [\Gamma_{s*}\mathbb{X}, \mathbb{Y}] + [\Gamma_{s*}\mathbb{X}, \mathbb{Y}] - [\mathbb{X}, \mathbb{Y}]\right) \\
&= \lim_{s\to 0}\left[\Gamma_{s*}\mathbb{X}, \frac{\Gamma_{s*}\mathbb{Y} - \mathbb{Y}}{s}\right] + \lim_{s\to 0}\left[\frac{\Gamma_{s*}\mathbb{X} - \mathbb{X}}{s}, \mathbb{Y}\right] \\
&= \left[\left.\frac{\partial}{\partial s}\right|_{s=0} \Gamma_{s*}\mathbb{X}, \mathbb{Y}\right] + \left[\mathbb{X}, \left.\frac{\partial}{\partial s}\right|_{s=0} \Gamma_{s*}\mathbb{Y}\right].\Bigg)
\end{aligned}$$

Equating (LHS) and (RHS) and cancelling the minus signs, we get

$$[\mathbb{Z}, [\mathbb{X}, \mathbb{Y}]] = [[\mathbb{Z}, \mathbb{X}], \mathbb{Y}] + [\mathbb{X}, [\mathbb{Z}, \mathbb{Y}]].$$

Proposition 1.8.8 (Jacobi identity)**.** Let $\mathbb{X}, \mathbb{Y}, \mathbb{Z} \in \mathcal{X}(U)$. Then

$$[\mathbb{Z}, [\mathbb{X}, \mathbb{Y}]] = [[\mathbb{Z}, \mathbb{X}], \mathbb{Y}] + [\mathbb{X}, [\mathbb{Z}, \mathbb{Y}]].$$

Proof. See preceding calculation. □

Using the antisymmetry of the Jacobi bracket (Proposition 1.8.5(a)), we can also write the Jacobi identity as

$$[[\mathbb{X}, \mathbb{Y}], \mathbb{Z}] + [[\mathbb{Y}, \mathbb{Z}], \mathbb{X}] + [[\mathbb{Z}, \mathbb{X}], \mathbb{Y}] = 0.$$

The Jacobi identity can also be proved directly using the explicit formula for the Jacobi bracket – see Problem Sheet 3.4.

Example 1.8.9 (Pushforward of the Jacobi bracket – Linear case)**.** See Problem Sheet 3.6(c).

Next, we have an extension of Proposition 1.8.2 – there we calculated the derivative with respect to s of the pushforward with respect to a flow, Ψ_{s*}, at $s = 0$. In the following, we evaluate the derivative for arbitrary s.

Proposition 1.8.10. Let $U \subset \mathbb{R}^n$ be open, let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$ and let Ψ_s denote the flow of $\mathbb{Y}$. Then

$$\frac{\partial}{\partial s}\Psi_{s*}\mathbb{X} = [\Psi_{s*}\mathbb{X}, \mathbb{Y}].$$

Note that Proposition 1.8.2 follows from the special case where $s = 0$ (note that $\Psi_{0*}\mathbb{X} = \mathbb{X}$ for any vector field $\mathbb{X}$).

Proof. We express the derivative as the limit of a quotient, and use the composition properties of flows and pushforward maps. We have that

$$\frac{\partial}{\partial s}\Psi_{s*}\mathbb{X} = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\Psi_{(s+\epsilon)*}\mathbb{X} - \Psi_{s*}\mathbb{X}\right).$$

From Proposition 1.6.9,

$$\Psi_{s+\epsilon} = \Psi_\epsilon \circ \Psi_s,$$

and from Proposition 1.7.5,

$$\Psi_{(s+\epsilon)*} = \Psi_{\epsilon*}\Psi_{s*}.$$

Therefore,

$$\frac{\partial}{\partial s}\Psi_{s*}\mathbb{X} = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\Psi_{\epsilon*}\Psi_{s*}\mathbb{X} - \Psi_{s*}\mathbb{X}\right) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\Psi_{\epsilon*}\tilde{\mathbb{X}} - \tilde{\mathbb{X}}\right),$$

where $\tilde{\mathbb{X}} = \Psi_{s*}\mathbb{X}$. But the last expression gives

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\Psi_{\epsilon*}\tilde{\mathbb{X}} - \tilde{\mathbb{X}}\right) = \left.\frac{\partial}{\partial \epsilon}\right|_{\epsilon=0}\Psi_{\epsilon*}\tilde{\mathbb{X}} = [\tilde{\mathbb{X}}, \mathbb{Y}],$$

where in the last equality we have used Proposition 1.8.2. Substituting above and recalling how we defined $\tilde{\mathbb{X}}$, we get that

$$\frac{\partial}{\partial s}\Psi_{s*}\mathbb{X} = [\Psi_{s*}\mathbb{X}, \mathbb{Y}].$$

□

The following proposition shows that a vector field is invariant under pushforward by its own flow.

Proposition 1.8.11. Let $\mathbb{Y} \in \mathcal{X}(U)$ and let Ψ_s be its flow. Then

$$\Psi_{s*}\mathbb{Y} = \mathbb{Y}.$$

Proof. From Proposition 1.7.3, for any diffeomorphism $F \in \mathrm{Diff}(U)$,

$$F_*\mathbb{Y} = \left.\frac{\partial}{\partial t}\right|_{t=0} F \circ \Psi_t \circ F^{-1}.$$

Taking $F = \Psi_s$, we get that

$$\Psi_{s*}\mathbb{Y} = \left.\frac{\partial}{\partial t}\right|_{t=0}\Psi_s \circ \Psi_t \circ \Psi_{-s} = \left.\frac{\partial}{\partial t}\right|_{t=0}\Psi_t.$$

since $\Psi_s \circ \Psi_t \circ \Psi_{-s} = \Psi_t$. Since

$$\left.\frac{\partial}{\partial t}\right|_{t=0}\Psi_t = \mathbb{Y},$$

the result follows. □

Theorem 1.8.12. Let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$ and let Φ_t, Ψ_s denote the flows of $\mathbb{X}$, and $\mathbb{Y}$. Then

$$\Phi_t \circ \Psi_s = \Psi_s \circ \Phi_t, \forall s,t \iff [\mathbb{X}, \mathbb{Y}] = 0.$$

Proof. We'll precede by constructing a chain of equivalent equations, starting with the left-hand side assertion. After composing with Ψ_{-s} on the left, we have that

$$\Phi_t \circ \Psi_s = \Psi_s \circ \Phi_t, \quad \forall s,t \iff \Phi_t = \Psi_s \circ \Phi_t \circ \Psi_{-s}, \quad \forall s,t.$$

Φ_t is the flow generated by $\mathbb{X}$, of course, and $\Psi_s \circ \Phi_t \circ \Psi_{-s}$, for fixed s, is the flow in t generated by $\Psi_{s*}\mathbb{X}$ (cf Proposition 1.7.3). Flows are the same if and only if their generating vector fields are the same (this is uniqueness of solutions to ODEs). Therefore,

$$\Phi_t = \Psi_s \circ \Phi_t \circ \Psi_{-s}, \quad \forall s, t \iff \mathbb{X} = \Psi_{s*}\mathbb{X}, \quad \forall s.$$

Since $\mathbb{X}$ does not depend on s, the right-hand side equation above implies that the derivative of $\Psi_{s*}\mathbb{X}$ with respect to s must vanish. Conversely, if we know that $\Psi_{s*}\mathbb{X}$ is independent of s, then it must be equal to $\Psi_{0*}\mathbb{X}$, which is just $\mathbb{X}$. Therefore,

$$\mathbb{X} = \Psi_{s*}\mathbb{X}, \quad \forall s \iff \frac{\partial}{\partial s}\Psi_{s*}\mathbb{X} = 0, \quad \forall s.$$

From Proposition 1.8.10,

$$\frac{\partial}{\partial s}\Psi_{s*}\mathbb{X} = 0, \quad \forall s \iff [\Psi_{s*}\mathbb{X}, \mathbb{Y}] = 0, \quad \forall s.$$

Since $\mathbb{Y} = \Psi_{s*}\mathbb{Y}$ (cf Proposition 1.8.11), the bracket in the right-hand side equation above may be written as $[\Psi_{s*}\mathbb{X}, \Psi_{s*}\mathbb{Y}]$, or, using Proposition 1.8.6, $\Psi_{s*}[\mathbb{X}, \mathbb{Y}]$. Therefore,

$$[\Psi_{s*}\mathbb{X}, \mathbb{Y}] = 0, \quad \forall s \iff \Psi_{s*}[\mathbb{X}, \mathbb{Y}] = 0, \quad \forall s.$$

But it is clear that

$$\Psi_{s*}[\mathbb{X}, \mathbb{Y}] = 0, \quad \forall s \iff [\mathbb{X}, \mathbb{Y}] = 0.$$

For if $[\mathbb{X}, \mathbb{Y}]$ vanishes, then its pushforward vanishes (any pushforward of the zero vector field is zero). And if $\Psi_{s*}[\mathbb{X}, \mathbb{Y}] = 0$ (for any s in fact, not necessarily all), then application of Ψ_{-s*} to both sides of this equation, and the fact that $\Psi_{-s*}\Psi_{s*}$ is the identity map on vector fields, yields $[\mathbb{X}, \mathbb{Y}] = 0$. Thus we come to the end of our chain of equivalences, and the result is proved. □

Example 1.8.13 (Simple illustration of Theorem 1.8.12)**.** Let $U = \mathbb{R}^2$ with coordinates (x, v), and let $\mathbb{X}(x, v) = (v, a)$ and $\mathbb{Y} = (b, c)$, where a, b and c are constants. We have that

$$[\mathbb{X}, \mathbb{Y}] = (-c, 0),$$

which vanishes if and only if $c = 0$.

The flow of $\mathbb{X}$ was found in Problem 2.8 (there we set $a = 1$). The result is

$$\Phi_t(x, v) = (x + vt + at^2/2, v + at),$$

and may be interpreted in terms of position x and velocity v under uniform acceleration a as a function of time. The flow for $\mathbb{Y}$ is even simpler; it's just a translation in x and v, as follows:

$$\Psi_s(x, v) = (x + bs, v + cs).$$

We have that

$$(\Phi_t \circ \Psi_s)(x, v) = \Phi_t(\Psi_s(x, v)) = \Phi_t(x + bs, v + cs) = (x + bs + (v + cs)t + at^2/2, v + cs + at).$$

On the other hand,

$$(\Psi_s \circ \Phi_t)(x, v) = \Psi_s(\Phi_t(x, v)) = \Psi_s(x + vt + at^2/2, v + at) = (x + vt + at^2/2 + bs, v + at + cs).$$

Comparison of the two shows that

$$(\Phi_t \circ \Psi_s)(x, v) - (\Psi_s \circ \Phi_t)(x, v) = (cst, 0).$$

Thus, the flows commute if and only if $c = 0$.

This can be interpreted as saying that motion under constant acceleration is invariant under translations in position (you can translate either at the beginning or the end of the motion) but not under translations in velocity (if you give an object a kick, it makes a difference whether you do it at the beginning or at the end of its motion; a kick at the beginning affects the final position, while a kick at the end doesn't).

1.9 The Pullback (functions only)

Definition 1.9.1 (Pullback on functions)**.** Let $U, V \subset \mathbb{R}^n$ be open, and let $F \in \mathrm{Diff}(U, V)$. The **pullback** by F, denoted by F^*, is a map from smooth functions on V to smooth functions on U, which takes a function g to the function $g \circ F$. That is,

$$F^* : C^\infty(V) \to C^\infty(U) : g \mapsto F^*g := g \circ F$$

As we will see in Section 3.3, the definition does not require that F be a diffeomorphism; indeed, F could be any smooth map, and the definition would still make sense.

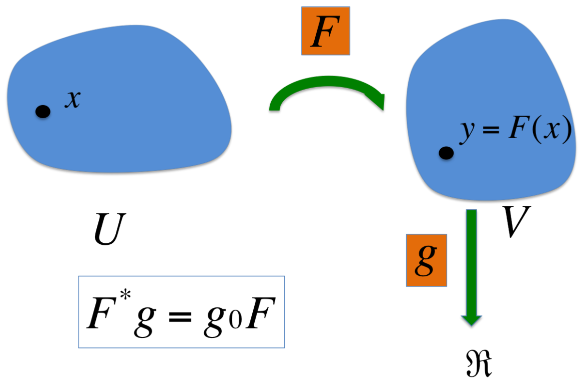


Figure 17: The pullback. F is a diffeomorphism from U to V, g a smooth function on V and $F^*g = g \circ F$ a smooth function on U.

The pullback F^* and the diffeomorphism F are equivalent ways of viewing the same basic object. To elaborate, it is clear from the definition that if we know F, then we know F^*. The converse is also true, namely that F^* determines F. To see this, let us define functions $\widehat{y}^j$ on V by

$$\widehat{y}^j(y) = y^j.$$

That is, the function $\widehat{y}^j$ picks out the jth component of its argument. Now consider the pullback of $\widehat{y}^j$. We have that

$$(F^*\widehat{y}^j)(x) = \widehat{y}^j(F(x)) = F^j(x);$$

the pullback of $\widehat{y}^j$ is the jth component of F. Thus, if we know the pullbacks just of the functions $\widehat{y}^j$ for $j = 1, \ldots, n$, then we have determined F.

The pullback F^* is defined on an infinite-dimensional function space (which makes it more complicated), but it is a linear map (which makes it simpler); that is, if $g_1, g_2 \in C^\infty(V)$ and $a_1, a_2 \in \mathbb{R}$, then

$$F^*(a_1g_1 + a_2g_2) = a_1F^*g_1 + a_2F^*g_2.$$

In contrast, F is defined as a map between finite-dimensional spaces, but it may be a nonlinear.

For some considerations, it is advantageous to consider one point of view or the other. For our discussion of non-commuting flows in the following section, the pullback point of view is useful. Hence we are developing it here.

Proposition 1.9.2 (Pullback by a composition)**.** Let $U, V, W \subset \mathbb{R}^n$ be open, and let $F \in \mathrm{Diff}(U, V)$, $G \in \mathrm{Diff}(V, W)$, so that $G \circ F \in \mathrm{Diff}(U, W)$. Then

$$(G \circ F)^* = F^*G^*.$$

Proof. Let $h \in C^\infty(W)$. Then

$$((G \circ F)^*h)(x) = h((G \circ F)(x)) = h(G(F(x))).$$

On the other side, we have that

$$(F^*G^*h)(x) = (G^*h)(F(x)) = h(G(F(x))).$$

$\square$

What we really want to consider here are pullbacks by flows. Let $\mathbb{X}$ be a smooth vector field on U with flow Φ_t, and let f be a function on U. Regarding Φ_t as a diffeomorphism from U to U, we can consider the pullback of f by Φ_t, namely $\Phi_t^* f$. Let us calculate the derivative of $\Phi_t^* f$ with respect to t. We have that

$$\frac{\partial}{\partial t}(\Phi_t^* f)(x) = \frac{\partial}{\partial t} f(\Phi_t(x)).$$

Using the Chain Rule (Proposition 1.3.12), we get that

$$\frac{\partial}{\partial t} f(\Phi_t(x)) = \frac{\partial f}{\partial x^j}(\Phi_t(x))\,\frac{\partial \Phi_t^j}{\partial t}(x).$$

From the relation between vectors fields and flows (see (**) and (***) on page 17), we have that

$$\frac{\partial \Phi_t^j}{\partial t}(x) = \mathbb{X}^j(\Phi_t(x)).$$

Thus,

$$\frac{\partial}{\partial t}(\Phi_t^* f)(x) = \mathbb{X}^j(\Phi_t(x))\frac{\partial f}{\partial x^j}(\Phi_t(x)) = (\mathbb{X}\cdot\nabla f)(\Phi_t(x)) = (\Phi_t^*(\mathbb{X}\cdot\nabla f))(x).$$

Omitting the argument x, this can be written compactly as

$$\frac{\partial}{\partial t}\Phi_t^* f = \Phi_t^*(\mathbb{X}\cdot\nabla f). \tag{8}$$

Let us take a second t-derivative. We get that

$$\frac{\partial^2}{\partial t^2}\Phi_t^* f = \frac{\partial}{\partial t}\Phi_t^*(\mathbb{X}\cdot\nabla f).$$

Letting $g = \mathbb{X}\cdot\nabla f$, we may write this as

$$\frac{\partial^2}{\partial t^2}\Phi_t^* f = \frac{\partial}{\partial t}\Phi_t^* g = \Phi_t^*(\mathbb{X}\cdot\nabla g),$$

where in the last equality we have used (8). Finally, we replace g back by $\mathbb{X}\cdot\nabla f$ to get

$$\frac{\partial^2}{\partial t^2}\Phi_t^* f = \Phi_t^*((\mathbb{X}\cdot\nabla)\mathbb{X}\cdot\nabla f) = \Phi_t^*((\mathbb{X}\cdot\nabla)^2 f).$$

It is clear that this can be generalised to higher t-derivatives, as follows:

$$\frac{\partial^j}{\partial t^j}\Phi_t^* f = \Phi_t^*((\mathbb{X}\cdot\nabla)^j f). \tag{9}$$

The expression

$$\mathbb{X}\cdot\nabla = \mathbb{X}^j\frac{\partial}{\partial x^j}$$

maps smooth functions into smooth functions, and is called a linear first-order differential operator ("operator" because it maps functions to functions, "first-order differential" because it involves first-order partial derivatives, and "linear" because $\mathbb{X}\cdot\boldsymbol{\nabla}(f+g) = \mathbb{X}\cdot\boldsymbol{\nabla} f + \mathbb{X}\cdot\boldsymbol{\nabla} g$). There is a actually a special name and notation for this operator, which we introduce next.

Definition 1.9.3 (Lie derivative of functions)**.** Let $U \subset \mathbb{R}^n$ be open, and let $\mathbb{X} \subset \mathcal{X}(U)$. The **Lie derivative** with respect to $\mathbb{X}$, denoted $L_{\mathbb{X}}$, is the mapping

$$L_{\mathbb{X}} : C^\infty(U) \to C^\infty(U); f \mapsto L_{\mathbb{X}} f = \mathbb{X} \cdot \nabla f.$$

The Lie derivative $L_{\mathbb{X}}$ and the vector field $\mathbb{X}$ stand in the same relation to each other as do the pullback F^* and the diffeomorphism F. $L_{\mathbb{X}}$ and F^* are linear maps (operators) defined on an infinite-dimensional function space, while $\mathbb{X}$ and F are nonlinear maps defined on a finite-dimensional space.

Using the Lie derivative notation, we may write (9) as

$$\frac{\partial^j}{\partial t^j} \Phi_t^* f = \Phi_t^* \left(L_{\mathbb{X}}^j f \right). \tag{10}$$

With these formulas for the t-derivatives of $\Phi_t^* f$, we can develop a power series for $\Phi_t^* f$.

Proposition 1.9.4 (Power series for pullback by a flow)**.** Suppose $\Phi_t^* f(x)$ is analytic in t; that is, suppose $\Phi_t^* f(x)$ has a convergent power series in t. Then that power series is given by

$$\Phi_t^* f = e^{tL_{\mathbb{X}}} f,$$

where

$$e^{tL_{\mathbb{X}}} := \sum_{j=0}^{\infty} \frac{t^j}{j!} L_{\mathbb{X}}^j.$$

Proof. By assumption, $\Phi_t^* f$ has a convergent power series, which we may write as

$$\Phi_t^* f = \sum_{j=0}^{\infty} \left(\frac{\partial^j}{\partial t^j} \Phi_t^* f \right) \bigg|_{t=0} \frac{t^j}{j!}.$$

From (10) and the fact that $\Phi_0^* f = f$ for all f (since Φ_0 is the identity map), we have that

$$\left(\frac{\partial^j}{\partial t^j} \Phi_t^* f \right) \bigg|_{t=0} = \Phi_t^* \left(L_{\mathbb{X}}^j f \right) \Big|_{t=0} = L_{\mathbb{X}}^j f.$$

Substituting this expression into the power series above, we get

$$\Phi_t^* f = \sum_{j=0}^{\infty} (L_{\mathbb{X}}^j f) \frac{t^j}{j!} = e^{tL_{\mathbb{X}}} f,$$

as required. $\square$

Example 1.9.5 (Pullback: Linear case)**.** Let us illustrate the concepts and results introduced in this section in the linear case. Throughout, $U = \mathbb{R}^n$ and $V = \mathbb{R}^p$.

A linear function $g(x)$ on $\mathbb{R}^p$ is a function of the form

$$g(y) = b \cdot y,$$

where $b \in \mathbb{R}^p$ is a fixed vector.

A linear map $F : \mathbb{R}^n \to \mathbb{R}^p$ is a map of the form

$$F(x) = M \cdot x,$$

where M is an $p \times n$ matrix.

We have that

$$(F^* g)(x) = g(F(x)) = b \cdot (F(x)) = b \cdot M \cdot x.$$

We can also write this as

$$(F^*g)(x) = (M^T b) \cdot x,$$

where M^T is the transpose of M, which makes it explicit that F^*g is also a linear function.

Next, consider the linear vector field $\mathbb{X}(x) = A \cdot x$ on $\mathbb{R}^n$, where A is an $n \times n$ matrix, and a linear function $f(x) = a \cdot x$ on $\mathbb{R}^n$, where $a \in \mathbb{R}^n$. Then

$$(L_{\mathbb{X}} f)(x) = \mathbb{X}(x) \cdot \nabla f(x).$$

It is clear that $\nabla f(x) = a$. Therefore,

$$(L_{\mathbb{X}} f)(x) = a \cdot \mathbb{X}(x) = a \cdot A \cdot x.$$

It follows that

$$(L_{\mathbb{X}}^j f)(x) = a \cdot A^j \cdot x.$$

Then

$$\sum_{j=0}^{\infty} \frac{t^j}{j!} L_{\mathbb{X}}^j f(x) = \sum_{j=0}^{\infty} \frac{t^j a \cdot A^j}{j!} \cdot x = a \cdot \left(\sum_{j=0}^{\infty} \frac{t^j A^j}{j!} \right) = a \cdot e^{At} \cdot x = \Phi_t^* f(x),$$

where $\Phi_t(x) = e^{At} \cdot x$ is the flow of $\mathbb{X}(x)$. Thus we verify Proposition 1.9.4 directly.

Just as the diffeomorphism F and pullback map F^* are different representations of the same object, so too the vector field $\mathbb{X}$ and $L_{\mathbb{X}}$ are different representations of the same object. (We won't elaborate on this relationship further, but for those of you who wish to pursue it, it can be understood in terms of the group of diffeomorphisms on an open set U and its natural representation on $C^\infty(U)$.) With this in mind, we recall that we have an operation, the Jacobi bracket, on vector fields. We would like to determine what the Jacobi bracket corresponds to in terms of the Lie derivative.

Proposition 1.9.6 (Jacobi bracket and Lie derivative)**.** Let $U \subset \mathbb{R}^n$ be open, and let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$. Then

$$L_{[\mathbb{X},\mathbb{Y}]} = L_{\mathbb{X}} L_{\mathbb{Y}} - L_{\mathbb{Y}} L_{\mathbb{X}}.$$

Proof. We will apply both sides of the preceding to a smooth function, f, and find we obtain the same result. The key point is that, on the right-hand side, terms involving second partial derivatives vanish, due to the equality of mixed partials.

On the left-hand side, we get

$$L_{[\mathbb{X},\mathbb{Y}]} f = [\mathbb{X}, \mathbb{Y}]^j \frac{\partial f}{\partial x^j} = \left(\mathbb{X}^k \frac{\partial \mathbb{Y}^j}{\partial x^k} - \mathbb{Y}^k \frac{\partial \mathbb{X}^j}{\partial x^k} \right) \frac{\partial f}{\partial x^j}, \tag{LHS}$$

where we have used the component expression for $[\mathbb{X}, \mathbb{Y}]^j$.

On the right-hand side, we get

$$(L_{\mathbb{X}} L_{\mathbb{Y}} - L_{\mathbb{Y}} L_{\mathbb{X}}) f = L_{\mathbb{X}}(L_{\mathbb{Y}} f) - L_{\mathbb{Y}}(L_{\mathbb{X}} f).$$

Writing the first term in terms of components, we get

$$L_{\mathbb{X}}(L_{\mathbb{Y}} f) = \mathbb{X}^k \frac{\partial}{\partial x^k} (L_{\mathbb{Y}} f) = \mathbb{X}^k \frac{\partial}{\partial x^k} \left(\mathbb{Y}^j \frac{\partial f}{\partial x^j} \right) = \mathbb{X}^k \frac{\partial \mathbb{Y}^j}{\partial x^k} \frac{\partial f}{\partial x^j} + \mathbb{X}^k \mathbb{Y}^j \frac{\partial^2 f}{\partial x^k \partial x^j}.$$

Note that second partial derivatives of f appear. Similarly, we have that

$$L_{\mathbb{Y}}(L_{\mathbb{X}} f) = \mathbb{Y}^k \frac{\partial \mathbb{X}^j}{\partial x^k} \frac{\partial f}{\partial x^j} + \mathbb{Y}^k \mathbb{X}^j \frac{\partial^2 f}{\partial x^k \partial x^j} = \mathbb{Y}^k \frac{\partial \mathbb{X}^j}{\partial x^k} \frac{\partial f}{\partial x^j} + \mathbb{X}^k \mathbb{Y}^j \frac{\partial^2 f}{\partial x^j \partial x^k},$$

where in the last equality we have interchanged the summation indices j and k (which we are free to do). Subtracting the preceding expression from the one before and using the equality of mixed partials (Proposition 1.3.16), we get that

$$\left(L_{\mathbb{X}}L_{\mathbb{Y}} - L_{\mathbb{Y}}L_{\mathbb{X}}\right) f = \mathbb{X}^k \frac{\partial \mathbb{Y}^j}{\partial x^k}\frac{\partial f}{\partial x^j} - \mathbb{Y}^k \frac{\partial \mathbb{X}^j}{\partial x^k}\frac{\partial f}{\partial x^j}. \qquad (RHS)$$

We see that (LHS) and (RHS) are the same, and the result follows.

□

1.10 Noncommuting flows

We want to extend our consideration of the relationship between the Jacobi bracket of vector fields and the commutativity of their corresponding flows, as in Theorem 1.8.12, to the case where the flows do not commute.

Theorem 1.10.1 (Jacobi bracket and non-commuting flows)**.** Let $U \subset \mathbb{R}^n$ be open. Let $\mathbb{X}, \mathbb{Y} \in \mathcal{X}(U)$ be vector fields on U and let Φ_t, Ψ_s be their respective flows. Let

$$\Gamma_{s,t} = \Psi_{-s} \circ \Phi_{-t} \circ \Psi_s \circ \Phi_t.$$

Regarding s and t as small, we have that

$$\Gamma^*_{s,t} = I + stL_{[\mathbb{X},\mathbb{Y}]} + O(3),$$

where $O(3)$ denotes terms of cubic or higher order in s and t (e.g., terms proportional to s^3, s^2t, or more generally, to $s^a t^b$, where $a + b > 2$).

Proof. Consider $\Gamma^*_{s,t} f$. By Proposition 1.9.2,

$$\Gamma^*_{s,t} f = \Phi^*_t \Psi^*_s \Phi^*_{-t} \Psi^*_{-s} f.$$

By Proposition 1.9.4, this is given by

$$e^{tL_{\mathbb{X}}} e^{sL_{\mathbb{Y}}} e^{-tL_{\mathbb{X}}} e^{-sL_{\mathbb{Y}}} f.$$

Let us expand the exponentials through terms second order in s and t. We obtain

$$\left(1 + tL_{\mathbb{X}} + \tfrac{1}{2}t^2L^2_{\mathbb{X}}\right)\left(1 + sL_{\mathbb{Y}} + \tfrac{1}{2}s^2L^2_{\mathbb{Y}}\right)\left(1 - tL_{\mathbb{X}} + \tfrac{1}{2}t^2L^2_{\mathbb{X}}\right)\left(1 - sL_{\mathbb{Y}} + \tfrac{1}{2}s^2L^2_{\mathbb{Y}}\right) f + O(3).$$

Let us compute the contributions order by order. At zeroth order in s and t, we obtain f (obtained from taking the term '1' from each of the four factors above). At first order in s and t, we obtain

$$tL_{\mathbb{X}}f + sL_{\mathbb{Y}}f - tL_{\mathbb{X}}f - sL_{\mathbb{Y}}f = 0,$$

so there is no first-order contribution. At second order, we obtain contributions for the terms s^2 and t^2 in each of the factors as well as from products of first-order terms from pairs of factors. The first set of terms yield

$$\tfrac{1}{2}\left(t^2L^2_{\mathbb{X}}f + s^2L^2_{\mathbb{Y}}f + t^2L^2_{\mathbb{X}}f + s^2L^2_{\mathbb{Y}}f\right) = t^2L^2_{\mathbb{X}}f + s^2Y^2_{\mathbb{Y}}f.$$

The second set yield

$$tsL_{\mathbb{X}}L_{\mathbb{Y}}f - t^2L^2_{\mathbb{X}}f - tsL_{\mathbb{X}}L_{\mathbb{Y}}f - stL_{\mathbb{Y}}L_{\mathbb{X}}f - s^2L^2_{\mathbb{Y}}f + tsL_{\mathbb{X}}L_{\mathbb{Y}}f.$$

Combining the preceding expressions and accounting for cancellations (in particular, terms in s^2 and t^2 cancel), we obtain the second-order contribution

$$tsL_{\mathbb{X}}L_{\mathbb{Y}}f - tsL_{\mathbb{X}}L_{\mathbb{Y}}f - stL_{\mathbb{Y}}L_{\mathbb{X}} + tsL_{\mathbb{X}}L_{\mathbb{Y}} = st\left(L_{\mathbb{X}}L_{\mathbb{Y}} - L_{\mathbb{Y}}L_{\mathbb{X}}\right) f.$$

By Proposition 1.9.6, this can be written as

$$stL_{[\mathbb{X},\mathbb{Y}]}f.$$

Combining the previous calculations, we obtain

$$\Gamma^*_{s,t}f = f + stL_{[\mathbb{X},\mathbb{Y}]}f + O(3),$$

as required. □

In the preceding, we worked with the pullback representation of flows. In keeping with the discussion following Definition 1.9.1, there is an equivalent statement in terms of flows. This statement is illustrated in Figure 18;

$$\Gamma_{s,t}(x) = x + st[\mathbb{X},\mathbb{Y}](x) + O(3).$$

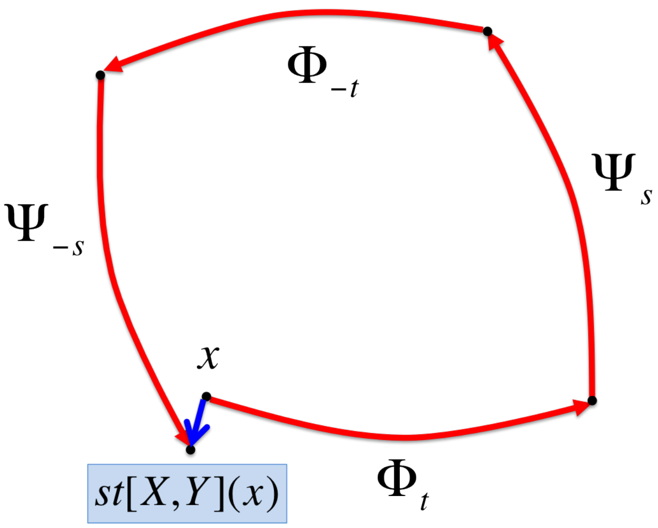


Figure 18: The flows Φ_t, Ψ_s, Φ_{-t} and Ψ_{-s} are applied in succession to an initial point x. For s and t small, the initial and final point differ by the displacement $st[\mathbb{X},\mathbb{Y}](x)$, up to higher-order corrections.

Example 1.10.2 (Noncommuting flows (following Example 1.8.13))**.** Let $\mathbb{X}(x,v)$, $\mathbb{Y}(x,v)$, Φ_t and Ψ_s be as in Example 1.8.13. There, we saw that

$$[\mathbb{X},\mathbb{Y}](x,v) = (-c,0).$$

We also saw that

$$\Phi_t(x,v) = (x + vt + at^2/2, v + at),$$
$$\Psi_s(x,v) = (x + bs, v + cs),$$

and

$$\Psi_s(\Phi_t(x,v)) = (x + bs + vt + \tfrac{1}{2}at^2, v + cs + at).$$

Then

$$\Phi_{-t}(\Psi_s(\Phi_t(x,v))) = (x + bs + vt + \tfrac{1}{2}at^2 - (v + cs + at)t + \tfrac{1}{2}at^2, v + cs) = (x + bs - cst, v + cs),$$

and

$$\Psi_{-s}(\Phi_{-t}(\Psi_s(\Phi_t(x,v)))) = (x - cst, v) = (x,v) + st[\mathbb{X},\mathbb{Y}](x,v),$$

in accord with Theorem 1.10.1.

1.11 The Frobenius Theorem

The Frobenius theorem states that, under certain conditions (and only under such conditions), certain systems of first-order PDEs have a unique solution, at least in a neighbourhood of the initial data. Moreover, when the necessary conditions are satisfied, the theorem provides an explicit formula for the solution in terms of flows of ODEs.

1.11.1 Some motivation.

Consider the following trivial one-dimensional first-order ordinary differential equation,

$$\dot{x}(t) = f(t), \quad x(0) = x_0..$$

The reason this equation is trivial is that, because the right-hand side does not depend on the unknown function $x(t)$, the solution can be written explicitly as an integral,

$$x(t) = x_0 + \int_0^t f(s)\, ds.$$

Let us consider an analogously simple partial differential equation for a function u of two variables, x and y,

$$\begin{aligned} \frac{\partial u}{\partial x} &= f(x, y), \\ \frac{\partial u}{\partial y} &= g(x, y), \\ u(x_0, y_0) &= u_0. \end{aligned}$$

Here, too, the right-hand side does not depend on the unknown function u. However, in this case, a solution does not automatically exist! A necessary condition for a solution to exist is the equality of mixed partials, namely $\partial^2 u/\partial y \partial x = \partial^2 u/\partial x \partial y$. For this to hold, we must have that

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x}.$$

If this condition is satisfied, then we can write down an explicit formula for $u(x, y)$ in the form of an integral,

$$u(x, y) = u_0 + \int_{(x_0, y_0)}^{(x,y)} (f, g) \cdot \mathbf{ds},$$

where the integral may be taken along any path joining (x_0, y_0) to (x, y).

In what follows, we generalise to systems of first-order partial differential equations in which, unlike the case above, the right-hand side depends on the unknown function(s).

1.11.2 Basic example.

We first consider the case of a system of two first-order partial differential equations for a single function $u(x, y)$ of two variables (later, we will consider the general case of p functions in q variables). Let $f = f(x, y, z)$ and $g = g(x, y, z)$ be functions on an open set $U \subset \mathbb{R}^3$. Consider the system

$$\begin{aligned} \frac{\partial u}{\partial x}(x, y) &= f(x, y, u(x, y)), \\ \frac{\partial u}{\partial y}(x, y) &= g(x, y, u(x, y)), \\ u(x_0, y_0) &= u_0. \end{aligned} \tag{11}$$

Note that the right-hand side is allowed to depend on the unknown function $u(x,y)$. The condition $u(x_0,y_0)=u_0$ is called the **initial data**.

Necessary condition. First, we derive a necessary condition in order that (11) has a solution for all x_0, y_0, u_0. Assuming a solution exists for all initial data, we equate the mixed partial derivatives of u with respect to x and y,

$$\frac{\partial}{\partial y}\frac{\partial u}{\partial x}=\frac{\partial}{\partial x}\frac{\partial u}{\partial y}.$$

On the left-hand side, we obtain

$$\begin{aligned}&\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}(x,y)\right)=\frac{\partial}{\partial y}\left(f(x,y,u(x,y))\right)\\ &=\frac{\partial f}{\partial y}(x,y,u(x,y))+\frac{\partial f}{\partial z}(x,y,u(x,y))\frac{\partial u}{\partial y}(x,y)=\frac{\partial f}{\partial y}(x,y,u(x,y))+\frac{\partial f}{\partial z}(x,y,u(x,y))g(x,y,u(x,y)),\end{aligned}$$

where in the last equality we have used the second PDE to replace $\partial u/\partial y$ by g. Similarly, on the right-hand side, we obtain

$$\begin{aligned}&\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}(x,y)\right)=\frac{\partial}{\partial x}\left(g(x,y,u(x,y))\right)\\ &=\frac{\partial g}{\partial x}(x,y,u(x,y))+\frac{\partial g}{\partial z}(x,y,u(x,y))\frac{\partial u}{\partial x}(x,y)=\frac{\partial g}{\partial x}(x,y,u(x,y))+\frac{\partial g}{\partial z}(x,y,u(x,y))f(x,y,u(x,y)),\end{aligned}$$

Equating the two expressions, we get that

$$\frac{\partial f}{\partial y}(x,y,u(x,y))+\frac{\partial f}{\partial z}(x,y,u(x,y))g(x,y,u(x,y))=\frac{\partial g}{\partial x}(x,y,u(x,y))+\frac{\partial g}{\partial z}(x,y,u(x,y))f(x,y,u(x,y)).$$

By assumption, a solution $u(x,y)$ exists for all initial data. Thus, for arbitrary $(x,y,z)\in U$, there exists a solution with $u(x,y)=z$. Since the preceding must hold for this particular solution (as it must for all solutions), we may conclude that on all of U,

$$\frac{\partial f}{\partial y}+\frac{\partial f}{\partial z}g=\frac{\partial g}{\partial x}+\frac{\partial g}{\partial z}f. \tag{12}$$

This is a necessary condition for a solution of (11) to exist for all initial data.

Geometrical setting. We continue to assume that a solution of (11) exists for all initial data. Consider the graph of a solution, i.e. the surface given by $z=u(x,y)$. We want to construct two vector fields, $\mathbb{X}$ and $\mathbb{Y}$, which are everywhere tangent to the graphs of all solutions (satisfying different initial data). Equivalenly, $\mathbb{X}$ and $\mathbb{Y}$ should be orthogonal to the normal to the surfaces. The normal may be determined as follows: Let $h(x,y,z)=z-u(x,y)$. Then the surface is given by $h=0$, and the normal is given by ∇h, where

$$\nabla h=\left(-\frac{\partial u}{\partial x},-\frac{\partial u}{\partial y},1\right)=(-f,-g,1).$$

Let $\mathbb{W}=(a,b,c)$ denote a vector field which is orthogonal to ∇h (here, a, b and c denote functions of x, y, z). Then $-af-bg+c=0$, or

$$c=af+bg.$$

There are two linearly independent solutions, parameterised by a and b. For the first, which we take to be $\mathbb{X}$, we take $a=1$ and $b=0$. For the second, which we take to be $\mathbb{Y}$, we take $a=0$ and $b=1$. Thus,

$$\mathbb{X}(x,y,z)=(1,0,f(x,y,z)),\quad \mathbb{Y}(x,y,z)=(0,1,g(x,y,z)).$$

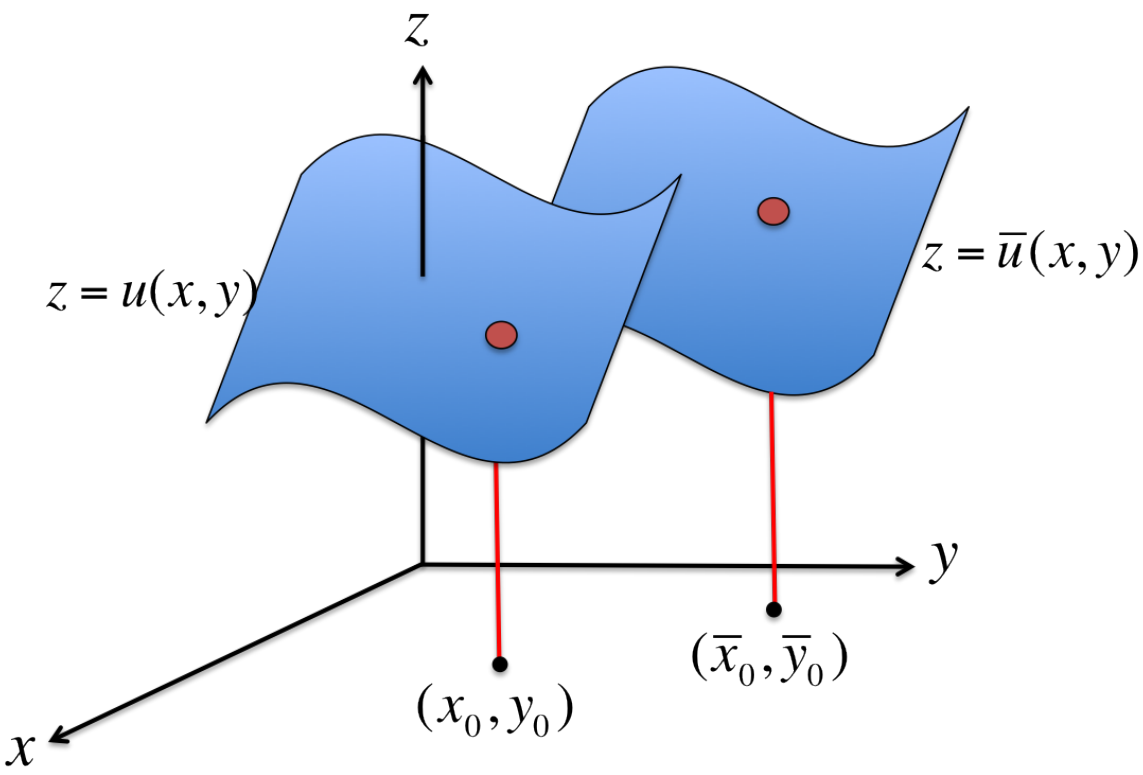


Figure 19: Graphs of solutions of (11) with different initial data.

Let us compute $[\mathbb{X}, \mathbb{Y}]$. We have that

$$\begin{aligned}
[\mathbb{X}, \mathbb{Y}] &= (\mathbb{X} \cdot \nabla)\mathbb{Y} - (\mathbb{Y} \cdot \nabla)\mathbb{X} \\
&= \left(\frac{\partial}{\partial x} + f\frac{\partial}{\partial z}\right)(0, 1, g) - \left(\frac{\partial}{\partial y} + g\frac{\partial}{\partial z}\right)(1, 0, f) \\
&= \left(0, 0, \frac{\partial g}{\partial x} + f\frac{\partial g}{\partial z}\right) - \left(0, 0, \frac{\partial f}{\partial y} + g\frac{\partial f}{\partial z}\right).
\end{aligned}$$

We observe that

$$[\mathbb{X}, \mathbb{Y}] = 0 \iff \frac{\partial f}{\partial y} + g\frac{\partial f}{\partial z} = \frac{\partial g}{\partial x} + f\frac{\partial g}{\partial z}.$$

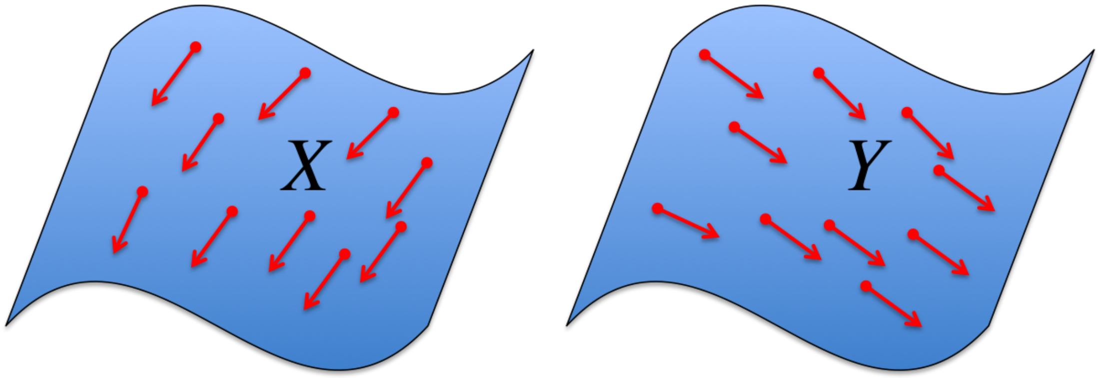
Thus, referring to (12), we see that $[\mathbb{X}, \mathbb{Y}] = 0$ is just the necessary condition for (11) to have a solution for all initial data. It turns out that the condition $[\mathbb{X}, \mathbb{Y}] = 0$ is also sufficient for a solution to exist.

Explicit solution. We assume that $[\mathbb{X}, \mathbb{Y}] = 0$ and proceed to construct the solution to (11). As shown in Figure 20, the vector fields $\mathbb{X}$ and $\mathbb{Y}$ lie tangent to the surface $z = u(x, y)$. This suggests that we can use the flows of $\mathbb{X}$ and $\mathbb{Y}$ to navigate from some given point on the surface, say (x_0, y_0, u_0), to an arbitrary point $(x, y, u(x, y))$. This is indeed how we will construct a solution.

Let Φ_t and Ψ_s denote the flows of $\mathbb{X} = (1, 0, f)$ and $\mathbb{Y} = (0, 1, g)$. Then Φ_t is determined by the solutions to the system

$$\dot{x} = 1, \quad \dot{y} = 0, \quad \dot{z} = f(x, y, z).$$

We can easily solve the first two equations to get $x(t) = x_0 + t$ and $y(t) = y_0$, where x_0 and y_0 are the initial values of x and y. We cannot determine $z(t)$ a priori; it will depend, of course,



(a) The vector field $\mathbb{X}$ (b) The vector field $\mathbb{Y}$

Figure 20

on the form of f. Therefore, we may write that

$$\Phi_t(x, y, z) = (x + t, y, *),$$

where $*$ denotes an undetermined component. Similarly, Ψ_s is determined by the solutions to the system

$$x' = 0, \quad y' = 1, \quad z' = g(x, y, z),$$

and is given by

$$\Psi_s(x, y, z) = (x, y + s, *).$$

We note that

$$\Psi_{y-y_0}(x_0, y_0, u_0) = (x_0, y, *),$$

and

$$\Phi_{x-x_0}\left(\Psi_{y-y_0}(x_0, y_0, u_0)\right) = \Phi_{x-x_0}(x_0, y, *) = (x, y, *).$$

Let

$$u(x, y) = \Phi^3_{x-x_0}\left(\Psi_{y-y_0}(x_0, y_0, u_0)\right),$$

so that

$$\Phi_{x-x_0}(\Psi_{y-y_0}(x_0, y_0, u_0)) = (x, y, u(x, y)).$$

That is, $u(x, y)$ is defined as the third component of the point reached by first applying the flow Φ_t to the initial data (x_0, y_0, u_0) for a time $t = x - x_0$, and then applying the flow Ψ_s for a time $s = y - y_0$. See Figure 21.

We claim that $u(x, y)$ satisfies (11).

First, let's check the initial data. For $x = x_0$ and $y = y_0$, we have that

$$u(x_0, y_0) = \Phi^3_0(\Psi_0(x_0, y_0, u_0)) = u_0,$$

as required.

Next, we check the equation for $\partial u/\partial x$. We have that

$$\frac{\partial u}{\partial x}(x, y) = \frac{\partial}{\partial x}\Phi^3_{x-x_0}(\Psi_{y-y_0}(x_0, y_0, u_0)) = \mathbb{X}^3(\Phi_{x-x_0}(\Psi_{y-y_0}(x_0, y_0, u_0))),$$

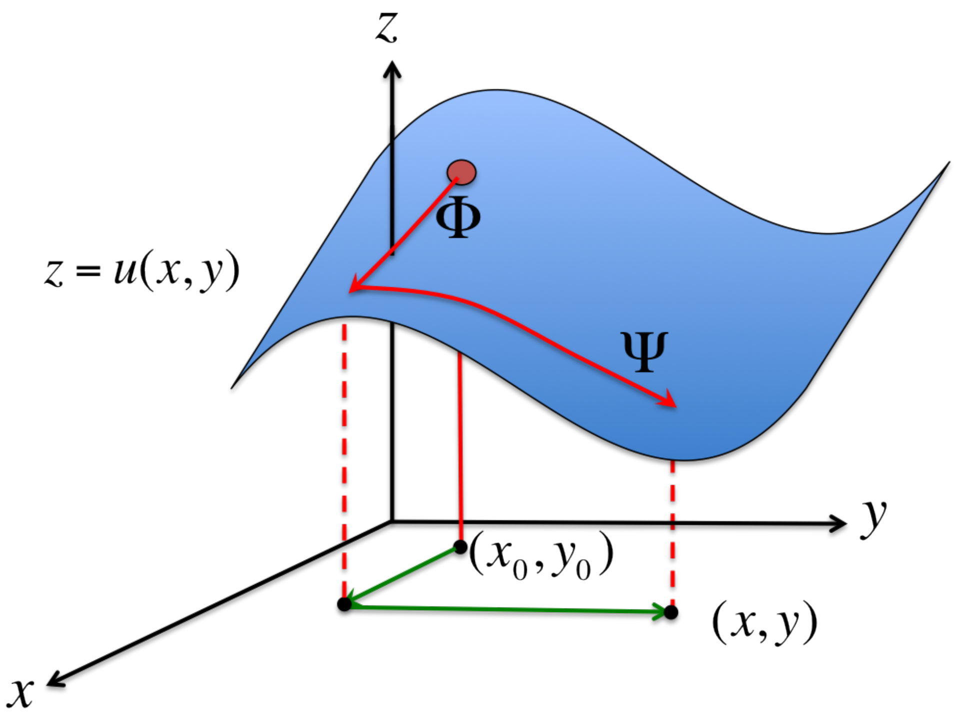


Figure 21: Construction of the solution of (11).

where we have used the fact that $\partial\Phi_t/\partial t = \mathbb{X} \circ \Phi_t$, and we have set t equal to $x - x_0$. But $\mathbb{X}^3 = f$, so we get that

$$\frac{\partial u}{\partial x}(x, y) = f\left(\Phi_{x-x_0}\left(\Psi_{y-y_0}(x_0, y_0, u_0)\right)\right) = f(x, y, u(x, y)),$$

as required.

Finally, we consider the equation for $\partial u/\partial y$. Rather than differentiate the equation

$$\Phi_{x-x_0}(\Psi_{y-y_0}(x_0, y_0, u_0)) = (x, y, u(x, y))$$

directly, we use the fact that necessary condition must be satisfied, i.e.

$$[\mathbb{X}, \mathbb{Y}] = 0.$$

By Theorem 1.8.12, it follows that

$$\Phi_t \circ \Phi_s = \Psi_s \circ \Phi_t.$$

Therefore, interchanging Φ_{x-x_0} and Ψ_{y-y_0}, we get that

$$\Psi_{y-y_0}(\Phi_{x-x_0}(x_0, y_0, u_0)) = (x, y, u(x, y)).$$

Repeating the calculation of the preceding paragraph but with roles of x and y interchanged, we get that

$$\frac{\partial u}{\partial y}(x, y) = \frac{\partial}{\partial y}\Psi^3_{y-y_0}(\Phi_{x-x_0}(x_0, y_0, u_0)) = \mathbb{Y}^3(\Psi_{y-y_0}(\Phi_{x-x_0}(x_0, y_0, u_0))),$$

where we have used the fact that $\partial\Psi_s/\partial s = \mathbb{Y} \circ \Psi_s$, and we have then set s equal to $y - y_0$. But $\mathbb{Y}^3 = g$, so we get that

$$\frac{\partial u}{\partial y}(x, y) = g\left(\Psi_{y-y_0}\left(\Phi_{x-x_0}(x_0, y_0, u_0)\right)\right) = g\left(\Phi_{x-x_0}\left(\Psi_{y-y_0}(x_0, y_0, u_0)\right)\right) = g(x, y, u(x, y)).$$

Uniqueness. (*nonexaminable) Finally, we show that the solution of (11) is unique. The argument is based on the uniqueness of solutions of ODE's, Theorem 1.6.3.

Suppose that $v(x,y)$ is another solution of (11). First, we show that

$$v(x,y_0) = u(x,y_0), \quad \forall x \text{ such that } (x,y_0) \in U.$$

Let $U(t) = u(x_0+t,y_0)$ and $V(t) = v(x_0+t,y_0)$. We have that

$$\begin{aligned}
\dot{U}(t) &= \frac{\partial u}{\partial x}(x_0+t,y_0) = f(x_0+t,y_0,U(t)) := F(U(t),t), \quad U(0) = u(x_0,y_0) = u_0,\\
\dot{V}(t) &= \frac{\partial v}{\partial x}(x_0+t,y_0) = f(x_0+t,y_0,V(t)) := F(V(t),t), \quad V(0) = v(x_0,y_0) = u_0.
\end{aligned}$$

Thus, $U(t)$ and $V(t)$ satisfy the same ODE and initial condition, and therefore must coincide.

A similar argument shows that

$$v(x,y) = u(x,y), \quad \forall (x,y) \in U.$$

Let $U(s) = u(x,y_0+s)$ and $V(s) = v(x,y_0+s)$. We have that

$$\begin{aligned}
\dot{U}(s) &= \frac{\partial u}{\partial y}(x,y_0+s) = g(x,y_0+s,U(s)) := G(U(s),s), \quad U(0) = u(x,y_0),\\
\dot{V}(s) &= \frac{\partial v}{\partial y}(x,y_0+s) = g(x,y_0+s,V(s)) := G(V(s),s), \quad V(0) = v(x,y_0).
\end{aligned}$$

Thus, $U(s)$ and $V(s)$ satisfy the same ODE and, since $u(x,y_0) = v(x,y_0)$ from above, the same initial condition. Therefore, they must coincide.

Example 1.11.1. Verify that the system

$$\begin{aligned}
\frac{\partial u}{\partial x}(x,y) &= y(u+1),\\
\frac{\partial u}{\partial y}(x,y) &= x(u+1),
\end{aligned} \tag{13}$$

has a solution, and find the solution satisfying the initial data

$$u(0,0) = 2.$$

First, we construct the vector fields $\mathbb{X}$ and $\mathbb{Y}$. Since $f = y(z+1)$ and $g = x(z+1)$, we have that

$$\mathbb{X} = (1,0,y(z+1)), \quad \mathbb{Y} = (0,1,x(z+1)).$$

Then

$$\mathbb{X}\cdot\nabla\mathbb{Y} = \left(\frac{\partial}{\partial x} + y(z+1)\frac{\partial}{\partial z}\right)(0,1,x(z+1)) = (0,0,z+1+y(z+1)x).$$

Similarly,

$$\mathbb{Y}\cdot\nabla\mathbb{X} = \left(\frac{\partial}{\partial y} + x(z+1)\frac{\partial}{\partial z}\right)(1,0,y(z+1)) = (0,0,z+1+x(z+1)y).$$

It follows that

$$[\mathbb{X},\mathbb{Y}] = \mathbb{X}\cdot\nabla\mathbb{Y} - \mathbb{Y}\cdot\nabla\mathbb{X} = 0,$$

so that the necessary condition is satisfied.

Next, let us compute $\Phi_x(\Psi_y(0,0,2))$, where Φ_t and Ψ_s are the flows of $\mathbb{X}$ and $\mathbb{Y}$. First, $\Psi_s(0,0,2)$ is determined from solutions of the system of ODE's given by

$$x' = 0, \quad y' = 1, \quad z' = x(z+1), \quad x(0) = y(0) = 0, \quad z(0) = 2.$$

Clearly $x(s) = 0$ and $y(s) = s$. Therefore, z satisfies $z'(s) = 0$, so that

$$z(s) = 2.$$

Setting $s = y$, we get that

$$\Psi_y(0,0,2) = (0, y, 2).$$

Next, we compute $\Phi_x(0, y, 2)$. This is determined from solutions of the system of ODE's given by

$$\dot{x} = 1, \quad \dot{y} = 0, \quad \dot{z} = y(z+1), \quad x(0) = 0, \quad y(0) = y, \quad z(0) = 2.$$

Clearly $x(t) = t$ and $y(t) = y$. Therefore, z satisfies

$$\dot{z} = y(z+1), \quad z(0) = 2.$$

The equation for $z(t)$ may be solved either by separation of variables or by using an integrating factor. We'll use separation of variables. We have that

$$\int_2^{z(t)} \frac{dz}{z+1} = \int_0^t y dt = yt,$$

or

$$\log(z(t)+1) - \log 3 = yt,$$

or

$$z(t) = 3e^{yt} - 1.$$

Setting $t = x$, we get that

$$\Phi_x(0, y, 2) = \Phi_x(\Psi_y(0,0,2) = (x, y, 3e^{yx} - 1) = (x, y, u(x, y)),$$

or

$$u(x, y) = 3e^{xy} - 1.$$

It is easy to confirm that u satisfies the required equation and initial data.

1.11.3 General statement and proof of the Frobenius Theorem

Statement of Frobenius Theorem. Let x denote coordinates on $\mathbb{R}^p$ and z coordinates on $\mathbb{R}^q$. Let $U \subset \mathbb{R}^p$ and $V \subset \mathbb{R}^q$ be open sets. Let f_i^α denote smooth functions on $U \times V$,

$$f_i^\alpha : U \times V \to \mathbb{R}; \quad (x, z) \mapsto f_i^\alpha(x, z),$$

where $1 \le i \le p$ and $1 \le \alpha \le q$. Consider the system of first-order partial differential equations for $u : U \to V$ given by

$$\begin{aligned} \frac{\partial u^\alpha}{\partial x^i}(x) &= f_i^\alpha(x, u(x)), \\ u(x_0) &= u_0, \quad x_0 \in U, u_0 \in V. \end{aligned} \tag{14}$$

Define p vector fields $\mathbb{X}_{(i)}$, $1 \le i \le p$, on $U \times V$ as follows:

$$\mathbb{X}_{(i)}^j(x, z) = \delta_i^j, 1 \le j \le p, \quad \mathbb{X}_{(i)}^{p+\alpha}(x, z) = f_i^\alpha(x, z), 1 \le \alpha \le q.$$

That is, among the first p components of $\mathbb{X}_{(i)}$, there is a single nonzero component, namely the ith, which is equal to one, while the last q components of $\mathbb{X}_{(i)}$ are given by $f_i^1, \ldots, f_i^q$. Suppose the vector fields $\mathbb{X}_{(i)}$ are complete.

Theorem 1.11.2 (Frobenius)**.** For all $(x_0, u_0) \in U \times V$, the system (14) has a solution $u(x)$ if and only if

$$[\mathbb{X}_{(i)}, \mathbb{X}_{(j)}] = 0, \quad 1 \leq i, j \leq p. \tag{15}$$

Moreover, if a solution exists, then it is unique.

Above, in (11) we considered the case $p = 2$ and $q = 1$. Instead of $x \in U \subset \mathbb{R}^p$, we wrote $(x, y) \in \mathbb{R}^2$, and instead of $f_i^\alpha : U \times V \to \mathbb{R}$, we had two functions $f(x, y, z)$ and $g(x, y, z)$ defined on $\mathbb{R}^3$. Instead of $\mathbb{X}_{(i)}$, we introduced two vector fields $\mathbb{X}$ and $\mathbb{Y}$ on $\mathbb{R}^3$ given by

$$\mathbb{X}(x, y, z) = (1, 0, f(x, y, z)), \quad \mathbb{Y}(x, y, z) = (0, 1, g(x, y, z)).$$

Proof.

Necessary condition. We show that condition (15) is necessary. Assume that a solution $u(x)$ of (14) exists for all initial data. It follows that the mixed partials of u are everywhere equal,

$$\frac{\partial}{\partial x^i}\left(\frac{\partial u^\alpha}{\partial x^j}\right) = \frac{\partial}{\partial x^j}\left(\frac{\partial u^\alpha}{\partial x^i}.\right) \tag{16}$$

We have that

$$\frac{\partial}{\partial x^i}\left(\frac{\partial u^\alpha}{\partial x^j}(x)\right) = \frac{\partial}{\partial x^i}\left(f_j^\alpha(x, u(x))\right) =$$
$$= \frac{\partial f_j^\alpha}{\partial x^i}(x, u(x)) + \frac{\partial f_j^\alpha}{\partial z^\beta}(x, u(x))\frac{\partial u^\beta}{\partial x^i}(x) = \left(\frac{\partial f_j^\alpha}{\partial x^i} + f_i^\beta \frac{\partial f_j^\alpha}{\partial z^\beta}\right)(x, u(x)),$$

where we have used the PDE (14). A similar expression is obtained for the RHS of (16). As solutions are assumed to exist for all initial data, $(x, u(x))$ may be taken to be arbitrary in $U \times V$. It follows that

$$\frac{\partial f_j^\alpha}{\partial x^i} + f_i^\beta \frac{\partial f_j^\alpha}{\partial z^\beta} = \frac{\partial f_i^\alpha}{\partial x^j} + f_j^\beta \frac{\partial f_i^\alpha}{\partial z^\beta} \tag{17}$$

on $U \times V$.

Consider next the Jacobi bracket $[\mathbb{X}_{(i)}, \mathbb{X}_{(j)}]$. Since the first p components of $\mathbb{X}_{(i)}$ and $\mathbb{X}_{(j)}$ are constant, it follows that the first p components of their bracket vanishes. We compute the remaining q components as follows: For $1 \leq \alpha \leq q$,

$$[\mathbb{X}_{(i)}, \mathbb{X}_{(j)}]^{p+\alpha} = \left(\frac{\partial}{\partial x^i} + f_i^\beta \frac{\partial}{\partial z^\beta}\right) f_j^\alpha - \left(\frac{\partial}{\partial x^j} + f_j^\beta \frac{\partial}{\partial z^\beta}\right) f_i^\alpha =$$
$$= \frac{\partial f_j^\alpha}{\partial x^i} + f_i^\beta \frac{\partial f_j^\alpha}{\partial z^\beta} - \frac{\partial f_i^\alpha}{\partial x^j} - f_j^\beta \frac{\partial f_i^\alpha}{\partial z^\beta}. \tag{18}$$

It is evident from (17) and (18) that (16) holds if and only if $[\mathbb{X}_{(i)}, \mathbb{X}_{(j)}] = 0$.

Explicit construction. Existence of solution. Let $\Phi_{(i)t}$ denote the flow of $\mathbb{X}_{(i)}$. The first p components of $\Phi_{(i)t}$ are given by

$$\Phi^j_{(i)t}(x_0) = x_0^j + \delta_i^j t, \quad 1 \leq j \leq p.$$

For simplicity, suppose $x_0 = 0$ in (14) (it is easy to generalise to $x_0 \neq 0$). We define $u(x)$ by

$$(x, u(x)) = \left(\Phi_{(1)x^1} \circ \cdots \circ \Phi_{(p)x^p}\right)(0, u_0). \tag{19}$$

It is clear that $u(0) = u_0$. Also,

$$\frac{\partial u^\alpha}{\partial x^1}(x) = \frac{\partial}{\partial x^1}\Phi^\alpha_{(1)x^1} \circ \cdots \circ \Phi_{(p)x^p}(0, u_0) =$$
$$= \mathbb{X}^\alpha_{(1)}(\Phi_{(1)x^1} \circ \cdots \circ \Phi_{(p)x^p}(0, u_0)) = \mathbb{X}^\alpha_{(1)}(x, u(x)) = f_1^\alpha(x, u(x)),$$

so that (14) is satisfied for $i = 1$. For $i > 1$, use the commutativity of the flows,

$$\Phi_{(i)s} \circ \Phi_{(j)t} = \Phi_{(j)t} \circ \Phi_{(i)s},$$

which follows from (15), to bring the factor $\Phi_{(i)x^i}$ to the front in (19); then calculate $\partial/\partial x^i$ as we calculated $\partial/\partial x^1$ above.

Uniqueness. (*Nonexaminable.)

The uniqueness of solutions to the system of partial differential equations (14) follows from the uniqueness of solutions to ordinary differential equations. Suppose that the necessary condition (15) holds, and let $u(x)$ and $v(x)$ be two solutions of (14). For simplicity, let $x_0 = 0$ and assume that the flows of the $\mathbb{X}_{(j)}$'s are complete.

We show that $u(x) = v(x)$ by induction. Clearly

$$u(0) = v(0),$$

since both u and v satisfy the initial data. Next, we show that if

$$u(x^1, x^2, \ldots, x^{k-1}, 0, \ldots, 0) = v(x^1, x^2, \ldots, x^{k-1}, 0, \ldots, 0), \tag{20}$$

then

$$u(x^1, x^2, \ldots, x^{k-1}, x^k, 0, \ldots, 0) = v(x^1, x^2, \ldots, x^{k-1}, x^k, 0, \ldots, 0). \tag{21}$$

Let $x^1, \ldots, x^{k-1}$ be fixed, and let

$$x(t) = (x^1, x^2, \ldots, x^{k-1}, t, 0, \ldots, 0) \in U \subset \mathbb{R}^p.$$

Then

$$\dot{x}^i(t) = \delta^i_k.$$

Let

$$U(t) = u(x(t)) \in V \subset \mathbb{R}^q.$$

Since u satisfies (14), it follows that

$$\dot{U}^\alpha(t) = \frac{\partial u^\alpha}{\partial x^i}(x(t))\dot{x}^i(t) = \frac{\partial u^\alpha}{\partial x^k}(x(t)) = f^\alpha_k(x(t), u(x(t)) = f^\alpha_k(x(t), U(t)).$$

Therefore, letting

$$F^\alpha(U, t) = f^\alpha_k(x(t), U),$$

we get that

$$\dot{U}(t) = F(U(t), t), \quad U(0) = u(x^1, x^2, \ldots, x^{k-1}, 0, \ldots, 0). \tag{22}$$

Similarly, letting

$$V(t) = V(x(t)) \in V \subset \mathbb{R}^q$$

we have that

$$\dot{V}(t) = F(V(t), t)), \quad V(0) = V(x^1, x^2, \ldots, x^{k-1}, 0, \ldots, 0). \tag{23}$$

From (22) and (23) and the induction hypothesis, $U(t)$ and $V(t)$ satisfy the same differential equation with the same initial condition. It follows that $U(t) = V(t)$, and in particular, letting $t = x^k$, that

$$u(x^1, \ldots, x^{k-1}, x^k, 0, \ldots, 0) = v(x^1, \ldots, x^{k-1}, x^k, 0, \ldots, 0),$$

as required. □

Example 1.11.3 (2011 Examination, Q3(b)).

Question: Show that the system

$$\frac{\partial u}{\partial x} = u,$$
$$\frac{\partial^2 u}{\partial y \partial y} = e^{-x}u^2$$

with initial data $u(0,0) = 1$ has a unique solution in a neighbourhood of $(0,0)$ (you don't need to find the solution explicitly). (*Hint: Introduce a second function* $v = \partial u/\partial y$.)

Solution: Let $v = \partial u/\partial y$. Then, using the given equation $\partial u/\partial x = u$, we get that

$$\frac{\partial v}{\partial x} = \frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial u}{\partial y} = v.$$

Thus, we get the following system of first-order PDE's for u and v:

$$\frac{\partial u}{\partial x} = u,$$
$$\frac{\partial v}{\partial x} = v,$$
$$\frac{\partial u}{\partial y} = v,$$
$$\frac{\partial v}{\partial y} = e^{-x}u^2.$$

The necessary and condition for a local solution to exist is that the vector fields $\mathbb{X}(x,y,u,v)$, $\mathbb{Y}(x,y,u,v)$, given by

$$\mathbb{X} = (1,0,u,v), \quad \mathbb{Y} = (0,1,v,e^{-x}u^2)$$

have vanishing Jacobi bracket (equivalently, the mixed second partial derivatives of u must be equal, and similarly for v). Calculation gives

$$\begin{aligned}[\mathbb{X},\mathbb{Y}] &= (\partial_x + u\partial_u + v\partial_v)(0,1,v,e^{-x}u^2) - (\partial_y + v\partial_u + e^{-x}u^2\partial_v)(1,0,u,v) \\ &= (0,0,v,e^{-x}u^2) - (0,0,v,e^{-x}u^2) = 0,\end{aligned}$$

as required.

1.12 *More general versions of the Frobenius Theorem [not in lectures, nonexaminable]

Consider the system

$$\sum_{j=1}^{p} A_{ij}(x,u(x))\frac{\partial u^\alpha}{\partial x^i}(x) = f_i^\alpha(x,u(x)),$$
$$u(x_0) = u_0. \tag{24}$$

We assume that the A_{ij}'s are smooth functions on $U \times V$, and moreover that

$$\det A_{ij}(x,z) \neq 0.$$

The original system (14) is recovered by taking $A_{ij} = \delta_{ij}$ (so that A_{ij} is constant in this case). We can formulate the following generalisation of the Frobenius theorem: Define p vector fields $\mathbb{Y}_{(i)}$, $1 \le i \le p$, on $U \times V$ as follows:

$$\mathbb{Y}^j_{(i)}(x, z) = (A_{i1}, \ldots, A_{ij}, \ldots, A_{ip}, f^1_i, \ldots, f^\alpha_i, \ldots, f^q_p)(x, z).$$

Theorem 1.12.1 (Generalised Frobenius)**.** For all $(x_0, u_0) \in U \times V$, the system (14) has a solution $u(x)$ if and only if

$$[\mathbb{Y}_{(i)}, \mathbb{Y}_{(j)}] = \sum_{k=1}^{p} c^k_{ij} \mathbb{Y}_{(k)} \tag{25}$$

for some smooth functions c^k_{ij} on $U \times V$. Moreover, if a solution exists, then it is unique.

This result can be deduced from the original version Theorem 1.11.2. The main fact that is needed is explained in the following section, namely that if $\{\mathbb{X}_{(1)}, \ldots \mathbb{X}_{(r)}\}$ and $\{\mathbb{Y}_{(1)}, \ldots \mathbb{Y}_{(r)}\}$ are equivalent distributions on $\mathbb{R}^n$, then $\{\mathbb{X}_{(1)}, \ldots \mathbb{X}_{(r)}\}$ is integrable if and only if $\{\mathbb{Y}_{(1)}, \ldots \mathbb{Y}_{(r)}\}$ is integrable.

1.12.1 Distributions

A **k-dimensional distribution** on an open set $U \subset \mathbb{R}^n$ is a set of k smooth linearly independent vector fields $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ on U. Linearly independent means that, at each point $x \in U$, the vectors $\mathbb{Y}_{(1)}(x), \ldots \mathbb{Y}_{(k)}(x)$ are linearly independent. In order for this to be the case, we must have that $k \le n$. (Note: you may have come across another mathematical definition of 'distribution' in analysis, namely as a linear functional on a suitable function space. The present definition is quite separate.)

A distribution $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ is said to be **integrable** if the Jacobi bracket of any two vectors $\mathbb{Y}_{(r)}$ and $\mathbb{Y}_{(s)}$ can be expressed as a linear combination of the $\mathbb{Y}_{(j)}$'s, ie

$$[\mathbb{Y}_{(r)}, \mathbb{Y}_{(s)}](x) = \sum_{j=1}^{k} c^j_{rs}(x) \mathbb{Y}_{(j)}(x)$$

for some coefficients $c^j_{rs}(x)$. Here are some examples:

1. A 1-dimensional distribution $\{\mathbb{Y}\}$ is trivially integrable, since $[\mathbb{Y}, \mathbb{Y}] = 0$ identically.

2. The 2-dimensional distribution $\{\mathbb{Y}_{(1)}, \mathbb{Y}_{(2)}\}$ in $\mathbb{R}^3$ given by

$$\mathbb{Y}_{(1)}(x, y, z) = (f_{(1)}, g_{(1)}, 0)(x, y, z),$$
$$\mathbb{Y}_{(2)}(x, y, z) = (f_{(2)}, g_{(2)}, 0)(x, y, z),$$

is integrable. To see why, first note that the third component of $[\mathbb{Y}_{(1)}, \mathbb{Y}_{(2)}]$ necessarily vanishes. Next, note that any vector $\mathbb{Z}$ with zero third component can be expressed as a linear combination of $\mathbb{Y}_{(1)}$ and $\mathbb{Y}_{(2)}$. This is because $\mathbb{Y}_{(1)}$, $\mathbb{Y}_{(2)}$ and $\mathbb{Z}$ are necessarily linearly dependent (they constitute 3 vectors in the xy-plane), so we can find coefficients a, b and c such that

$$a\mathbb{Y}_{(1)} + b\mathbb{Y}_{(2)} + c\mathbb{Z} = 0.$$

But c cannot vanish (otherwise, we would have $a\mathbb{Y}_{(1)} + b\mathbb{Y}_{(2)} = 0$), so that

$$\mathbb{Z} = -(a/c)\mathbb{Y}_{(1)} - (b/c)\mathbb{Y}_{(2)}.$$

3. The 2-dimensional distribution given by

$$\mathbb{Y}_{(1)}(x,y,z) = (1,0,0),$$
$$\mathbb{Y}_{(2)}(x,y,z) = (0,1,f(x)),$$

is not integrable. To see this, we note that

$$[\mathbb{Y}_{(1)}, \mathbb{Y}_{(2)}] = (0,0,f'(x)),$$

which cannot be expressed as a linear combination of $\mathbb{Y}_{(1)}$ and $\mathbb{Y}_{(2)}$.

4. The 2-dimensional distribution $\{S, D\}$, where the vector fields S and D are given in Problem Sheet 4, is not integrable.

Let $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ be another k-dimensional distribution. We say that $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ and $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ are **equivalent** if one set of vectors can be expressed as a linear combination of the other, ie

$$\mathbb{X}_{(i)} = \sum_{j=1}^{k} a_{ij}(x)\mathbb{Y}_{(j)}$$

for some coefficients $a_{ij}(x)$. If this is the case, it follows that the matrix $a_{ij}(x)$ is invertible. To see why, suppose it weren't. If $a_{ij}(x)$ were not invertible, we could find a k-dimensional vector $(v_1, \cdots, v_k)$ such that

$$\sum_{i=1}^{k} v_i a_{ij}(x) = 0.$$

But this would imply that

$$\sum_{i=1}^{k} v_i \mathbb{X}_{(i)} = \sum_{i=1}^{k} v_i \sum_{j=1}^{k} a_{ij}(x)\mathbb{Y}_{(j)} = \sum_{j=1}^{k} \left(\sum_{i=1}^{k} v_i a_{ij}(x) \right) \mathbb{Y}_{(j)} = 0,$$

contradicting the assumption that the $\mathbb{X}_{(i)}$'s are linearly independent everywhere.

Lemma 1.12.2. Suppose $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ and $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ are equivalent distributions. Then $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ is integrable if and only if $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ is integrable.

Proof. For definiteness, suppose that $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ is integrable. We will prove that $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ is integrable. Let

$$\mathbb{X}_{(i)} = \sum_{j=1}^{k} a_{ij}(x)\mathbb{Y}_{(j)},$$

and recall the product rule for the Jacobi bracket (Proposition 1.8.5),

$$[\mathbb{Y}, f\mathbb{X}] = \mathbb{Y}(f)\mathbb{X} + f[\mathbb{Y}, \mathbb{X}].$$

Now calculate:

$$[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}] = \sum_{t=1}^{k}\sum_{u=1}^{k} [a_{rt}\mathbb{Y}_{(r)}, a_{su}\mathbb{Y}_{(u)}] \text{(expanding the } \mathbb{X}\text{'s in terms of } \mathbb{Y}\text{'s)}$$

$$= \sum_{t=1}^{k}\sum_{u=1}^{k} \left(a_{rt}\mathbb{Y}_{(r)}(a_{su})\mathbb{Y}_{(u)} - a_{su}\mathbb{Y}_{(u)}(a_{rt})\mathbb{Y}_{(r)} + a_{rt}a_{su}[\mathbb{Y}_{(r)}, \mathbb{Y}_{(u)}] \right) \text{ (using product rule)}$$

$$= \sum_{t=1}^{k}\sum_{u=1}^{k} \left(a_{rt}\mathbb{Y}_{(r)}(a_{su})\mathbb{Y}_{(u)} - a_{su}\mathbb{Y}_{(u)}(a_{rt})\mathbb{Y}_{(r)} + \sum_{v=1}^{k} a_{rt}a_{su}c^{v}_{ru}\mathbb{Y}_{(v)} \right) \text{ (using integrability of } \mathbb{Y}\text{'s).}$$

The last expression is clearly a linear combination of the $\mathbb{Y}$'s, which in turn can be expressed as a linear combination of the $\mathbb{X}$'s, since

$$\mathbb{Y}_{(j)} = \sum_{l=1}^{k} a_{jl}^{-1}(x)\mathbb{X}_{(l)}.$$

Thus, $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ is indeed integrable, as claimed. □

1.12.2 Alternative formulation of Frobenius' Theorem

Theorem 1.12.3. Let $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ be an integrable k-dimensional distibution in U. Then in a neighbourhood of a point $x_0 \in U$, there exists a k-dimensional surface S which is tangent to each of the vector fields $\mathbb{Y}_{(j)}$.

Proof. We show in the Lemma below that in a neighbourhood of x_0, $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ is equivalent to a distribution $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ for which

$$[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}] = 0$$

for all r and s. Let $\Psi_{(j)s}$ denote the flow of $\mathbb{X}_{(j)}$. Define the surface S to be the set of points

$$x(s_1, \ldots, s_k) = \left(\Psi_{(1)s_1} \circ \cdots \Psi_{(k)s_k}\right)(x_0)$$

obtained by applying the flows of the $\mathbb{X}_{(j)}$'s in succession. The flow times $s_1, \ldots, s_k$ serve as parameters on S. Since the flows $\Psi_{(i)s_i}$ and $\Psi_{(j)s_k}$ commute, it is clear that

$$\Psi_{(j)t}(x(s_1, \ldots, s_k)) = x(s_1, \ldots, s_{j-1}, s_j + t, s_{j+1}, \ldots, s_k).$$

That is, the effect of applying the jth flow to a point of S for a time t is to shift the value of jth parameter s_j by an amount t. Therefore, under the flow, points of S remain on S. Thus S is tangent to each of the $\mathbb{X}_{(j)}$'s. □

Lemma 1.12.4. An integrable k-dimensional distibution $\{\mathbb{Y}_{(1)}, \ldots, \mathbb{Y}_{(k)}\}$ in U is equivalent to a distribution $\{\mathbb{X}_{(1)}, \ldots, \mathbb{X}_{(k)}\}$ for which

$$[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}] = 0$$

for all r and s.

Proof. By applying a suitable linear transformation, we can choose coordinates so that, at x_0,

$$\mathbb{Y}_{(j)}(x_0) = (0, \ldots, 0, 1, 0, \ldots, 0),$$

ie $\mathbb{Y}_{(j)}(x_0)$ is the unit vector $e_{(j)}$ in the jth direction. Elsewhere, we write the $\mathbb{Y}_{(j)}$'s in the form

$$\mathbb{Y}_{(j)}(x) = \sum_{l=1}^{k} a_{jl}(x)e_{(j)} + \sum_{m=k+1}^{n} b_{jm}e_{(m)},$$

so that

$$a(x_0) = I$$

(here we regard $a(x)$ as a $k \times k$ matrix). Since I is invertible and a is continuous, we may conclude that, in a neighbourhood U of x_0, a is invertible. Let $d(x)$ denote the inverse of $a(x)$, ie

$$\sum_{j=1}^{k} d_{ij}(x)a_{jl}(x) = \delta_{il}$$

for x in U. Then define

$$\mathbb{X}_{(i)} = \sum_{j=1}^{n} d_{ij}(x)\mathbb{Y}_{(j)}.$$

It follows that, for $x \in \mathcal{U}$,

$$\mathbb{X}_{(i)}(x) = e_{(i)} + \sum_{m=k+1}^{n} f_{im}(x)e_{(m)},$$

where $f_{im} = \sum_{j=1}^{k} d_{ij}b_{jm}$. By assumption, $\{\mathbb{Y}_{(1)}, \dots, \mathbb{Y}_{(k)}\}$ and therefore $\{\mathbb{X}_{(1)}, \dots, \mathbb{X}_{(k)}\}$ is integrable. Thus,

$$[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}] = \sum_{t=1}^{k} g_{rs}^{t}\mathbb{X}_{(t)}$$

for some coefficients g_{rs}^t.

From the form of the $\mathbb{X}_{(i)}$'s, it is clear that the first k components of $[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}]$ are given by the f_{rs}^t's, ie

$$[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}]^t = f_{rs}^t, \text{ for } 1 \le t \le k.$$

It is also clear that

$$[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}]^t = 0, \text{ for } 1 \le t \le k$$

because the first k components of the $\mathbb{X}_{(j)}$'s are either zero or a constant. Comparing the two preceding equations, we may conclude that

$$f_{rs}^t = 0,$$

which implies that

$$[\mathbb{X}_{(r)}, \mathbb{X}_{(s)}] = 0.$$

□

2 Algebraic k-forms

2.1 Dual space

In what follows, V denotes an n-dimensional vector space.

Let $\mathcal{F}(V)$ denote the set of functions $f : V \to \mathbb{R}$ on V. $\mathcal{F}(V)$ can be regarded as a vector space. The zero element is the function, denoted 0, which is equal to zero everywhere. Addition of functions and multiplication of functions by scalars is defined in the obvious way. For example, if f and g are functions, we define the function $f + g$ by $(f + g)(v) = f(v) + g(v)$. Similarly, if $\alpha \in \mathbb{R}$, we define the function αf by $(\alpha f)(v) = \alpha f(v)$. It is straightforward to verify that the usual vector-space properties are satisfied (e.g., commutativity, associativity, distributive law, etc, for addition and scalar multiplication).

As a vector space, $\mathcal{F}(V)$ is infinite dimensional. We can identify various subspaces of $\mathcal{F}(V)$, for example the space of continuous functions $C^0(V)$, or the space of smooth functions $C^\infty(V)$. These are also infinite dimensional.

A function $f : V \to \mathbb{R}$ is linear if for all $u, v \in V$ and for all $\alpha, \beta \in \mathbb{R}$,

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v).$$

Definition 2.1.1. [Dual space] The *dual space* of V, denoted V^*, is the subspace of $\mathcal{F}(V)$ consisting of *linear functions* on V.

It is easy to verify that V^* is a vector space. That is, the sum of two linear functions is a linear function, and a scalar multiple of a linear function is a linear function (we won't give an explicit verification here).

Example 2.1.2 (Dual space of $\mathbb{R}^3$)**.** Let $V = \mathbb{R}^3$. Let $\mathbf{a} \in \mathbb{R}^3$, and define $f \in \mathbb{R}^{3*}$ by

$$f(\mathbf{r}) = \mathbf{a} \cdot \mathbf{r}.$$

Using the familiar properties of the dot product, it is easy to verify that f is indeed a linear function (i.e., $\mathbf{a} \cdot (\alpha\mathbf{r} + \beta\mathbf{s}) = \alpha\mathbf{a} \cdot \mathbf{r} + \beta\mathbf{a} \cdot \mathbf{s}$.) Indeed, every element of $\mathbb{R}^{3*}$ can be represented in this way. To see that this is the case, let $\hat{\boldsymbol{\imath}}$, $\hat{\boldsymbol{\jmath}}$ and $\hat{\mathbf{k}}$ denote the standard basis in $\mathbb{R}^3$. Given $f \in \mathbb{R}^{3*}$, let

$$f_x = f(\hat{\boldsymbol{\imath}}), \quad f_y = f(\hat{\boldsymbol{\jmath}}), \quad f_z = f(\hat{\mathbf{k}}).$$

Then it is easy to check that

$$f(\mathbf{r}) = (f_x\hat{\boldsymbol{\imath}} + f_y\hat{\boldsymbol{\jmath}} + f_z\hat{\mathbf{k}}) \cdot \mathbf{r}.$$

Thus, by means of the dot product, we can associate to every vector in $\mathbb{R}^3$ a linear function in $\mathbb{R}^{3*}$, and vice versa.

In fact, this construction of the dual space of $\mathbb{R}^3$ generalises to an arbitrary finite-dimensional vector space V. Let $e_{(1)}, \dots, e_{(n)}$ denote a basis for V. For $v \in V$, we write $v = v^i e_{(i)}$ to denote its expansion in terms of this basis. We define a set of n elements of V^*, denoted $f^{(1)}, \dots, f^{(n)}$, by the relation

$$f^{(j)}(v^i e_{(i)}) = v^j.$$

That is, $f^{(j)}$ picks out the jth component of v when v is expanded in terms of the $e_{(i)}$'s. In what follows we will use the following notation. Let $a \in V^*$, and let

$$a_i = a(e_{(i)}).$$

Thus,

$$f_i^{(j)} = \delta_i^j.$$

It is easy to see that $f^{(j)}$ is a linear function on V, and hence an element of V^*. In fact, we have the following:

Proposition 2.1.3 (Dual basis)**.** $f^{(1)}, \dots, f^{(n)}$ constitute a basis for V^*.

Proof. We need to show two things, namely that i) the $f^{(j)}$'s are linearly independent, and ii) every element of V^* can be expressed as a linear combination of the $f^{(j)}$'s.

First, we show that the $f^{(j)}$'s are linearly independent. Suppose that

$$\alpha_j f^{(j)} = 0.$$

Evaluate both sides on $e_{(i)}$ to obtain $\alpha_j \delta_i^j = 0$, or

$$\alpha_i = 0.$$

As i is arbitrary, linear independence follows.

Next, let $a \in V^*$. We claim that

$$a = a_j f^{(j)}.$$

To verify, let us apply both sides to $v \in V$. On the left-hand side, we get $a(v)$. On the right-hand side, we get

$$a_j f^{(j)}(v) = a_j v^j = a(e_{(j)})v^j = a(v^j e_{(j)}) \text{ (by linearity) } = a(v).$$

$\square$

Collectively, $f^{(1)}, \ldots, f^{(n)}$ is called the *dual basis* of $e_{(1)}, \ldots, e_{(n)}$. It follows from Proposition 2.1.3 that V^* is n-dimensional, and that there is a vector space isomorphism between V and V^* which maps $u = u^i e_{(i)}$ into $\sum_{i=1}^n u^i f^{(i)}$. In this way, to every $u \in V$, we can associate a unique element $f \in V^*$ by $f(v) = \sum_{i=1}^n u^i v^i$. We could regard this last expression as defining a dot product, and write $f(v) = u \cdot v$.

At this point, you might wonder why we introduce the dual space in the first place. If V and V^* are isomorphic, why not regard them as being the same? The rest of this section will be devoted to answering this question.

The **first** answer is that V might not have a natural, or "built-in", dot product. Of course, one can always introduce a dot product, or inner product, as it is also called, on a vector space, by declaring that a particular basis $e_{(1)}, \ldots, e_{(n)}$ is orthonormal (in effect, this is what we did above), so that if $u = u^i e_{(i)}$ and $v = v^j e_{(j)}$, then $u \cdot v = \sum_{i=1}^n u^i v^i$. But this raises the question, does this inner product have any intrinsic meaning? If we were to choose a different basis $\bar{e}_{(j)}$, would the dot product change? (The answer is yes, unless the $n \times n$ matrix M defined by $e_{(i)} = \sum_{j=1}^n M_{ij} \bar{e}_{(j)}$ happens to satisfy $M^T = M^{-1}$.)

Certain vector spaces do have an intrinsically defined inner product. An example is n-dimensional Euclidean space, which is endowed with, in addition to its vector space properties, an intrinsic notion of geometrical distance, or length. The inner product can be expressed geometrically in terms of length by the expression

$$u \cdot v = \tfrac{1}{4}\left(||u+v||^2 - ||u-v||^2\right),$$

independently of any particular choice of basis.

Other vector spaces do not have an intrinsically defined inner product. An example from mechanics is 2-dimensional phase space $P = \{(q,p)\}$, where q represents the position of a particle moving along a line, and p represents its momentum. P is a perfectly good vector space; vector addition and scalar multiplication make sense, and have an intrinsic meaning within mechanics. However, there is no intrinsic inner product. Note that an expression such as $q_1 q_2 + p_1 p_2$ is incoherent from the point of view of mechanics, as positions and momenta have different physical dimensions.

To summarise, if V has an intrinsically defined inner product, then one can identify V and V^*, and ignore the distinction between them. Otherwise, it is better to regard V and V^* as being distinct.

The **second** answer to the question is that, under a linear transformation, vectors in V and V^* transform differently. Let $e_{(1)}, \ldots, e_{(n)}$ and $\bar{e}_{(1)}, \ldots, \bar{e}_{(n)}$ be two bases for V. Then one set of basis vectors can be expressed as linear combinations of the others, e.g.

$$\bar{e}_{(i)} = \sum_{j=1}^n M_{ij} e_{(j)},$$

where M is an $n \times n$ matrix. Let $f^{(j)}$ and $\bar{f}^{(j)}$ denote the dual bases of $e_{(i)}$ and $\bar{e}_{(i)}$ respectively. Here, too, one set of basis vectors can be expressed as linear combinations of the others,

$$\bar{f}^{(i)} = \sum_{j=1}^n N_{ij} f^{(j)}.$$

Given the definition of the dual basis, one can calculate the matrix N. It turns out that N is not equal to M, but rather, we have the following:

Proposition 2.1.4 (Transformation of dual basis)**.**

$$N = M^{T^{-1}}.$$

Proof. See Problem Sheet 6.1(b). □

The **third** answer to the question is that elements of V and V^* have different geometrical meanings. A vector $v \in V$ corresponds to a directed displacement, or an "arrow". More precisely, v can be thought of as an instantaneous velocity $\dot{\mathbf{r}}(0)$ along a curve $\mathbf{r}(t)$ in V. A vector $a \in V^*$ in the dual space corresponds to a function on V, and as a function it can be represented geometrically by sets of contours (sets on which the function takes a constant value). Since a is a linear function, it can be completely described by just two contours, namely the 0-contour, the set on which a vanishes, and the 1-contour, where a is equal to 1. Again, because a is linear, the 0-contour is an $(n-1)$-dimensional plane through the origin, and the 1-contour is an $(n-1)$-dimensional plane which is parallel to the 0-contour. The real number $a(v)$ can be expressed purely geometrically (i.e., independently of choice of basis, units, etc) as a ratio of lengths. See Figure 22. This geometric picture also provides some insight as to why vectors in V and V^* transform differently under a linear transformation. See Figure 23.

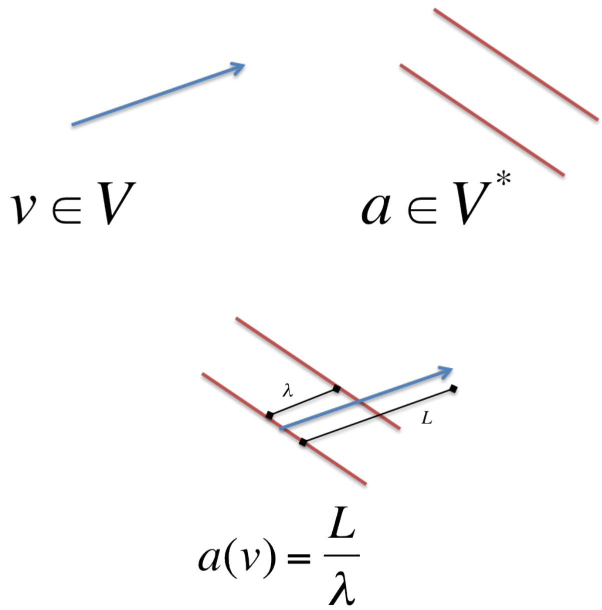


Figure 22: $v \in V$ is represented by an arrow, and represents a velocity, or a displacement per unit parameter. $a \in V^*$ is represented by a pair of parallel hyperplanes which correspond to its zero and unit level sets. The value of $a(v)$ is given purely geometrically by the ratio of the length of v to the length of the component of v which lies between the planes of a.

2.2 Permutations

A *permutation of n objects*, or permutation for short, is a bijection

$$\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}; j \mapsto \sigma(j)$$

A standard way to display a permutation is as a table,

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}.$$

Let S_n denote the set of all permutations of n objects. The identity map e, given by $e(j) = j$, is a permutation. The composition $\sigma \circ \tau$ of two permutations σ and τ, given by $\sigma \circ \tau(j) = \sigma(\tau(j))$, is a permutation. Composition is usually denoted by $\sigma\tau$, ie without the $\circ$ symbol. To each permutation σ corresponds a unique permutation σ^{-1} such that $\sigma\sigma^{-1} = \sigma^{-1}\sigma = e$. In this way,

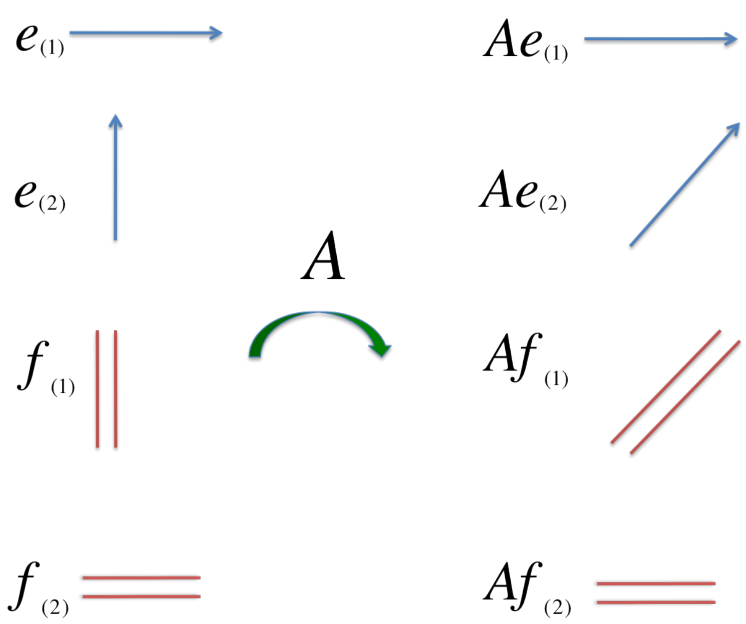


Figure 23: The standard basis for V and the dual basis for V^*. Under a transformation A which sends $e_{(1)}$ to $e_{(1)}$ and $e_{(2)}$ to $e_{(1)} + e_{(2)}$, the dual basis vectors $f^{(1)}$ and $f^{(2)}$ are sent to $f^{(1)} - f^{(2)}$ and $f^{(2)}$ respectively.

S_n forms a group, called the *symmetric group*, or the *permutation group*. The transposition $\tau_{rs} \in S_n$ is the permutation given by

$$\tau_{rs}(r) = s, \quad \tau_{rs}(s) = r, \quad \tau_{rs}(j) = j, \ \forall j \neq r, s.$$

Proposition 2.2.1 (Permutations and transpositions)**.** Every permutation can be written as a product (composition) of transpositions.

Proof. A general argument as well as a specific example is given in Problem 6.2. □

To each $\sigma \in S_n$ we may associate an $n \times n$ *permutation matrix* $P(\sigma)$ given by

$$P_{ij}(\sigma) = \delta_{i,\sigma(j)}.$$

Proposition 2.2.2 (Permutations and permutation matrices)**.** For all $\sigma, \tau \in S_n$,

$$P(\sigma\tau) = P(\sigma)P(\tau).$$

Proof.

$$[P(\sigma)P(\tau)]_{ik} = \sum_{j=1}^{n} P_{ij}(\sigma)P_{jk}(\tau) = \sum_{j=1}^{n} \delta_{i,\sigma(j)}\delta_{j,\tau(k)} = \delta_{i,\sigma(\tau(k))} = P_{ik}(\sigma\tau).$$

□

It follows from Proposition 2.2.2 that $P(e) = I$, where I is the $n \times n$ identity matrix, and that $P(\sigma^{-1}) = P^{-1}(\sigma)$. In Problem 6.5 it is shown that $P(\sigma)$ is an orthogonal matrix, i.e. $P^{-1}(\sigma) = P^T(\sigma)$.

Definition 2.2.3 (Sign of a permutation)**.** The *sign* of a permutation, denoted $\operatorname{sgn}\sigma$, is defined by

$$\operatorname{sgn}\sigma = \operatorname{sgn}\det P(\sigma).$$

In fact, since $\det P(\sigma)$ is either 1 or -1 (Problem 6.5 again), we could also write $\operatorname{sgn}\sigma = \det P(\sigma)$.

We have the following results about the sign of a permutation:

Proposition 2.2.4 (sgn is multiplicative)**.** For all $\sigma, \tau \in S_n$,

$$\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau).$$

Proof.

$$\begin{aligned}\operatorname{sgn}(\sigma\tau) &= \operatorname{sgn}\det P(\sigma\tau) = \operatorname{sgn}\det(P(\sigma)P(\tau)) \text{ (from Proposition 2.2.2)} \\ &= \operatorname{sgn}(\det P(\sigma)\det P(\tau)) = \operatorname{sgn}\det P(\sigma)\operatorname{sgn}\det P(\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau).\end{aligned}$$

□

Proposition 2.2.5 (sgn of inverse)**.**

$$\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma).$$

Proof. Since $P(\sigma^{-1}) = P^{-1}(\sigma)$ from Proposition 2.2.2, it follows that

$$\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}\det P(\sigma^{-1}) = \operatorname{sgn}\det P^{-1}(\sigma) = \operatorname{sgn}(\det P(\sigma))^{-1} = \operatorname{sgn}\det P(\sigma) = \operatorname{sgn}\sigma.$$

□

Proposition 2.2.6 (sgn of transposition)**.** If τ_{rs} is a transposition, then $\operatorname{sgn}\tau_{rs} = -1$

Proof. First consider τ_{12}. We have that

$$\tau_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & I \end{pmatrix},$$

where I denotes the $(n-2)$-dimensional identity matrix. It is easy to calculate that $\det P(\tau_{12}) = -1$, so that $\operatorname{sgn}\tau_{12} = -1$. For a general transposition, it is shown in Problem 6.3 that $\tau_{rs} = \sigma\tau_{12}\sigma^{-1}$, where σ is any permutation for which $\sigma(1) = r$ and $\sigma(2) = s$. From Propositions 2.2.2, 2.2.5 and 2.2.4, $\operatorname{sgn}\tau_{rs} = (\operatorname{sgn}\sigma)^2\operatorname{sgn}\tau_{12} = -1$. □

Proposition 2.2.7 (sgn of general permutation)**.** If σ is a product of k transpositions, then $\operatorname{sgn}\sigma = (-1)^k$.

Proof. By Proposition 2.2.1 and 2.2.4, $\operatorname{sgn}\sigma$ is the product of the signs of the k transpositions. By Proposition 2.2.6, the sign of each transposition is -1. Therefore, $\operatorname{sgn}\sigma = (-1)^k$. □

It follows that $\operatorname{sgn}(\sigma)$ is 1 or -1 according to whether σ is given by the product of an even or odd number of transpositions.

Proposition 2.2.8 (Averaging over permutations)**.** Let $f : S_n \to \mathbb{R}$ be a function on S_n. Then for all $\alpha, \beta \in S_n$,

$$\sum_{\sigma\in S_n} f(\sigma) = \sum_{\sigma\in S_n} f(\alpha\sigma\beta).$$

Also,

$$\sum_{\sigma\in S_n} f(\sigma) = \sum_{\sigma\in S_n} f(\sigma^{-1}).$$

Proof. The mappings $\sigma \mapsto \alpha\sigma\beta$ and $\sigma \mapsto \sigma^{-1}$ are bijections on S_n. Therefore, the sums above contain precisely the same terms, albeit in a different order. □

2.3 Algebraic k-forms

2.3.1 *Tensors [nonexaminable]

Algebraic k-forms are particular examples of more general objects, called tensors. While in this course we will not deal with tensors generally, you are likely to come across them elsewhere in your studies, at least particular instances. Therefore, we will begin with a brief discussion of tensors in general (which can be omitted, and is non-examinable).

As motivation for the definition of the dual space V^*, we first considered general functions on V. As motivation for the definition of tensors, we first consider general functions on $V^k \times V^{*l}$. That is, we consider (real-valued) functions whose arguments are k vectors from V and l vectors from V^*. The set of such functions forms a vector space, with addition and scalar multiplication defined in the usual way. Tensors are a subset of this space of functions. A *tensor* of type (k,l) is a map $t : V^k \times V^{*l} \to \mathbb{R}$ which is linear with respect to each of its arguments. That is,

$$\begin{aligned}
&t(v_{(1)}, \ldots, \alpha u + \beta w, \ldots, v_{(k)}, a^{(1)}, \ldots, a^{(l)}) \\
&\qquad = \alpha t(v_{(1)}, \ldots, u, \ldots, v_{(k)}, a^{(1)}, \ldots, a^{(l)}) + \beta t(v_{(1)}, \ldots, w, \ldots, v_{(k)}, a^{(1)}, \ldots, a^{(l)}),
\end{aligned}$$

$$\begin{aligned}
&t(v_{(1)}, \ldots, v_{(k)}, a^{(1)}, \ldots, \alpha b + \beta c, \ldots, a^{(l)}) \\
&\qquad = \alpha t(v_{(1)}, \ldots, v_{(k)}, a^{(1)}, \ldots, b, \ldots, a^{(l)}) + \beta t(v_{(1)}, \ldots, v_{(k)}, a^{(1)}, \ldots, c, \ldots, a^{(l)}).
\end{aligned}$$

It is easy to show that the space of tensors of type (k,l) forms a vector space (that is, the property of being linear with respect to each argument is preserved by addition and scalar multiplication). It is also easy to show that the dimension of this vector space is n^{k+l}.

Some examples: The dual space V^* is the space of tensors of type $(1,0)$. V itself can be regarded as the space of tensors of type $(0,1)$. Linear maps from V to V can be regarded as tensors of type $(1,1)$, as can linear maps from V^* to V^*. Linear maps from V to V^* are tensors of type $(2,0)$, and linear maps from V^* to V are tensors of type $(0,2)$.

Next, we define an operation called the *tensor product*. Suppose t_1 and t_2 are functions on sets U_1 and U_2, respectively. Then we can construct a function $t_1 \otimes t_2$ on $U_1 \times U_2$ by taking $t_1 \otimes t_2(u_1, u_2)$ to be the product $t_1(u_1)t_2(u_2)$. Now suppose that U_1 and U_2 are vector spaces and that t_1 and t_2 are linear functions (so that $t_1 \in U_1^*$ and $t_2 \in U_2^*$). Then $t_1 \otimes t_2$ is linear in each of its arguments, i.e.

$$t_1 \otimes t_2(\alpha v_1 + \beta w_1, u_2) = \alpha t_1 \otimes t_2(v_1, u_2) + \beta t_1 \otimes t_2(w_1, u_2),$$

and similarly

$$t_1 \otimes t_2(u_1, \alpha v_2 + \beta w_2) = \alpha t_1 \otimes t_2(u_1, v_2) + \beta t_1 \otimes t_2(u_1, w_2).$$

We can generalise this operation as follows. Suppose that t_1 is a tensor of type (k_1, l_1) and t_2 is a tensor of type (k_2, l_2). We can construct a tensor of type $(k_1 + k_2, l_1 + l_2)$, denoted $t_1 \otimes t_2$, by taking

$$\begin{aligned}
&t_1 \otimes t_2(v_{(1)}, \ldots v_{(k_1+k_2)}, a^{(1)}, \ldots, a^{(l_1+l_2)}) \\
&\qquad = t_1(v_{(1)}, \ldots v_{(k_1)}, a^{(1)}, \ldots, a^{(l_1)})\, t_2(v_{(k_1+1)}, \ldots v_{(k_1+k_2)}, a^{(l_1+1)}, \ldots, a^{(l_1+l_2)}).
\end{aligned}$$

For example, if t_1 is a tensor of type $(2,0)$, and t_2 is a tensor of type $(1,1)$, then

$$t_1 \otimes t_2(u, v, w, a) = t_1(u, v)t_2(w, a).$$

$t_1 \otimes t_2$ is called the tensor product of t_1 and t_2. It can be shown that an arbitrary tensor of type (k,l) can be expressed as a sum of tensor products of k tensors of type $(1,0)$ and l tensors of type $(0,1)$.

2.3.2 Algebraic k-forms

An *algebraic k-form on V* is a function on V^k that is linear in each argument and which changes sign if two arguments are interchanged. That is, letting a denote an algebraic k-form, we have that

$$a : V^k \to \mathbb{R}; \left(v_{(1)}, \ldots, v_{(k)}\right) \mapsto a\left(v_{(1)}, \ldots, v_{(k)}\right).$$

Linearity with respect to each argument means that, for $u, w \in V$ and $\alpha, \beta \in \mathbb{R}$,

$$a\left(v_{(1)}, \ldots, \alpha u + \beta w, \ldots, v_{(k)}\right) = \alpha a\left(v_{(1)}, \ldots, u, \ldots, v_{(k)}\right) + \beta a\left(v_{(1)}, \ldots, w, \ldots, v_{(k)}\right).$$

Changing sign under the interchange of two arguments means that, for any j, l with $1 \le j < l \le k$,

$$a\left(v_{(1)}, \ldots, , v_{(j)}, \ldots, v_{(l)}, \ldots, v_{(k)}\right) = -a\left(v_{(1)}, \ldots, , v_{(l)}, \ldots, v_{(j)}, \ldots, v_{(k)}\right).$$

Denote the set of algebraic k-forms by $\Lambda^k(V)$. By convention, $\Lambda^0(V)$ is given by $\mathbb{R}$. Also, $\Lambda^1(V)$ is identified with the dual space V^* (the antisymmetry requirement is empty for $k = 1$).

With regard to the discussion in Section 2.3.1, algebraic k-forms are a special type of tensor of type $(k, 0)$. What makes them special is the antisymmetry property under interchange of two arguments.

Proposition 2.3.1. $\Lambda^k(V)$ is a vector space.

Proof. Straightforward, and hence omitted. The point is that one can define the sum of two algebraic k-forms and multiplication of an algebraic k-form by a scalar in the obvious way (that is, respectively, as the sum of two functions on V^k, and the multiplication of a function on V^k by a scalar). You can then show that the required properties of vector addition and scalar multiplication are satisfied. □

The following shows that, under an arbitrary permutation of its arguments, the value of an algebraic k-form either remains the same or changes sign according to the sign of the permutation.

Proposition 2.3.2. For $a \in \Lambda^k(V)$ and for $\sigma \in S_k$,

$$a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) = \operatorname{sgn} \sigma \, a(v_{(1)}, \ldots, v_{(k)}).$$

Proof. From Proposition 2.2.1, σ can be expressed as a product of transpositions, let's say m of them. From the antisymmetry property of algebraic k-forms, each transposition produces a change of sign. Therefore,

$$a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) = (-1)^m \, a(v_{(1)}, \ldots, v_{(k)}).$$

From Proposition 2.2.7, $(-1)^m = \operatorname{sgn} \sigma$. □

Example 2.3.3. Given $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w} \in \mathbb{R}^3$, let

$$a(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}.$$

Show that a is an algebraic 3-form on $\mathbb{R}^3$. See Problem Sheet 6.

2.4 Basis k-forms

Let $e_{(1)}, \ldots, e_{(n)}$ denote a basis for V. Given $v \in V$, we write

$$v = v^i e_{(i)}.$$

We introduce some notation. Let $I = (i_1, \ldots, i_k)$ denote an ordered k-tuple of indices, where $1 \le i_r \le n$ (I is also called a multi-index). We introduce a Kronecker delta for pairs of k-tuples of indices, defined by

$$\delta(I, J) = \begin{cases} 1, & i_1 = j_1, \ldots, i_k = j_k, \\ 0, & \text{otherwise.} \end{cases}$$

Given $\sigma \in S_k$, define

$$\sigma(I) = \left(i_{\sigma^{-1}(1)}, \ldots, i_{\sigma^{-1}(k)}\right).$$

That, is $\sigma(I)$ is a permutation of the indices comprising I. For example, if

$$I = (2, 4, 7, 6) \text{ and } \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix},$$

then

$$\sigma(I) = (7, 2, 4, 6).$$

Proposition 2.4.1. Let $\sigma, \tau \in S_k$. Then

$$\sigma(\tau(I)) = (\sigma\tau)(I).$$

Proof. Let $J = (j_1, \ldots, j_k) = \tau(I)$ and $K = (k_1, \ldots, k_k) = \sigma(J)$. Then $j_r = i_{\tau^{-1}(r)}$, and $k_r = j_{\sigma^{-1}(r)}$, so that

$$k_r = i_{\tau^{-1}(\sigma^{-1}(r))} = i_{(\sigma\tau)^{-1}(r)}.$$

Therefore, K, which we defined to be $\sigma(\tau(I)$, is also given by $(\sigma\tau)(I)$, which is what we wanted to show. □

Let $E_{(I)} \in V^k$ denote the k-tuple of basis vectors given by

$$E_{(I)} = \left(e_{(i_1)}, \ldots, e_{(i_k)}\right).$$

Given $a \in \Lambda^k(V)$, we write

$$a_I = a(E_{(I)}).$$

We will call the a_I's the *coefficients* of a with respect to the basis $e_{(i)}$. An alternative notation for the coefficients is

$$a_{i_1 \cdots i_k} = a\left(e_{(i_1)}, \ldots, e_{(i_k)}\right).$$

Note that, by the antisymmetry property,

$$a_{i_1, \cdots, i_k} = 0 \text{ if any two of the indices } i_1, \ldots, i_k \text{ are the same.}$$

The notation a_I has the advantage of being more concise.

The linearity property implies that an algebraic k-form is completely determined by its coefficients. To see this, note that

$$\begin{aligned} a\left(v_{(1)}, \ldots, v_{(k)}\right) &= a\left({v_{(1)}}^{i_1} e_{(i_1)}, \ldots, {v_{(k)}}^{i_k} e_{(i_k)}\right) \\ &= {v_{(1)}}^{i_1} \cdots {v_{(k)}}^{i_k} a(e_{(i_1)}, \ldots, e_{(i_k)}) = a_{i_1 \cdots i_k} {v_{(1)}}^{i_1} \cdots {v_{(k)}}^{i_k}. \end{aligned}$$

For example, if a is an algebraic 2-form, then

$$a(u, v) = a_{ij} u^i v^j.$$

Definition 2.4.2. Let $J = (j_1, \ldots, j_k)$. The *basis k-form $F^{(J)}$* is the algebraic k-form on V defined by

$$F^{(J)}_I := F^{(J)}(E_{(I)}) = \begin{cases} 0, & \text{if } j_r = j_s \text{ for some } r \neq s, \\ \operatorname{sgn}\sigma, & J = \sigma(I), \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

That is, $F^{(J)}(E_{(I)})$ vanishes if J contains repeated indices, regardless of what I is. Otherwise, $F^{(J)}(E_{(I)})$ vanishes if I is not a permutation of J, while if I is a permutation of J, then $F^{(J)}(E_{(I)})$ is equal to the sign of that permutation. An equivalent way to write the preceding formula for $F^{(J)}(E_{(I)})$, which will be useful, is given by the following:

Proposition 2.4.3. Suppose J consists of distinct indices. Then

$$F^{(J)}(E_{(I)}) = \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \ \delta(\sigma(I), J).$$

Proof. This follows by comparison with (26). If I is not a permutation of J, then $\delta(\sigma(I), J)$ vanishes for all σ, and the sum vanishes. If $J = \sigma_*(I)$ for some permutation σ_* (note that there can be only one such σ_*), then the sum vanishes for all terms but $\sigma = \sigma_*$. □

We should verify that the $F^{(J)}$'s really are algebraic k-forms. This is done in the following:

Proposition 2.4.4.

$$F^{(J)} \in \Lambda^k(V).$$

Proof. $F^{(J)}$ is defined by its values on k-tuples of basis vectors; that is,

$$F^{(J)}(v_{(1)}, \ldots, v_{(k)}) := {v_{(1)}}^{i_1} \cdots {v_{(1)}}^{i_k} F^{(J)}(e_{(i_1)}, \ldots, e_{(i_k)}).$$

Therefore, $F^{(J)}$ is automatically linear in each argument.

We need to check that $F^{(J)}$ satisfies the antisymmetry property. It is enough to consider the case where the arguments of $F^{(J)}$ are basis vectors. To this end, we must show that

$$F^{(J)}(E_{(\tau(I))}) = \operatorname{sgn}\tau F^{(J)}(E_{(I)})$$

for all $\tau \in S_k$. From Proposition 2.4.3,

$$F^{(J)}(E_{(\tau(I))}) = \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \ \delta(\sigma(\tau(I)), J).$$

From Proposition 2.4.1,

$$\delta(\sigma(\tau(I)), J) = \delta((\sigma\tau)(I), J).$$

Since $\operatorname{sgn}\sigma = \operatorname{sgn}\tau \operatorname{sgn}(\sigma\tau)$, we may write that

$$F^{(J)}(E_{(\tau(I))} = \operatorname{sgn}\tau \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma\tau) \ \delta((\sigma\tau)(I)), J).$$

By Proposition 2.2.8,

$$\sum_{\sigma \in S_k} \operatorname{sgn}(\sigma\tau) \ \delta((\sigma\tau)(I)), J) = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) \ \delta(\sigma(I), J) = F^{(J)}(E_{(I)}).$$

Therefore,

$$F^{(J)}(E_{(\tau(I))}) = \operatorname{sgn}\tau F^{(J)}(E_{(I)}),$$

as required.

□

This leads to the following determinant formula for the basis k-forms:

Proposition 2.4.5.

$$F^{(J)}(v_{(1)}, \dots v_{(k)}) = \det \begin{pmatrix} {v_{(1)}}^{j_1} & {v_{(2)}}^{j_1} & \cdots & {v_{(k)}}^{j_1} \\ {v_{(1)}}^{j_2} & {v_{(2)}}^{j_2} & \cdots & {v_{(k)}}^{j_2} \\ \vdots & \vdots & \cdots & \vdots & . \\ {v_{(1)}}^{j_k} & {v_{(2)}}^{j_k} & \cdots & {v_{(k)}}^{j_k} \end{pmatrix}.$$

Proof. Using linearity and Proposition 2.4.3, we have that

$$\begin{aligned} F^{(J)}(v_{(1)}, \dots v_{(k)}) &= \sum_I {v_{(1)}}^{i_1} \cdots {v_{(k)}}^{i_k} F^{(J)}(E_{(I)}) \\ &= \sum_{\sigma \in S_k} \sum_I \operatorname{sgn}(\sigma) {v_{(1)}}^{i_1} \cdots {v_{(k)}}^{i_k} \, \delta(\sigma(I), J) = \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) {v_{(1)}}^{j_{\sigma(1)}} \cdots {v_{(k)}}^{j_{\sigma(k)}}. \end{aligned}$$

This last expression can be recognised as the formula for the determinant of the matrix given in the statement of the proposition. Note that for the sake of being very explicit, we have put in the summation over I, which really means

$$\sum_I = \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n},$$

although it would be implied by the summation convention. The point – rather obvious, but perhaps worth spelling out – is that for given σ and J, there is precisely one choice of indices I for which $\sigma(I) = J$.

$\square$

Example 2.4.6 (Examples of basis k-forms)**.**

a) $F^{(2)}(u) = f^{(2)}(u) = u^2$. In general, for $k = 1$, $F^{(j)}$ coincides with the dual basis vector $f^{(j)}$.

b)

$$F^{(1,3)}(u, v) = u^1 v^3 - u^3 v^1 = \det \begin{pmatrix} u^1 & v^1 \\ u^3 & v^3 \end{pmatrix}$$

c)

$$F^{(1,4,3)}(u, v, w) = u^1v^4w^3 + u^3v^1w^4 + u^4v^3w^1 - u^4v^1w^3 - u^3v^4w^1 - u^1v^3w^4 = \det \begin{pmatrix} u^1 & v^1 & w^1 \\ u^4 & v^4 & w^4 \\ u^3 & v^3 & w^3 \end{pmatrix}$$

?

The following shows that the basis k-forms are highly redundant; basis k-forms $F^{(J)}$ and $F^{(\sigma(J))}$ differ by at most a sign.

Proposition 2.4.7 (Basis k-forms and permutations)**.** For $\alpha \in S_k$,

$$F^{(\alpha(J))} = \operatorname{sgn} \alpha \, F^{(J)}.$$

Proof. From Proposition 2.4.3,

$$F^{(\alpha(J))}(E_{(I)}) = \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \delta(\sigma(I), \alpha(J)). \tag{27}$$

But

$$\delta(\sigma(I), \alpha(J)) = \delta(\alpha^{-1}(\sigma(I)), J) = \delta((\alpha^{-1}\sigma)(I), J),$$

using Proposition 2.4.1, and

$$\operatorname{sgn} \sigma = \operatorname{sgn} \alpha \operatorname{sgn} (\alpha^{-1}\sigma).$$

Substituting the preceding into (27), we get

$$F^{(\alpha(J))}(E_{(I)}) = \operatorname{sgn} \alpha \sum_{\sigma \in S_k} \operatorname{sgn} (\alpha^{-1}\sigma)\, \delta(\alpha^{-1}\sigma(I), J) = \operatorname{sgn} \alpha \sum_{\sigma \in S_k} \operatorname{sgn} (\sigma)\, \delta(\sigma(I), J) = \operatorname{sgn} \alpha\, F^{(J)}(E_{(I)}),$$

where we have used Proposition 2.2.8. As this holds for all I, it follows that $F^{(\alpha(J))} = \operatorname{sgn} \alpha\, F^{(J)}$, as required. □

Example 2.4.8.

a) $F^{(2,4)} = -F^{(4,2)}$

b) $F^{(2,4,2)} = 0$

c) $F^{(1,3,5,7)} =$ -$F^{(7,3,5,1)} = F^{(7,5,3,1)}$

Proposition 2.4.9. [Expansion of general k-form] Let $a \in \Lambda^k(V)$. Then

$$a = \frac{1}{k!} a_J F^{(J)}, \tag{28}$$

where we use the summation convention for J (that is, there is a sum over each index j_r in $J = (j_1, \dots, j_k)$).

Proof. We evaluate both sides on $E_{(I)}$. On the left-hand side, we have

$$a(E_{(I)}) = a_I. \tag{29}$$

On the right-hand side, the only terms which contribute are those for which the indices in J are distinct (otherwise, $F^{(J)} = 0$). Using Proposition 2.4.3, we get

$$\frac{1}{k!} a_J F^{(J)}(E_{(I)}) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn} (\sigma) a_J \delta(\sigma(I), J) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn} (\sigma) a_{\sigma(I)}. \tag{30}$$

From Proposition 2.3.2,

$$a_{\sigma(I)} = \operatorname{sgn} \sigma a_I.$$

Therefore, (30) becomes

$$\frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}^2(\sigma) a_I = \frac{1}{k!} \sum_{\sigma \in S_k} a_I = a_I,$$

which is the same as (29). □

Let J_* denote a k-tuple of indices which are distinct and in ascending order. That is, $J_* = (j_1, \dots, j_k)$ with $j_1 < j_2 < \dots < j_k$.

Proposition 2.4.10. The $F^{(J_*)}$'s form a basis for $\Lambda^k(V)$, and for all $a \in \Lambda^k(V)$,

$$a = \sum_{J_*} a_{J_*} F^{(J_*)}. \tag{31}$$

[Note that in contrast to (28), there no factor of $1/k!$, whose absence is balanced by the fact that the sum over J_* in (31) is restricted to ascending k-tuples].

Proof. First, we show that the F^{J_*}'s are linearly independent. Suppose

$$\sum_{J_*} c_{J_*} F^{(J_*)} = 0.$$

We want to show that $c_{J_*} = 0$. To do this, we apply both sides to $E_{(I_*)}$. We have that

$$F^{(J_*)}(E_{(I_*)}) = \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \;\delta(\sigma(I_*), J_*) = \delta(I_*, J_*),$$

since $\sigma(I_*)$ and J_* can coincide only if $\sigma = e$ (both I_* and J_* are in ascending order). Therefore,

$$\sum_{J_*} c_{J_*} F^{(J_*)}(E_{(I_*)}) = c_{I_*} = 0,$$

as required.

Next, we show that (31) holds, which in particular implies that the $F^{(J_*)}$'s span $\Lambda^k(V)$. We start with Proposition 2.4.10,

$$a = \frac{1}{k!} a_J F^{(J)}.$$

The only terms which contribute to the sum over J are those in which the indices in J are distinct. Every such J is related to a unique J_* by a unique permutation σ. Therefore, we can replace the (implicit) sum over J by sums over J_* and σ to get

$$a = \frac{1}{k!} \sum_{J_*} \sum_{\sigma \in S_k} a_{\sigma(J_*)} F^{(\sigma(J_*))}.$$

But $a_{\sigma(J_*)} = \operatorname{sgn}\sigma a_{J_*}$ from Proposition 2.3.2, and $F^{(\sigma(J_*))} = \operatorname{sgn}\sigma F^{(J_*)}$ from Proposition 2.4.7. Thus,

$$a = \frac{1}{k!} \sum_{J_*} \sum_{\sigma \in S_k} (\operatorname{sgn}\sigma)^2 a_{J_*} F^{(J_*)} = \sum_{J_*} a_{J_*} F^{(J_*)},$$

as required. $\square$

The number of distinct k-tuples J_* where the indices j_r are in strictly ascending order is the number of ways of choosing k distinct things from n things, which is given by the binomial coefficient $\binom{n}{k} = n!/(k!(n-k)!)$. It follows from Proposition 2.3.1 that

$$\dim \Lambda^k(V) = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Note that, for $k = 0$, this is consistent with taking $\Lambda^0 = \mathbb{R}$, since $\binom{n}{0} = 1$. Note too that $\Lambda^n(V)$ is also one-dimensional, while $\Lambda^k(V)$ for $k > n$ is zero-dimensional, and consists of a single element, namely the trivial function which maps everything to zero.

2.5 Wedge product

Definition 2.5.1. Let $a \in \Lambda^k(V)$ be an algebraic k-form and $b \in \Lambda^l(V)$ be an algebraic l-form. Their *wedge product*, denoted $a \wedge b$, is the algebraic $(k+l)$-form defined by

$$a \wedge b\,(v_{(1)}, \dots, v_{(k+l)}) = \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, a(v_{(\sigma(1))}, \dots, v_{(\sigma(k))}) b(v_{(\sigma(k+1))}, \dots, v_{(\sigma(k+l))}). \quad (32)$$

In words, the value of $a \wedge b$ for a set of $(k+l)$ vectors is obtained by permuting the vectors and evaluating a on the first k and b on the last l of them, multiplying the values of a and b so obtained, and then summing the result over all permutations counted with sign, and then dividing by $k!\,l!$. Note that permutations which differ only by permutations among the first k vectors and/or among the last l vectors produce the same contribution to the sum. The factor of $k!\,l!$ compensates for this.

The wedge product appears naturally in many problems in geometry, topology and physics (field theories). For us, the wedge product will provide us with a means to express k-forms, including basis k-forms, in terms of 1-forms, and in particular basis 1-forms. Many general results and calculations can be simplified by carrying them out for one-forms, and then extending the results to general k-forms by means of the wedge product. The wedge product will also be important when we come to discuss differential forms.

Example 2.5.2 (Examples of wedge product)**.**

a) $k = 0$. If a is an algebraic zero-form, i.e. $a \in \mathbb{R}$, then the wedge product reduces to scalar multiplication, i.e.
$$a \wedge b = ab.$$

b) If a and b are algebraic one-forms, then $a \wedge b$ is an algebraic two-form, and
$$a \wedge b\,(u, v) = a(u)b(v) - a(v)b(u).$$

c) If a is an algebraic one-form and b is an algebraic two-form, then $a \wedge b$ is an algebraic three-form and
$$a \wedge b\,(u, v, w) = a(u)b(v, w) + a(v)b(w, u) + a(w)b(u, v).$$
Note that in the formula (32), the sum over $\sigma \in S_3$ produces six terms, and there is also a factor of $1/2$ in front. However, each term in the sum appears, in effect, twice; for example $a(u)b(v, w)$ is accompanied by $-a(u)b(w, v)$, which is equal to $a(u)b(v, w)$, so that the six terms can be reduced to three, and the factor of $1/2$ cancels.

The wedge product satisfies the following properties:

Proposition 2.5.3 (Properties of the wedge product)**.**

i) $a \wedge b$ is an algebraic $(k+l)$-form. That is, $a \wedge b$, as defined by (32), is linear in each argument and changes sign under interchange of any pair of arguments.

ii) Linearity. If a is an algebraic k-form and b and c are algebraic l-forms, then
$$a \wedge (b + c) = a \wedge b + a \wedge c.$$

iii) (Anti)commutativity. If a is an algebraic k-form and b is an algebraic l-form, then

$$a \wedge b = (-1)^{kl} b \wedge a.$$

In other words, if either k or l is even, then $a \wedge b = b \wedge a$. If both k and l are odd, then $a \wedge b = -b \wedge a$.

iv) Associativity. If a is an algebraic k-form, b an algebraic l-form, and c an algebraic m-form, then

$$a \wedge (b \wedge c) = (a \wedge b) \wedge c.$$

v) Basis k-forms. Let $J = (j_1, \ldots, j_k)$. Then

$$F^{(J)} = f^{(j_1)} \wedge \cdots \wedge f^{(j_k)}.$$

The proofs of these results are a bit involved. They are not hard to understand, and the results are easy to verify in examples. The difficulty is the book-keeping required for a general argument. For this reason, the proofs are given separately in the following section, Section 2.6. The proofs are non-examinable.

As an illustration, let us verify Proposition 2.5.3 v) for a simple case by showing that

$$F^{(i,j)} = f^{(i)} \wedge f^{(j)}.$$

Let us apply both sides to vectors u and v. From Proposition 2.4.5,

$$F^{(i,j)}(u, v) = u^i v^j - u^j v^i,$$

while from (32),

$$f^{(i)} \wedge f^{(j)}(u, v) = f^{(i)}(u) f^{(j)}(v) - f^{(i)}(v) f^{(i)}(u) = u^i v^j - u^j v^i,$$

as claimed. A consequence of Proposition 2.5.3 v) is that a general algebraic k-form can be written as a sum of wedge products of algebraic one-forms.

We may derive the following component formula for the wedge product: Let a be an algebraic k-form and b an algebraic l-form given by

$$a = \frac{1}{k!} a_J F^{(J)}, \quad b = \frac{1}{l!} b_M F^{(M)}.$$

Then

$$a \wedge b = \frac{1}{k!\, l!} a_J b_M F^{(J,M)},$$

where $J = (j_1, \ldots, j_k)$, $M = (m_1, \ldots, m_l)$, $(J, M) = (j_1, \ldots, j_k, m_1, \ldots, m_l)$, and J and M are summed over. Sometimes we'll write this with the indices explicitly written out, as follows:

$$a \wedge b = \frac{1}{k!\, l!} a_{j_1,\ldots,j_k} b_{m_1,\ldots,m_l} f^{(j_1)} \wedge \cdots \wedge f^{(j_k)} \wedge f^{(m_1)} \wedge \cdots \wedge f^{(m_l)}.$$

Here is an example of this formula:

$$\begin{aligned}\left(f^{(1)} \wedge f^{(3)} - 2f^{(4)} \wedge f^{(1)}\right) &\wedge \left(3f^{(2)} - 4f^{(3)} + 5f^{(4)}\right) \\ &= -3f^{(1)} \wedge f^{(2)} \wedge f^{(3)} - 6f^{(1)} \wedge f^{(2)} \wedge f^{(4)} + 13f^{(1)} \wedge f^{(3)} \wedge f^{(4)}.\end{aligned}$$

2.6 *Proof of properties of the wedge product [nonexaminable]

These notes contain proofs for the properties of the wedge product stated without proof in the previous section. This material is not examinable. Throughout, let V denote an n-dimensional vector space, and $e_{(1)}, \ldots, e_{(n)}$ a basis on V. Let $f^{(1)}, \ldots, f^{(n)}$ denote the corresponding basis on the dual space V^*.

Definition 2.6.1. Let $a \in \Lambda^k(V)$ and $b \in \Lambda^l(V)$. Their wedge product, denoted $a \wedge b$, is given by

$$a \wedge b(v_{(1)}, \ldots, v_{(k+l)}) = \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}).$$

Proposition 2.6.2. *$a \wedge b$ is a $(k+l)$-form.*

Proof.
We need to show that $a \wedge b$ is linear and antisymmetric in its arguments.

Antisymmetry: Given $k+l$ vectors $v_{(1)}, \ldots, v_{(k+l)}$, reorder them by a permutation τ to define $k+l$ vectors $w_{(1)}, \ldots, w_{(k+l)}$ given by

$$w_{(r)} = v_{(\tau(r))}.$$

Then

$$\begin{aligned} a \wedge b(v_{(\tau(1))}, \ldots, v_{(\tau(k+l))}) &= a \wedge b(w_{(1)}, \ldots, w_{(k+l)}) \\ &= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, a(w_{(\sigma(1))}, \ldots, w_{(\sigma(k))}) b(w_{(\sigma(k+1))}, \ldots, w_{(\sigma(k+l))}) \\ &= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, a(v_{(\tau\sigma(1))}, \ldots, v_{(\tau\sigma(k))}) b(v_{(\tau\sigma(k+1))}, \ldots, v_{(\tau\sigma(k+l))}) \\ &= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\tau^{-1}\sigma)\, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}), \end{aligned}$$

where in the last line we have used Proposition 2.2.8 to replace σ by $\tau^{-1}\sigma$ in the summand. Since $\operatorname{sgn}(\tau^{-1}\sigma) = \operatorname{sgn}\tau \operatorname{sgn}\sigma$, we get that

$$\begin{aligned} a \wedge b(v_{(\tau(1))}, \ldots, v_{(\tau(k+l))}) &= a \wedge b(w_{(1)}, \ldots, w_{(k+l)}) \\ &= \operatorname{sgn}\tau \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}) \\ &= \operatorname{sgn}\tau\, a \wedge b(v_{(1)}, \ldots, v_{(k+l)}), \end{aligned}$$

as required.

Linearity: From the antisymmetry property just established, it suffices to prove linearity with respect to the $(k+l)$th argument. The idea is that linearity for $a \wedge b$ follows from the fact that a and b are linear. However, a little book-keeping is needed to keep track of whether, under a permutation σ, the $(k+l)$th argument belongs to a or b. Let $v_{(1)}, \ldots v_{(k+l-1)}$ be $k+l-1$ vectors in V, w and x two additional vectors in V, and $\alpha, \beta \in \mathbb{R}$.

Define three sets of $k+l$ vectors, $A_{(i)}$, $B_{(i)}$ and $C_{(i)}$ as follows: For $i < k+l$,

$$A_{(i)} = B_{(i)} = C_{(i)} = v_{(i)},$$

while for $i = k+l$,

$$A_{(k+l)} = w, \quad B_{(k+l)} = x, \quad C_{(k+l)} = \alpha w + \beta x.$$

We will show that

$$a \wedge b(C_{(1)}, \ldots, C_{(k+l)}) = \alpha a \wedge b(A_{(1)}, \ldots, A_{(k+l)}) + \beta a \wedge b(B_{(1)}, \ldots, B_{(k+l)}),$$

which amounts to linearity in the $(k+l)$th argument.

We have that

$$a \wedge b(C_{(1)}), \ldots, C_{(k+l)}) = \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \, a(C_{(\sigma(1))}, \ldots, C_{(\sigma(k))}) b(C_{(\sigma(k+1))}, \ldots, C_{(\sigma(k+l))}).$$

For given $\sigma \in S_{k+l}$, define r by

$$\sigma(r) = k + l,$$

or, equivalently, $r = \sigma^{-1}(k+l)$. If $r \leq k$, we have that

$$a(C_{(\sigma(1))}, \ldots, C_{(\sigma(k))}) = a(C_{(\sigma(1))}, \ldots, \alpha w + \beta x, \ldots, C_{(\sigma(k))}),$$

where, on the right-hand side, $\alpha w + \beta x$ is the rth argument of a. By the linearity of a,

$$\begin{aligned} &a(C_{(\sigma(1))}, \ldots, \alpha w + \beta x, \ldots, C_{(\sigma(k))}) \\ &\qquad = \alpha a(C_{(\sigma(1))}, \ldots, w, \ldots, C_{(\sigma(k))}) + \beta a(C_{(\sigma(1))}, \ldots, x, \ldots, C_{(\sigma(k))}) \\ &\qquad\qquad = \alpha a(A_{(\sigma(1))}, \ldots, A_{(\sigma(k))}) + \beta a(B_{(\sigma(1))}, \ldots, B_{(\sigma(k))}). \end{aligned}$$

For this same σ,

$$b(C_{(\sigma(k+1))}, \ldots, C_{(\sigma(k+l))}) = b(A_{(\sigma(k+1))}, \ldots, A_{(\sigma(k+l))}) = b(B_{(\sigma(k+1))}, \ldots, B_{(\sigma(k+l))}).$$

It follows that, for this same σ,

$$\begin{aligned} &a(C_{(\sigma(1))}, \ldots, C_{(\sigma(k))}) b(C_{(\sigma(k+1))}, \ldots, C_{(\sigma(k+l))}) \\ &\qquad = \alpha a(A_{(\sigma(1))}, \ldots, A_{(\sigma(k))}) b(A_{(\sigma(k+1))}, \ldots, A_{(\sigma(k+l))}) \\ &\qquad\qquad + \beta a(B_{(\sigma(1))}, \ldots, B_{(\sigma(k))}) b(B_{(\sigma(k+1))}, \ldots, B_{(\sigma(k+l))}). \end{aligned}$$

In case $r > k$, a similar argument (which we won't repeat) establishes that the preceding formula still holds. Therefore

$$\begin{aligned} a \wedge b(C_{(1)}), \ldots, C_{(k+l)}) &= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \\ &\qquad \times \Big(\alpha a(A_{(\sigma(1))}, \ldots, A_{(\sigma(k))}) b(A_{(\sigma(k+1))}, \ldots, A_{(\sigma(k+l))}) \\ &\qquad + \beta a(B_{(\sigma(1))}, \ldots, B_{(\sigma(k))}) b(B_{(\sigma(k+1))}, \ldots, B_{(\sigma(k+l))}) \Big) \\ &= \alpha a \wedge b(A_{(1)}, \ldots, A_{(k+l)}) + \beta a \wedge b(B_{(1)}, \ldots, B_{(k+l)}), \end{aligned}$$

as required.

$\square$

The following is a formula for the wedge product of more than two forms. It wasn't given in lectures, but is used below to establish that the wedge product is associative as well as to express the basis k-forms $F^{(J)}$ as wedge products of the $f^{(j)}$'s.

Proposition 2.6.3. Let $a_{(1)}, a_{(2)}, \ldots$ be a sequence of algebraic forms on V of degrees $k_1, k_2, \ldots$ Let $b_{(1)}, b_{(2)}, \ldots$ denote the successive wedge products of the $a_{(i)}$'s, defined by

$$\begin{aligned} b_{(1)} &= a_{(1)}, \\ b_{(j)} &= b_{(j-1)} \wedge a_{(j)}, \quad j > 1. \end{aligned}$$

That is, $b_{(2)} = a_{(1)} \wedge a_{(2)}$, $b_{(3)} = (a_{(1)} \wedge a_{(2)}) \wedge a_{(3)}$, etc. $b_{(q)}$ is an l_q-form, where

$$l_q = k_1 + \cdots + k_q.$$

Then $b_{(q)}$ is given by

$$b_{(q)}(v_{(1)}, \ldots, v_{(l_q)}) = \frac{1}{k_1! \cdots k_q!} \times \Bigg(\sum_{\sigma \in S_{l_q}} \operatorname{sgn} \sigma \, a_{(1)}(v_{(\sigma(1))}, \ldots, v_{(\sigma(k_1))}) a_{(2)}(v_{(\sigma(l_1+1))}, \ldots, v_{(\sigma(l_1+k_2))}) \cdots \cdots a_{(q)}(v_{(\sigma(l_{q-1}+1))}, \ldots, v_{(\sigma(l_q))}) \Bigg).$$

Proof. We proceed by induction on q, the number of forms in the product. If $q = 2$, the formula above for $b_{(2)}$ coincides with the definition of the wedge product. For $q > 2$, assume the formula holds for $q - 1$, and prove that it holds for q, as follows. From the definition of the wedge product,

$$\begin{aligned} b_{(q)}(v_{(1)}, \ldots, v_{(l_q)}) &= b_{(q-1)} \wedge a_{(q)}(v_{(1)}, \ldots, v_{(l_q)}) \\ &= \frac{1}{l_{q-1}! k_q!} \times \Bigg(\sum_{\sigma \in S_{l_q}} \operatorname{sgn} \sigma \, b_{(q-1)}(v_{(\sigma(1))}, \ldots, v_{(\sigma(l_{q-1}))}) a_{(q)}(v_{(\sigma(l_{q-1}+1))}, \ldots, v_{(\sigma(l_q))}) \Bigg). \end{aligned}$$

From the induction hypothesis,

$$b_{(q-1)}(v_{(\sigma(1))}, \ldots, v_{(\sigma(l_{q-1}))}) = \frac{1}{k_1! \cdots k_{q-1}!} \sum_{\tau \in S_{l_{q-1}}} \operatorname{sgn} \tau \times \Bigg(a_{(1)}(v_{(\sigma(\tau(1)))}, \ldots, v_{(\sigma(\tau(k_1)))}) a_{(2)}(v_{(\sigma(\tau(l_1+1)))}, \ldots, v_{(\sigma(\tau(l_1+k_2)))}) \cdots \cdots a_{(q-1)}(v_{(\sigma(\tau(l_{q-2}+1)))}, \ldots, v_{(\sigma(\tau(l_{q-1})))}) \Bigg).$$

Substitute above to get

$$b_{(q)}(v_{(1)}, \ldots, v_{(l_q)}) = \frac{1}{k_1! \cdots k_{q-1}!} \frac{1}{l_{q-1}! k_q!} \sum_{\tau \in S_{l_{q-1}}} \operatorname{sgn} \tau \sum_{\sigma \in S_{l_q}} \operatorname{sgn} \sigma \times \Bigg(a_{(1)}(v_{(\sigma(\tau(1)))}, \ldots, v_{(\sigma(\tau(k_1)))}) a_{(2)}(v_{(\sigma(\tau(l_1+1)))}, \ldots, v_{(\sigma(\tau(l_1+k_2)))}) \cdots \cdots a_{(q-1)}(v_{(\sigma(\tau(l_{q-2}+1)))}, \ldots, v_{(\sigma(\tau(l_{q-1})))}) \cdot \cdot a_{(q)}(v_{(\sigma(l_{q-1}+1))}, \ldots, v_{(\sigma(l_q))}) \Bigg).$$

Given $\tau \in S_{l_{q-1}}$, define $\hat{\tau} \in S_{l_q}$ by

$$\begin{aligned} \hat{\tau}(r) &= \tau(r), && 1 \leq r \leq l_{q-1}, \\ \hat{\tau}(r) &= r, && l_{q-1} < r \leq l_q. \end{aligned}$$

That is, $\hat{\tau}$ permutes the numbers 1 through l_{q-1} in the same way as τ does while leaving the numbers l_{q-1} through l_q alone (note that l_q's are an increasing sequence of numbers). Then the preceding expression can be re-written as

$$\begin{aligned} b_{(q)}(v_{(1)}, \ldots, v_{(l_q)}) = \frac{1}{k_1! \cdots k_{q-1}!} \frac{1}{l_{q-1}! k_q!} \sum_{\tau \in S_{l_{q-1}}} \operatorname{sgn} \tau \sum_{\sigma \in S_{l_q}} \operatorname{sgn} \sigma \times \\ \Big(a_{(1)}(v_{(\sigma\hat{\tau}(1))}, \ldots, v_{(\sigma\hat{\tau}(k_1))}) a_{(2)}(v_{(\sigma\hat{\tau}(l_1+1))}, \ldots, v_{(\sigma\hat{\tau}(l_1+k_2))}) \cdots \\ \cdots a_{(q-1)}(v_{(\sigma\hat{\tau}(l_{q-2}+1))}, \ldots, v_{(\sigma\hat{\tau}(l_{q-1}))}) \cdot a_{(q)}(v_{(\sigma\hat{\tau}(l_{q-1}+1))}, \ldots, v_{(\sigma\hat{\tau}(l_q))}) \Big). \end{aligned}$$

Use Proposition 2.2.8 to replace σ by $\sigma\hat{\tau}^{-1}$ in the sum over σ. Then $\operatorname{sgn} \sigma$ is replaced by $\operatorname{sgn} \sigma \operatorname{sgn} \hat{\tau}^{-1} = \operatorname{sgn} \sigma \operatorname{sgn} \hat{\tau}$. It is clear that $\operatorname{sgn} \tau = \operatorname{sgn} \hat{\tau}$ (both permutations can be realised by the same set of transpositions), so that the factors of $\operatorname{sgn} \tau$ cancel from the sum above. As the summand no longer depends on τ, the sum over $\tau \in S_{l_{q-1}}$ yields a factor of $l_{q-1}!$, which cancels the same factor appearing in the denominator. We get that

$$\begin{aligned} b_{(q)}(v_{(1)}, \ldots, v_{(l_q)}) = \frac{1}{k_1! \cdots k_q!} \sum_{\sigma \in S_{l_q}} \operatorname{sgn} \sigma \times \\ \Big(a_{(1)}(v_{(\sigma(1))}, \ldots, v_{(\sigma(k_1))}) a_{(2)}(v_{(\sigma(l_1+1))}, \ldots, v_{(\sigma(l_1+k_2))}) \cdots a_{(q)}(v_{(\sigma(l_{q-1}+1))}, \ldots, v_{(\sigma(l_q))}) \Big), \end{aligned}$$

as required. □

Proposition 2.6.4. The wedge product satisfies the following properties:

$$\begin{aligned} a \wedge (\alpha b + \beta c) &= \alpha a \wedge b + \beta a \wedge c, \quad \alpha, \beta \in \mathbb{R}, a \in \Lambda^k(V), b, c \in \Lambda^l(V) \quad \text{(Linearity)}, \\ a \wedge b &= (-1)^{kl} b \wedge a, \quad a \in \Lambda^k(V), b \in \Lambda^l(V) \quad \text{((Anti)commutativity)}, \\ a \wedge (b \wedge c) &= (a \wedge b) \wedge c \quad \alpha \in \Lambda^k(V), b \in \Lambda^l(V), c \in \Lambda^p(V) \quad \text{(Associativity)}. \end{aligned}$$

Proof. **Linearity** follows immediately from a direct calculation:

$$\begin{aligned} &(a \wedge (\alpha b + \beta c))(v_{(1)}, \ldots, v_{(k+l)}) \\ &= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) \Big(\alpha b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}) \\ &\qquad + \beta c(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}) \Big) \\ &= \alpha \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}) \\ &+ \beta \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) c(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}) \\ &= \alpha a \wedge b(v_{(1)}, \ldots, v_{(k+l)}) + \beta a \wedge c(v_{(1)}, \ldots, v_{(k+l)}). \end{aligned}$$

For **(anti)commutativity**, define the cyclic permutation $\tau \in S_{k+l}$ by

$$\begin{aligned} &\tau(1) = l+1, \quad \tau(2) = l+2, \quad \ldots, \tau(k) = l+k, \\ &\tau(k+1) = 1, \quad \tau(k+2) = 2, \quad \ldots, \tau(k+l) = l. \end{aligned}$$

We have that

$$a \wedge b(v_{(1)}, \ldots, v_{(k+l)}) = \frac{1}{k!\, l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}).$$

Replace σ by $\sigma\tau$ in the summand above (Proposition 2.2.8 again) to get that

$$\begin{aligned} &a \wedge b(v_{(1)}, \ldots, v_{(k+l)}) \\ &\qquad = \operatorname{sgn} \tau \frac{1}{k!\, l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \, a(v_{(\sigma(l+1))}, \ldots, v_{(\sigma(l+k))}) b(v_{(\sigma(1))}, \ldots, v_{(\sigma(l))}) \\ &\qquad = \operatorname{sgn} \tau \, b \wedge a(v_{(1)}, \ldots, v_{(k+l)}), \end{aligned}$$

ie

$$a \wedge b = \operatorname{sgn} \tau b \wedge a.$$

It remains to calculate $\operatorname{sgn} \tau$. We note that $\tau = \rho^k$, where ρ is the cyclic permutation given by

$$\rho(i) = i+1, \quad i < k+l, \qquad \rho(k+l) = 1.$$

From Problem 6.4,

$$\operatorname{sgn} \rho = (-1)^{k+l-1},$$

so that

$$\operatorname{sgn} \tau = (-1)^{l(k+l-1)}.$$

But $l^2 - l = l(l-1)$, so $l^2 - l$ is always even. Therefore,

$$\operatorname{sgn} \tau = (-1)^{kl},$$

as required.

For **associativity**, we note that Proposition 2.6.3 gives an explicit formula for $(a \wedge b) \wedge c$, namely

$$\begin{aligned} &(a \wedge b) \wedge c\,(v_{(1)}, \ldots, v_{(k+l+p)}) = \frac{1}{k!} \frac{1}{l!} \frac{1}{p!} \times \\ &\times \sum_{\sigma \in S_{k+l+p}} \operatorname{sgn} \sigma \, a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}) b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}) c(v_{(\sigma(k+l+1))}, \ldots, v_{(\sigma(k+l+p))}). \end{aligned} \tag{33}$$

For $a \wedge (b \wedge c)$, use (anti)commutativity to bring it into a form where we can apply Proposition 2.6.3 That is,

$$a \wedge (b \wedge c) = (-1)^{k(l+p)} (b \wedge c) \wedge a,$$

so that

$$\begin{aligned} &a \wedge (b \wedge c)(v_{(1)}, \ldots, v_{(k+l+p)}) = (-1)^{k(l+p)}) (b \wedge c) \wedge a\,(v_{(1)}, \ldots, v_{(k+l+p)}) \\ &= (-1)^{k(l+p)} \frac{1}{k!\, l!\, p!} \times \sum_{\sigma \in S_{k+l+p}} \operatorname{sgn} \sigma \, b(v_{(\sigma(1))}, \ldots, v_{(\sigma(l))}) c(v_{(\sigma(l+1))}, \ldots, v_{(\sigma(l+p))}) a(v_{(\sigma(l+p+1))}, \ldots, v_{(\sigma(l+p+k))}). \end{aligned}$$

Arguing as we did for (anti)commutativity, let τ denote the permutation which shifts every number up by k, ie

$$\tau(r) = \begin{cases} r+k, & 1 \le r \le l+p, \\ r-(l+p), & l+p+1 \le r \le l+p+k. \end{cases}$$

Replace σ by $\sigma\tau$ in the preceding (using Proposition 2.2.8) to obtain

$$\begin{aligned} a \wedge (b \wedge c)(v_{(1)}, \ldots, v_{(k+l+p)}) = (-1)^{k(l+p)})(b \wedge c) \wedge a(v_{(1)}, \ldots, v_{(k+l+p)}) = \\ = (-1)^{k(l+p)})\frac{1}{k!}\frac{1}{l!}\frac{1}{p!} \times \\ \times \sum_{\sigma \in S_{k+l+p}} \operatorname{sgn}(\sigma\tau) b(v_{(\sigma\tau(1))}, \ldots, v_{(\sigma\tau(l))}) c(v_{(\sigma\tau(l+1))}, \ldots, v_{(\sigma\tau(l+p))}) a(v_{(\sigma\tau(l+p+1))}, \ldots, v_{(\sigma\tau(l+p+k))}) \\ = (-1)^{k(l+p)}\frac{1}{k!}\frac{1}{l!}\frac{1}{p!} \operatorname{sgn}\tau \times \\ \times \sum_{\sigma \in S_{k+l+p}} \operatorname{sgn}\sigma\, b(v_{(\sigma(k+1))}, \ldots, v_{(\sigma(k+l))}) c(v_{(\sigma(k+l+1))}, \ldots, v_{(\sigma(k+l+p))}) a(v_{(\sigma(1))}, \ldots, v_{(\sigma(k))}). \end{aligned} \tag{34}$$

Equations (33) and (34) agree apart from the sign factor $(-1)^{k(l+p)} \operatorname{sgn}\tau$. But $\tau = \rho^k$, where $\rho \in S_{k+l+p}$ is the cyclic permutation that shifts every number up by 1, with $k+l+p$ mapped back to 1. From Problem 6.4, $\operatorname{sgn}\rho = (-1)^{k+l+p-1}$. Therefore,

$$\operatorname{sgn}\tau = (\operatorname{sgn}\rho)^k = (-1)^{k(k+l+p-1)} = (-1)^{k(l+p)},$$

since, as we noted above, $k^2 - k = k(k-1)$ is always even. Therefore,

$$(-1)^{k(l+p)} \operatorname{sgn}\tau = 1,$$

so that equations (33) and (34) agree, and associativity is confirmed.

□

Proposition 2.6.5. Let $J = (j_1, \ldots, j_k)$. Then

$$F^{(J)} = f^{(j_1)} \wedge \cdots \wedge f^{(j_k)}.$$

Proof. It is enough to show that both sides agree when their arguments are basis vectors. Let $I = (i_1, \ldots, i_k)$ and $E_{(I)} = (e_{(i_1)}, \ldots, e_{(i_k)})$. Then from Proposition 2.6.3,

$$\begin{aligned} f^{(j_1)} \wedge \cdots \wedge f^{(j_k)}(E_{(I)}) &= \sum_{\sigma \in S_k} \operatorname{sgn}\sigma\, f^{(j_1)}(e_{(\sigma(i_1))}) \cdots f^{(j_k)}(e_{(\sigma(i_k))}) \\ &= \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \delta^{j_1}_{\sigma(i_1)} \cdots \delta^{j_k}_{\sigma(i_k)} = \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \delta(\sigma(I), J). \end{aligned}$$

But this last expression coincides with the definition of $F^{(J)}(E_{(I)})$. □

2.7 Contraction

Definition 2.7.1. Let $v \in V$. We define $i_v : \Lambda^k(V) \to \Lambda^{k-1}(V)$, the contraction with v, by

$$i_v c = 0, \text{ for } c \in \Lambda^0(V),$$
$$i_v a(w_{(1)}, \ldots, w_{(k-1)}) = a(v, w_{(1)}, \ldots, w_{(k-1)}), \text{ for } a \in \Lambda^k(V),\ k > 0,$$

where $w_{(1)}, \ldots, w_{(k-1)} \in V$. Thus, i_v maps k-forms to $(k-1)$-forms by fixing the first argument of the k-form to be v. The contraction on any zero-form is defined to be zero.

Proposition 2.7.2 (Properties of contraction)**.**

$$i_v(a+b) = i_v a + i_v b, \quad a, b \in \Lambda^k(V),$$
$$i_v(a \wedge b) = (i_v a) \wedge b + (-1)^k a \wedge (i_v b), \quad a \in \Lambda^k(V), b \in \Lambda^l(V).$$

The proof (which is non-examinable) is given in the following section.

Corollary 2.7.3. Let $a_1 \in \Lambda^{k_1}(V)$ be an algebraic k_1-form on V, $a_2 \in \Lambda^{k_2}(V)$ an algebraic k_2-form on V, and $\dots$ $a_r \in \Lambda^{k_r}(V)$ an algebraic k_r-form on V. Then

$$\begin{aligned} i_v\,(a_1 \wedge a_2 \wedge a_3 \cdots \wedge a_r) &= \\ = (i_v a_1) \wedge a_2 \wedge \cdots \wedge a_r + (-1)^{k_1} a_1 \wedge (i_v a_2) \wedge a_3 \wedge \cdots \wedge a_r &+ (-1)^{k_1+k_2} a_1 \wedge a_2 \wedge (i_v a_3) \wedge \cdots \wedge a_r \\ &+ \cdots + (-1)^{k_1+k_2+\cdots+k_{r-1}} a_1 \wedge a_2 \wedge \cdots \wedge a_{r-1} \wedge i_v a_r. \end{aligned}$$

Proof. This follows by induction on r. For $r = 2$, it is just the second formula in Proposition 2.7.2. For $r = 3$, we have that

$$i_v\,(a_1 \wedge a_2 \wedge a_3) = (i_v a_1) + (-1)^{k_1} a_1 \wedge i_v(a_2 \wedge a_3),$$

and

$$i_v(a_2 \wedge a_3) = (i_v a_2) \wedge a_3 + (-1)^{k_2} a_2 \wedge i_v a_3.$$

Combining the preceding results, we get that

$$i_v\,(a_1 \wedge a_2 \wedge a_3) = (i_v a_1) \wedge a_2 \wedge a_3 + (-1)^{k_1} a_1 \wedge (i_v a_2) \wedge a_3 + (-1)^{k_1+k_2} a_1 \wedge a_2 \wedge i_v a_3,$$

which yields the result for $r = 3$. The result for general r follows similarly; we omit the explicit argument. $\square$

Example 2.7.4.

a)
$$i_v f^{(j)} = v^j = f^{(j)}(v).$$

b)
$$i_v f^{(1)} \wedge f^{(2)} = v^1 f^{(2)} - v^2 f^{(1)}.$$

c)
$$i_v f^{(2)} \wedge f^{(4)} \wedge f^{(3)} = v^2 f^{(4)} \wedge f^{(3)} - v^4 f^{(2)} \wedge f^{(3)} + v^3 f^{(2)} \wedge f^{(4)}.$$

Proposition 2.7.5 (Coordinate formula for contraction)**.**

Let
$$a = \frac{1}{k!}\, a_{i_1 \cdots i_k}(x)\, f^{(i_1)} \wedge \cdots \wedge\, f^{(i_k)} \in \Lambda^k(V).$$

Then
$$i_v a = \frac{1}{(k-1)!}\, v^j a_{j i_2 \cdots i_k}(x)\, f^{(i_2)} \cdots \wedge\, f^{(i_k)}.$$

The proof, which is nonexaminable, is given in the following section.

2.8 *Proof of properties of contraction [nonexaminable]

These notes contain a proof of Proposition 2.7.2, which was omitted from the lectures.

Let V be an n-dimensional vector space, and let $v \in V$. We define $i_v : \Lambda^k(V) \to \Lambda^{k-1}(V)$, the contraction with v, by

$$
\begin{aligned}
i_v c &= 0, \text{ for } c \in \Lambda^0(V),\\
i_v a(w_{(1)}, \dots, w_{(k-1)}) &= a(v, w_{(1)}, \dots, w_{(k-1)}), \text{ for } a \in \Lambda^k(V),\ k > 0,
\end{aligned}
$$

where $w_{(1)}, \dots, w_{(k-1)} \in V$. Thus, i_v maps k-forms to $(k-1)$-forms by fixing the first argument of the k-form to be v.

We want to prove the following (Proposition 2.7.2):

$$
\begin{aligned}
i_v(a+b) &= i_v a + i_v b, \quad a, b \in \Lambda^k(V),\\
i_v(a \wedge b) &= (i_v a) \wedge b + (-1)^k a \wedge (i_v b), \quad a \in \Lambda^k(V), b \in \Lambda^l(V).
\end{aligned}
$$

Proof. The first property, which states the contraction is a linear map, is easy to verify. Let a and b be algebraic k -forms. Then $i_v(a+b)$ is a $(k-1)$-form, which is determined by its values on an arbitrary set of $k-1$ vector fields, which we denote by $w_{(1)}, \dots, w_{(k-1)}$. We have that

$$
\begin{aligned}
(i_v(a+b))(w_{(1)}, \dots, w_{(k-1)}) &= (a+b)(v, w_{(1)}, \dots, w_{(k-1)})\\
&= a(v, w_{(1)}, \dots, w_{(k-1)}) + b(v, w_{(1)}, \dots, w_{(k-1)})\\
&= i_v a(w_{(1)}, \dots, w_{(k-1)}) + i_v b(w_{(1)}, \dots, w_{(k-1)}).
\end{aligned}
$$

Therefore, if we omit the arguments $w_{(1)}, \dots, w_{(k-1)}$, we get the (quite obvious) relation between $(k-1)$-forms,

$$i_v(a+b) = i_v a + i_v b.$$

Next we consider the contraction acting on a wedge product. The argument here is more involved. Let a be an algebraic k-form and b an algebraic l-form. Then $i_v(a \wedge b)$ is a $(k+l-1)$-form, which is determined by its values on an arbitrary set of $k+l-1$ vector fields, which we denote by $w_{(1)}, \dots, w_{(k+l-1)}$. We have that

$$(i_v(a \wedge b))(w_{(1)}, \dots, w_{(k+l-1)}) = (a \wedge b)(v, w_{(1)}, \dots, w_{(k+l-1)}).$$

It will be convenient to introduce an alternative notation for the vector fields $v, w_{(1)}, \dots, w_{(k+l-1)}$, so that they can all be referred to by a single index. Thus we write

$$x_{(1)} = v, \quad x_{(2)} = w_{(1)}, \quad \dots, \quad x_{(k+l)} = w_{(k+l-1)}.$$

Then from the definition of the wedge product,

$$(a \wedge b)(x_{(1)}, \dots, x_{(k+l)}) = \frac{1}{k!}\frac{1}{l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, a(x_{(\sigma(1))}, \dots, x_{(\sigma(k))}) b(x_{(\sigma(k+1))}, \dots, x_{(\sigma(k+l))}).$$

Given $1 \le r \le k+l$, let S^r_{k+l} denote the subset of S_{k+l} consisting of permutations σ for which $\sigma(r) = 1$. Then every permutation $\sigma \in S_{k+l}$ belongs to a unique S^r_{k+l}, namely the one with $r = \sigma^{-1}(1)$. We have that

$$\sum_{\sigma \in S_{k+l}} = \sum_{r=1}^{k+l} \sum_{\sigma \in S^r_{k+l}},$$

so that

$$(a \wedge b)(x_{(1)}, \ldots, x_{(k+l)}) = \sum_{r=1}^{k+l} T_r, \tag{35}$$

where

$$T_r = \frac{1}{k!}\frac{1}{l!} \sum_{\sigma \in S^r_{k+l}} \operatorname{sgn}\sigma\, a(x_{(\sigma(1))}, \ldots, x_{(\sigma(k))}) b(x_{(\sigma(k+1))}, \ldots, x_{(\sigma(k+l))}).$$

Consider first the case where $r \le k$. Given $\sigma \in S^r_{k+l}$, it follows that $\sigma\tau_{r1}(1) = \sigma(r) = 1$, so that $\sigma\tau_{r1}$ is a permutation which leaves 1 invariant, and therefore permutes the numbers 2 through $k+l$ amongst themselves. By shifting these numbers down by one, we can construct a permutation $\hat\sigma \in S_{k+l-1}$ given by

$$\hat\sigma(s) = (\sigma\tau_{r1})(s+1) - 1.$$

It is then straightforward to check that

$$\begin{aligned}
x_{(\sigma(1))} &= w_{(\hat\sigma(r-1))},\\
x_{(\sigma(r))} &= v,\\
x_{(\sigma(s))} &= w_{(\hat\sigma(s-1))} \text{ for } s \ne 1, r.
\end{aligned}$$

Note that since $r \le k$, $x_{(\sigma(r))}$ appears in T_r as an argument of a. Also,

$$\operatorname{sgn}\sigma = \operatorname{sgn}\tau_{r1} \operatorname{sgn}\hat\sigma = (-1)^{\delta_{r,1}-1}\operatorname{sgn}\hat\sigma.$$

Therefore,

$$\begin{aligned}
&a(x_{(\sigma(1))}, \ldots, x_{(\sigma(k))}) b(x_{(\sigma(k+1))}, \ldots, x_{(\sigma(k+l))})\\
&= (-1)^{\delta_{r,1}-1} a(w_{(\hat\sigma(r-1))}, w_{(\hat\sigma(1))}, \ldots, w_{(\hat\sigma(r-2))}, v, w_{(\hat\sigma(r))}, \ldots, w_{(\hat\sigma(k-1))}) b(w_{(\hat\sigma(k))}, \ldots, w_{(\hat\sigma(k+l-1))})\\
&\qquad = a(v, w_{(\hat\sigma(1))}, \ldots, w_{(\hat\sigma(k-1))}) b(w_{(\hat\sigma(k))}, \ldots, w_{(\hat\sigma(k+l-1))})\\
&\qquad\qquad = (i_v a)(w_{(\hat\sigma(1))}, \ldots, w_{(\hat\sigma(k-1))}) b(w_{(\hat\sigma(k))}, \ldots, w_{(\hat\sigma(k+l-1))}),
\end{aligned}$$

where in the second-to-last equality an extra sign factor of $(-1)^{\delta_{r,1}-1}$ comes from interchanging the vector fields v and $w_{(\hat\sigma(r-1))}$ in a (if $r = 1$, no exchange is necessary). Substituting into the expression for T_r, we get

$$\begin{aligned}
T_r &= \frac{1}{k!}\frac{1}{l!} \sum_{\hat\sigma \in S_{k+l-1}} \operatorname{sgn}\hat\sigma (i_v a)(w_{(\hat\sigma(1)(}, \ldots, w_{(\hat\sigma(k-1))}) b(w_{(\hat\sigma(k))}, \ldots, w_{(\hat\sigma(k+l-1))})\\
&= \frac{1}{k}\frac{1}{(k-1)!}\frac{1}{l!} \sum_{\hat\sigma \in S_{k+l-1}} \operatorname{sgn}\hat\sigma (i_v a)(w_{(\hat\sigma(1)}, \ldots, w_{(\hat\sigma(k-1))}) b(w_{(\hat\sigma(k))}, \ldots, w_{(\hat\sigma(k+l-1))})\\
&\qquad = \frac{1}{k}((i_v a) \wedge b)(w_{(1)}, \ldots, w_{(k+l-1)}).
\end{aligned}$$

Since T_r does not depend on r, summing over r between 1 and k just eliminates the factor of $1/k$ above. Therefore,

$$\sum_{r=1}^{k} T_r = ((i_v a) \wedge b)(w_{(1)}, \ldots, w_{(k+l-1)}). \tag{36}$$

The case where $r > k$ is treated similarly, but takes a bit more work. Let $C_{1,r,k+1} \in S_{k+l}$ denote the cyclic permutation which maps 1 to r, r to $k+1$, and $k+1$ to 1, leaving all other numbers between 1 and $k+l$ unchanged. Given $\sigma \in S^r_{k+l}$, it follows that $\sigma C_{1,r,k+1}(1) = \sigma(r) = 1$, so that $\sigma C_{1,r,k+1}$ is a permutation which leaves 1 invariant, and therefore permutes the numbers

2 through $k+l$ amongst themselves. By shifting these numbers down by one, we can construct a permutation $\hat{\sigma} \in S_{k+l-1}$ by

$$\hat{\sigma}(s) = (\sigma C_{1,r,k+1})(s+1) - 1.$$

It is then straightforward to check that

$$\begin{aligned} x_{(\sigma(1))} &= w_{(\hat{\sigma}(k))}, \\ x_{(\sigma(k+1))} &= w_{(\hat{\sigma}(r-1))}, \\ x_{(\sigma(r))} &= v, \\ x_{(\sigma(s))} &= w_{(\hat{\sigma}(s-1))}, \text{ for } s \neq 1, r, k+1. \end{aligned}$$

Note that since $r > k$, $x_{(\sigma(r))} = v$ appears in T_r as an argument of b. Also,

$$\operatorname{sgn} \sigma = \operatorname{sgn} C_{1,r,k+1} \operatorname{sgn} \hat{\sigma} = (-1)^{\delta_{r,k+1}} \operatorname{sgn} \hat{\sigma}$$

(note that if $r = k+1$, then $C_{1,r,k+1}$ reduces to the transposition τ_{r1}). Therefore,

$$\begin{aligned} &a(x_{(\sigma(1))}, \ldots, x_{(\sigma(k))}) b(x_{(\sigma(k+1))}, \ldots, x_{(\sigma(k+l))}) \\ &= (-1)^{\delta_{r,k+1}} a(w_{(\hat{\sigma}(k))}, w_{(\hat{\sigma}(1))}, \ldots, w_{(\hat{\sigma}(k-1))}) b(w_{(\hat{\sigma}(r-1))}, w_{(\hat{\sigma}(k))}, \ldots, w_{(\hat{\sigma}(r-2))}, v, w_{(\hat{\sigma}(r))}, \ldots, w_{(\hat{\sigma}(k+l-1))}). \end{aligned}$$

We have that

$$a(w_{(\hat{\sigma}(k))}, w_{(\hat{\sigma}(1))}, \ldots, w_{(\hat{\sigma}(k-1))}) = (-1)^{k-1} a(w_{(\hat{\sigma}(1))}, w_{(\hat{\sigma}(2))}, \ldots, w_{(\hat{\sigma}(k))}).$$

since a cyclic permutation of the arguments of a k-form produces a sign factor $(-1)^{k-1}$. Also,

$$\begin{aligned} &b(w_{(\hat{\sigma}(r-1))}, w_{(\hat{\sigma}(k))}, \ldots, w_{(\hat{\sigma}(r-2))}, v, w_{(\hat{\sigma}(r))}, \ldots, w_{(\hat{\sigma}(k+l-1))}) \\ &\quad = (-1)^{\delta_{r,k+1}-1} b(v, w_{(\hat{\sigma}(k))}, \ldots, w_{(\hat{\sigma}(r-2))}, w_{(\hat{\sigma}(r-1))}, w_{(\hat{\sigma}(r))}, \ldots, w_{(\hat{\sigma}(k+l-1))}) \\ &\quad = (-1)^{\delta_{r,k+1}-1} (i_v b)(w_{(\hat{\sigma}(k))}, \ldots, w_{(\hat{\sigma}(k+l-1))}), \end{aligned}$$

where in the second-to-last equality an extra sign factor of $(-1)^{\delta_{r,k+1}}$ comes from interchanging the vector fields v and $w_{(\hat{\sigma}(r-1))}$ in b (if $r = k+1$, no exchange is necessary). Substituting the preceding into the expression for T_r, we get

$$\begin{aligned} &T_r = \\ &(-1)^{k-1}(-1)^{\delta_{r,k+1}-1}(-1)^{\delta_{r,k+1}} \frac{1}{k!}\frac{1}{l!} \sum_{\hat{\sigma} \in S_{k+l-1}} \operatorname{sgn} \hat{\sigma} a(w_{(\hat{\sigma}(1))}, \ldots, w_{(\hat{\sigma}(k))})(i_v b)(w_{(\hat{\sigma}(k+1))}, \ldots, w_{(\hat{\sigma}(k+l-1))})) \\ &\quad = (-1)^k \frac{1}{l} \frac{1}{(k)!} \frac{1}{(l-1)!} \sum_{\hat{\sigma} \in S_{k+l-1}} \operatorname{sgn} \hat{\sigma} a(w_{(\hat{\sigma}(1))}, \ldots, w_{(\hat{\sigma}(k))})(i_v b)(w_{(\hat{\sigma}(k+1))}, \ldots, w_{(\hat{\sigma}(k+l-1))})) \\ &\quad = (-1)^k \frac{1}{l} (a \wedge i_v b)(w_{(1)}, \ldots, w_{(k+l-1)}). \end{aligned}$$

Since T_r does not depend on r, summing over r between $k+1$ and $k+l$ just eliminates the factor of $1/l$ above. Therefore,

$$\sum_{r=k+1}^{k+l} T_r = (a \wedge i_v b)(w_{(1)}, \ldots, w_{(k+l-1)}). \tag{37}$$

Finally, we substitute (36) and (37) into (35) to obtain

$$(i_v(a \wedge b))(w_{(1)}, \ldots, w_{(k+l-1)}) = \Big((i_v a) \wedge b + (-1)^k a \wedge i_v b\Big)(w_{(1)}, \ldots, w_{k+l-1}),$$

as required.

□

Proof of Proposition 2.7.5. Since $i_v(a+b) = i_v a + i_v b$, it follows that

$$i_v a = \frac{1}{k!} a_{i_1 \cdots i_k}(x) i_v \left(f^{(i_1)} \wedge \cdots \wedge f^{(i_k)} \right).$$

Next, using $i_v(a \wedge b) = (i_v a) \wedge b + a \wedge (i_v b)$ repeatedly, and noting that $i_v a$ is a 0-form, i.e. a real number, if a is a 1-form, we get that

$$\begin{aligned} i_v \left(f^{(i_1)} \wedge \cdots \wedge f^{(i_k)} \right) &= (i_v f^{(i_1)}) f^{(i_2)} \wedge \cdots \wedge f^{(i_k)} - f^{(i_1)} \wedge i_v \left(f^{(i_2)} \wedge \cdots \wedge f^{(i_k)} \right) \\ &= v^{i_1} f^{(i_2)} \wedge \cdots \wedge f^{(i_k)} - f^{(i_1)} \wedge i_v \left(f^{(i_2)} \wedge \cdots \wedge f^{(i_k)} \right). \end{aligned}$$

Repeated application of this formula yields

$$i_v \left(f^{(i_1)} \wedge \cdots \wedge f^{(i_k)} \right) = \sum_{j=1}^{k} (-1)^{j-1} v^{i_j} f^{(i_1)} \wedge \cdots \widehat{f^{(i_j)}} \wedge \cdots \wedge f^{(i_k)},$$

where $\widehat{f^{(i_j)}}$ means that the factor of $f^{(i_j)}$ is to be *omitted* from the product. Therefore,

$$i_v a = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{j-1} a_{i_1 \cdots i_j \cdots i_k} v^{i_j} f^{(i_1)} \wedge \cdots \widehat{f^{(i_j)}} \wedge \cdots \wedge f^{(i_k)}.$$

We have that

$$a_{i_1 i_2 \cdots i_j \cdots i_k} = (-1)^{j-1} a_{i_j i_1 i_2 \cdots \widehat{i_j} \cdots i_k}.$$

Substituting above and relabeling the indices in the sums, we get

$$i_v a = \frac{1}{k!} \sum_{j=1}^{k} a_{j i_2 \cdots i_k} v^j f^{(i_2)} \wedge \cdots \wedge f^{(i_k)} = \frac{1}{(k-1)!} a_{j i_2 \cdots i_k} v^j f^{(i_2)} \wedge \cdots \wedge f^{(i_k)}.$$

□

2.9 Algebraic forms on $\mathbb{R}^3$

We have that

$$\dim \Lambda^0(\mathbb{R}^3) = 1, \quad \dim \Lambda^1(\mathbb{R}^3) = \dim \Lambda^2(\mathbb{R}^3) = 3, \quad \dim \Lambda^3(\mathbb{R}^3) = 1.$$

Thus, algebraic zero-forms and three-forms on $\mathbb{R}^3$ can be identified with scalars, while algebraic one-forms and two-forms can be identified with vectors. By making these identifications, the wedge product can be seen to correspond with the familiar dot product and cross product of vector algebra.

A **zero-form** is by definition a scalar.

A **one-form** $a \in \Lambda^1(\mathbb{R}^3)$ can be written as $a = a_1 f^{(1)} + a_2 f^{(2)} + a_3 f^{(3)}$. We may associate a vector $\mathbf{A} = (A_1, A_2, A_3)$ to a one-form, and vice versa, by making the (rather obvious) identifications

$$a_1 = A_1, \quad a_2 = A_2, \quad a_3 = A_3.$$

A **two-form** $b \in \Lambda^2(\mathbb{R}^3)$ can be written as

$$b = b_{12} f^{(1)} \wedge f^{(2)} + b_{23} f^{(2)} \wedge f^{(3)} + b_{31} f^{(3)} \wedge f^{(1)}.$$

We may associate a vector $\mathbf{B} = (B_1, B_2, B_3)$ to a two-form, and vice versa, by making the (somewhat less obvious) identifications

$$b_{23} = B_1, \quad b_{31} = B_2, \quad b_{12} = B_3.$$

A **three-form** $c \in \Lambda^3(\mathbb{R}^3)$ can be written as $c = c_{123} f^{(1)} \wedge f^{(2)} \wedge f^{(3)}$. We may associate a scalar C to a three-form, and vice versa, by making the (fairly obvious) identification

$$c_{123} = C.$$

Let us translate the wedge product of algebraic forms on $\mathbb{R}^3$ into relations between vectors and scalars. A wedge product involving a 0-form is just scalar multiplication, so we won't consider it further. The wedge product of two one-forms yields a two-form, and the wedge product of a one-form and a two-form is a three-form. These are the only two cases to consider, as all other wedge products (involving, for example, two two-forms or a one-form and a three-form) produce forms of degree greater than three, and therefore vanish automatically.

Let a and d be two one-forms, and let $b = a \wedge d$. Then b is a two-form. Let $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{B}$ denote the vectors associated with a, d and b, respectively. As we show below,

$$\mathbf{B} = \mathbf{A} \times \mathbf{D}.$$

Let's show this explicitly. We have that

$$\begin{aligned}
a \wedge d &= \left(a_1 f^{(1)} + a_2 f^{(2)} + a_3 f^{(3)}\right) \wedge \left(d_1 f^{(1)} + d_2 f^{(2)} + d_3 f^{(3)}\right) \\
&= a_1 d_2 f^{(1)} \wedge f^{(2)} + a_1 d_3 f^{(1)} \wedge f^{(3)} + a_2 d_1 f^{(2)} \wedge f^{(1)} + a_2 d_3 f^{(2)} \wedge f^{(3)} + a_3 d_1 f^{(3)} \wedge f^{(1)} + a_3 d_2 f^{(3)} \wedge f^{(2)} \\
&= (a_1 d_2 - a_2 d_1) f^{(1)} \wedge f^{(2)} + (a_2 d_3 - a_3 d_2) f^{(2)} \wedge f^{(3)} + (a_3 d_1 - a_1 d_3) f^{(3)} \wedge f^{(1)}.
\end{aligned}$$

This two-form may be associated with the vector $\mathbf{B}$ as follows:

$$\mathbf{B} = (a_2 d_3 - a_3 d_2, a_3 d_1 - a_1 d_3, a_1 d_2 - a_2 d_1) = (A_2 D_3 - A_3 D_2, A_3 D_1 - A_1 D_3, A_1 D_2 - A_2 D_1) = \mathbf{A} \times \mathbf{D}.$$

Thus, the wedge product of two one-forms corresponds to the cross-product of the associated vectors. Note that the anticommutativity of the wedge product of one-forms, i.e. the fact that $a \wedge d = -d \wedge a$, corresponds to the antisymmetry of the cross-product,

$$\mathbf{A} \times \mathbf{D} = -\mathbf{D} \times \mathbf{A}.$$

Next, let a be a one-form, b be a two-form, and $c = a \wedge b$. Then c is a three-form. Let $\mathbf{A}$ and $\mathbf{B}$ be the vectors associated with a and b, and C the scalar associated with c. As we show below,

$$C = \mathbf{A} \cdot \mathbf{B}.$$

Let's show this explicitly. We have that

$$\begin{aligned}
a \wedge b &= \left(a_1 f^{(1)} + a_2 f^{(2)} + a_3 f^{(3)}\right) \wedge \left(b_{12} f^{(1)} \wedge f^{(2)} + b_{23} f^{(2)} \wedge f^{(3)} + b_{31} f^{(3)} \wedge f^{(1)}.\right) \\
&= a_1 b_{23} f^{(1)} \wedge f^{(2)} \wedge f^{(3)} + a_2 b_{31} f^{(2)} \wedge f^{(3)} \wedge f^{(1)} + a_3 b_{12} f^{(3)} \wedge f^{(1)} \wedge f^{(2)} \\
&= (a_1 b_{23} + a_2 b_{31} + a_3 b_{12}) f^{(1)} \wedge f^{(2)} \wedge f^{(3)}.
\end{aligned}$$

The coefficient of the basis three-form is given

$$a_1 b_{23} + a_2 b_{31} + a_3 b_{12} = A_1 B_1 + A_2 B_2 + A_3 B_3 = \mathbf{A} \cdot \mathbf{B}.$$

Thus, the wedge product of a one-form and a two-form corresponds to the dot product of the associated vectors. Note that the commutativity of the wedge product of a one-form and a two-form, i.e. the fact that $a \wedge b = b \wedge a$, corresponds to the symmetry of the dot product,

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}.$$

Let a, d and e be one-forms, and let $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{E}$ be the associated vectors. The associativity of the wedge product, i.e. the fact that $a \wedge (d \wedge e) = (a \wedge d) \wedge e$, corresponds to the triple product identity,

$$\mathbf{A} \cdot (\mathbf{D} \times \mathbf{E}) = (\mathbf{A} \times \mathbf{D}) \cdot \mathbf{E}.$$

Let us also consider the contraction for algebraic forms on $\mathbb{R}^3$. This, too, turns out to correspond to familiar vector-algebra operations. Let $\mathbf{v} = v^1 e_{(1)} + v^2 e_{(2)} + v^3 e_{(3)} \in \mathbb{R}^3$ be a vector in $\mathbb{R}^3$. Let $a = a_1 f^{(1)} + a_2 f^{(2)} + a_3 f^{(3)} \in \Lambda^1(\mathbb{R}^3)$, and let $\mathbf{A}$ be the associated vector in $\mathbb{R}^3$. Then $i_{\mathbf{v}} a$ is a scalar, and is given by

$$i_{\mathbf{v}} a = \mathbf{A} \cdot \mathbf{v}.$$

Thus, contraction with a one-form corresponds to the dot product.

Let $b = b_{12} f^{(1)} \wedge f^{(2)} + b_{23} f^{(2)} \wedge f^{(3)} + b_{31} f^{(3)} \wedge f^{(1)} \in \Lambda^2(\mathbb{R}^3)$ be an algebraic two-form on $\mathbb{R}^3$, and let $\mathbf{B}$ be the associated vector in $\mathbb{R}^3$. Then $i_{\mathbf{v}} b$ is a one-form, and is given by

$$i_{\mathbf{v}} b = \left(v^3 b_{31} - v^2 b_{12}\right) f^{(1)} + \left(v^1 b_{12} - v^3 b_{23}\right) f^{(2)} + \left(v^2 b_{23} - v^1 b_{31}\right) f^{(3)},$$

which corresponds to the vector $\mathbf{B} \times \mathbf{v}$. Thus, contraction with a two-form corresponds to the cross product.

Let $c = c_{123} f^{(1)} \wedge f^{(2)} \wedge f^{(3)} \in \Lambda^3(\mathbb{R}^3)$ be an algebraic three-form on $\mathbb{R}^3$ with associated scalar c_{123}. Then $i_{\mathbf{v}} c$ is a two-form, and is given by

$$i_{\mathbf{v}} c = c_{123} v^3 f^{(1)} \wedge f^{(2)} + c_{123} v^1 f^{(2)} \wedge f^{(3)} + c_{123} v^2 f^{(3)} \wedge f^{(1)},$$

which corresponds to the vector $C\mathbf{v}$. Thus, contraction with a three-form corresponds to scalar multiplication.

3 Differential forms

3.1 Definition of differential forms

Definition 3.1.1 (Differential forms)**.** Let U be an open subset of $\mathbb{R}^n$. A **differential k-form** on U, or **k-form** for short, is a smooth map

$$\alpha : U \to \Lambda^k(\mathbb{R}^n); \quad x \mapsto \alpha(x) = \frac{1}{k!} \alpha_J(x) F^{(J)}.$$

Here, "smooth" means that the coefficient functions are smooth, i.e. $\alpha_J(x) \in \mathbb{C}^\infty(\infty)$.

Let $\Omega^k(U)$ denote the space of differential k-forms on U. For $k = 0$, 0-forms are smooth functions on U, i.e. $\Omega^0(U) = C^\infty(U)$.

Given $v_{(1)}, \ldots, v_{(k)} \in \mathbb{R}^n$, we write $\alpha(x; v_{(1)}, \ldots, v_{(k)})$ to denote $\alpha(x)$ (which is an algebraic k-form) evaluated on the k vectors $v_{(1)}, \ldots, v_{(k)}$.

Addition, wedge product and contraction of differential forms with vectors are defined pointwise, in analogy with the corresponding operators on algebraic k-forms.

Definition 3.1.2 (Addition of differential forms)**.** Given $\alpha, \beta \in \Omega^k(U)$, then $\alpha(x)$ and $\beta(x)$ are both algebraic k-forms, and it makes sense to add them. Thus, we define $\alpha + \beta \in \Omega^k(U)$ by

$$(\alpha + \beta)(x) := \alpha(x) + \beta(x).$$

Definition 3.1.3 (Wedge product of differential forms)**.** If α is a k-form on U and β is an l-form, then we define the $(k+l)$-form $\alpha \wedge \beta$ by

$$(\alpha \wedge \beta)(x) := \alpha(x) \wedge \beta(x).$$

It follows from Proposition 2.5.3 that

$$\begin{aligned}
\alpha \wedge \beta &= (-1)^{kl} \beta \wedge \alpha, \\
\alpha \wedge (\beta + \gamma) &= \alpha \wedge \beta + \alpha \wedge \gamma, \text{ where } \gamma \text{ is a differential } l\text{-form, like } \beta, \\
\alpha \wedge (\beta \wedge \gamma) &= (\alpha \wedge \beta) \wedge \gamma, \text{ where } \gamma \text{ is a differential } m\text{-form.}
\end{aligned}$$

For example, if we regard differential 1-forms on $\mathbb{R}^3$ as vector fields, then we are just defining the sum and cross product of vector fields by the rather obvious formula $(\mathbf{A} + \mathbf{B})(\mathbf{r}) = \mathbf{A}(\mathbf{r}) + \mathbf{B}(\mathbf{r})$, and $(\mathbf{A} \times \mathbf{B})(\mathbf{r}) = \mathbf{A}(\mathbf{r}) \times \mathbf{B}(\mathbf{r})$.

Definition 3.1.4 (Contraction of differential forms)**.** Let $U \subset \mathbb{R}^n$ be open. Given a differential k-form $\omega \in \Omega^k$ and a vector field $\mathbb{X} \in \mathcal{X}(U)$, the **contraction** of ω with $\mathbb{X}$, denoted $i_{\mathbb{X}}\omega$, is the differential $(k-1)$-form defined by

$$(i_{\mathbb{X}}\omega)(x) := i_{\mathbb{X}(x)}\omega(x).$$

Note that $\mathbb{X}(x) \in \mathbb{R}^n$ is a vector in $\mathbb{R}^n$, and $\omega(x) \in \Lambda^k(\mathbb{R}^n)$ is an algebraic k-form. Hence, $i_{\mathbb{X}(x)}\omega(x) \in \Lambda^{k-1}(\mathbb{R}^n)$, where i_v denotes the contraction map of Section 2.7, is an algebraic $(k-1)$-form that depends smoothly on x (since $\mathbb{X}(x)$ and $\omega(x)$ depend smoothly on x), which therefore determines a differential $(k-1)$-form.

Note that the contraction of a 0-form is always zero.

The properties of the contraction map for algebraic forms carry over to differential forms; specifically, from Proposition 2.7.2, we have the following:

$$i_{\mathbb{X}}(\alpha + \beta) = i_{\mathbb{X}}\alpha + i_{\mathbb{X}}\beta, \tag{38}$$

$$i_{\mathbb{X}}(\alpha \wedge \beta) = (i_{\mathbb{X}}\alpha) \wedge \beta + (-1)^k \alpha \wedge i_{\mathbb{X}}\beta. \tag{39}$$

Example 3.1.5. Let α be the differential 1-form and β the differential 2-form on $\mathbb{R}^3$ given by

$$\begin{aligned}
\alpha &= (x+y) f^{(2)} + (x^2 - y^2) f^{(3)}, \\
\beta &= z\, f^{(1)} \wedge f^{(2)} + xz\, f^{(1)} \wedge f^{(3)}
\end{aligned}$$

and let $\mathbb{X}$ be the vector field on $\mathbb{R}^3$ given by

$$\mathbb{X} = xe_{(2)} + e_{(3)} = (0, -x, 1).$$

Then

$$\alpha \wedge \beta = (x+y)xz\, f^{(2)} \wedge f^{(1)} \wedge f^{(3)} + (x^2 - y^2) z f^{(3)} \wedge f^{(1)} \wedge f^{(2)} = -(x+y)yz\, f^{(1)} \wedge f^{(2)} \wedge f^{(3)} = -(x+y)yz\, F^{(1,2,3)},$$

and

$$(i_{\mathbb{X}}\beta)(x, y, z) = i_{(0,-x,1)}\left(z\, f^{(1)} \wedge f^{(2)}\right) + i_{(0,-x,1)}\left(xz\, f^{(1)} \wedge f^{(3)}\right) = xzf^{(1)} - xzf^{(1)} = 0.$$

3.2 The exterior derivative

Definition 3.2.1 (Exterior derivative)**.** Let $U \subset \mathbb{R}^n$ be open. The **exterior derivative**, denoted d: $\Omega^k(U) \to \Omega^{k+1}(U)$, is a map from differential k-forms to differential $(k+1)$-forms defined as follows:

- $k = 0$. For $g \in \Omega^0(U) = C^\infty(U)$,
$$dg := \frac{\partial g}{\partial x^i} f^{(i)}.$$
- $k > 0$. For $\omega \in \Omega^k(U)$, we may write that
$$\omega = \frac{1}{k!}\omega_{i_1 \dots i_k} f^{(i_1)} \wedge \cdots \wedge f^{(i_k)},$$
where $\omega_{i_1 \dots i_k}(x) \in C^\infty(U)$ (cf Proposition 2.4.10). Then
$$d\omega(x) := \frac{1}{k!} d\omega_{i_1 \dots i_k} \wedge f^{(i_1)} \wedge \cdots \wedge f^{(i_k)}.$$
Equivalently,
$$d\omega(x) := \frac{1}{k!}\frac{\partial \omega_{i_1 \dots i_k}}{\partial x^j} f^{(j)} \wedge f^{(i_1)} \wedge \cdots \wedge f^{(i_k)}.$$

Thus, the exterior derivative of a function corresponds to the gradient of the function; indeed, dg is often called a gradient one-form. In particular, let us consider the ith coordinate function, which we denote by x^i. We regard x^i as a function on U given by

$$x^i(x_*) = x^i_*.$$

That is, x^i, regarded as a function, picks out the ith component of its argument. Then

$$dx^i = \frac{\partial x^i}{\partial x^j} f^{(j)} = \delta^i_j f^{(j)} = f^{(i)}.$$

Thus, dx^i is equal to the constant one-form $f^{(i)}$; we can think of $f^{(i)}$ as a differential one-form $\alpha_j f^{(j)}$ with a single nonzero coefficient function, α_i, which is equal to 1.

Notation. It is conventional to write dx^i instead of $f^{(i)}$. Similarly, we write

$$dx^{i_1} \wedge \cdots \wedge dx^{i_k} \text{ instead of } f^{(i_1)} \wedge \cdots \wedge f^{(i_k)},$$

and sometimes

$$dx^I \text{ instead of } F^{(I)}.$$

Thus, from now on, we will write differential one-forms as

$$\alpha = \alpha_i \, dx^i.$$

In particular, a gradient one-form will be written as

$$dg = \frac{\partial g}{\partial x^i} dx^i,$$

which conforms to usage in standard calculus. A general k-form will be written as

$$\omega = \frac{1}{k!}\omega_{i_1 \dots i_k} dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$

or sometimes as

$$\omega = \frac{1}{k!}\omega_I dx^I.$$

Proposition 3.2.2 (d and addition)**.** Let $\alpha, \beta \in \Omega^k(U)$. Then

$$d(\alpha + \beta) = d\alpha + d\beta.$$

Proof. We have that

$$d(\alpha+\beta) = d\left(\frac{1}{k!}\,(\alpha_{i_1\cdots i_k} + \beta_{i_1\cdots i_k})\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) = \left(\frac{1}{k!}\,(d\alpha_{i_1\cdots i_k} + d\beta_{i_1\cdots i_k}) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right)$$
$$= \frac{1}{k!}\,d\alpha_{i_1\cdots i_k} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} + \frac{1}{k!}\,d\beta_{i_1\cdots i_k} \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} = d\alpha + d\beta.$$

□

Proposition 3.2.3 (d and wedge product)**.** Let $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^l(U)$. Then

$$d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^k \alpha \wedge d\beta.$$

Proof. Since d is linear (c.f. Proposition 3.2.2), it suffices to consider the case where α and β consist of a single term, rather than a sum of terms, i.e.

$$\alpha = f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$
$$\beta = g\,dx^{j_1} \wedge \cdots \wedge dx^{j_l}.$$

Then

$$\alpha \wedge \beta = fg\,dx^{i_1} \wedge \cdots \wedge dx^{i_k} \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_l},$$

so that

$$d(\alpha\wedge\beta) = d\left(fg\,dx^{i_1} \wedge \cdots \wedge dx^{i_k} \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_l}\right) = d(fg)\wedge dx^{i_1} \wedge\cdots\wedge dx^{i_k} \wedge dx^{j_1} \wedge\cdots\wedge dx^{j_l}.$$

But

$$d(fg) = \frac{\partial}{\partial x^j}(fg)\,dx^j = \left(f\frac{\partial g}{\partial x^j} + g\frac{\partial f}{\partial x^j}\right)\,dx^j = f\,dg + g\,df.$$

Therefore,

$$d(\alpha \wedge \beta) = (g\,df + f\,dg) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_l}.$$

We have that

$$g\,df \wedge dx^{i_1} \wedge\cdots\wedge dx^{i_k} \wedge dx^{j_1} \wedge\cdots\wedge dx^{j_l} = \left(df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right)\wedge\left(g\,dx^{j_1} \wedge \cdots \wedge dx^{j_l}\right) = d\alpha\wedge\beta.$$

On the other hand, since dg is a 1-form, we have that

$$dg \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} = (-1)^k dx^{i_1} \wedge \cdots \wedge dx^{i_k} \wedge dg.$$

Therefore,

$$f\,dg\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_k}\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l} = (-1)^k\left(f\,dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right)\wedge(dg\wedge dx^{j_1}\wedge\cdots\wedge dx^{j_l}) = (-1)^k\alpha\wedge d\beta.$$

Combining the preceding results, we obtain

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta,$$

as required. □

Proposition 3.2.4 ($d^2 = 0$)**.**

For all $\omega \in \Omega^k(U)$,

$$d^2\omega = 0.$$

Proof. First, we show $d^2 = 0$ for 0-forms. Let $f \in \Omega^0(U)$. From Definition 3.2.1 for d applied to a function, we have that

$$df = \frac{\partial f}{\partial x^j}\, dx^j.$$

Then from Definition 3.2.1 for d of a one-form, we have that

$$d(df) = d\left(\frac{\partial f}{\partial x^j}\right) \wedge dx^j.$$

But

$$d\left(\frac{\partial f}{\partial x^j}\right) = \frac{\partial^2 f}{\partial x^j \partial x^i}\, dx^i.$$

Then

$$d(df) = \frac{\partial^2 f}{\partial x^j \partial x^i}\, dx^i \wedge dx^j.$$

This expression vanishes, since the first factor is symmetric in i and j (by the equality of mixed partials) while the second factor is antisymmetric (since $dx^i \wedge dx^j = -dx^j \wedge dx^i$).

Let us show this last fact explicitly, writing in the summations over i and j. By the equality of mixed partials (Proposition 1.3.16), we may write that

$$\begin{aligned}
d(df) &= \sum_{i,j=1}^{n} \tfrac{1}{2}\left(\frac{\partial^2 f}{\partial x^j \partial x^i} + \frac{\partial^2 f}{\partial x^i \partial x^j}\right) dx^i \wedge dx^j \\
&= \sum_{i,j=1}^{n} \tfrac{1}{2}\frac{\partial^2 f}{\partial x^i \partial x^j}\, dx^i \wedge dx^j + \sum_{i,j=1}^{n} \tfrac{1}{2}\frac{\partial^2 f}{\partial x^j \partial x^i}\, dx^i \wedge dx^j \\
&= \sum_{i,j=1}^{n} \tfrac{1}{2}\frac{\partial^2 f}{\partial x^i \partial x^j}\, dx^i \wedge dx^j + \sum_{i,j=1}^{n} \tfrac{1}{2}\frac{\partial^2 f}{\partial x^i \partial x^j}\, dx^j \wedge dx^i \quad \text{(swapping } i \text{ and } j \text{ in the second sum)} \\
&= \sum_{i,j=1}^{n} \tfrac{1}{2}\frac{\partial^2 f}{\partial x^i \partial x^j}\left(dx^i \wedge dx^j + dx^j \wedge dx^i\right) = 0.
\end{aligned}$$

Next, we show that $d^2 = 0$ for k-forms. Since d is linear, it suffices to consider k-forms of the form

$$\alpha = f\, dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$

where f is a smooth function. Then

$$d\alpha = df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

From Proposition 3.2.3,

$$d(d\alpha) = (d(df)) \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} - df \wedge d\left(dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right).$$

Since we have already shown that $d(df) = 0$, it follows that the first term vanishes. The second term also vanishes, since

$$d\left(dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) = d\left(1 \cdot dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) = 0,$$

by Definition 3.2.1 (the coefficient of the basis k-form is the constant function $f = 1$, whose exterior derivative $d(1)$ vanishes). $\square$

Example 3.2.5 (Exterior derivative on $\mathbb{R}^3$)**.**

- 0-forms. Under the correspondence between scalar and vector fields on $\mathbb{R}^3$, on the one hand, and differential forms on $\mathbb{R}^3$ on the other (as described in Section 2.9), a 0-form f corresponds to a function, and the 1-form df corresponds to a vector field. We have that

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

Thus, df corresponds to the *gradient* of f.

- 1-forms. Under the correspondence between scalar and vector fields and differential forms on $\mathbb{R}^3$, a 1-form α on $\mathbb{R}^3$, given by

$$\alpha = \alpha_1\, dx + \alpha_2\, dy + \alpha_3\, dz,$$

corresponds to a vector field $\mathbf{A} = (A_1, A_2, A_3)$ with components given by $A_j = \alpha_j$. The 2-form $d\alpha$ also corresponds to a vector field. We have that

$$\begin{aligned} d\alpha &= d\alpha_1 \wedge dx + d\alpha_2 \wedge dy + d\alpha_3 \wedge\, dz \\ &= \left(\frac{\partial \alpha_1}{\partial x}\, dx + \frac{\partial \alpha_1}{\partial y}\, dy + \frac{\partial \alpha_1}{\partial z}\, dz\right) \wedge dx + \left(\frac{\partial \alpha_2}{\partial x}\, dx + \frac{\partial \alpha_2}{\partial y}\, dy + \frac{\partial \alpha_2}{\partial z}\, dz\right) \wedge dy \\ &\quad + \left(\frac{\partial \alpha_3}{\partial x}\, dx + \frac{\partial \alpha_3}{\partial y}\, dy + \frac{\partial \alpha_3}{\partial z}\, dz\right) \wedge dz \\ &= \left(\frac{\partial \alpha_2}{\partial x} - \frac{\partial \alpha_1}{\partial y}\right) dx \wedge dy + \left(\frac{\partial \alpha_1}{\partial z} - \frac{\partial \alpha_3}{\partial x}\right) dz \wedge dx + \left(\frac{\partial \alpha_3}{\partial y} - \frac{\partial \alpha_2}{\partial z}\right) dy \wedge dz. \end{aligned}$$

Thus, $d\alpha$ corresponds to a vector field $\mathbf{B}$ with components

$$\begin{aligned} B_1 &= \frac{\partial \alpha_3}{\partial y} - \frac{\partial \alpha_2}{\partial z}, \\ B_2 &= \frac{\partial \alpha_1}{\partial z} - \frac{\partial \alpha_3}{\partial x}, \\ B_3 &= \frac{\partial \alpha_2}{\partial x} - \frac{\partial \alpha_1}{\partial y}. \end{aligned}$$

$\mathbf{B}$ may be recognised as the curl of $\mathbf{A}$. Thus, $d\alpha$ corresponds to the *curl* of $\mathbf{A}$.

- 2-forms. Under the correspondence between scalar and vector fields and differential forms on $\mathbb{R}^3$, a 2-form β on $\mathbb{R}^3$, given by

$$\beta = \beta_{12}\, dx \wedge dy + \beta_{23}\, dy \wedge dz + \beta_{31}\, dz \wedge dx,$$

corresponds to a vector field $\mathbf{B} = (B_1, B_2, B_3)$ with components given by

$$\begin{aligned} B_1 &= \beta_{23}, \\ B_2 &= \beta_{31}, \\ B_3 &= \beta_{12}. \end{aligned}$$

The 3-form $d\beta$ corresponds to a scalar field. We have that

$$\begin{aligned} d\beta &= d\beta_{12} \wedge dx \wedge dy + d\beta_{23} \wedge dy \wedge dz + d\beta_{31} \wedge dz \wedge dx \\ &= \left(\frac{\partial \beta_{12}}{\partial x}\, dx + \frac{\partial \beta_{12}}{\partial y}\, dy + \frac{\partial \beta_{12}}{\partial z}\, dz\right) \wedge dx \wedge dy + \left(\frac{\partial \beta_{23}}{\partial x}\, dx + \frac{\partial \beta_{23}}{\partial y}\, dy + \frac{\partial \beta_{23}}{\partial z}\, dz\right) \wedge dy \wedge dz \\ &\quad + \left(\frac{\partial \beta_{31}}{\partial x}\, dx + \frac{\partial \beta_{31}}{\partial y}\, dy + \frac{\partial \beta_{31}}{\partial z}\, dz\right) \wedge dz \wedge dx \\ &= \frac{\partial \beta_{12}}{\partial z}\, dz \wedge dx \wedge dy + \frac{\partial \beta_{23}}{\partial x}\, dx \wedge dy \wedge dz + \frac{\partial \beta_{31}}{\partial y}\, dy \wedge dz \wedge dx \\ &= \left(\frac{\partial \beta_{12}}{\partial z} + \frac{\partial \beta_{23}}{\partial x} + \frac{\partial \beta_{31}}{\partial y}\right) dx \wedge dy \wedge dz. \end{aligned}$$

The 3-form $d\beta$ corresponds to the scalar field

$$\frac{\partial \beta_{23}}{\partial x} + \frac{\partial \beta_{31}}{\partial y} + \frac{\partial \beta_{12}}{\partial z},$$

which corresponds to

$$\frac{\partial B_1}{\partial x} + \frac{\partial B_2}{\partial y} + \frac{\partial B_3}{\partial z} = \boldsymbol{\nabla} \cdot \mathbf{B}.$$

Thus, $d\beta$ corresponds to the divergence of $\mathbf{B}$.

As discussed in Problem Sheet 8, the fact that $d^2 = 0$ corresponds to the identities $\nabla \times (\nabla f) = 0$ and $\nabla \cdot (\nabla \times \mathbf{A}) = 0$.

Example 3.2.6. More on contraction.

We give some further examples of the contraction map, using the new notation in which basis one-forms are denoted by dx^i instead of $f^{(i)}$.

a) $i_{\mathbb{X}} dx^i = dx^i(\mathbb{X}) = \mathbb{X}^i$

b) $i_{\mathbb{X}}(dx^i \wedge dx^j) = \mathbb{X}^i dx^j - \mathbb{X}^j dx^i$. This makes use of the second property in Proposition 2.7.2 concerning the contraction of a wedge product $\alpha \wedge \beta$ in the case that both α and β are 1-forms.

c) $i_{\mathbb{X}}(dx^i \wedge dx^j \wedge dx^k) = \mathbb{X}^i dx^j \wedge dx^k - \mathbb{X}^j dx^i \wedge dx^k + \mathbb{X}^k dx^i \wedge dx^j$. This makes use of Corollary 2.7.3 concerning the contraction of a wedge product of forms, where each form is of degree 1.

d) $i_{\mathbb{X}} \left(dx^{j_1} \wedge \cdots \wedge dx^{j_k} \right) = \sum_{r=1}^{k} (-1)^{r-1} \mathbb{X}^{j_r} dx^{j_1} \wedge \cdots \wedge \widehat{dx^{j_r}} \wedge \cdots \wedge dx^{j_k}$, where $\widehat{dx^{j_r}}$ indicates that the factor dx^{j_r} is to be omitted. This makes use of Corollary 2.7.3 concerning the contraction of a wedge product of forms, where each form is of degree 1.

3.3 The Pullback

Definition 3.3.1 (Pullback)**.** Let $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$ be open, and let $F : U \to V$ be a smooth map. The **pullback**, denoted F^*, is a map $F^* : \Omega^k(V) \to \Omega^k(U)$ – that is, the pullback maps differential forms on V back to differential forms on U. Given $\beta \in \Omega^k(V)$, $F^*\beta$ is defined as follows. We note that as $F^*\beta$ is a differential k-form on U, $F^*\beta(x)$ is an algebraic k-form on $\mathbb{R}^m$, which may defined by specifying its value when applied to k arbitrary vectors in $\mathbb{R}^m$. Denoting these vectors by $u_{(1)}, \ldots, u_{(k)}$, the definition is given by

$$(F^*\beta)(x; \mathbf{u_{(1)}}, \ldots, \mathbf{u_{(n)}}) = \beta(F(x); F'(x)\mathbf{u_{(1)}}, \ldots, F'(x)\mathbf{u_{(n)}}).$$

For 0-forms, i.e. functions, $F^*f = f \circ F$, in accord with Definition 1.9.1.

Exercise 3.3.2.

i) Show that F^* is well defined, i.e. $F^*\beta \in \Omega^k(U)$.

ii) For $\beta, \gamma \in \Omega^k(V)$, show that $F^*(\beta + \gamma) = F^*\beta + F^*\gamma$.

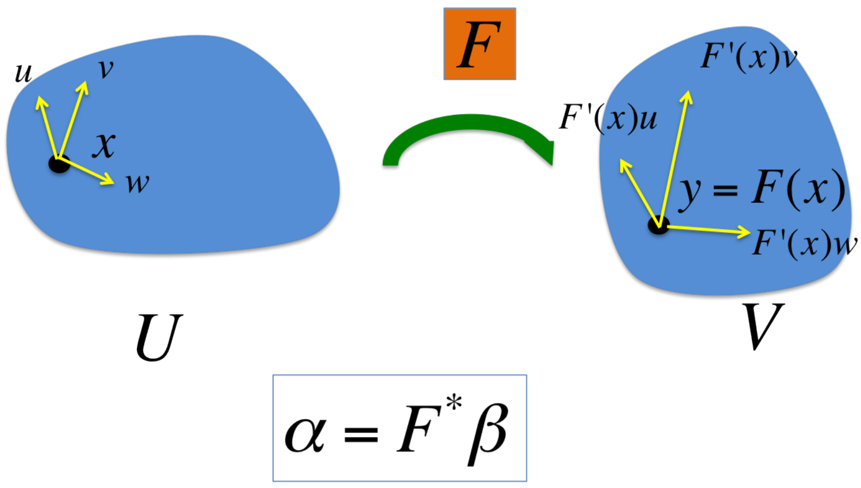


Figure 24: $\alpha(x)$ is an algebraic 3-form, which is to be evaluated on three vectors, u, v and w, to produce a number. This number is given by evaluating $\beta(F(x))$, an algebraic 3-form, on the three vectors $F'(x)u$, $F'(x)v$ and $F'(x)w$.

Example 3.3.3 (Component formula for pullback of one-forms)**.** Let $\beta \in \Omega^1(V)$. Then

$$\beta(y) = \beta_j(y)\, dy^j.$$

Let $\alpha = F^*\beta$. We may write that

$$\alpha(x) = \alpha_i(x)\, dx^i.$$

We want to express the α_i's in terms of the β_j's.

To proceed, we recall the definition of $\alpha_i(x)$, namely

$$\alpha_i(x) = \alpha(x; e_{(i)}).$$

That is, $\alpha(x)$ is an algebraic one-form on $\mathbb{R}^n$, and its ith component is obtained by applying it to the basis vector $e_{(i)}$. From Definition 3.3.1 for the pullback,

$$\alpha(x; e_{(i)}) = F^*\beta(x; e_{(i)}) = \beta(F(x); F'(x)\, e_{(i)}) = \beta_j(F(x))\big(F'(x)e_{(i)}\big)^j.$$

In general,

$$\big(F'(x)\, v\big)^j = \frac{\partial F^j}{\partial x^k}(x)v^k.$$

Therefore,

$$\big(F'(x)e_{(i)}\big)^j = \frac{\partial F^j}{\partial x^k}(x){e_{(i)}}^k = \frac{\partial F^j}{\partial x^k}(x)\delta_i^k = \frac{\partial F^j}{\partial x^i}(x).$$

Substituting above, we get that

$$\alpha_i(x) = \frac{\partial F^j}{\partial x^i}(x)\beta_j(F(x)),$$

or

$$\alpha = F^*\beta = \left(\frac{\partial F^j}{\partial x^i}\beta_j \circ F\right)\, dx^i.$$

In particular,

$$F^* dy^j = \frac{\partial F^j}{\partial x^i}\, dx^i. \tag{40}$$

Proposition 3.3.4 (Pullback of wedge product)**.**

$$F^*(\beta \wedge \gamma) = F^*\beta \wedge F^*\gamma.$$

Proof. This follows directly from the definition of the wedge product, Eq. (32), and Definition 3.3.1 of the pullback.

(*) For completeness, we'll demonstrate this explictly. Let β be a k-form and γ an l-form on $V \subset \mathbb{R}^n$. Let $\mathbf{u_{(1)}}, \dots, \mathbf{u_{(k+l)}}$ be $k+l$ vectors in $\mathbb{R}^m$. From the definition of the pullback, Definition 3.3.1, we have that

$$(F^*(\beta \wedge \gamma))(x; \mathbf{u_{(1)}}, \dots, \mathbf{u_{(k+l)}}) = (\beta \wedge \gamma)(F(x); F'(x)\mathbf{u_{(1)}}, \dots, F'(x)\mathbf{u_{(k+l)}}).$$

From the definition of the wedge product, Definition 2.5.1, we get that

$$\begin{aligned}
&(F^*(\beta \wedge \gamma))(x; \mathbf{u_{(1)}}, \dots, \mathbf{u_{(k+l)}}) \\
&= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, \beta(F(x); F'(x)\mathbf{u_{((\sigma(1)))}}, \dots, F'(x)\mathbf{u_{((\sigma(k)))}})\gamma(F(x); F'(x)\mathbf{u_{((\sigma(k+1)))}}, \dots, F'(x)\mathbf{u_{((\sigma(k+l)))}}).
\end{aligned}$$

Referring again to the definition of the pullback, we have that

$$\begin{aligned}
\beta(F(x); F'(x)\mathbf{u_{((\sigma(1)))}}, \dots, F'(x)\mathbf{u_{((\sigma(k)))}}) &= (F^*\beta)(x; \mathbf{u_{(\sigma(1))}}, \dots, \mathbf{u_{(\sigma(k))}}), \\
\gamma(F(x); F'(x)\mathbf{u_{((\sigma(k+1)))}}, \dots, F'(x)\mathbf{u_{((\sigma(k+l)))}}) &= (F^*\gamma)(x; \mathbf{u_{(\sigma(k+1))}}, \dots, \mathbf{u_{(\sigma(k+l))}}).
\end{aligned}$$

Combining these results, we get that

$$\begin{aligned}
&(F^*(\beta \wedge \gamma))(x; \mathbf{u_{(1)}}, \dots, \mathbf{u_{(n)}}) \\
&= \frac{1}{k!\,l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}\sigma\, (F^*\beta)(x; \mathbf{u_{(\sigma(1))}}, \dots, \mathbf{u_{(\sigma(k))}})(F^*\gamma)(x; \mathbf{u_{(\sigma(k+1))}}, \dots, \mathbf{u_{(\sigma(k+l))}}) \\
&= (F^*\beta \wedge F^*\gamma)(x; \mathbf{u_{(1)}}, \dots, \mathbf{u_{(n)}}),
\end{aligned}$$

as required. □

Example 3.3.5 (Component formula for pullback of k-forms)**.** Let $\beta \in \Omega^k(V)$. We may write that

$$\beta = \frac{1}{k!}\beta_{j_1 \dots j_k} dy^{j_1} \wedge \dots \wedge dy^{j_k}$$

Then $\alpha = F^*\beta$ is given by

$$\alpha = \frac{1}{k!}\frac{\partial F^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial F^{j_k}}{\partial x^{i_k}} (\beta_{j_1 \dots j_k} \circ F)\; dx^{i_1} \wedge \dots \wedge dx^{i_k}.$$

Remark 3.3.6. A k-form can be expressed as a sum of terms of the form $f\, dx^{i_1} \wedge \dots \wedge dx^{i_k}$, where f is a function. Moreover, the basis k-form $dx^{i_1} \wedge \dots \wedge dx^{i_k}$ can be written as $d\alpha$, where α is the $(k-1)$-form given by

$$\alpha = x^{i_1} dx^{i_2} \wedge \dots \wedge dx^{i_k}.$$

Therefore, a k-form can be written as a sum of terms of the form $f d\alpha$, where α is a $(k-1)$-form. This observation will be the basis for a number of proofs to follow involving k-forms, where we will use induction on k.

Proposition 3.3.7 (Pullback and d)**.** Let $\beta \in \Omega^k(V)$. Then

$$F^*d\beta = dF^*\beta.$$

Proof. We proceed by induction on k.

First, we show this is true for 0-forms, i.e. functions. Let $g \in \Omega^0(V) = C^\infty(V)$. We compute $F^* dg$ as follows. We have that dg is given by

$$dg = \frac{\partial g}{\partial y^j}\, dy^j.$$

Then from Example 3.3.3,

$$F^* dg(x) = \frac{\partial F^j}{\partial x^i}(x) \frac{\partial g}{\partial y^j}(F(x))\, dx^i.$$

Next, we compute $d(F^* g)$, as follows:

$$d(F^* g)(x) = \frac{\partial}{\partial x^i} g(F(x))\, dx^i = \frac{\partial g}{\partial y^j}(F(x)) \frac{\partial F^j}{\partial x^i}(x)\, dx^i.$$

Thus,

$$F^* dg = dF^* g. \tag{41}$$

Next, we assume the result is true for $(k-1)$-forms. Let β be a k-form on V. By Remark 3.3.6, it suffices to take $\beta = g\, d\gamma$, where γ is a $(k-1)$-form on V (note that the pullback is linear, by Exercise 3.3.2). We have that

$$\begin{aligned} F^* d(g\, d\gamma) &= F^*(dg \wedge d\gamma) \text{ (since } d^2\gamma = 0) \ = (F^* dg) \wedge (F^* d\gamma) \text{ (by Proposition 3.3.4)} \\ &= d(F^* g) \wedge dF^* \gamma \text{ (by induction hypothesis)} \ = d(F^* g\, dF^* \gamma) \text{ (by Proposition 3.2.3 and since } d^2 F^* \gamma = 0) \\ &= d(F^* g\, F^* d\gamma) \text{ (by induction hypothesis)} \ = dF^*(g\, d\gamma) \text{ (by Proposition 3.3.4)} . \end{aligned}$$

□

Example 3.3.8 (Polar coordinates on $\mathbb{R}^2$)**.** Let $U = \{(r, \theta)\} = \mathbb{R}^2$, $V = \{(x, y)\} = \mathbb{R}^2$. Let $F : U \to V$ be given by

$$F(r, \theta) = (r \cos\theta, r \sin\theta).$$

Then $F^* : \Omega^k(V) \to \Omega^k(U)$ maps Cartesian-coordinate forms to polar-coordinate forms as follows:

- 0-forms.
  Let $g(x, y)$ be a function on V. Let $f = F^* g$. Then

  $$f(r, \theta) = g(F(r, \theta)) = g(r \cos\theta, r \sin\theta).$$

- 1-forms.
  Let

  $$\beta(x, y) = \beta_1(x, y)\, dx + \beta_2(x, y)\, dy.$$

  First, we'll compute $F^* dx$ and $F^* dy$. We could use the formulas in Example 3.3.3, but here is an easier way:

  $$F^* dx = dF^* x = d(r \cos\theta) = \cos\theta\, dr - r \sin\theta\, d\theta,$$

  where we have used Proposition 3.3.7. Similarly,

  $$F^* dy = dF^* y = d(r \sin\theta) = \sin\theta\, dr + r \cos\theta\, d\theta.$$

  Then, letting $\alpha = F^* \beta$, we have that

  $$\begin{aligned} \alpha &= \beta_1(r\cos\theta, r\sin\theta)(\cos\theta\, dr - r\sin\theta\, d\theta) + \beta_2(r\cos\theta, r\sin\theta)(\sin\theta\, dr + r\cos\theta\, d\theta) \\ &= (\cos\theta \beta_1(r\cos\theta, r\sin\theta) + \sin\theta \beta_2(r\cos\theta, r\sin\theta))\, dr + \\ &\quad (r\cos\theta \beta_2(r\cos\theta, r\sin\theta) - r\sin\theta \beta_1(r\cos\theta, r\sin\theta))\, d\theta \end{aligned}$$

- 2-forms.
  A general two-form on V is given by
  $$\gamma(x,y) = c(x,y)\,dx \wedge dy.$$
  First we pull back $dx \wedge dy$. Using Proposition 3.3.4, $F^*(dx \wedge dy) = (F^*dx) \wedge F^*dy$, and we have already computed F^*dx and F^*dy above. Therefore,
  $$\begin{aligned} F^*(dx \wedge dy) &= (F^*dx) \wedge F^*dy = (\cos\theta\,dr - r\sin\theta\,d\theta) \wedge (\sin\theta\,dr + r\cos\theta\,d\theta) \\ &= r\cos^2\theta\,dr \wedge d\theta - r\sin^2\theta\,d\theta \wedge dr = (r\cos^2\theta + r\sin^2\theta)\,dr \wedge d\theta = r\,dr \wedge d\theta. \end{aligned}$$
  You might recognise this as the transformation of the area element from Cartesian to polar coordinates – indeed, this is no coincidence, as we shall see later. Finally, we have that
  $$F^*\gamma(r,\theta) = c(r\cos\theta, r\sin\theta)r\,dr \wedge d\theta.$$

Proposition 3.3.9 (Pullback and composition)**.** Let $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ and $W \subset \mathbb{R}^p$ be open. Let $F : U \to V$ and $G : V \to W$ be smooth maps. Then
$$(G \circ F)^* = F^*G^*.$$

Proof. Proceed by induction on k. For 0-forms, i.e. functions, the result is clear, and was shown in Proposition 1.9.2. Suppose it holds for $(k-1)$-forms. By Remark 3.3.6 it suffices to consider k-forms of the form $f d\alpha$, where f is a function and α is a $(k-1)$-form. Then
$$\begin{aligned} &(G \circ F)^* f\,d\alpha = (G \circ F)^* f\,(G \circ F)^* d\alpha \text{ (by Proposition 3.3.4)} \\ &= ((G\circ F)^* f)\,d(G\circ F)^*\alpha \text{ (by Proposition 3.3.7)} = (F^*G^*f)\,dF^*G^*\alpha \text{ (by induction hypothesis)} \\ &= F^*G^*f\,F^*G^*d\alpha \text{ (by Proposition 3.3.7)} = F^*G^*(f\,d\alpha) \text{ (by Proposition 3.3.4) .} \end{aligned}$$

□

3.4 The Lie derivative

Definition 3.4.1 (Lie derivative of forms)**.** Let $U \subset \mathbb{R}^n$ be open, and let $\mathbb{X} \in \mathcal{X}(U)$ be a smooth vector field on U with flow Φ_t. Let $\omega \in \Omega^k(U)$ be a differential k-form. The **Lie derivative** of ω with respect to $\mathbb{X}$, denoted $L_\mathbb{X}\omega$, is given by
$$L_\mathbb{X}\omega = \left.\frac{\partial}{\partial t}\right|_{t=0} \Phi_t^*\omega.$$

Proposition 3.4.2 (Simple properties of the Lie derivative forms)**.**

i) $L_\mathbb{X}(\alpha + \beta) = L_\mathbb{X}\alpha + L_\mathbb{X}\beta$

ii) $L_\mathbb{X}(\alpha \wedge \beta) = (L_\mathbb{X}\alpha) \wedge \beta + \alpha \wedge L_\mathbb{X}\beta$

iii) $L_\mathbb{X}d = dL_\mathbb{X}$.

Proof. These properties follow directly from corresponding properties of the pullback.

i) We have that

$$L_{\mathbb{X}}(\alpha+\beta) = \left.\frac{\partial}{\partial t}\right|_{t=0} \Phi_t^*\left(\alpha+\beta\right).$$

Since the pullback is linear (Exercise 3.3.2) it follows that

$$\Phi_t^*\left(\alpha+\beta\right) = \Phi_t^*\alpha + \Phi_t^*\beta.$$

Then

$$L_{\mathbb{X}}(\alpha+\beta) = \left.\frac{\partial}{\partial t}\right|_{t=0} \left(\Phi_t^*\alpha + \Phi_t^*\beta\right) = L_{\mathbb{X}}\alpha + L_{\mathbb{X}}\beta.$$

ii) From Proposition 3.3.4, we have that

$$\Phi_t^*(\alpha\wedge\beta) = (\Phi_t^*\alpha)\wedge(\Phi_t^*\beta).$$

From the product rule, it follows that

$$\begin{aligned}\left.\frac{\partial}{\partial t}\right|_{t=0} \Phi_t^*(\alpha\wedge\beta) &= \left.\frac{\partial}{\partial t}\right|_{t=0} (\Phi_t^*\alpha)\wedge(\Phi_t^*\beta)\\ &= \left.\left(\frac{\partial}{\partial t}\Phi_t^*\alpha\right)\right|_{t=0} \wedge (\Phi_t^*\beta)|_{t=0} + (\Phi_t^*\beta)|_{t=0}\wedge \left.\left(\frac{\partial}{\partial t}\Phi_t^*\alpha\right)\right|_{t=0}\\ &= (L_{\mathbb{X}}\alpha)\wedge\beta + \alpha\wedge(L_{\mathbb{X}}\beta).\end{aligned}$$

iii) From Proposition 3.3.7, we have that

$$d\Phi_t^*\omega = \Phi_t^* d\omega.$$

It follows that

$$L_{\mathbb{X}}d\omega = \left.\frac{\partial}{\partial t}\right|_{t=0} \Phi_t^* d\omega = \left.\frac{\partial}{\partial t}\right|_{t=0} d\Phi_t^*\omega = d\left.\frac{\partial}{\partial t}\right|_{t=0} \Phi_t^*\omega = dL_{\mathbb{X}}\omega.$$

□

Proposition 3.4.3 (Homotopy formula)**.** For $\omega\in\Omega^k(U)$,

$$L_{\mathbb{X}}\omega = i_{\mathbb{X}}d\omega + di_{\mathbb{X}}\omega.$$

Proof. We proceed by induction. First, we show that the formula holds for 0-forms. For a 0-form f, the left-hand side of the homotopy formula gives

$$LHS := L_{\mathbb{X}}f = \left.\frac{\partial}{\partial t}\right|_{t=0} \Phi_t^* f = \left.\frac{\partial}{\partial t}\right|_{t=0} f\circ\Phi_t = \mathbb{X}\cdot\boldsymbol{\nabla} f = \mathbb{X}^i\frac{\partial f}{\partial x^i},$$

in accord with Definition 1.9.3. On the right-hand side, we have that

$$RHS := i_{\mathbb{X}}df + di_{\mathbb{X}}f = i_{\mathbb{X}}df = \mathbb{X}^i\frac{\partial f}{\partial x^i},$$

since $i_{\mathbb{X}}f = 0$. Thus, LHS and RHS coincide, and the homotopy formula holds for 0-forms.

Next, we assume that the homotopy formula holds for $(k-1)$-forms, and we show that it holds for k-forms. Since all the operations we are considering – namely $L_{\mathbb{X}}$, $i_{\mathbb{X}}$ and d – are linear, by Remark 3.3.6 it suffices to take

$$\omega = f\,d\alpha,$$

where α is a $(k-1)$-form.

Applied to $\omega = f\,d\alpha$, the left-hand side of the homotopy formula gives

$$LHS := L_{\mathbb{X}}(f\,d\alpha) = (L_{\mathbb{X}}f)\,d\alpha + f L_{\mathbb{X}}\,d\alpha,$$

by Proposition 3.4.2 ii). From the preceding,

$$L_{\mathbb{X}}f = i_{\mathbb{X}}df,$$

while from Proposition 3.4.2 iii),

$$L_{\mathbb{X}}d\alpha = dL_{\mathbb{X}}\alpha.$$

By the induction hypothesis, the homotopy formula applies to α, since α is a $(k-1)$-form. Therefore,

$$L_{\mathbb{X}}\alpha = i_{\mathbb{X}}d\alpha + di_{\mathbb{X}}\alpha,$$

so that

$$d\,L_{\mathbb{X}}\alpha = di_{\mathbb{X}}d\alpha,$$

since $d^2 = 0$ by Proposition 3.2.4. Combining the previous results, we get that

$$LHS = (L_{\mathbb{X}}f)\,d\alpha + f L_{\mathbb{X}}\,d\alpha = (i_{\mathbb{X}}df)\,d\alpha + f\,di_{\mathbb{X}}d\alpha.$$

The right-hand side of the homotopy formula applied to $\omega = f\,d\alpha$ gives

$$RHS := i_{\mathbb{X}}d(f\,d\alpha) + di_{\mathbb{X}}(f\,d\alpha).$$

Let us consider the first term on the right-hand side of the preceding. Since

$$d(f\,d\alpha) = df \wedge d\alpha + f d(d\alpha) = df \wedge d\alpha,$$

it follows from (39) that

$$i_{\mathbb{X}}d(f\,d\alpha) = i_{\mathbb{X}}(df \wedge d\alpha) = (i_{\mathbb{X}}df)d\alpha - df \wedge (i_{\mathbb{X}}d\alpha).$$

Next, we consider the second term on the right-hand side of the preceding. We have that

$$di_{\mathbb{X}}(f\,d\alpha) = d\,(f i_{\mathbb{X}}d\alpha) = df \wedge (i_{\mathbb{X}}d\alpha) + f di_{\mathbb{X}}d\alpha.$$

Combining the preceding expressions, we get

$$RHS := (i_{\mathbb{X}}df)d\alpha - df \wedge (i_{\mathbb{X}}d\alpha) + df \wedge (i_{\mathbb{X}}d\alpha) + f\,di_{\mathbb{X}}d\alpha = (i_{\mathbb{X}}df)d\alpha + f\,di_{\mathbb{X}}d\alpha.$$

Thus, LHS and RHS coincide, and the homotopy formula holds for k-forms. □

Example 3.4.4. Let β be the 2 -form on $\mathbb{R}^3$ given by

$$\beta = -x\,dx \wedge dy + y\,dy \wedge dz,$$

and let $\mathbb{X}$ be the vector field on $\mathbb{R}^3$ given by

$$X = (y, 0, z).$$

We compute $L_{\mathbb{X}}\beta$ using the homotopy formula. We have that $d\beta = 0$, so that $i_{\mathbb{X}}d\beta = 0$. On the other hand,

$$i_{\mathbb{X}}\beta = -x\,i_{\mathbb{X}}(dx \wedge dy) + y\,i_{\mathbb{X}}(dy \wedge dz) = -xy\,dy - yz\,dy = -y(x+z)\,dy.$$

Then

$$L_{\mathbb{X}}\beta = d\,i_{\mathbb{X}}\beta = d(-y(x+z)\,dy) = -y(dx \wedge dy + dz \wedge dy) = -y\,dx \wedge dy + y\,dy \wedge dz.$$

3.5 The Poincaré Lemma

From Proposition 3.2.4, we know that $d(d\alpha) = 0$ for any differential form α. This fact motivates the following question: Suppose ω is a k-form such that $d\omega = 0$. Can we find a differential $(k-1)$-form α such that $\omega = d\alpha$? Two specific examples of this question occur in vector calculus: i) if $\boldsymbol{\nabla} \times \mathbf{F} = 0$, can we write $\mathbf{F} = \boldsymbol{\nabla}\phi$ for some scalar function ϕ, and ii) if $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$, can we write $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$ for some vector field $\mathbf{A}$?

Definition 3.5.1 (Closed and exact forms.)**.** A differential k-form ω is **closed** if $d\omega = 0$. ω is **exact** if $\omega = d\alpha$ for some $(k-1)$-form α.

In this language, Proposition 3.2.4 says that every exact form is closed. The question we are asking is whether every closed form is exact.

The Poincaré Lemma is a partial answer to this question. It provides a sufficient condition for every closed form to be exact. It turns out that this condition is related to the topology of the space on which the forms are defined. Examples of closed forms which are not exact occur in spaces which are, in a sense that can be made precise, topologically nontrivial. In physics applications, these often correspond to singularities, e.g. point charges or lines of current. The general area for these questions is called differential topology. It is a means of studying the topology of manifolds through analytic methods, specifically through certain special differential forms, which are characterised by their satisfying some simple ("natural") partial differential equation.

3.5.1 Time-dependent vector fields

We will need a generalisation of vector fields and one-parameter subgroups. This generalisation corresponds to going from autonomous to non-autonomous systems of ODEs.

Definition 3.5.2 (One-parameter *family* of diffeomorphisms)**.** Let $U, V \subset \mathbb{R}^n$ be open, and let I be an open interval in $\mathbb{R}$. A **one-parameter family of diffeomorphisms** is a smooth family of maps

$$\hat{\Phi} : I \times U \to V; \quad (t, x) \mapsto \hat{\Phi}_t(x)$$

such that $\hat{\Phi}_t$ is a diffeomorphism onto its image. That is, letting $U_t = \hat{\Phi}_t(U) \subset V$, then $\hat{\Phi}_t : U \to U_t$ is a diffeomorphism.

Note that a one-parameter subgroup of diffeomorphisms, given by Definition 1.6.11, is a special case of a one-parameter family of diffeomorphisms. In particular, for a family, as opposed to a subgroup, we do *not* assume that $\hat{\Phi}_s \circ \hat{\Phi}_t = \hat{\Phi}_{s+t}$ nor that $\hat{\Phi}_0 = \mathrm{Id}_U$.

Given the one-parameter family $\hat{\Phi}_t$, we define a one-parameter family of vector fields $\hat{\mathbb{X}}_t$ on U_t by

$$\frac{\partial}{\partial t}\hat{\Phi}_t(x) = \hat{\mathbb{X}}_t(\hat{\Phi}_t(x)), \text{ or } \frac{\partial}{\partial t}\hat{\Phi}_t = \hat{\mathbb{X}}_t \circ \hat{\Phi}_t. \tag{42}$$

In the language of mechanics, $\hat{\mathbb{X}}_t(\hat{\Phi}_t(x))$ is the velocity at time t along the trajectory $x(t) = \hat{\Phi}_t(x)$. Since $\hat{\Phi}_t$ is a diffeomorphism and therefore invertible, we can rearrange (42) to obtain the following expression for $\hat{\mathbb{X}}_t$ evaluated at x (rather than at $\hat{\Phi}_t(x)$),

$$\hat{\mathbb{X}}_t(x) = \left(\frac{\partial}{\partial t}\hat{\Phi}_t\right)(\hat{\Phi}_t^{-1}(x)). \tag{43}$$

Note that as $\hat{\Phi}_t$ is not necessarily a one-parameter subgroup, we cannot assume that $\hat{\Phi}_t^{-1} = \hat{\Phi}_{-t}$.

Associated to $\hat{\mathbb{X}}_t$ is a system of first-order differential equations,

$$\dot{x}(t) = \hat{\mathbb{X}}_t(x(t)). \tag{44}$$

This system is *nonautonomous*, as $\hat{\mathbb{X}}_t$ depends explicitly on t. (As in exercises from Problem Sheet 2, we could recast (44) as an autonomous system on $U \times \mathbb{R}$, but we won't do this here.) Then $\hat{\Phi}_t(x)$ is the solution of (44).

Example 3.5.3 (Time-dependent flow)**.** Let $U = V = \mathbb{R}^n$ and $I = (0, \infty)$. Define a one-parameter family of diffeomorphisms $\hat{\Phi}_t$ for $t > 0$ by

$$\hat{\Phi}_t(x) = tx,$$

ie $\hat{\Phi}_t$ just dilates or contracts by a scalar factor t. Let's compute $\hat{\mathbb{X}}_t$ using (43). By inspection,

$$\hat{\Phi}_t^{-1}(x) = \frac{1}{t}x,$$

and

$$\frac{\partial}{\partial t}\hat{\Phi}_t(x) = x.$$

Therefore,

$$\hat{\mathbb{X}}_t(x) = \left(\frac{\partial}{\partial t}\hat{\Phi}_t\right)(\hat{\Phi}_t^{-1}(x)) = \left(\frac{\partial}{\partial t}\hat{\Phi}_t\right)\left(\frac{1}{t}x\right) = \frac{1}{t}x.$$

Given a k-form ω, we want to evaluate the t-derivative of its pullback with respect to $\hat{\Phi}_t$. A formula is given by the following:

Proposition 3.5.4 (Lie derivative with respect to time-dependent vector field)**.**

$$\frac{\partial}{\partial t}\hat{\Phi}_t^*\bigg|_{t=t_0}\omega = \hat{\Phi}_{t_0}^* L_{\hat{\mathbb{X}}_{t_0}}\omega.$$

Some comments: t_0 is just the value of t where we evaluate the derivative. Sometimes, we'll omit t_0, and write the formula more concisely as

$$\frac{\partial}{\partial t}\hat{\Phi}_t^*\omega = \hat{\Phi}_t^* L_{\hat{\mathbb{X}}_t}\omega.$$

Note that, in the case where $t_0 = 0$ and $\hat{\Phi}_t$ is the flow of a fixed vector field $\hat{\mathbb{X}}$, Proposition 3.5.4 is just the definition of the Lie derivative, ie Definition 3.4.1.

Proof. From a first-order Taylor expansion about $t = t_0$, we have that

$$\begin{aligned}\hat{\Phi}_{t_0+\epsilon} &= \hat{\Phi}_{t_0} + \epsilon\, \frac{\partial}{\partial t}\hat{\Phi}_t\bigg|_{t=t_0} + O(\epsilon^2)\\ &= \hat{\Phi}_{t_0} + \epsilon\hat{\mathbb{X}}_{t_0} \circ \hat{\Phi}_{t_0} + O(\epsilon^2)\\ &= (\mathrm{Id} + \epsilon\hat{\mathbb{X}}_{t_0}) \circ \hat{\Phi}_{t_0} + O(\epsilon^2), \end{aligned} \tag{45}$$

where we have used the definition (42) of $\hat{\mathbb{X}}_{t_0}$.

Let Ψ_s denote the flow (one-parameter *subgroup* of differmorphisms) of the fixed (i.e., t-independent) vector field $\hat{\mathbb{X}}_{t_0}$, so that

$$\hat{\mathbb{X}}_{t_0} = \frac{\partial}{\partial s}\Psi_s\bigg|_{s=0}. \tag{46}$$

Then a first-order Taylor expansion of Ψ_s about $s = 0$ gives

$$\Psi_\epsilon = \Psi_0 + \epsilon\hat{\mathbb{X}}_{t_0} + O(\epsilon^2) = (\mathrm{Id} + \epsilon\hat{\mathbb{X}}_{t_0}) + O(\epsilon^2), \tag{47}$$

since $\Psi_0 = \mathrm{Id}$. Substituting (47) into (45), we get that

$$\hat{\Phi}_{t_0+\epsilon} = \Psi_\epsilon \circ \hat{\Phi}_{t_0} + O(\epsilon^2). \tag{48}$$

From (48) and the composition property of the pullback (Proposition 3.3.9), we have that

$$\hat{\Phi}^*_{t_0+\epsilon}\omega = (\Psi_\epsilon \circ \hat{\Phi}_{t_0})^*\omega + O(\epsilon^2) = \hat{\Phi}^*_{t_0}\Psi^*_\epsilon\omega + O(\epsilon^2). \tag{49}$$

Then we compute

$$\begin{aligned}
\frac{\partial}{\partial t}\hat{\Phi}^*_t\omega\bigg|_{t=t_0} &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\hat{\Phi}^*_{t_0+\epsilon}\omega - \hat{\Phi}^*_{t_0}\omega\right) \quad \text{(definition of derivative)}\\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\hat{\Phi}^*_{t_0}\Psi^*_\epsilon\omega - \hat{\Phi}^*_{t_0}\Psi^*_0\omega + O(\epsilon^2)\right) \quad \text{(using (49) and } \Psi_0 = \mathrm{Id})\\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\hat{\Phi}^*_{t_0}\Psi^*_\epsilon\omega - \hat{\Phi}^*_{t_0}\Psi^*_0\omega\right) \quad (O(\epsilon^2) \text{ term vanishes in the limit})\\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\hat{\Phi}^*_{t_0}\left(\Psi^*_\epsilon\omega - \Psi^*_0\omega\right) \quad \text{(using linearity of pullback)}\\
&= \hat{\Phi}^*_{t_0}\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\Psi^*_\epsilon\omega - \Psi^*_0\omega\right) \quad (\text{as } \hat{\Phi}^*_{t_0} \text{ is independent of } \epsilon)\\
&= \hat{\Phi}^*_{t_0}\frac{\partial}{\partial t}\Psi^*_t\omega\bigg|_{t=0} \quad \text{definition of derivative}\\
&= \hat{\Phi}^*_{t_0}L_{\hat{\mathbb{X}}_{t_0}}\omega \quad \text{(using Definition 3.4.1)},
\end{aligned} \tag{50}$$

□

3.5.2 Poincaré Lemma

Theorem 3.5.5 (Poincaré Lemma). Let $\hat{\Phi}_t : U \to U$ be a one-parameter family of diffeomorphisms defined for $0 < t \le 1$. Let $\beta \in \Omega^k(U)$ be a closed k-form. Suppose that

$$\hat{\Phi}^*_1\beta = \beta, \quad \lim_{t\to 0}\hat{\Phi}^*_t\beta = 0.$$

Then

$$\beta = d\alpha,$$

where

$$\alpha = \int_0^1 \hat{\Phi}^*_t\left(i_{\hat{\mathbb{X}}_t}\beta\right)\,dt,$$

and $\hat{\mathbb{X}}_t$ is defined as above by

$$\frac{\partial}{\partial t}\hat{\Phi}_t(x) = \hat{\mathbb{X}}_t(\hat{\Phi}_t(x)).$$

Remark. In many applications, $\hat{\Phi}_t$ is not invertible for $t = 0$, so that $\hat{\Phi}_0$ is not a diffeomorphism (cf. Example 3.5.3). In this case, $\hat{\mathbb{X}}_0$ is not defined (cf. Eq. (43)).

Proof. From the assumptions above, we have that

$$\beta = \lim_{\epsilon\to 0}\hat{\Phi}^*_t\beta\Big|_\epsilon^1 = \lim_{\epsilon\to 0}\int_\epsilon^1\frac{\partial}{\partial t}\hat{\Phi}^*_t\beta\,dt = \int_0^1\frac{\partial}{\partial t}\hat{\Phi}^*_t\beta\,dt.$$

From Proposition 3.5.4,

$$\frac{\partial}{\partial t}\hat{\Phi}^*_t\beta = \hat{\Phi}^*_t L_{\hat{\mathbb{X}}_t}\beta.$$

From Proposition 3.4.3,

$$L_{\hat{\mathbb{X}}_t}\beta = di_{\hat{\mathbb{X}}_t}\beta + i_{\hat{\mathbb{X}}_t}d\beta.$$

But since β is closed by assumption, the second term vanishes, and we have that

$$L_{\hat{\mathbb{X}}_t}\beta = di_{\hat{\mathbb{X}}_t}\beta.$$

It follows that

$$\hat{\Phi}_t^* L_{\hat{\mathbb{X}}_t}\beta = \hat{\Phi}_t^* di_{\hat{\mathbb{X}}_t}\beta.$$

Since d commutes with the pullback (Proposition 3.3.7), we may write that

$$\hat{\Phi}_t^* L_{\hat{\mathbb{X}}_t}\beta = d\hat{\Phi}_t^* i_{\hat{\mathbb{X}}_t}\beta.$$

Substituting into the equation for β above, we get that

$$\beta = \int_0^1 d\left(\hat{\Phi}_t^* i_{\hat{\mathbb{X}}_t}\beta\right)\, dt.$$

Next, we note that d can be taken outside the t-integral. This is essentially due to the fact that d is linear; that is, $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$ – and the integral over t can be understood as a limit of sums. Or else, one can write down the explicit formula for $d\left(\hat{\Phi}_t^* i_{\hat{\mathbb{X}}_t}\beta\right)$, and verify that it may be taken outside the integral. (Below, in Proposition 4.1.6, we'll write out both of these arguments in more detail for completeness). It follows that

$$\beta = d\left(\int_0^1 \hat{\Phi}_t^*\left(i_{\hat{\mathbb{X}}_t}\beta\right)\, dt\right),$$

as required. □

Definition 3.5.6 (Contractible set)**.** An open set $U \subset \mathbb{R}^n$ is said to be *contractible* if there exists a one-parameter family of diffeomorphisms $\hat{\Phi}_t : U \to U$, with $0 < t \leq 1$ such that

$$\hat{\Phi}_1(x) = x, \quad \lim_{t\to 0}\hat{\Phi}_t(x) = x_*,$$

for some fixed $x_* \in U$. That is, $\hat{\Phi}_t$ interpolates between the identity map for $t = 1$ and, as t approaches 0, a map which maps every x in U to a single point x_*.

Clearly, for any differential form β,

$$\hat{\Phi}_1^*\beta = \beta.$$

Also, the fact that $\hat{\Phi}_0(x) = x_*$ implies that $\hat{\Phi}_0'(x) = 0$. Therefore, from Definition 3.3.1 for the pullback,

$$\lim_{\epsilon\to 0}\hat{\Phi}_\epsilon^*\beta = 0.$$

Therefore, the hypothesis of the Poincaré Lemma are satisfied. It follows that, on a contractible space, *any* closed k-form β can be expressed as $d\alpha$ for some $(k-1)$-form α. From Example 3.5.3, we see that $\mathbb{R}^n$ is contractible. Therefore, on $\mathbb{R}^n$, every closed form is exact.

On noncontractible spaces, it is no longer necessarily the case that every closed form is exact.

Example 3.5.7. Let $\rho \in \Omega^3(\mathbb{R}^3)$. Show that $\rho = d\omega$, where ω is a two-form.

We note that every three-form on $\mathbb{R}^3$ is closed, and the Poincaré Lemma may be applied, with $\Phi_t(x, y, z) = (tx, ty, tz)$, as in Example 3.5.3. Thus,

$$\rho = d\omega,$$

where

$$\omega = \int_0^1 \hat{\Phi}_t^* \left(i_{\hat{\mathbb{X}}_t} \rho \right) \, dt,$$

and $\hat{\mathbb{X}}_t$ is given as in Example 3.5.3 by

$$\hat{\mathbb{X}}_t(x, y, z) = \frac{1}{t}(x, y, z).$$

We may write that

$$\rho(x, y, z) = c(x, y, z) \, dx \wedge dy \wedge dz.$$

We have that

$$i_{\hat{\mathbb{X}}_t} \rho = c i_{\hat{\mathbb{X}}_t} \left(dx \wedge dy \wedge dz \right).$$

From the formula for the contraction with a wedge product (39), we get that

$$i_{\hat{\mathbb{X}}_t} \left(dx \wedge dy \wedge dz \right) = \left(i_{\hat{\mathbb{X}}_t} dx \right) dy \wedge dz - \left(i_{\hat{\mathbb{X}}_t} dy \right) dx \wedge dz + \left(i_{\hat{\mathbb{X}}_t} dz \right) dx \wedge dy.$$

In general, $i_{\hat{\mathbb{X}}_t} dx = dx(\hat{\mathbb{X}}_t)$ is the first component of $\hat{\mathbb{X}}_t$, $i_{\hat{\mathbb{X}}_t} dy$ is the second component of $\hat{\mathbb{X}}_t$, etc. Therefore,

$$i_{\hat{\mathbb{X}}_t} \left(dx \wedge dy \wedge dz \right) = \frac{1}{t} \left(x \, dy \wedge dz - y \, dx \wedge dz + z \, dx \wedge dy \right) = \frac{1}{t} \left(x \, dy \wedge dz + y \, dz \wedge dx + z \, dx \wedge dy \right),$$

and

$$i_{\hat{\mathbb{X}}_t} \rho = \frac{c(x, y, z)}{t} \left(x \, dy \wedge dz + y \, dz \wedge dx + z \, dx \wedge dy \right).$$

Next, we compute the pullback of $i_{\hat{\mathbb{X}}_t} \rho$ by $\hat{\Phi}_t^*$. We have that

$$\hat{\Phi}_t^* x = tx, \quad \hat{\Phi}_t^* y = ty, \quad \hat{\Phi}_t^* z = tz.$$

Therefore, since $\hat{\Phi}_t^* dx = d\hat{\Phi}_t^* x$, and similarly for dy and dz, we have that

$$\hat{\Phi}_t^* dx = t \, dx, \quad \hat{\Phi}_t^* dy = t \, dy, \quad \hat{\Phi}_t^* dz = t \, dz.$$

Therefore (since the pullback of a wedge product is the wedge product of the pullbacks), we have that

$$\hat{\Phi}_t^* \left(dx \wedge dy \right) = t^2 \, dx \wedge dy, \quad \hat{\Phi}_t^* \left(dy \wedge dz \right) = t^2 \, dy \wedge dz, \quad \hat{\Phi}_t^* \left(dz \wedge dx \right) = t^2 \, dz \wedge dx.$$

Therefore, the pullback of one of the terms in $i_{\hat{\mathbb{X}}_t} \rho$ (the first) is given by

$$\hat{\Phi}_t^* \left(\frac{c(x, y, z)}{t} x \, dy \wedge dz \right) = \frac{c(tx, ty, tz)}{t} tx \, tdy \wedge tdz = t^2 c(tx, ty, tz) x \, dy \wedge dz.$$

The two other terms are treated similarly, and we get that

$$\hat{\Phi}_t^* \left(i_{\hat{\mathbb{X}}_t} \rho \right) = t^2 c(tx, ty, tz) \left(x \, dy \wedge dz + y \, dz \wedge dx + z \, dz \wedge dx \right).$$

Then

$$\omega(x, y, z) = \left(\int_0^1 t^2 c(tx, ty, tz) \, dt \right) \left(x \, dy \wedge dz + y \, dz \wedge dx + z \, dz \wedge dx \right).$$

Under the correspondence between forms on $\mathbb{R}^3$ and scalar and vectors (cf Section 2.9 and Example 3.2.5), a differential three-form ρ corresponds to a scalar field c, and a differential two-form ω corresponds to a vector field $\mathbf{E}$. $d\omega$, which is a differential three-form, corresponds to $\boldsymbol{\nabla} \cdot \mathbf{E}$, which is a scalar field. The previous result, translated into the language of vector calculus, says that if $\mathbf{E}$ is the vector field given by

$$\mathbf{E}(x, y, z) = \left(\int_0^1 t^2 c(tx, ty, tz) \, dt \right) \mathbf{r},$$

where $\mathbf{r} = (x, y, z)$, then

$$\boldsymbol{\nabla} \cdot \mathbf{E} = c.$$

4 Integration of Differential Forms

We begin with some preliminary remarks.

Let $U \subset \mathbb{R}^n$ be an open set in $\mathbb{R}^n$. Let $\rho \in \Omega^n(U)$ be an n-form on U. ρ can be written as

$$\rho(x) = f(x)\, dx^1 \wedge \cdots dx^n,$$

where $f(x)$ is the x-dependent coefficient (a function) of the standard basis n-form. As the notation suggests, differential forms are meant to be integrated. The integral of ρ over U will be defined as follows:

$$\int_U \rho \text{ “:=” } \int_U f(x)\, dx^1 \cdots dx^n.$$

This proposed definition raises two questions:

1. Is the value of the integral dependent on the choice of coordinates? That is, suppose $F : U \to V$ is a diffeomorphism between open sets $U, V \subset \mathbb{R}^n$, which we might think of as a change of coordinates (cf Example 3.3.8). Let $\rho \subset \Omega^n(V)$. Is it the case that

$$\int_V \rho = \int_U F^* \rho \ ?$$

   The following simple example shows that the answer is “No”, in general. Let U and V be the unit squares in $\mathbb{R}^2$ given by

$$U \subset \mathbb{R}^2 = \{(x^1, x^2) \,|\, 0 \le x^1, x^2 \le 1\}, \quad V \subset \mathbb{R}^2 = \{(y^1, y^2) \,|\, 0 \le y^1, y^2 \le 1\}.$$

   Note that we are thinking of U and V as being different sets; V has coordinates y, and U has coordinates x. Let ρ be the two-form on V given by

$$\rho = dy^1 \wedge dy^2$$

   (that is, the coefficient function is constant, and is equal to 1). Then from the provisional definition above,

$$\int_V \rho = \int_V dy^1\, dy^2 = 1.$$

   Let $F : U \to V$ be given by

$$(y^1, y^2) = F(x^1, x^2) = (x^2, x^1).$$

   That is, F effectively interchanges the coordinates.

$$F^* \rho = (F^* dy^1) \wedge (F^* dy^2) = dx^2 \wedge dx^1 = -dx^1 \wedge dx^2.$$

   Then from the provisional definition,

$$\int_U F^* \rho = -1.$$

   Thus, it appears that the sign of $\int_U \rho$ can change under a change of coordinates. It will turn out that this is the only ambiguity.

2. How do we integrate k-forms on U if $k < n$?

4.1 Singular k-cubes and integration of differential forms

We will address the Question 2 above first, that is how to integrate k-forms on U. In this discussion, we let $t = (t^1, \ldots, t^k)$ denote coordinates on $\mathbb{R}^k$.

Let
$$I^k = \left\{ t \in \mathbb{R}^k \,|\, 0 \le t^i \le 1, 1 \le i \le k \right\}$$
denote the closed unit k-cube in $\mathbb{R}^k$.

Definition 4.1.1 (Singular k-cube)**.** Let $U \subset \mathbb{R}^n$ be open. A **singular k-cube** on U is a smooth map
$$c : I^k \to U.$$

(Note: since I^k is not open, we have not really defined what it means for a map on I^k to be smooth. We will say that $c : I^k \to U$ is smooth if there exists an open set $B \subset \mathbb{R}^k$ such that $I^k \subset B$ and a smooth map $\tilde{c} : B \to U$ such that $c(t) = \tilde{c}(t)$ for all $t \in I^k$. One says that $\tilde{c}$ is a *smooth extension* of c. We shall not give too much attention to this point, although in rigorous treatments it requires care.)

You can think of a singular k-cube as being a parameterisation of a region of U. Typically you can think of this region as being k-dimensional, but this need not be the case (and we haven't really defined what a k-dimensional region is). You should note that c need not be 1-1 nor onto. For example, the constant map $c(t) = x_0$, where $x_0 \in U$ is fixed, whose image is a single point, is a singular k-cube.

Let $c : I^k \to U$ be a singular k-cube on U. Let $\omega \in \Omega^k(U)$ be a k-form on U. Then $c^*\omega$ is a k-form on I^k, and we may write that
$$(c^*\omega)\,(t) = f(t)\, dt^1 \wedge \cdots \wedge dt^k,$$
for some smooth function $f(t)$ defined on I^k.

Definition 4.1.2. The **integral of a k-form** ω **over a singular k-cube** c, denoted $\int_c \omega$, is defined by
$$\int_c w := \int_{I^k} c^*\omega := \int_{I^k} f(t)\, dt^1 \cdots dt^k.$$

Example 4.1.3.
We consider the integration of a two-form over a singular two-cube on $U = \mathbb{R}^3$ (so $k = 2$ and $n = 3$). For convenience, denote coordinates on I^2 by (u, v) rather than (t^1, t^2), and coordinates on $U = \mathbb{R}^3$ by (x, y, z). Let $c : I^2 \to \mathbb{R}^3$ be given by
$$c(u, v) = (u, v, 2 - u^2 - v^2).$$

Let
$$\omega(x, y, z) = x\, dy \wedge dz.$$
We compute $\int_c \omega$ as follows: First, we need to compute $c^*\omega$. We have that
$$c^*x(u, v) = x(c(u, v)) = u, \quad c^*y(u, , v) = y(c(u, v)) = v, \quad c^*z(u, , v) = z(c(u, v)) = 2 - u^2 - v^2,$$
and
$$c^*dy = dc^*y = dv, \quad c^*dz = dc^*z = -2u\, du - 2v\, dv.$$

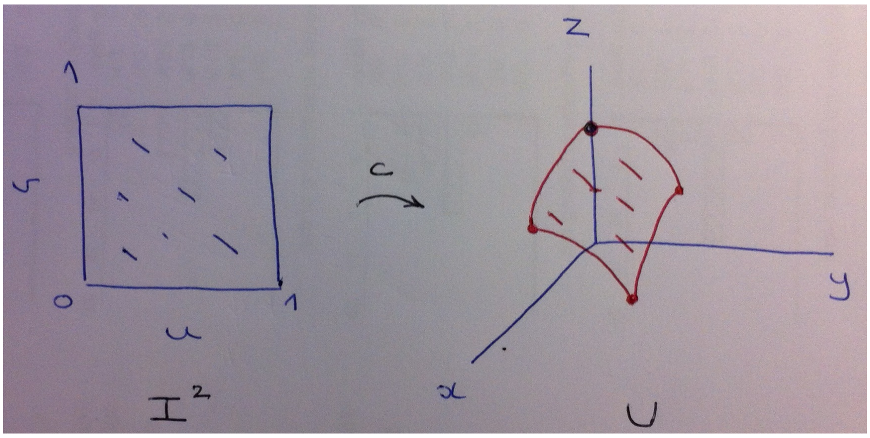


Figure 25

Then

$$c^*\omega = u\,dv \wedge (-2u\,du - 2v\,dv) = 2u^2\,du \wedge dv.$$

Now we integrate:

$$\int_c \omega = \int_{I^2} c^*\omega = \int_0^1 \int_0^1 2u^2\,du\,dv = \frac{2}{3}.$$

Next, we address the first question in the preamble above: to what extent does the integral of a k-form depend on the choice of coordinates? We need the following definition:

Definition 4.1.4 (Orientation-preserving maps)**.** Let $U, V \subset \mathbb{R}^n$ be open. Let $G : U \to V$ be a diffeomorphism. Then $\det G'(x) \neq 0$ for all $x \in U$. We say that G is **orientation-preserving** if $\det G' > 0$ on U.

*The basic result we need is the following change-of-variables formula from multidimensional integration, which we will state without proof (a heuristic explanation is given below).

Theorem 4.1.5 (Change of variables formula)**.** Let $G : U \subset \mathbb{R}^n \to V \subset \mathbb{R}^n$ be an orientation-preserving diffeomorphism, and let $f \in C^\infty(V)$ be integrable. "Integrable" in this context means that the integral $\int_V |f(y)|\,d^n y$ is finite. Since f is smooth, this is automatically the case if V is a bounded set. If V is not bounded, then $f(y)$ must vanish for large $|y|$ sufficiently quickly in order for the integral to converge. Then

$$\int_V f(y)\,dy^1 \cdots dy^n = \int_U f(G(x)) \det G'(x)\,dx^1 \cdots dx^n.$$

Proof. See Spivak and Hubbard. A rough sketch is given in the rough sketch below.

The idea is that the small area element B based in y in V, as shown in the figure, is the image under G of the square area element A based at x in U. The area of B (denoted "vol" for volume in the figure) is given approximately by $\det G'(x)\mathrm{vol}\,(A)$. The integral of f over V is obtained by partitioning V into small area elements, multiplying the area of each element by the value of the function at a point of the area element, summing, and then taking the limit as the partition is made finer. The change of variables formula is given by this limit.

$\square$

Proposition 4.1.6 (Independence of parameterisation)**.**

Let $B \subset \mathbb{R}^k$ be an open set which contains the k-cube I^k. Let $G : B \to B$ be an orientation-preserving diffeomorphism, and suppose that $G(I^k) = I^k$. That is, G maps the k-cube into

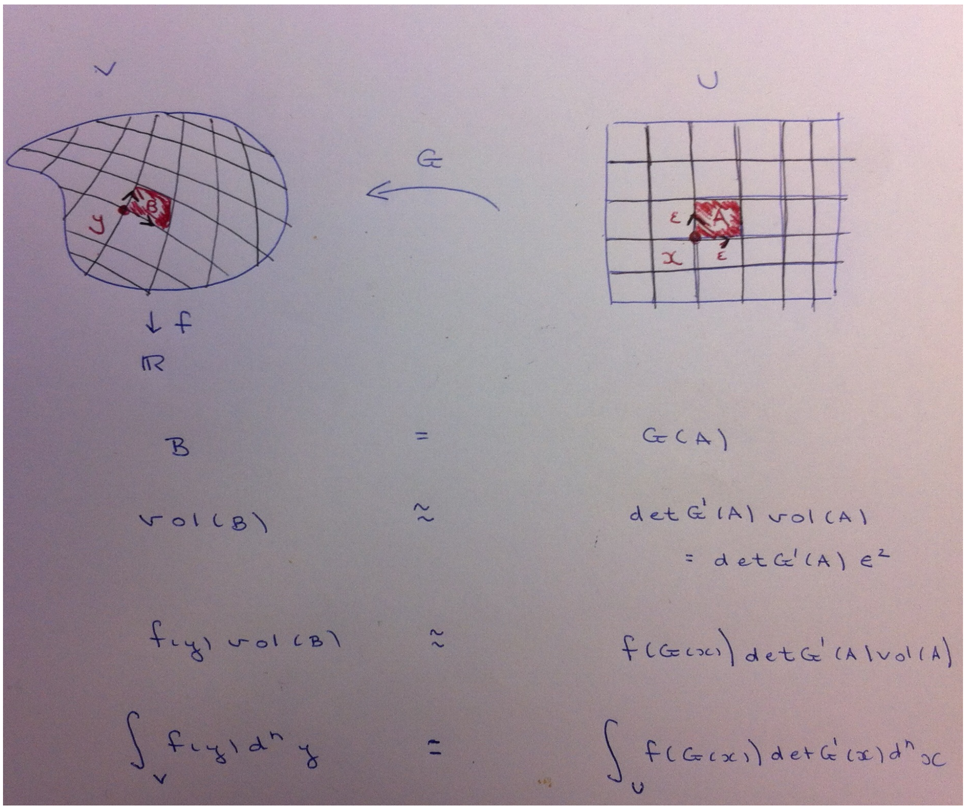


Figure 26

itself, and may be thought of as a smooth change of variables on I^k. Let $c : I^k \to U$ be a singular k-cube, and $\omega \in \Omega^k(U)$. Then

$$\int_c \omega = \int_{c \circ G} \omega.$$

Note that $c \circ G : I^k \to U$ is a singular k-cube.

Proof. *Let

$$c^*\omega(t) = f(t)\, dt^1 \wedge \cdots \wedge dt^k.$$

Then the left-hand side of the assertion reads

$$\int_c \omega = \int_{I^k} f(t)\, dt^1 \cdots dt^k.$$

On the right-hand side, we have that

$$\int_{c \circ G} \omega = \int_{I^k} (c \circ G)^*\omega,$$

from the definition of integration of a k-form. By Proposition 3.3.9,

$$(c \circ G)^*\omega = G^*c^*\omega = G^*\left(f\, dt^1 \wedge \cdots \wedge dt^k\right).$$

We have that

$$G^*f = f \circ G,$$

and

$$G^* dt^i = d(G^* t^i) = dG^i = \frac{\partial G^i}{\partial t^{j_i}}\, dt^{j_i}.$$

Then

$$\begin{aligned} G^*(f\, dt^1 \wedge \cdots \wedge dt^k) &= (f \circ G) \frac{\partial G^1}{\partial t^{j_1}} \cdots \frac{\partial G^k}{\partial t^{j_k}}\, dt^{j_1} \wedge \cdots \wedge dt^{j_k} \\ &= (f \circ G) \sum_{\sigma \in S_k} \frac{\partial G^1}{\partial t^{\sigma(1)}} \cdots \frac{\partial G^k}{\partial t^{\sigma(k)}}\, dt^{\sigma(1)} \wedge \cdots \wedge dt^{\sigma(k)}, \end{aligned}$$

since the only nonvanishing contributions to the sum come from terms where the indices $j_1, \ldots, j_k$ are all distinct, and therefore are given by a permutation; that is, $j_i = \sigma(i)$ for some $\sigma \in S_k$. From the anticommutativity of the wedge product, it follows that

$$dt^{\sigma(1)} \wedge \cdots \wedge dt^{(\sigma(k)} = \operatorname{sgn} \sigma\, dt^1 \wedge \cdots \wedge dt^k.$$

Therefore,

$$G^*(f\, dt^1 \wedge \cdots \wedge dt^k) = (f \circ G) \left(\sum_{\sigma \in S_k} \operatorname{sgn} \sigma \frac{\partial G^1}{\partial t^{\sigma(1)}} \cdots \frac{\partial G^k}{\partial t^{\sigma(k)}} \right) dt^1 \wedge \cdots \wedge dt^k.$$

But we recognise the combinatorial formula for the determinant (cf Proposition 2.4.5),

$$\sum_{\sigma \in S_k} \operatorname{sgn} \sigma \frac{\partial G^1}{\partial t^{\sigma(1)}} \cdots \frac{\partial G^k}{\partial t^{\sigma(k)}} = \det G'.$$

Therefore,

$$G^*(f\, dt^1 \wedge \cdots \wedge dt^k) = (f \circ G) \det G' dt^1 \wedge dt^k.$$

It follows that

$$\int_{c \circ G} \omega = \int_{I^k} f(G(t)) \det G'(t)\, dt^1 \cdots dt^k.$$

By the Change of Variables formula (Theorem 4.1.5) and the fact that $G(I^k) = I^k$ by assumption, we have that

$$\int_{I^k} f(G(t)) \det G'(t)\, dt^1 \cdots dt^k = \int_{I^k} f(t) dt^1 \cdots dt^k.$$

Therefore,

$$\int_{c \circ G} \omega = \int_{I^k} f(t) dt^1 \cdots dt^k = \int_c \omega,$$

as required.

□

Example 4.1.7. Here is a simple example of change of parameterisation in a one-dimensional integral. Really, it's just how you would integrate $1/(1+t^2)$ by making the trigonometric substitution $t = \tan s$. Below we carry this through, with some small changes of notation, to fit with the statement of Proposition 4.1.6.

Let ω be the one-form on $U = \mathbb{R}$ given by

$$\omega(x) = \frac{1}{1+x^2}\, dx.$$

Let $c : I \to U$ be the singular one-cube on U given by

$$c(t) = t.$$

Then

$$c^*\omega = \frac{1}{1+t^2}\, dt,$$

and

$$\int_c \omega = \int_0^1 \frac{dt}{1+t^2}.$$

Let $G : I^1 \to I^1$ be given by

$$G(t) = \tan \frac{\pi}{4} t.$$

We want to compute

$$\int_{c\circ G} \omega = \int_0^1 (c \circ G)^* \omega.$$

We have that

$$(c \circ G)^*\omega = G^* \left(\frac{1}{1+t^2}\, dt \right).$$

Also,

$$G^* t = \tan \frac{\pi}{4} t,$$

so that

$$G^* dt = \frac{\pi}{4} \sec^2 \frac{\pi}{4} t\, dt.$$

Then

$$G^* \left(\frac{1}{1+t^2}\, dt \right) = \frac{(\pi/4) \sec^2((\pi/4)t)}{1 + \tan^2((\pi/4)t)}\, dt = \frac{\pi}{4}\, dt.$$

Then

$$\int_{c\circ G} \omega = \int_0^{\pi/4} dt = \frac{\pi}{4}.$$

Finally, we want to consider singular k-cubes for $k = 0$, and the analogue of integration for 0-forms. We define

$$I^0 = \{0\}.$$

That is, I^0 consists of a single point, namely 0. A singular 0-form on U is a map $c : I^0 \to U; 0 \mapsto c(0) \in U$. That is, a 0-form maps 0 to a single point $c(0) \in U$. Given a zero-form $f \in \Omega^0(U)$, i.e. a function, we make the following definition:

$$\int_c f := f(c(0)).$$

That is, "integration" of a 0-form over a singular 0-cube is really just evaluation, i.e. evaluating a function at point.

4.2 Boundaries

We want to consider boundaries of singular k-cubes. Roughly, this means looking at the map c restricted to the different faces of I^k. This leads us to the more general notion of singular k-chains.

Definition 4.2.1 (Singular k-chain)**.** A **singular k-chain** on U, denoted $\mathcal{C}$, is a formal sum of a finite number of singular k-cubes $c_r : I^k \to U$ with integer coefficients, i.e.

$$\mathcal{C} = a_1 c_1 + \cdots a_s c_s, \quad a_r \in \mathbb{Z}.$$

k-chains may be added. Here is an example: Let

$$\mathcal{C} = c_1 + 2c_3, \quad \mathcal{C}' = -2c_1 + c_3 + c_4.$$

Then

$$\mathcal{C} + \mathcal{C}' = -c_1 + 3c_3 + c_4.$$

k-chains may also be multiplied by integers. Here is an example: Given $\mathcal{C}$ as above,

$$2\mathcal{C} = 2c_1 + 4c_3.$$

In general, let

$$\mathcal{C} = \sum_{r=1}^{s} a_r c_r, \quad \mathcal{C}' = \sum_{r=1}^{s} a'_r c_r.$$

Note that some of the a_r's and a'_r's can be zero. Then

$$\mathcal{C} + \mathcal{C}' = \sum_{r=1}^{s} (a_r + a'_r) c_r, \quad m\mathcal{C} = \sum_{r=1}^{s} m a_r c_r, \quad m \in \mathbb{Z}.$$

Given a singular k-chain $\mathcal{C}$ on U and a k-form $\omega \in \Omega^k(U)$, we define

$$\int_{\mathcal{C}} \omega := \sum_{r=1}^{s} a_r \int_{c_r} \omega. \tag{51}$$

That is, the integral over a k-chain is just the sum of the integrals over the k-chains it comprises weighted by the integer coefficients.

Definition 4.2.2 (Faces of k-cubes)**.**

Let $c : I^k \to U$ be a singular k-cube on U. Take j such that $1 \leq j \leq k$ and $\alpha = 0$ or 1. The **(j, α)th face** of c, denoted $c_{(j,\alpha)}$, is the singular $(k-1)$-cube given by

$$c_{(j,\alpha)} : I^{k-1} \to U,$$

where

$$c_{(j,\alpha)}(t^1, \ldots, t^{k-1}) = c(t^1, \ldots, t^{j-1}, \alpha, t^j, \ldots, t^{k-1}).$$

That is, $c_{(j,\alpha)}$ is obtained by fixing the jth argument of c to be α. This restricts c to one of the faces of I^k.

Definition 4.2.3 (Boundary of singular k-cube)**.** Let $c : I^k \to U$ be a singular k-cube. The **boundary** of a singular k-cube c, denoted ∂c, is the singular $(k-1)$-chain given by

$$\partial c = \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} c_{(j,\alpha)}.$$

For $k = 0$, the boundary of a singular 0-cube is taken be 0;

$$\partial c = 0.$$

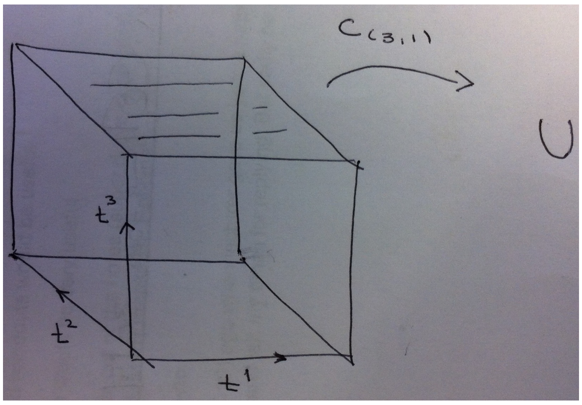


Figure 27: For $j = 3$ and $\alpha = 1$, the (j, α)th face of the singular 3-cube c is obtained by restricting c to the top face of the cube, as shown.

Example 4.2.4. We consider again the singular 2-cube on $U = \mathbb{R}^3$ from Example 4.1.3,

$$c : I^2 \to \mathbb{R}^3; (u, v) \mapsto c(u, v) = (u, v, 2 - u^2 - v^2).$$

The faces are given by

$$\begin{aligned}
c_{(1,0)}(t) &= c(0, t) = (0, t, 2 - t^2),\\
c_{(1,1)}(t) &= c(1, t) = (1, t, 1 - t^2),\\
c_{(2,0)}(t) &= c(t, 0) = (t, 0, 2 - t^2),\\
c_{(2,1)}(t) &= c(t, 1) = (t, 1, 1 - t^2).
\end{aligned}$$

The boundary is given by

$$\partial c = -c_{(1,0)} + c_{(1,1)} + c_{(2,0)} - c_{(2,1)}.$$

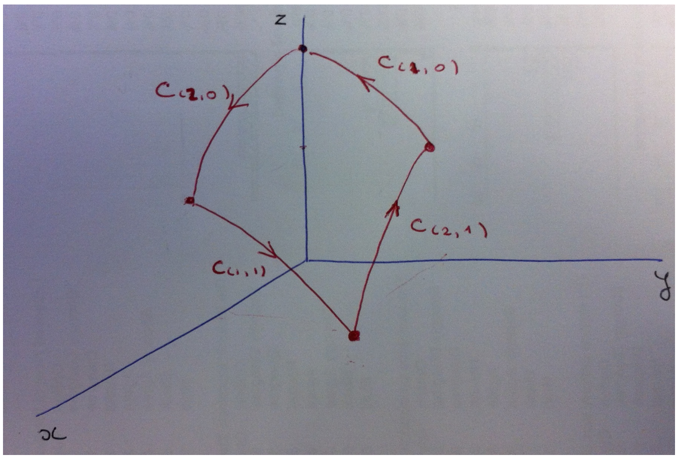


Figure 28

Arrows are drawn according to the sign $(-1)^{j+\alpha}$. If $(-1)^{j+\alpha} = 1$, then arrows point in the direction of increasing t. If $(-1)^{j+\alpha} = -1$, then arrows point in the direction of decreasing t.

Let $\omega \in \Omega^{k-1}(U)$ be a $(k-1)$-form, and $c : I^k \to U$ a singular k-cube. Then

$$\int_{\partial c} \omega = \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} \int_{c_{(j,\alpha)}} \omega.$$

Let us discuss the sign factor, $(-1)^{j+\alpha} = (-1)^{\alpha}(-1)^{j}$, which appears in the definition of the boundary. The factor of $(-1)^{\alpha}$ means that opposite faces (corresponding to faces with the same j but different α's) come with opposite signs. The factor $(-1)^{j}$ is not as easily motivated, but is justified by, and can be derived from, the following two results.

Given a k-chain

$$\mathcal{C} = \sum_{r=1}^{s} a_r c_r,$$

we define its **boundary**, denoted $\partial\mathcal{C}$, by

$$\partial\mathcal{C} = \sum_{r=1}^{s} a_r \partial c_r.$$

$\partial\mathcal{C}$ is a $(k-1)$-chain.

Proposition 4.2.5 ($\partial^2 = 0$)**.**

Let $\mathcal{C}$ be a singular k-chain on U. Then

$$\partial^2\mathcal{C} = \partial(\partial C) = 0.$$

Proof. See Section 4.4 for the (purely combinatorial) proof.

□

Example 4.2.6. Let c be a singular two-cube. Then

$$\partial^2 c = -\partial c_{(1,0)} + \partial c_{(1,1)} + \partial c_{(2,0)} - \partial c_{(2,1)}.$$

We have that

$$\partial c_{(1,0)} = -\left(c_{(1,0)}\right)_{(1,0)} + \left(c_{(1,0)}\right)_{(1,1)} = -c(0,0) + c(0,1).$$

The other faces of faces are computed similarly. We get

$$\begin{aligned} -\partial c_{(1,0)} &= c(0,0) - c(0,1), \\ \partial c_{(1,1)} &= -c(1,0) + c(1,1), \\ \partial c_{(2,0)} &= -c(0,0) + c(1,0), \\ -\partial c_{(2,1)} &= c(0,1) + -c_{(1,1)}. \end{aligned}$$

Adding up the singular 0-cells on the right-hand side, we see that the terms cancel pairwise.

*The following result gives some further insight into the signs which appear in the definition of the boundary map. Let us consider a generalised boundary-like map, denoted $\bar{\delta}$, defined on k-chains as follows. For a singular 0-cube c on $U \subset \mathbb{R}^n$,

$$\bar{\delta} c = 0.$$

For a singular k-cube with $k \geq 1$,

$$\bar{\delta} c = \sum_{j=1}^{k} \sum_{\alpha=0,1} n_k(j,\alpha) c_{(j,\alpha)}, \tag{52}$$

where $n_k(j,\alpha)$ are integers. For $\mathcal{C} = \sum_r b_r c_r$, where the c_r's are singular k-cubes on $U \subset \mathbb{R}^n$,

$$\bar{\delta}\mathcal{C} = \sum_r b_r \bar{\delta} c_r.$$

In addition, for any 1-cube c with $c(1) = c(0)$, we require that $\bar{\delta} c = 0$. This is equivalent to

$$n_1(1,1) = -n_1(1,0). \tag{53}$$

The standard boundary map, ∂, corresponds to taking

$$n_k(j,\alpha) = (-1)^{j+\alpha}.$$

Proposition 4.2.7.

Let $U \subset \mathbb{R}^n$ be open, and let $\mathcal{C}$ be a singular k-chain on U. Then, with $\bar{\delta}$ as above, $\bar{\delta}^2 = 0$ if and only if

$$n_k(j,\alpha) = \sigma_k(-1)^{j+\alpha}, \tag{54}$$

where σ_k is an integer.

Proof. See Section 4.4 □

4.3 Stokes theorem

We begin with a calculation. Let $U \subset \mathbb{R}^n$ be open, $c : I^k \to U$ a singular k-cube, and $\omega \in \Omega^{k-1}(U)$ a $(k-1)$-form. Then $c^*\omega$ is a $(k-1)$-form on I^k, and may be written in the form

$$c^*\omega = \sum_{i=1}^{k} g_i(t)\, dt^1 \wedge \cdots \wedge \widehat{dt^i} \wedge \cdots \wedge dt^k, \tag{55}$$

where the notation $\widehat{dt^i}$ means that dt^i is omitted from the wedge product.

The following gives an expression for $c^*_{(j,\alpha)}\omega$ in terms of $c^*\omega$. For the sake of clarity, we'll use coordinates $s = (s^1, \ldots, s^{k-1})$ on I^{k-1} and coordinates $t = (t^1, \ldots, t^k)$ on I^k.

Proposition 4.3.1. Let $\omega \in \Omega^{k-1}(U)$ be a $(k-1)$-form on U, and let $c^*\omega$ be given by (55). Then

$$\left(c^*_{(j,\alpha)}\omega\right)(s) = g_j(s^1, \ldots, s^{j-1}, \alpha, s^j, \ldots, s^{k-1})\, ds^1 \wedge \cdots \wedge ds^{k-1}.$$

Proof. This is a straightforward calculation. The only difficulty is being careful about which coordinate goes where. It is helpful to introduce the following map: Let

$$e_{(j,\alpha)} : I^{k-1} \to I^k$$

be given by

$$e_{(j,\alpha)}(s) = (s^1, \ldots, s^{j-1}, \alpha, s^j, \ldots, s^k).$$

Thus, $e_{(j,\alpha)}$ maps the $(k-1)$-cube I^{k-1} onto one of the faces of I^k, namely the face obtained by setting the jth coordinate equal to α. In other words, if we let $t^r(s) := e^r_{(j,\alpha)}(s)$, then

$$t^r(s) = \begin{cases} s^r, & r < j, \\ \alpha, & r = j, \\ s^{r-1}, & r > j. \end{cases}$$

Then

$$c_{(j,\alpha)} = c \circ e_{(j,\alpha)}.$$

Now we may compute $c^*_{(j,\alpha)}\omega$ as follows:

$$c^*_{(j,\alpha)}\omega = (c \circ e_{(j,\alpha)})^*\omega = e^*_{(j,\alpha)} c^*\omega \text{ (by Proposition 3.3.9)} = e^*_{(j,\alpha)} \left(\sum_{i=1}^{k} g_i(t) dt^1 \wedge \cdots \wedge \widehat{dt^i} \wedge \cdots \wedge dt^k \right).$$

We pullback each factor in turn. We have that

$$\left(e^*_{(j,\alpha)} g_i \right)(s) = g_i(e_{(j,\alpha)}(s)) = g_i(s^1, \ldots, s^{j-1}, \alpha, s^j, \ldots, s^{k-1}).$$

Also,

$$e^*_{(j,\alpha)} dt^r = d e^*_{(j,\alpha)} t^r = \begin{cases} ds^r, & r < j, \\ d\alpha, & r = j, \\ ds^{r-1}, & r > j. \end{cases} = \begin{cases} ds^r, & r < j, \\ 0, & r = j, \\ ds^{r-1}, & r > j. \end{cases}$$

Then

$$e^*_{(j,\alpha)} \left(dt^1 \wedge \cdots \wedge \widehat{dt^i} \wedge \cdots \wedge dt^k \right) = \begin{cases} 0, & j \neq i \text{ (since } e^*_{(j,\alpha)} dt^j = 0), \\ ds^1 \wedge \cdots \wedge ds^{k-1}, & j = i. \end{cases}$$

It follows that

$$\left(c^*_{(j,\alpha)} \omega \right)(s) = g_j(s^1, \ldots, s^{j-1}, \alpha, s^j, \ldots, s^{k-1})\, ds^1 \wedge \cdots \wedge ds^{k-1},$$

as required. □

Theorem 4.3.2 (Stokes theorem)**.**

Let $U \subset \mathbb{R}^n$ be open, and let $c : I^k \to U$ be a singular k-cube on U. Let $\omega \in \Omega^{k-1}(U)$ be a $(k-1)$-form on U. Then

$$\int_c d\omega = \int_{\partial c} \omega.$$

Proof. As in (55), we write

$$c^*\omega = \sum_{i=1}^{k} g_i(t) dt^1 \wedge \cdots \wedge \widehat{dt^i} \wedge \cdots \wedge dt^k.$$

On the right-hand side of the assertion, we have

$$\begin{aligned} \int_{\partial c} \omega &= \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} \int_{I^{k-1}} c^*_{(j,\alpha)} \omega \\ &= \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} \int_0^1 \cdots \int_0^1 g_j(s^1, \ldots, s^{j-1}, \alpha, s^j, \ldots, s^{k-1})\, ds^1 \ldots ds^{k-1}, \end{aligned} \tag{56}$$

from Proposition 4.3.1.

On the left-hand side, we have

$$\int_c d\omega = \int_{I^k} c^* d\omega = \int_{I^k} dc^*\omega.$$

Then

$$dc^*\omega = d\left(\sum_{i=1}^{k} g_i\, dt^1 \wedge \cdots \wedge \widehat{dt^i} \wedge \cdots \wedge dt^k\right)$$

$$= \sum_{i=1}^{k}\left(\sum_{j=1}^{k} \frac{\partial g_i}{\partial t^j}\, dt^j\right) \wedge dt^1 \wedge \cdots \wedge \widehat{dt^i} \wedge \cdots \wedge dt^k$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{k} \frac{\partial g_i}{\partial t^j}\, dt^j \wedge dt^1 \wedge \cdots \wedge \widehat{dt^i} \wedge \cdots \wedge dt^k \text{ (reversing the order of the sums over } i \text{ and } j)$$

$$= \sum_{j=1}^{k} \frac{\partial g_j}{\partial t^j}\, dt^j \wedge dt^1 \wedge \cdots \wedge \widehat{dt^j} \wedge \cdots \wedge dt^k,$$

since the only terms in the sum over i which contribute have $i = j$ (otherwise, the fact that $dt^i \wedge dt^i = 0$ means the term vanishes). It follows that

$$dc^*\omega = \sum_{j=1}^{k}(-1)^{j-1}\frac{\partial g_j}{\partial t^j}\, dt^1 \wedge \cdots \wedge dt^k.$$

Then

$$\int_c d\omega = \sum_{j=1}^{k}(-1)^{j-1}\int_0^1 \cdots \int_0^1 \frac{\partial g_j}{\partial t^j} dt^1 \cdots dt^k.$$

Consider the jth term in the sum, and integrate with respect to t^j:

$$\int_0^1 \frac{\partial g_j}{\partial t^j}(t)\, dt^j = g_j(t^1, \ldots, t^{j-1}, 1, t^{j+1}, \ldots, t^k) - g_j(t^1, \ldots, t^{j-1}, 0, t^{j+1}, \ldots, t^k)$$

$$= \sum_{\alpha=0,1} (-1)^{\alpha+1} g_j(t^1, \ldots, t^{j-1}, \alpha, t^{j+1}, \ldots, t^k).$$

Let us changes variables, replacing t^r by s^r for $r < j$ and by s^{r-1} for $s > j$. Then

$$\int_c d\omega = \sum_{j=1}^{k}(-1)^{j-1}(-1)^{\alpha+1}\int_0^1 \cdots \int_0^1 g_j(s^1, \ldots, s^{j-1}, \alpha, s^j, \ldots, s^{k-1})\, ds^1 \cdots ds^{k-1}.$$

This last expression coincides with (56).

□

Example 4.3.3.

a) For $k = 1$ and take $U = \mathbb{R}$, Stokes' theorem is just the Fundamental Theorem of Calculus. Consider a 0-form, i.e. a function $f(x)$ on $\mathbb{R}$. Let $c : [0, 1] \to \mathbb{R}$ be the singular 1-cube given by $c(t) = t$. Then Theorem 4.3.2 says that

$$\int_c df = \int_0^1 \frac{df}{dt} dt = f(1) - f(0) = \int_{\partial c} f.$$

b) We continue with the example discussed in Examples 4.1.3 and 4.2.4. In this case, $k = 2$. Consider the singular 2-cube on $\mathbb{R}^3$ given by

$$c(u, v) = (u, v, 2 - u^2 - v^2).$$

Let ω be the 1-form given by

$$\omega = y^2\, dz.$$

We verify Stokes theorem by calculating $\int_{\partial c} \omega$ and $\int_c d\omega$ explicitly.

First, we compute $\int_{\partial c} \omega$. We computed ∂c in Example 4.2.4. We compute $c^*_{(j,\alpha)}\omega$ for each face in turn.

$\underline{j = 1, \alpha = 0}$. We have that $c_{(1,0)}(t) = (0, t, 2 - t^2)$, so that

$$c^*_{(1,0)}x = 0, \quad c^*_{(1,0)}y = t, \quad c^*_{(1,0)}z = 2 - t^2,$$

and

$$c^*_{(1,0)}dx = 0, \quad c^*_{(1,0)}dy = dt, \quad c^*_{(1,0)}dz = -2t\, dt,$$

Then

$$c^*_{(1,0)}\omega = t^2(-2t)\, dt = -2t^3\, dt.$$

$\underline{j = 1, \alpha = 1}$. We have that $c_{(1,1)}(t) = (1, t, 1 - t^2)$, so that

$$c^*_{(1,1)}x = 1, \quad c^*_{(1,1)}y = t, \quad c^*_{(1,1)}z = 1 - t^2,$$

and

$$c^*_{(1,1)}dx = 0, \quad c^*_{(1,1)}dy = dt, \quad c^*_{(1,1)}dz = -2t\, dt,$$

Then

$$c^*_{(1,1)}\omega = t^2(-2t)\, dt = -2t^3\, dt.$$

$\underline{j = 2, \alpha = 0}$. We have that $c_{(2,0)}(t) = (t, 0, 2 - t^2)$, so that

$$c^*_{(2,0)}x = t, \quad c^*_{(2,0)}y = 0, \quad c^*_{(2,0)}z = 2 - t^2,$$

and

$$c^*_{(2,0)}dx = dt, \quad c^*_{(2,0)}dy = 0, \quad c^*_{(2,0)}dz = -2t\, dt,$$

Then

$$c^*_{(2,1)}\omega = 0.$$

$\underline{j = 2, \alpha = 1}$. We have that $c_{(2,1)}(t) = (t, 1, 1 - t^2)$, so that

$$c^*_{(2,1)}x = t, \quad c^*_{(2,1)}y = 1, \quad c^*_{(2,1)}z = 1 - t^2,$$

and

$$c^*_{(2,1)}dx = dt, \quad c^*_{(2,1)}dy = 0, \quad c^*_{(2,1)}dz = -2t\, dt,$$

Then

$$c^*_{(2,1)}\omega = -2t\, dt.$$

It follows that

$$\sum_{j=1}^{2} \sum_{\alpha=0,1} (-1)^{j+\alpha} c^*_{(j,\alpha)}\omega = 2t\, dt,$$

as the contributions from $c_{(1,0)}$ and $c_{(1,1)}$ cancel. We obtain

$$\int_{\partial c} \omega = \int_0^1 2t\, dt = 1.$$

Next we compute $\int_c d\omega$. We have that

$$c^* d\omega = dc^*\omega = d(-2v^2(u\,du + v\,dv)) = -4vu\,dv \wedge du = 4uv\,du \wedge dv.$$

Then

$$\int_c d\omega = \int_0^1 \int_0^1 4uv\,du\,dv = 1.$$

4.4 *Proofs of results for the boundary map [nonexaminable]

Proof of Proposition 4.2.5. It suffices to show that $\partial^2 c = 0$ for a singular k-cube c (the argument for a general k-chain then follows by linearity). As above,

$$\partial c = \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} c_{(j,\alpha)},$$

so that

$$\partial^2 c = \partial(\partial c) = \sum_{j=1}^{k} \sum_{\alpha=0,1} (-1)^{j+\alpha} \partial c_{(j,\alpha)} = \sum_{j=1}^{k} \sum_{l=1}^{k-1} \sum_{\alpha,\beta=0,1} (-1)^{j+l+\alpha+\beta} (c_{(j,\alpha)})_{(l,\beta)}.$$

Here, $(c_{(j,\alpha)})_{(l,\beta)}$, the (l,β)th face of $c_{(j,\alpha)}$, is a singular $(k-2)$-cube. Let us divide the preceding expression into two contributions according to whether $j > l$ or $j \le l$. We write

$$\partial^2 c = S_1 + S_2, \tag{57}$$

where

$$S_1 = \sum_{j=1}^{k} \sum_{l=1}^{j-1} \sum_{\alpha,\beta=0,1} (-1)^{j+l+\alpha+\beta} (c_{(j,\alpha)})_{(l,\beta)},$$

and

$$S_2 = \sum_{j=1}^{k} \sum_{l=j}^{k-1} \sum_{\alpha,\beta=0,1} (-1)^{j+l+\alpha+\beta} (c_{(j,\alpha)})_{(l,\beta)}.$$

We will change summation variables in S_1 so that the domain of summation coincides with that of S_2. First, note that we can start the j-sum in S_1 at $j = 2$, since there are no contributions from $j = 1$ (the l-sum is empty in this case). Thus,

$$S_1 = \sum_{j=2}^{k} \sum_{l=1}^{j-1} \sum_{\alpha,\beta=0,1} (-1)^{j+l+\alpha+\beta} (c_{(j,\alpha)})_{(l,\beta)}.$$

Let $n = j - 1$, $m = l$ and then replace j, l by m, n and interchange α and β to get

$$S_1 = \sum_{n=1}^{k-1} \sum_{m=1}^{n} \sum_{\alpha,\beta=0,1} (-1)^{m+n+1+\alpha+\beta} (c_{(n+1,\beta)})_{(m,\alpha)} = -\sum_{n=1}^{k-1} \sum_{m=1}^{n} \sum_{\alpha,\beta=0,1} (-1)^{m+n+\alpha+\beta} (c_{(n+1,\beta)})_{(m,\alpha)}.$$

Interchange the m and n sums to get

$$S_1 = -\sum_{m=1}^{k-1} \sum_{n=m}^{k-1} \sum_{\alpha,\beta=0,1} (-1)^{m+n+\alpha+\beta} (c_{(n+1,\beta)})_{(m,\alpha)}.$$

Finally, rename (m,n) by (j,l) to get

$$S_1 = -\sum_{j=1}^{k-1}\sum_{l=j}^{k-1}\sum_{\alpha,\beta=0,1}(-1)^{j+l+\alpha+\beta}(c_{(l+1,\beta)})_{(j,\alpha)}.$$

Substitute into (57) to get

$$\partial^2 C = \sum_{j=1}^{k-1}\sum_{l=j}^{k-1}\sum_{\alpha,\beta=0,1}(-1)^{j+l+\alpha+\beta}\left(-(C_{(l+1,\beta)})_{(j,\alpha)} + (c_{(j,\alpha)})_{(l,\beta)}\right). \tag{58}$$

For $j \leq l$, we claim that

$$(c_{(l+1,\beta)})_{(j,\alpha)} = (c_{(j,\alpha)})_{(l,\beta)}. \tag{59}$$

From (58), this establishes that $\partial^2 C = 0$. Verifying (59) involves carefully applying the definition of faces,

$$c_{(j,\alpha)}(t^1,\ldots,t^{k-1}) = c(t^1,\ldots,t^{j-1},\alpha,t^j,\ldots,t^{k-1}), \tag{60}$$

in the appropriate order. Starting on the right-hand side, we have that

$$(c_{(j,\alpha)})_{(l,\beta)}(t^1,\ldots,t^{k-2}) = c_{(j,\alpha)}(t^1,\ldots,t^{l-1},\beta,t^l,t^{l+1},\ldots,t^{k-2}).$$

Apply (60) a second time to express $c_{(j,\alpha)}$ in terms of c. Since $j \leq l$, we get that

$$(c_{(j,\alpha)})_{(l,\beta)}(t^1,\ldots,t^{k-2}) = c(t^1,\ldots,t^{j-1},\alpha,t^j,\ldots,t^{l-1},\beta,t^l,t^{l+1},\ldots,t^{k-2}). \tag{61}$$

On the left-hand side of (59), we have that

$$(c_{(l+1,\beta)})_{(j,\alpha)}(t^1,\ldots,t^{k-2}) = c_{(l+1,\beta)}(t^1,\ldots,t^{j-1},\alpha,t^j,t^{j+1},\ldots,t^{k-2}).$$

Apply (60) a second time to express $c_{(l+1,\beta)}$ in terms of c, and note that, since $j \leq l$, the lth argument in $(t^1,\ldots,t^{j-1},\alpha,t^j,t^{j+1},\ldots,t^{k-2})$ is t^{l-1}. Then

$$(c_{(l+1,\beta)})_{(j,\alpha)}(t^1,\ldots,t^{k-2}) = c(t^1,\ldots,t^{j-1},\alpha,t^j,\ldots,t^{l-1},\beta,t^l,t^{l+1},\ldots,t^{k-2}). \tag{62}$$

The claim (59) follows from (61) and (62). □

Proof of Proposition 4.2.7.

If it holds, then, for $\mathcal{C}$ a singular k-chain, we have that $\bar{\delta}^2\mathcal{C} = \sigma_k\sigma_{k-1}\partial^2\mathcal{C}$, which vanishes by Proposition 4.2.5.

Next, assume that $\bar{\delta}^2\mathcal{C} = 0$ for all k-chains $\mathcal{C}$. From (53), (54) holds for $k = 1$ if we define

$$\sigma_1 = n_1(1,1).$$

We proceed by induction. We assume (54) holds for k and show it holds for $k+1$.

Let C be a singular $k+1$-cube on U. Then

$$\bar{\delta}^2 C = \bar{\delta}\left(\sum_{i=1}^{k+1}\sum_{\alpha=0,1} n_{k+1}(i,\alpha)C_{(i,\alpha)}\right) =$$

$$\sum_{\substack{j=1\\ j\neq i}}^{k}\sum_{\beta=0,1} n_k(j,\beta)\sum_{i=1}^{k+1}\sum_{\alpha=0,1} n_{k+1}(i,\alpha)\left(C_{(i,\alpha)}\right)_{(j,\beta)} =$$

$$\left(\sum_{i<j} + \sum_{i>j}\right)\sum_{\alpha,\beta=0,1} n_{k+1}(i,\alpha)n_k(j,\beta)\left(C_{(i,\alpha)}\right)_{(j,\beta)}, \tag{63}$$

where in the last expression i and j are summed over pairs of distinct integers between 1 and $k+1$, which we separate according to whether $i<j$ or $i>j$.

From the definition of the faces of k-cubes, it follows that, for $i<j$,

$$\left(C_{(i,\alpha)}\right)_{(j,\beta)} = \left(C_{(j+1,\beta)}\right)_{(i,\alpha)},$$

as both sides of the preceding give the $k-1$-cube

$$C'(t^1,\ldots,t^{k-1}) = C(t^1,\ldots,t^{i-1},\alpha,t^i,\ldots,t^{j-1},\beta,t^j,\ldots,t^{k-1}).$$

Therefore, (63) can be written as

$$\begin{aligned}\bar{\delta}^2 C = \sum_{i>j}\sum_{\alpha,\beta=0,1}\Big(n_{k+1}(i,\alpha)n_k(j,\beta)\left(C_{(j+1,\beta)}\right)_{(i,\alpha)} + n_{k+1}(i,\alpha)n_k(j,\beta)\left(C_{(i,\alpha)}\right)_{(j,\beta)}\Big)\\ = \sum_{i>j}\sum_{\alpha,\beta=0,1}\left(n_{k+1}(j+1,\beta)n_k(i,\alpha) + n_{k+1}(i,\alpha)n_k(j,\beta)\right)\left(C_{(i,\alpha)}\right)_{(j,\beta)},\end{aligned}$$

where the last line follows by combining terms involving the same "edges" $\left(C_{(i,\alpha)}\right)_{(j,\beta)}$ (an edge is a face of a face). $\left(C_{(i,\alpha)}\right)_{(j,\beta)}$. In order that $\bar{\delta}^2 C = 0$ for all C, we must have

$$n_{k+1}(j+1,\beta)n_k(i,\alpha) + n_{k+1}(i,\alpha)n_k(j,\beta) = 0.$$

From the induction hypothesis, it follows that

$$n_{k+1}(j+1,\beta)(-1)^{i+\alpha} = -n_{k+1}(i,\alpha)(-1)^{j+\beta}, \tag{64}$$

where we have canceled a factor of σ_k. Setting $i=\alpha=1$, we get that

$$n_{k+1}(j+1,\beta) = -\sigma_{k+1}(-1)^{j+\beta}, \tag{65}$$

where we define σ_{k+1} as

$$\sigma_{k+1} = n_{k+1}(1,1).$$

Setting $l=j+1$ in (65), we get

$$n_{k+1}(l,\beta) = \sigma_{k+1}(-1)^{l+\beta},$$

which establishes (54) for $2\le l\le k+1$. For $l=1$, set $i=1$ in (64) and use (54) for the lhs to conclude that

$$\sigma_{k+1}(-1)^{j+1+\beta}(-1)^{1+\alpha} = -n_{k+1}(1,\alpha)(-1)^{j+\beta},$$

or

$$n_{k+1}(1,\alpha) = \sigma_{k+1}(-1)^{1+\alpha}.$$

This establishes (54) for $l=1$. □

5 Bibliography

The course does not follow a particular textbook. There are many books and online resources that cover all or parts of the syllabus, but there is no, single recommended text. Below is a list of some standard texts, but you are encouraged to look for yourselves.

1. JH and BB Hubbard, "Vector calculus, linear algebra and differential forms: A unified approach", 2 ed, Prentice Hall

   This is a very good rigorous treatment of multivariable calculus. The main overlap with the syllabus is in Chapter 6, which deals with algebraic and differential forms. The text takes a different point of view to ours on several concepts, in particular the definition of the exterior derivative. Otherwise, it contains proofs of many results referred to in the lectures, including the inverse function theorem, multidimensional integrals and the change of variables formula. Recommended for supplementary material, but should not be regarded as a textbook in lieu of the course notes. There is little on part 1 of the syllabus (vector fields, flows, Jacobi bracket, Frobenius theorem, etc).

2. B Schutz, Geometrical methods in mathematical physics, Cambridge University Press

   This text provides a more informal introduction to much of the material in the course, and much more that isn't in the course, in the style of applied mathematics and theoretical physics. The text introduces differentiable manifolds early on, whereas in the course we work mainly with open subsets of $\mathbb{R}^n$. Therefore, the presentation requires some translation from the general setting of manifolds to $\mathbb{R}^n$. The text covers vector fields, Jacobi bracket, the Frobenius theorem, algebraic and differential forms, exterior derivative, Lie derivative and Stokes' theorem. It also discusses Lie groups and Riemannian geometry (not covered in the course) and a number of physical applications. The presentation and notation is occasionally nonstandard, differing from what we use in the course. Recommended for supplementary reading but not as a main text.

3. W Darling, Differential forms and Connections, Cambridge University Press

   This is a mathematical but elementary introduction to differentiable manifolds and differential forms. Algebraic forms on $\mathbb{R}^n$ are covered in Chapter 1, differential forms on $\mathbb{R}^n$ in Chapter 2, and integration of differential forms in Chapter 8; hence there is substantial intersection with parts 2, 3 and 4 of the syllabus. The text introduces differentiable manifolds in a simplified, slightly cheating way (avoiding the notion of topological space) which nevertheless enables a shortcut into the general theory. It also covers Riemannian geometry and vector bundles, along with applications – all these topics are outside the course syllabus. Recommended for supplementary reading and a different approach, but not as a substitute for the course notes. There is some overlap with part 1 of the syllabus (vector fields, flows, Jacobi bracket, Frobenius theorem, etc).

4. M Spivak, Calculus on Manifolds, Westview Press

   Similar to Hubbard and Hubbard, Spivak is a rigorous treatment of calculus on $\mathbb{R}^n$, and covers the material in parts 2, 3 and 4 of the syllabus, providing proofs of all the main results. The style is different; this is a short text written in the form of extended notes. This is a classic text. There is very little on part 1 of the syllabus (vector fields, flows, Jacobi bracket, Frobenius theorem, etc).

5. M Spivak, A comprehensive introduction to differential geometry, vol 1, Publish or Perish, Berkeley

   This is the first of a five-volume set, and follows on from Spivak's Calculus on Manifolds. This is definitely at a more advanced level than our course, but interested students may want to try it. There is an extensive treatment of part 1 of the syllabus (vector fields, flows, Jacobi bracket, Frobenius theorem, etc), all in the context of manifolds, as well as the rest of the syllabus. Subsequent volumes treat Riemmanian geometry, vector bundles and characteristic classes in some detail. Postgraduate level, for supplementary reading and future study.

6. V Arnold, Mathematical methods of classical mechanics, Springer-Verlag.

   As the name suggests, this is a text on classical mechanics, and a classic at that. Chapters 4, 7 and 8 provide a condensed introduction to much of the material we cover in the course. f differentiable manifolds and the main topics in the course. Postgraduate level. Good for supplementary reading.