Manifolds and Differential Forms

Reyer Sjamaar

DEPARTMENT OF MATHEMATICS, CORNELL UNIVERSITY, ITHACA, NEW YORK 14853-4201

E-mail address: sjamaar@math.cornell.edu

URL: http://www.math.cornell.edu/~sjamaar

Last updated: 2006-08-26T01:13-05:00

Contents

Preface

These are the lecture notes for Math 321, Manifolds and Differential Forms, as taught at Cornell University since the Fall of 2001. The course covers manifolds and differential forms for an audience of undergraduates who have taken a typical calculus sequence at a North American university, including basic linear algebra and multivariable calculus up to the integral theorems of Green, Gauß and Stokes. With a view to the fact that vector spaces are nowadays a standard item on the undergraduate menu, the text is not restricted to curves and surfaces in three-dimensional space, but treats manifolds of arbitrary dimension. Some prerequisites are briefly reviewed within the text and in appendices. The selection of material is similar to that in Spivak's book [**Spi65**] and in Flanders' book [**Fla89**], but the treatment is at a more elementary and informal level appropriate for sophomores and juniors.

A large portion of the text consists of problem sets placed at the end of each chapter. The exercises range from easy substitution drills to fairly involved but, I hope, interesting computations, as well as more theoretical or conceptual problems. More than once the text makes use of results obtained in the exercises.

Because of its transitional nature between calculus and analysis, a text of this kind has to walk a thin line between mathematical informality and rigour. I have tended to err on the side of caution by providing fairly detailed definitions and proofs. In class, depending on the aptitudes and preferences of the audience and also on the available time, one can skip over many of the details without too much loss of continuity. At any rate, most of the exercises do not require a great deal of formal logical skill and throughout I have tried to minimize the use of point-set topology.

This revised version of the notes is still a bit rough at the edges. Plans for improvement include: more and better graphics, an appendix on linear algebra, a chapter on fluid mechanics and one on curvature, perhaps including the theorems of Poincaré-Hopf and Gauß-Bonnet. These notes and eventual revisions can be downloaded from the course website at

`http://www.math.cornell.edu/~sjamaar/classes/321/index.html`.

Corrections, suggestions and comments will be received gratefully.

Ithaca, NY, 2006-08-26

CHAPTER 1

Introduction

We start with an informal, intuitive introduction to manifolds and how they arise in mathematical nature. Most of this material will be examined more thoroughly in later chapters.

1.1. Manifolds

Recall that Euclidean n-space $\mathbf{R}^n$ is the set of all column vectors with n real entries

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

which we shall call *points* or *n-vectors* and denote by lower case boldface letters. In $\mathbf{R}^2$ or $\mathbf{R}^3$ we often write

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \text{resp.} \qquad \mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

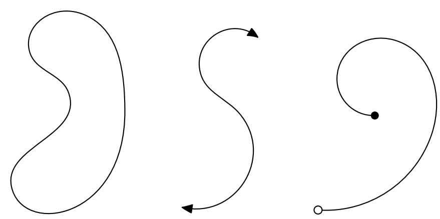
For reasons having to do with matrix multiplication, column vectors are not to be confused with row vectors $(x_1 \ x_2 \ \ldots \ x_n)$. For clarity, we shall usually separate the entries of a row vector by commas, as in $(x_1, x_2, \ldots, x_n)$. Occasionally, to save space, we shall represent a column vector $\mathbf{x}$ as the *transpose* of a row vector,

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)^T.$$

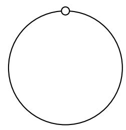
A *manifold* is a certain type of subset of $\mathbf{R}^n$. A precise definition will follow in Chapter 6, but one important consequence of the definition is that a manifold has a well-defined tangent space at every point. This fact enables us to apply the methods of calculus and linear algebra to the study of manifolds. The *dimension* of a manifold is by definition the dimension of its tangent spaces. The dimension of a manifold in $\mathbf{R}^n$ can be no higher than n.

Dimension 1. A one-dimensional manifold is, loosely speaking, a curve without kinks or self-intersections. Instead of the tangent "space" at a point one usually speaks of the tangent *line*. A curve in $\mathbf{R}^2$ is called a *plane* curve and a curve in $\mathbf{R}^3$ is a *space* curve, but you can have curves in any $\mathbf{R}^n$. Curves can be closed (as in the first picture below), unbounded (as indicated by the arrows in the second picture), or have one or two endpoints (the third picture shows a curve with an endpoint, indicated by a black dot; the white dot at the other end indicates that

that point does not belong to the curve; the curve "peters out" without coming to an endpoint). Endpoints are also called *boundary points*.

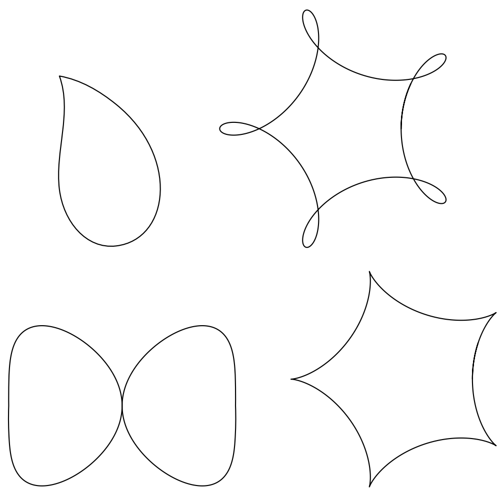


A circle with one point deleted is also an example of a manifold. Think of a torn elastic band.



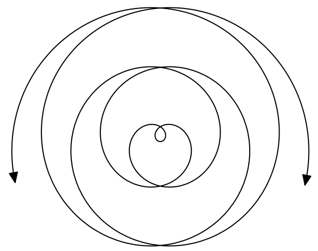
By straightening out the elastic band we see that this manifold is really the same as an open interval.

The four plane curves below are not manifolds. The teardrop has a kink, where two distinct tangent lines occur instead of a single well-defined tangent line; the five-fold loop has five points of self-intersection, at each of which there are two distinct tangent lines. The bow tie and the five-pointed star have well-defined tangent lines everywhere. Still they are not manifolds: the bow tie has a self-intersection and the cusps of the star have a jagged appearance which is proscribed by the definition of a manifold (which we have not yet given). The points where these curves fail to be manifolds are called *singularities*. The "good" points are called *smooth*.

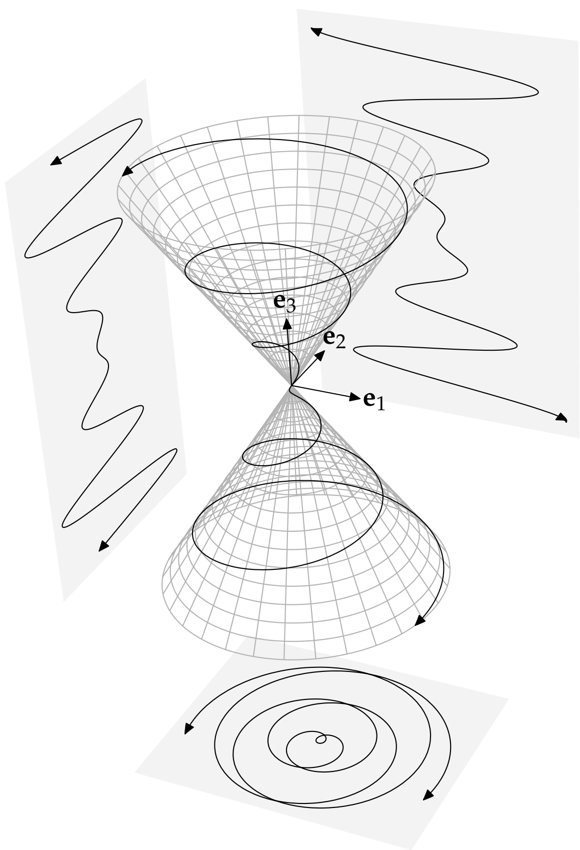


Singularities can sometimes be "resolved". For instance, the self-intersections of the Archimedean spiral, which is given in polar coordinates by r is a constant times

θ, where r is allowed to be negative,

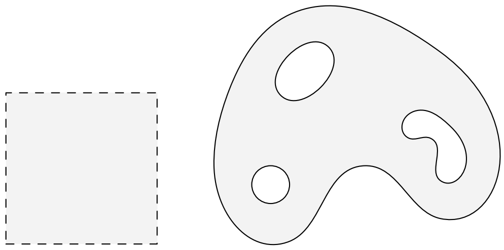


can be got rid of by uncoiling the spiral and wrapping it around a cone. You can convince yourself that the resulting space curve has no singularities by peeking at it along the direction of the x-axis or the y-axis. What you will see are the smooth curves shown in the yz-plane and the xz-plane.

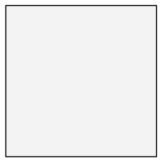


(The three-dimensional models in these notes are drawn in central perspective. They are best viewed facing the origin, which is usually in the middle of the picture, from a distance of 30 cm with one eye shut.) Singularities are extremely interesting, but in this course we shall focus on gaining a thorough understanding of the smooth points.

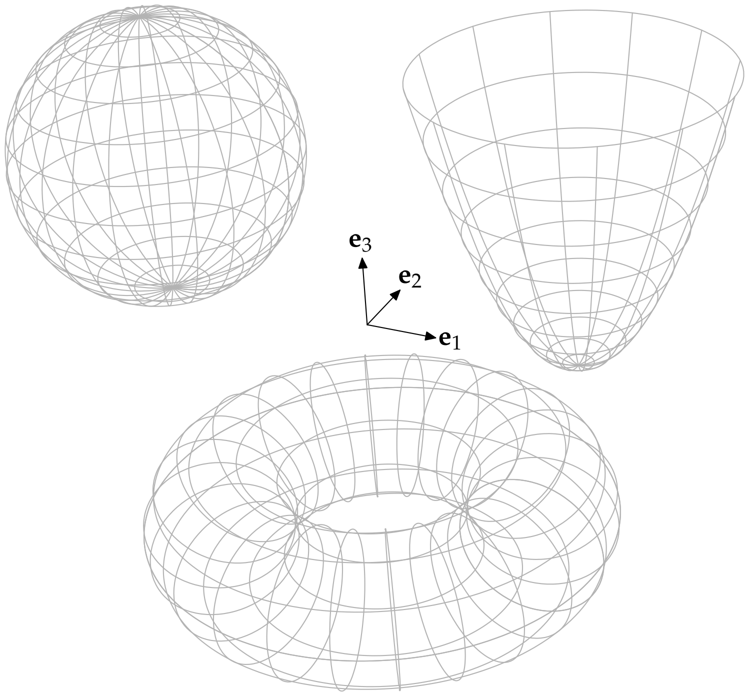
Dimension 2. A two-dimensional manifold is a smooth surface without self-intersections. It may have a boundary, which is always a one-dimensional manifold. You can have two-dimensional manifolds in the plane $\mathbf{R}^2$, but they are relatively boring. Examples are: an arbitrary open subset of $\mathbf{R}^2$, such as an open square, or a closed subset with a smooth boundary.



A closed square is not a manifold, because the corners are not smooth.[1]



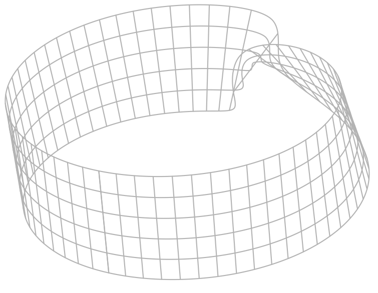
Two-dimensional manifolds in three-dimensional space include a sphere, a paraboloid and a torus.

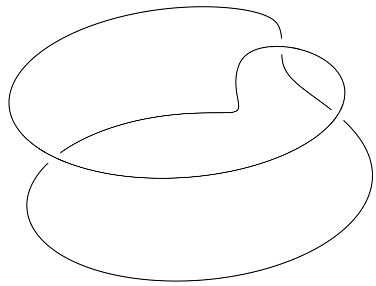


The famous *Möbius band* is made by pasting together the two ends of a rectangular strip of paper giving one end a half twist. The boundary of the band consists of

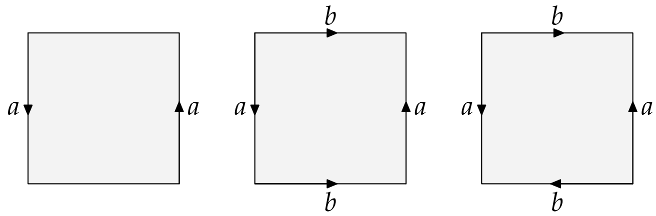
[1] To be strictly accurate, the closed square is a *topological* manifold with boundary, but not a *smooth* manifold with boundary. In these notes we will consider only smooth manifolds.

two boundary edges of the rectangle tied together and is therefore a single closed curve.



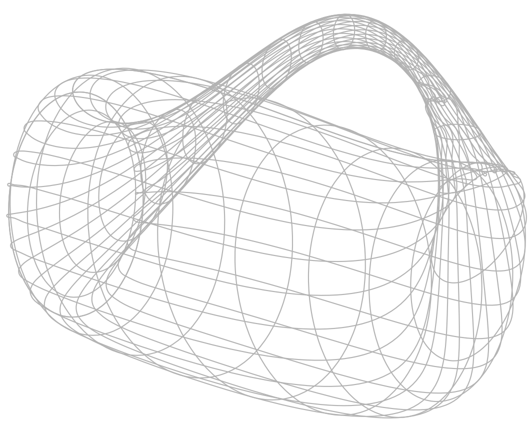


Out of the Möbius band we can create in two different ways a manifold without boundary by closing it up along the boundary edge. According to the direction in which we glue the edge to itself, we obtain the *Klein bottle* or the *projective plane*. A simple way to represent these three surfaces is by the following diagrams. The labels tell you which edges to glue together and the arrows tell you in which direction.

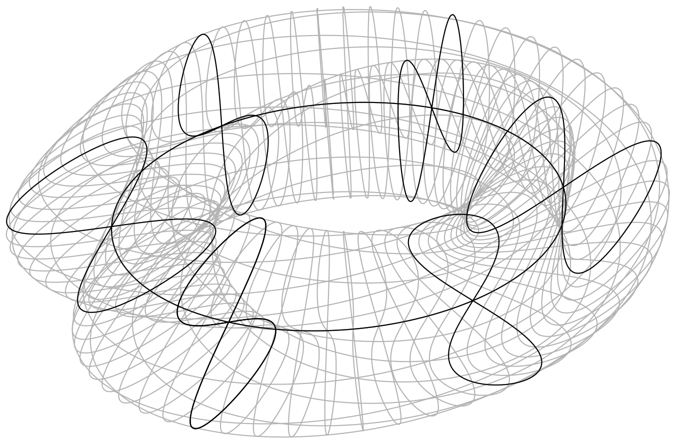


Möbius band Klein bottle projective plane

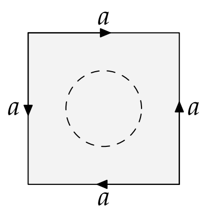
Perhaps the easiest way to make a Klein bottle is first to paste the top and bottom edges of the square together, which gives a tube, and then to join the resulting boundary circles, making sure the arrows match up. You will notice this cannot be done without passing one end through the wall of the tube. The resulting surface intersects itself along a circle and therefore is not a manifold.



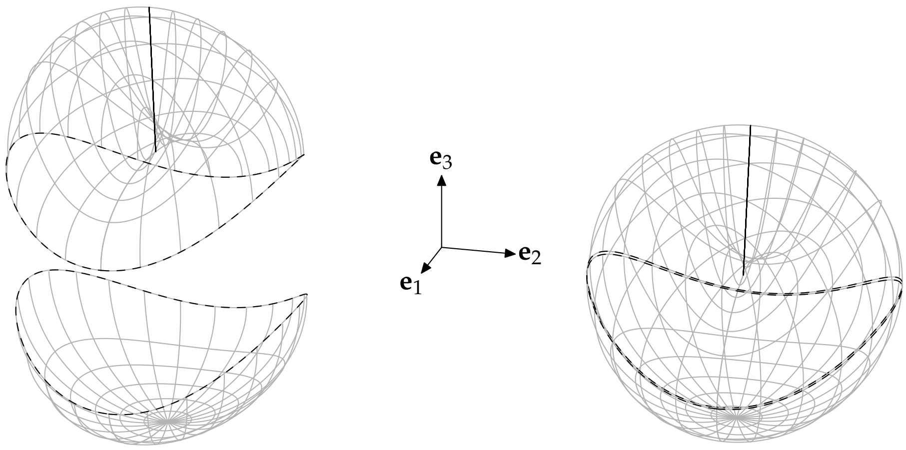
A different model of the Klein bottle is found by folding over the edge of a Möbius band until it touches the central circle. This creates a Möbius type band with a figure eight cross-section. Equivalently, take a length of tube with a figure eight cross-section and weld the ends together giving one end a half twist. Again the resulting surface has a self-intersection, namely the central circle of the original Möbius band. The self-intersection locus as well as a few of the cross-sections are shown in black in the following wire mesh model.



To represent the Klein bottle without self-intersections you need to embed it in four-dimensional space. The projective plane has the same peculiarity, and it too has self-intersecting models in three-dimensional space. Perhaps the easiest model is constructed by merging the edges a and b shown in the gluing diagram for the projective plane, which gives the following diagram.



First fold the lower right corner over to the upper left corner and seal the edges. This creates a pouch like a cherry turnover with two seams labelled a which meet at a corner. Now fuse the two seams to create a single seam labelled a. Below is a wire mesh model of the resulting surface. It is obtained by welding together two pieces along the dashed wires. The lower half shaped like a bowl corresponds to the dashed circular disc in the middle of the square. The upper half corresponds to the complement of the disc and is known as a *cross-cap*. The wire shown in black corresponds to the edge a. The interior points of the black wire are ordinary self-intersection points. Its two endpoints are qualitatively different singularities known as *pinch points*, where the surface is crinkled up.



1.2. Equations

Very commonly manifolds are given "implicitly", namely as the solution set of a system

$$\begin{aligned}\phi_1(x_1,\dots,x_n) &= c_1,\\ \phi_2(x_1,\dots,x_n) &= c_2,\\ &\vdots\\ \phi_m(x_1,\dots,x_n) &= c_m,\end{aligned}$$

of m equations in n unknowns. Here ϕ_1, $\phi_2,\dots,$ ϕ_m are functions, c_1, $c_2,\dots,$ c_m are constants and x_1, $x_2,\dots,$ x_n are variables. By introducing the useful shorthand

$$\mathbf{x} = \begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{pmatrix}, \qquad \phi(\mathbf{x}) = \begin{pmatrix} \phi_1(\mathbf{x})\\ \phi_2(\mathbf{x})\\ \vdots\\ \phi_m(\mathbf{x})\end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} c_1\\ c_2\\ \vdots\\ c_n\end{pmatrix},$$

we can represent this system as a single equation

$$\phi(\mathbf{x}) = \mathbf{c}.$$

It is in general difficult to find explicit solutions of such a system. (On the positive side, it is usually easy to decide whether any given point is a solution by plugging it into the equations.) Manifolds defined by linear equations (i.e. where ϕ is a matrix) are called *affine subspaces* of $\mathbf{R}^n$ and are studied in linear algebra. More interesting manifolds arise from nonlinear equations.

1.1. EXAMPLE. Consider the system of two equations in three unknowns,

$$\begin{aligned} x^2 + y^2 &= 1,\\ y + z &= 0.\end{aligned}$$

Here

$$\phi(\mathbf{x}) = \begin{pmatrix} x^2 + y^2\\ y + z\end{pmatrix} \qquad \text{and} \qquad \mathbf{c} = \begin{pmatrix} 1\\ 0\end{pmatrix}.$$

The solution set of this system is the intersection of a cylinder of radius 1 about the z-axis (given by the first equation) and a plane cutting the x-axis at a 45° angle (given by the second equation). Hence the solution set is an ellipse. It is a manifold of dimension 1.

1.2. EXAMPLE. The *sphere* of radius r about the origin in $\mathbf{R}^n$ is the set of all $\mathbf{x}$ in $\mathbf{R}^n$ satisfying the single equation $\|\mathbf{x}\| = r$. Here

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

is the *norm* or *length* of $\mathbf{x}$ and

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

is the *inner product* or *dot product* of $\mathbf{x}$ and $\mathbf{y}$. The sphere of radius r is an $n-1$-dimensional manifold in $\mathbf{R}^n$. The sphere of radius 1 is called the *unit sphere* and is denoted by S^{n-1}. What is a one-dimensional sphere? And a zero-dimensional sphere?

The solution set of a system of equations may have singularities and is therefore not necessarily a manifold. A simple example is $xy = 0$, the union of the two coordinate axes in the plane, which has a singularity at the origin. Other examples of singularities can be found in Exercise 1.5.

Tangent spaces. Let us use the example of the sphere to introduce the notion of a tangent space. Let

$$M = \{\, \mathbf{x} \in \mathbf{R}^n \mid \|\mathbf{x}\| = r \,\}$$

be the sphere of radius r about the origin in $\mathbf{R}^n$ and let $\mathbf{x}$ be a point in M. There are two reasonable, but inequivalent, views of how to define the tangent space to M at $\mathbf{x}$. The first view is that the tangent space at $\mathbf{x}$ consists of all vectors $\mathbf{y}$ such that $(\mathbf{y} - \mathbf{x}) \cdot \mathbf{x} = \mathbf{0}$, i.e. $\mathbf{y} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{x} = r^2$. In coordinates: $y_1 x_1 + \cdots + y_n x_n = r^2$. This is an inhomogeneous linear equation in $\mathbf{y}$. In this view, the tangent space at $\mathbf{x}$ is an affine subspace of $\mathbf{R}^n$, given by the single equation $\mathbf{y} \cdot \mathbf{x} = r^2$.

However, for most practical purposes it is easier to translate this affine subspace to the origin, which turns it into a linear subspace. This leads to the second view of the tangent space at $\mathbf{x}$, namely as the set of all $\mathbf{y}$ such that $\mathbf{y} \cdot \mathbf{x} = 0$, and this is the definition that we shall espouse. The standard notation for the tangent space to M at $\mathbf{x}$ is $T_{\mathbf{x}}M$. Thus

$$T_{\mathbf{x}}M = \{\, \mathbf{y} \in \mathbf{R}^n \mid \mathbf{y} \cdot \mathbf{x} = 0 \,\},$$

a linear subspace of $\mathbf{R}^n$. (In Exercise 1.6 you will be asked to find a basis of $T_{\mathbf{x}}M$ for a particular $\mathbf{x}$ and you will see that $T_{\mathbf{x}}M$ is $n-1$-dimensional.)

Inequalities. Manifolds with boundary are often presented as solution sets of a system of equations together with one or more inequalities. For instance, the *closed ball* of radius r about the origin in $\mathbf{R}^n$ is given by the single inequality $\|x\| \leq r$. Its boundary is the sphere of radius r.

1.3. Parametrizations

A dual method for describing manifolds is the "explicit" way, namely by parametrizations. For instance,

$$x = \cos\theta, \quad y = \sin\theta$$

parametrizes the unit circle in $\mathbf{R}^2$ and

$$x = \cos\theta\cos\phi, \quad y = \sin\theta\cos\phi, \quad z = \sin\phi$$

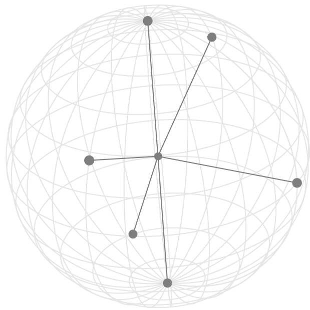
parametrizes the unit sphere in $\mathbf{R}^3$. (Here ϕ is the angle between a vector and the xy-plane and θ is the polar angle in the xy-plane.) The explicit method has various merits and demerits, which are complementary to those of the implicit method. One obvious advantage is that it is easy to find points lying on a parametrized manifold simply by plugging in values for the parameters. A disadvantage is that it can be hard to decide if any given point is on the manifold or not, because this involves solving for the parameters. Parametrizations are often harder to come by than a system of equations, but are at times more useful, for example when one wants to integrate over the manifold. Also, it is usually impossible to parametrize a manifold in such a way that every point is covered exactly once. Such is the case for the two-sphere. One commonly restricts the polar coordinates (θ, ϕ) to the rectangle $[0, 2\pi] \times [-\pi/2, \pi/2]$ to avoid counting points twice. Only the meridian $\theta = 0$ is then hit twice, but this does not matter for many purposes, such as computing the surface area or integrating a continuous function.

We will use parametrizations to give a formal definition of the notion of a manifold in Chapter 6. Note however that not every parametrization describes a manifold. Examples of parametrizations with singularities are given in Exercises 1.1 and 1.2.

1.4. Configuration spaces

Frequently manifolds arise in more abstract ways that may be hard to capture in terms of equations or parametrizations. Examples are solution curves of differential equations (see e.g. Exercise 1.10) and configuration spaces. The *configuration* of a mechanical system (such as a pendulum, a spinning top, the solar system, a fluid, or a gas etc.) is its state or position at any given time. (The configuration ignores any motions that the system may be undergoing. So a configuration is like a snapshot or a movie still. When the system moves, its configuration changes.) In practice one usually describes a configuration by specifying the coordinates of suitably chosen parts of the system. The *configuration space* or *state space* of the system is an abstract space, the points of which are in one-to-one correspondence to all physically possible configurations of the system. Very often the configuration space turns out to be a manifold. Its dimension is called the *number of degrees of freedom* of the system. The configuration space of even a fairly small system can be quite complicated.

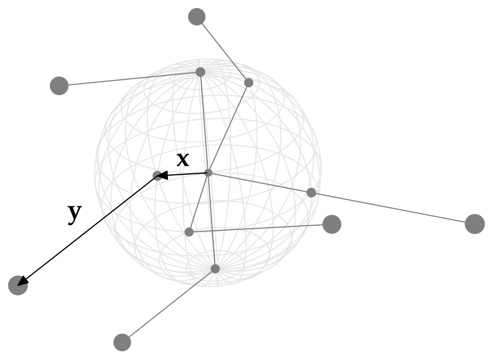
1.3. EXAMPLE. A *spherical pendulum* is a weight or bob attached to a fixed centre by a rigid rod, free to swing in any direction in three-space.



The state of the pendulum is entirely determined by the position of the bob. The bob can move from any point at a fixed distance (equal to the length of the rod) from the centre to any other. The configuration space is therefore a two-dimensional sphere.

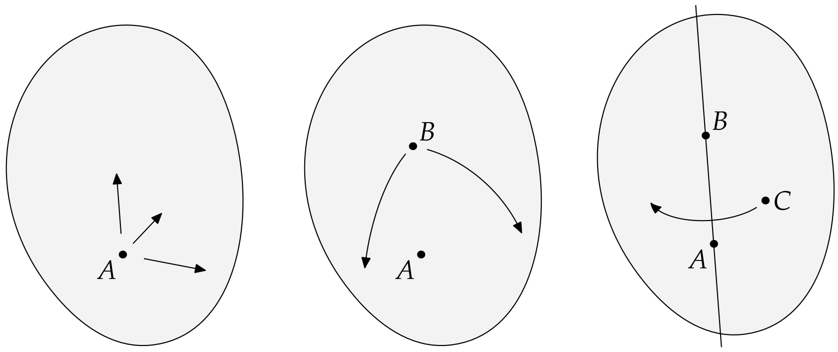
Some believe that only spaces of dimension ≤ 3 (or 4, for those who have heard of relativity) can have a basis in physical reality. The following two examples show that this is not true.

1.4. EXAMPLE. Take a spherical pendulum of length r and attach a second one of length s to the moving end of the first by a universal joint. The resulting system is a *double spherical pendulum*. The state of this system can be specified by a pair of vectors $(\mathbf{x}, \mathbf{y})$, $\mathbf{x}$ being the vector pointing from the centre to the first weight and $\mathbf{y}$ the vector pointing from the first to the second weight.



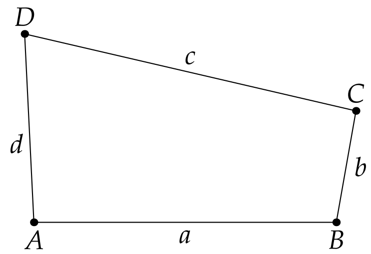
The vector $\mathbf{x}$ is constrained to a sphere of radius r about the centre and $\mathbf{y}$ to a sphere of radius s about the head of $\mathbf{x}$. Aside from this limitation, every pair of vectors can occur (if we suppose the second rod is allowed to swing completely freely and move "through" the first rod) and describes a distinct configuration. Thus there are four degrees of freedom. The configuration space is a four-dimensional manifold, known as the *(Cartesian) product* of two two-dimensional spheres.

1.5. EXAMPLE. What is the number of degrees of freedom of a rigid body moving in $\mathbf{R}^3$? Select any triple of points A, B, C in the solid that do not lie on one line. The point A can move about freely and is determined by three coordinates, and so it has three degrees of freedom. But the position of A alone does not determine the position of the whole solid. If A is kept fixed, the point B can perform two independent swivelling motions. In other words, it moves on a sphere centred at A, which gives two more degrees of freedom. If A and B are both kept fixed, the point C can rotate about the axis AB, which gives one further degree of freedom.

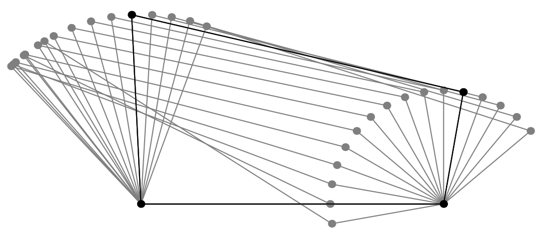


The positions of A, B and C determine the position of the solid uniquely, so the total number of degrees of freedom is $3 + 2 + 1 = 6$. Thus the configuration space of a rigid body is a six-dimensional manifold.

1.6. EXAMPLE (the space of quadrilaterals). Consider all quadrilaterals $ABCD$ in the plane with fixed sidelengths a, b, c, d.

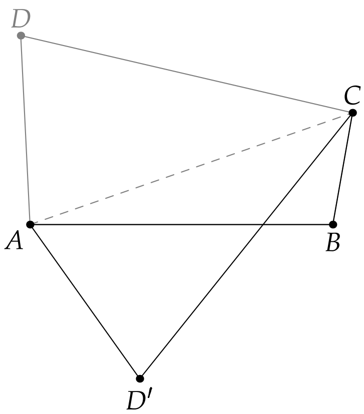


(Think of four rigid rods attached by hinges.) What are all the possibilities? For simplicity let us disregard translations by keeping the first edge AB fixed in one place. Edges are allowed to cross each other, so the short edge BC can spin full circle about the point B. During this motion the point D moves back and forth on a circle of radius d centred at A. A few possible positions are shown here.

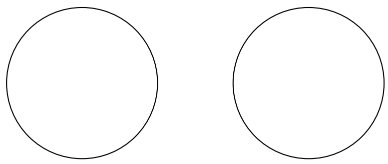


(As C moves all the way around, where does the point D reach its greatest left- or rightward displacement?) Arrangements such as this are commonly used in engines for converting a circular motion to a pumping motion, or vice versa. The position of the "pump" D is wholly determined by that of the "wheel" C. This means that the configurations are in one-to-one correspondence with the points on the circle of radius b about the point B, i.e. the configuration space is a circle.

Actually, this is not completely accurate: for every choice of C, there are *two* choices D and D' for the fourth point! They are interchanged by reflection in the diagonal AC.

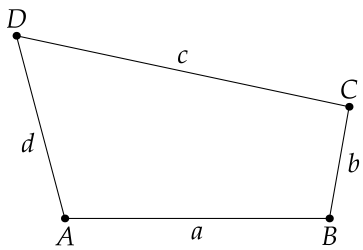


So there is in fact another circle's worth of possible configurations. It is not possible to move continuously from the first set of configurations to the second; in fact they are each other's mirror images. Thus the configuration space is a disjoint union of two circles.

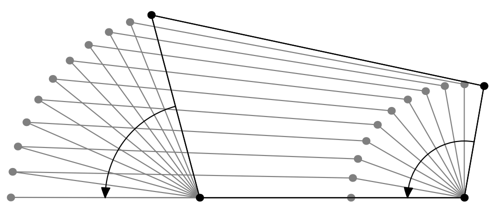


This is an example of a *disconnected* manifold consisting of two *connected components*.

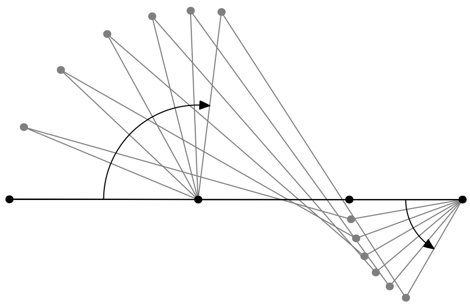
1.7. EXAMPLE (quadrilaterals, continued). Even this is not the full story: it *is* possible to move from one circle to the other when $b + c = a + d$ (and also when $a + b = c + d$).



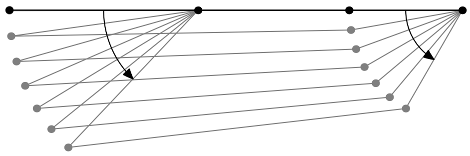
In this case, when BC points straight to the left, the quadrilateral collapses to a line segment:



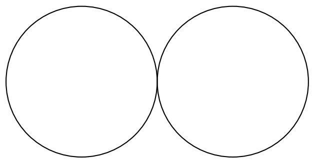
and when C moves further down, there are two possible directions for D to go, back up:



or further down:



This means that when $b + c = a + d$ the two components of the configuration space are merged at a point.



The juncture represents the collapsed quadrilateral. This configuration space is not a manifold, but most configuration spaces occurring in nature are (and an engineer designing an engine wouldn't want to use this quadrilateral to make a piston drive a flywheel). More singularities appear in the case of a parallelogram ($a = c$ and $b = d$) and in the equilateral case ($a = b = c = d$).

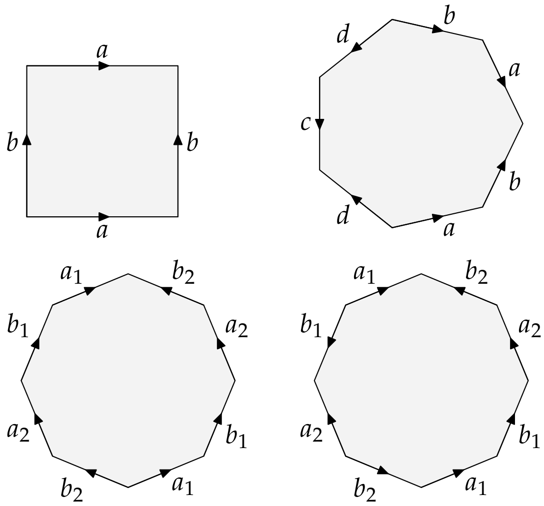
Exercises

1.1. The formulas $x = t - \sin t$, $y = 1 - \cos t$ ($t \in \mathbf{R}$) parametrize a plane curve. Graph this curve as carefully as you can. You may use software and turn in computer output. Also include a few tangent lines at judiciously chosen points. (E.g. find all tangent lines with slope 0, ± 1, and ∞.) To compute tangent lines, recall that the tangent vector at a point (x, y) of the curve has components dx/dt and dy/dt. In your plot, identify all points where the curve is not a manifold.

1.2. Same questions as in Exercise 1.1 for the curve $x = 3at/(1 + t^3)$, $y = 3at^2/(1 + t^3)$.

1.3. Parametrize the space curve wrapped around the cone shown in Section 1.1.

1.4. Sketch the surfaces defined by the following gluing diagrams.



(Proceed in stages, first gluing the a's, then the b's, etc., and try to identify what you get at every step. One of these surfaces cannot be embedded in $\mathbf{R}^3$, so use a self-intersection where necessary.)

1.5. For the values of n indicated below graph the surface in $\mathbf{R}^3$ defined by $x^n = y^2 z$. Determine all the points where the surface does not have a well-defined tangent plane. (Computer output is OK, but bear in mind that few drawing programs do an adequate job of plotting these surfaces, so you may be better off drawing them by hand. As a preliminary step, determine the intersection of each surface with a general plane parallel to one of the coordinate planes.)

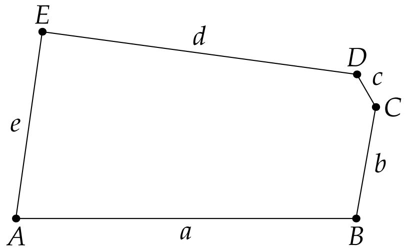
(i) $n = 0$.
(ii) $n = 1$.
(iii) $n = 2$.
(iv) $n = 3$.

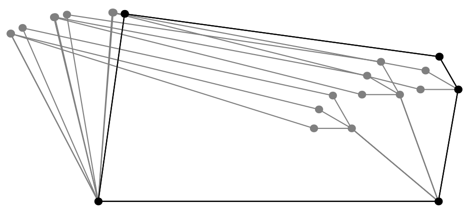
1.6. Let M be the sphere of radius $\sqrt{n}$ about the origin in $\mathbf{R}^n$ and let $\mathbf{x}$ be the point $(1, 1, \dots, 1)$ on M. Find a basis of the tangent space to M at $\mathbf{x}$. (Use that $T_{\mathbf{x}}M$ is the set of all $\mathbf{y}$ such that $\mathbf{y} \cdot \mathbf{x} = 0$. View this equation as a homogeneous linear equation in the entries $y_1, y_2, \dots, y_n$ of $\mathbf{y}$ and find the general solution by means of linear algebra.)

1.7. What is the number of degrees of freedom of a bicycle? (Imagine that it moves freely through empty space and is not constrained to the surface of the earth.)

1.8. Choose two distinct positive real numbers a and b. What is the configuration space of all parallelograms $ABCD$ such that AB and CD have length a and BC and AD have length b? What happens if $a = b$? (As in Examples 1.6 and 1.7 assume that the edge AB is kept fixed in place so as to rule out translations.)

1.9. What is the configuration space of all pentagons $ABCDE$ in the plane with fixed sidelengths a, b, c, d, e? (As in the case of quadrilaterals, for certain choices of sidelengths singularities may occur. You may ignore these cases. To reduce the number of degrees of freedom you may also assume the edge AB to be fixed in place.)





1.10. The Lotka-Volterra system is an early (ca. 1925) predator-prey model. It is the pair of differential equations

$$\frac{dx}{dt} = -rx + sxy,$$
$$\frac{dy}{dt} = py - qxy,$$

where $x(t)$ represents the number of prey and $y(t)$ the number of predators at time t, while p, q, r, s are positive constants. In this problem we will consider the solution curves (also called trajectories) $(x(t), y(t))$ of this system that are contained in the positive quadrant ($x > 0$, $y > 0$) and derive an implicit equation satisfied by these solution curves. (The Lotka-Volterra system is exceptional in this regard. Usually it is impossible to write down an equation for the solution curves of a differential equation.)

(i) Show that the solutions of the system satisfy a single differential equation of the form $dy/dx = f(x)g(y)$, where $f(x)$ is a function that depends only on x and $g(y)$ a function that depends only on y.

(ii) Solve the differential equation of part (i) by separating the variables, i.e. by writing $\dfrac{1}{g(y)}\,dy = f(x)\,dx$ and integrating both sides. (Don't forget the integration constant.)

(iii) Set $p = q = r = s = 1$ and plot a number of solution curves. Indicate the direction in which the solutions move. Be warned that solving the system may give better results than solving the implicit equation! You may use computer software such as Maple, Mathematica or MATLAB. A useful Java applet, `pplane`, can be found at `http://www.math.rice.edu/~dfield/dfpp.html`.

CHAPTER 2

Differential forms on Euclidean space

The notion of a differential form encompasses such ideas as elements of surface area and volume elements, the work exerted by a force, the flow of a fluid, and the curvature of a surface, space or hyperspace. An important operation on differential forms is exterior differentiation, which generalizes the operators div, grad and curl of vector calculus. The study of differential forms, which was initiated by E. Cartan in the years around 1900, is often termed the *exterior differential calculus*. A mathematically rigorous study of differential forms requires the machinery of multilinear algebra, which is examined in Chapter 7. Fortunately, it is entirely possible to acquire a solid working knowledge of differential forms without entering into this formalism. That is the objective of this chapter.

2.1. Elementary properties

A *differential form of degree k* or a *k-form* on $\mathbf{R}^n$ is an expression

$$\alpha = \sum_I f_I \, dx_I.$$

(If you don't know the symbol α, look up and memorize the Greek alphabet in the back of the notes.) Here I stands for a *multi-index* $(i_1, i_2, \dots, i_k)$ of *degree k*, that is a "vector" consisting of k integer entries ranging between 1 and n. The f_I are smooth functions on $\mathbf{R}^n$ called the *coefficients* of α, and dx_I is an abbreviation for

$$dx_{i_1} \, dx_{i_2} \cdots dx_{i_k}.$$

(The notation $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$ is also often used to distinguish this kind of product from another kind, called the tensor product.)

For instance the expressions

$$\alpha = \sin(x_1 + e^{x_4}) \, dx_1 \, dx_5 + x_2 x_5^2 \, dx_2 \, dx_3 + 6 \, dx_2 \, dx_4 + \cos x_2 \, dx_5 \, dx_3,$$
$$\beta = x_1 x_3 x_5 \, dx_1 \, dx_6 \, dx_3 \, dx_2,$$

represent a 2-form on $\mathbf{R}^5$, resp. a 4-form on $\mathbf{R}^6$. The form α consists of four terms, corresponding to the multi-indices $(1,5)$, $(2,3)$, $(2,4)$ and $(5,3)$, whereas β consists of one term, corresponding to the multi-index $(1,6,3,2)$.

Note, however, that α could equally well be regarded as a 2-form on $\mathbf{R}^6$ that does not involve the variable x_6. To avoid such ambiguities it is good practice to state explicitly the domain of definition when writing a differential form.

Another reason for being precise about the domain of a form is that the coefficients f_I may not be defined on all of $\mathbf{R}^n$, but only on an open subset U of $\mathbf{R}^n$. In such a case we say α is a k-form on U. Thus the expression $\ln(x^2 + y^2) z \, dz$ is not a 1-form on $\mathbf{R}^3$, but on the open set $U = \mathbf{R}^3 - \{ (x, y, z) \mid x^2 + y^2 \neq 0 \}$, i.e. the complement of the z-axis.

You can think of dx_i as an infinitesimal increment in the variable x_i and of dx_I as the volume of an infinitesimal k-dimensional rectangular block with sides dx_{i_1}, $dx_{i_2}, \dots, dx_{i_k}$. (A precise definition will follow in Section 7.2.) By volume we here mean *oriented* volume, which takes into account the order of the variables. Thus, if we interchange two variables, the sign changes:

$$dx_{i_1}\,dx_{i_2}\cdots dx_{i_q}\cdots dx_{i_p}\cdots dx_{i_k} = -dx_{i_1}\,dx_{i_2}\cdots dx_{i_p}\cdots dx_{i_q}\cdots dx_{i_k}, \tag{2.1}$$

and so forth. This is called *anticommutativity*, *graded commutativity*, or the *alternating property*. In particular, this rule implies $dx_i\,dx_i = -dx_i\,dx_i$, so $dx_i\,dx_i = 0$ for all i.

Let us consider k-forms for some special values of k.

A 0-form on $\mathbf{R}^n$ is simply a smooth function (no dx's).

A general 1-form looks like

$$f_1\,dx_1 + f_2\,dx_2 + \cdots + f_n\,dx_n.$$

A general 2-form has the shape

$$\begin{aligned}\sum_{i,j} f_{i,j}\,dx_i\,dx_j = f_{1,1}\,dx_1\,dx_1 + f_{1,2}\,dx_1\,dx_2 + \cdots + f_{1,n}\,dx_1\,dx_n \\ + f_{2,1}\,dx_2\,dx_1 + f_{2,2}\,dx_2\,dx_2 + \cdots + f_{2,n}\,dx_2\,dx_n + \cdots \\ + f_{n,1}\,dx_n\,dx_1 + f_{n,2}\,dx_n\,dx_2 + \cdots + f_{n,n}\,dx_n\,dx_n.\end{aligned}$$

Because of the alternating property (2.1) the terms $f_{i,i}\,dx_i\,dx_i$ vanish, and a pair of terms such as $f_{1,2}\,dx_1\,dx_2$ and $f_{2,1}\,dx_2\,dx_1$ can be grouped together: $f_{1,2}\,dx_1\,dx_2 + f_{2,1}\,dx_2\,dx_1 = (f_{1,2} - f_{2,1})\,dx_1\,dx_2$. So we can write any 2-form as

$$\begin{aligned}\sum_{1\le i<j\le n} g_{i,j}\,dx_i\,dx_j = g_{1,2}\,dx_1\,dx_2 + \cdots + g_{1,n}\,dx_1\,dx_n \\ + g_{2,3}\,dx_2\,dx_3 + \cdots + g_{2,n}\,dx_2\,dx_n + \cdots + g_{n-1,n}\,dx_{n-1}\,dx_n.\end{aligned}$$

Written like this, a 2-form has at most

$$n + n - 1 + n - 2 + \cdots + 2 + 1 = \frac{1}{2}n(n-1)$$

components.

Likewise, a general $n-1$-form can be written as a sum of n components,

$$\begin{aligned}f_1\,dx_2\,dx_3\cdots dx_n + f_2\,dx_1\,dx_3\cdots dx_n + \cdots + f_n\,dx_1\,dx_2\cdots dx_{n-1} \\ = \sum_{i=1}^{n} f_i\,dx_1\,dx_2\cdots \widehat{dx_i}\cdots dx_n,\end{aligned}$$

where $\widehat{dx_i}$ means "omit the factor dx_i".

Every n-form on $\mathbf{R}^n$ can be written as $f\,dx_1\,dx_2\cdots dx_n$. The special n-form $dx_1\,dx_2\cdots dx_n$ is also known as the *volume form*.

Forms of degree $k > n$ on $\mathbf{R}^n$ are always 0, because at least one variable has to repeat in any expression $dx_{i_1}\cdots dx_{i_k}$. By convention forms of negative degree are 0.

In general a form of degree k can be expressed as a sum

$$\alpha = \sum_I f_I\,dx_I,$$

where the I are *increasing* multi-indices, $1 \le i_1 < i_2 < \cdots < i_k \le n$. We shall almost always represent forms in this manner. The maximum number of terms occurring in α is then the number of increasing multi-indices of degree k. An increasing multi-index of degree k amounts to a choice of k numbers from among the numbers $1, 2, \dots, n$. The total number of increasing multi-indices of degree k is therefore equal to the binomial coefficient "n choose k",

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

(Compare this to the number of *all* multi-indices of degree k, which is n^k.) Two k-forms $\alpha = \sum_I f_I \, dx_I$ and $\beta = \sum_I g_I \, dx_I$ (with I ranging over the increasing multi-indices of degree k) are considered equal if and only if $f_I = g_I$ for all I. The collection of all k-forms on an open set U is denoted by $\Omega^k(U)$. Since k-forms can be added together and multiplied by scalars, the collection $\Omega^k(U)$ constitutes a vector space.

A form is *constant* if the coefficients f_I are constant functions. The set of constant k-forms is a linear subspace of $\Omega^k(U)$ of dimension $\binom{n}{k}$. A basis of this subspace is given by the forms dx_I, where I ranges over all increasing multi-indices of degree k. (The space $\Omega^k(U)$ itself is infinite-dimensional.)

The *(exterior) product* of a k-form $\alpha = \sum_I f_I \, dx_I$ and an l-form $\beta = \sum_J g_J \, dx_J$ is defined to be the $k+l$-form

$$\alpha\beta = \sum_{I,J} f_I g_J \, dx_I \, dx_J.$$

Usually many terms in a product cancel out or can be combined. For instance,

$$(y\,dx + x\,dy)(x\,dx\,dz + y\,dy\,dz) = y^2\,dx\,dy\,dz + x^2\,dy\,dx\,dz = (y^2 - x^2)\,dx\,dy\,dz.$$

As an extreme example of such a cancellation, consider an arbitrary form α of degree k. Its p-th power α^p is of degree kp, which is greater than n if $k > 0$ and $p > n$. Therefore

$$\alpha^{n+1} = 0$$

for any form α on $\mathbf{R}^n$ of positive degree.

The alternating property combines with the multiplication rule to give the following result.

2.1. PROPOSITION (graded commutativity).

$$\boxed{\beta\alpha = (-1)^{kl}\alpha\beta}$$

for all k-forms α and all l-forms β.

PROOF. Let $I = (i_1, i_2, \dots, i_k)$ and $J = (j_1, j_2, \dots, j_l)$. Successively applying the alternating property we get

$$\begin{aligned}
dx_I \, dx_J &= dx_{i_1} \, dx_{i_2} \cdots dx_{i_k} \, dx_{j_1} \, dx_{j_2} \, dx_{j_3} \cdots dx_{j_l} \\
&= (-1)^k \, dx_{j_1} \, dx_{i_1} \, dx_{i_2} \cdots dx_{i_k} \, dx_{j_2} \, dx_{j_3} \cdots dx_{j_l} \\
&= (-1)^{2k} \, dx_{j_1} \, dx_{j_2} \, dx_{i_1} \, dx_{i_2} \cdots dx_{i_k} \, dx_{j_3} \cdots dx_{j_l} \\
&\;\;\vdots \\
&= (-1)^{kl} \, dx_J \, dx_I.
\end{aligned}$$

For general forms $\alpha = \sum_I f_I \, dx_I$ and $\beta = \sum_J g_J \, dx_J$ we get from this

$$\beta\alpha = \sum_{I,J} g_J f_I \, dx_J \, dx_I = (-1)^{kl} \sum_{I,J} f_I g_J \, dx_I \, dx_J = (-1)^{kl}\alpha\beta,$$

which establishes the result. QED

A noteworthy special case is $\alpha = \beta$. Then we get $\alpha^2 = (-1)^{k^2}\alpha^2 = (-1)^k\alpha^2$. This equality is vacuous if k is even, but tells us that $\alpha^2 = 0$ if k is odd.

2.2. COROLLARY. *$\alpha^2 = 0$ if α is a form of odd degree.*

2.2. The exterior derivative

If f is a 0-form, that is a smooth function, we define df to be the 1-form

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \, dx_i.$$

Then we have the *product* or *Leibniz* rule:

$$d(fg) = f \, dg + g \, df.$$

If $\alpha = \sum_I f_I \, dx_I$ is a k-form, each of the coefficients f_I is a smooth function and we define $d\alpha$ to be the $k+1$-form

$$d\alpha = \sum_I df_I \, dx_I.$$

The operation d is called *exterior differentiation*. An operator of this sort is called a first-order partial differential operator, because it involves the first partial derivatives of the coefficients of a form.

2.3. EXAMPLE. If $\alpha = f \, dx + g \, dy$ is a 1-form on $\mathbf{R}^2$, then

$$d\alpha = f_y \, dy \, dx + g_x \, dx \, dy = (g_x - f_y) \, dx \, dy.$$

(Recall that f_y is an alternative notation for $\partial f / \partial y$.) More generally, for a 1-form $\alpha = \sum_{i=1}^n f_i \, dx_i$ on $\mathbf{R}^n$ we have

$$\begin{aligned} d\alpha &= \sum_{i=1}^{n} df_i \, dx_i = \sum_{i,j=1}^{n} \frac{\partial f_i}{\partial x_j} \, dx_j \, dx_i \\ &= \sum_{1 \le i < j \le n} \frac{\partial f_i}{\partial x_j} \, dx_j \, dx_i + \sum_{1 \le j < i \le n} \frac{\partial f_i}{\partial x_j} \, dx_j \, dx_i \\ &= - \sum_{1 \le i < j \le n} \frac{\partial f_i}{\partial x_j} \, dx_i \, dx_j + \sum_{1 \le i < j \le n} \frac{\partial f_j}{\partial x_i} \, dx_i \, dx_j \qquad (2.2) \\ &= \sum_{1 \le i < j \le n} \left(\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} \right) dx_i \, dx_j, \end{aligned}$$

where in line (2.2) in the first sum we used the alternating property and in the second sum we interchanged the roles of i and j.

2.4. EXAMPLE. If $\alpha = f\,dx\,dy + g\,dx\,dz + h\,dy\,dz$ is a 2-form on $\mathbf{R}^3$, then

$$d\alpha = f_z\,dz\,dx\,dy + g_y\,dy\,dx\,dz + h_x\,dx\,dy\,dz = (f_z - g_y + h_x)\,dx\,dy\,dz.$$

For a general 2-form $\alpha = \sum_{1\le i<j\le n} f_{i,j}\,dx_i\,dx_j$ on $\mathbf{R}^n$ we have

$$\begin{aligned}
d\alpha &= \sum_{1\le i<j\le n} df_{i,j}\,dx_i = \sum_{1\le i<j\le n}\sum_{k=1}^{n} \frac{\partial f_{i,j}}{\partial x_k}\,dx_k\,dx_i\,dx_j \\
&= \sum_{1\le k<i<j\le n} \frac{\partial f_{i,j}}{\partial x_k}\,dx_k\,dx_i\,dx_j + \sum_{1\le i<k<j\le n} \frac{\partial f_{i,j}}{\partial x_k}\,dx_k\,dx_i\,dx_j \\
&\qquad + \sum_{1\le i<j<k\le n} \frac{\partial f_{i,j}}{\partial x_k}\,dx_k\,dx_i\,dx_j \\
&= \sum_{1\le i<j<k\le n} \frac{\partial f_{j,k}}{\partial x_i}\,dx_i\,dx_j\,dx_k + \sum_{1\le i<j<k\le n} \frac{\partial f_{i,k}}{\partial x_j}\,dx_j\,dx_i\,dx_k \\
&\qquad + \sum_{1\le i<j<k\le n} \frac{\partial f_{i,j}}{\partial x_k}\,dx_k\,dx_i\,dx_j \qquad (2.3) \\
&= \sum_{1\le i<j<k\le n} \Big(\frac{\partial f_{i,j}}{\partial x_k} - \frac{\partial f_{i,k}}{\partial x_j} + \frac{\partial f_{j,k}}{\partial x_i}\Big)\,dx_i\,dx_j\,dx_k. \qquad (2.4)
\end{aligned}$$

Here in line (2.3) we rearranged the subscripts (for instance, in the first term we relabelled $k \longrightarrow i$, $i \longrightarrow j$ and $j \longrightarrow k$) and in line (2.4) we applied the alternating property.

An obvious but quite useful remark is that if α is an n-form on $\mathbf{R}^n$, then $d\alpha$ is of degree $n+1$ and so $d\alpha = 0$.

The operator d is linear and satisfies a generalized Leibniz rule.

2.5. PROPOSITION. (i) *$d(a\alpha + b\beta) = a\,d\alpha + b\,d\beta$ for all k-forms α and β and all scalars a and b.*
(ii) *$d(\alpha\beta) = (d\alpha)\beta + (-1)^k\alpha\,d\beta$ for all k-forms α and l-forms β.*

PROOF. The linearity property (i) follows from the linearity of partial differentiation:

$$\frac{\partial(af+bg)}{\partial x_i} = a\frac{\partial f}{\partial x_i} + b\frac{\partial g}{\partial x_i}$$

for al smooth functions f, g and constants a, b.

Now let $\alpha = \sum_I f_I\,dx_I$ and $\beta = \sum_J g_J\,dx_J$. The Leibniz rule for functions and Proposition 2.1 give

$$\begin{aligned}
d(\alpha\beta) &= \sum_{I,J} d(f_I g_J)\,dx_I\,dx_J = \sum_{I,J}(f_I\,dg_J + g_J\,df_I)\,dx_I\,dx_J \\
&= \sum_{I,J}\big(df_I\,dx_I(g_J\,dx_J) + (-1)^k f_I\,dx_I(dg_J\,dx_J)\big) \\
&= (d\alpha)\beta + (-1)^k\alpha\,d\beta,
\end{aligned}$$

which proves part (ii). QED

Here is one of the most curious properties of the exterior derivative.

2.6. PROPOSITION. $d(d\alpha) = 0$ *for any form* α. *In short,*

$$\boxed{d^2 = 0.}$$

PROOF. Let $\alpha = \sum_I f_I \, dx_I$. Then

$$d(d\alpha) = d\Bigl(\sum_I \sum_{i=1}^n \frac{\partial f_I}{\partial x_i} \, dx_i \, dx_I\Bigr) = \sum_I \sum_{i=1}^n d\Bigl(\frac{\partial f_I}{\partial x_i}\Bigr) \, dx_i \, dx_I.$$

Applying the formula of Example 2.3 (replacing f_i with $\partial f_I / \partial x_i$) we find

$$\sum_{i=1}^n d\Bigl(\frac{\partial f_I}{\partial x_i}\Bigr) \, dx_i = \sum_{1 \le i < j \le n} \Bigl(\frac{\partial^2 f_I}{\partial x_i \partial x_j} - \frac{\partial^2 f_I}{\partial x_j \partial x_i}\Bigr) \, dx_i \, dx_j = 0,$$

because for any smooth (indeed, C^2) function f the mixed partials $\partial^2 f / \partial x_i \partial x_j$ and $\partial^2 f / \partial x_j \partial x_i$ are equal. Hence $d(d\alpha) = 0$. QED

2.3. Closed and exact forms

A form α is *closed* if $d\alpha = 0$. It is *exact* if $\alpha = d\beta$ for some form β (of degree one less).

2.7. PROPOSITION. *Every exact form is closed.*

PROOF. If $\alpha = d\beta$ then $d\alpha = d(d\beta) = 0$ by Proposition 2.6. QED

2.8. EXAMPLE. $-y \, dx + x \, dy$ is not closed and therefore cannot be exact. On the other hand $y \, dx + x \, dy$ is closed. It is also exact, because $d(xy) = y \, dx + x \, dy$. For a 0-form (function) f on $\mathbf{R}^n$ to be closed all its partial derivatives must vanish, which means it is constant. A nonzero constant function is not exact, because forms of degree -1 are 0.

Is every closed form of positive degree exact? This question has interesting ramifications, which we shall explore in Chapters 4, 5 and 10. Amazingly, the answer depends strongly on the topology, that is the qualitative "shape", of the domain of definition of the form.

Let us consider the simplest case of a 1-form $\alpha = \sum_{i=1}^n f_i \, dx_i$. Determining whether α is exact means solving the equation $dg = \alpha$ for the function g. This amounts to

$$\frac{\partial g}{\partial x_1} = f_1, \qquad \frac{\partial g}{\partial x_2} = f_2, \qquad \dots, \qquad \frac{\partial g}{\partial x_n} = f_n, \tag{2.5}$$

a system of *first-order partial differential equations*. Finding a solution is sometimes called *integrating* the system. By Proposition 2.7 this is not possible unless α is closed. By the formula in Example 2.3 α is closed if and only if

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}$$

for all $1 \le i < j \le n$. These identities must be satisfied for the system (2.5) to be solvable and are therefore called the *integrability conditions* for the system.

2.9. EXAMPLE. Let $\alpha = y \, dx + (z \cos yz + x) \, dy + y \cos yz \, dz$. Then

$$\begin{aligned} d\alpha = dy \, dx + \bigl(z(-y \sin yz) + \cos yz\bigr) \, dz \, dy + dx \, dy \\ + \bigl(y(-z \sin yz) + \cos yz\bigr) \, dy \, dz = 0, \end{aligned}$$

so α is closed. Is α exact? Let us solve the equations

$$\frac{\partial g}{\partial x} = y, \quad \frac{\partial g}{\partial y} = z\cos yz + x, \quad \frac{\partial g}{\partial z} = y\cos yz$$

by successive integration. The first equation gives $g = yx + c(y,z)$, where c is a function of y and z only. Substituting into the second equation gives $\partial c/\partial y = z\cos yz$, so $c = \sin yz + k(z)$. Substituting into the third equation gives $k' = 0$, so k is a constant. So $g = yx + \sin yz$ is a solution and therefore α is exact.

This method works always for a 1-form defined on all of $\mathbf{R}^n$. (See Exercise 2.6.) Hence every closed 1-form on $\mathbf{R}^n$ is exact.

2.10. EXAMPLE. The 1-form on $\mathbf{R}^2 - \{\mathbf{0}\}$ defined by

$$\alpha = -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy = \frac{-y\,dx + x\,dy}{x^2+y^2}.$$

is called the *angle form* for reasons that will become clear in Section 4.3. From

$$\frac{\partial}{\partial x}\frac{x}{x^2+y^2} = \frac{y^2-x^2}{(x^2+y^2)^2}, \qquad \frac{\partial}{\partial y}\frac{y}{x^2+y^2} = \frac{x^2-y^2}{(x^2+y^2)^2}$$

it follows that the angle form is closed. This example is continued in Examples 4.1 and 4.6, where we shall see that this form is *not* exact.

For a 2-form $\alpha = \sum_{1\le i<j\le n} f_{i,j}\,dx_i\,dx_j$ and a 1-form $\beta = \sum_{i=1}^n g_i\,dx_i$ the equation $d\beta = \alpha$ amounts to the system

$$\frac{\partial g_j}{\partial x_i} - \frac{\partial g_i}{\partial x_j} = f_{i,j}. \tag{2.6}$$

By the formula in Example 2.4 the integrability condition $d\alpha = 0$ comes down to

$$\frac{\partial f_{i,j}}{\partial x_k} - \frac{\partial f_{i,k}}{\partial x_j} + \frac{\partial f_{j,k}}{\partial x_i} = 0$$

for all $1 \le i < j < k \le n$. We shall learn how to solve the system (2.6), and its higher-degree analogues, in Example 10.18.

2.4. The Hodge star operator

The binomial coefficient $\binom{n}{k}$ is the number of ways of selecting k (unordered) objects from a collection of n objects. Equivalently, $\binom{n}{k}$ is the number of ways of partitioning a pile of n objects into a pile of k objects and a pile of $n-k$ objects. Thus we see that

$$\binom{n}{k} = \binom{n}{n-k}.$$

This means that in a certain sense there are as many k-forms as $n-k$-forms. In fact, there is a natural way to turn k-forms into $n-k$-forms. This is the *Hodge star operator*. Hodge star of α is denoted by $*\alpha$ (or sometimes α^*) and is defined as follows. If $\alpha = \sum_I f_I\,dx_I$, then

$$*\alpha = \sum_I f_I(*dx_I),$$

with

$$*dx_I = \varepsilon_I\,dx_{I^c}.$$

Here, for any increasing multi-index I, I^c denotes the *complementary* increasing multi-index, which consists of all numbers between 1 and n that do not occur in I. The factor ε_I is a sign,

$$\varepsilon_I = \begin{cases} 1 & \text{if } dx_I\, dx_{I^c} = dx_1\, dx_2 \cdots dx_n, \\ -1 & \text{if } dx_I\, dx_{I^c} = -dx_1\, dx_2 \cdots dx_n. \end{cases}$$

In other words, $*dx_I$ is the product of all the dx_j's that do not occur in dx_I, times a factor ± 1 which is chosen in such a way that $dx_I(*dx_I)$ is the volume form:

$$dx_I(*dx_I) = dx_1\, dx_2 \cdots dx_n.$$

2.11. EXAMPLE. Let $n = 6$ and $I = (2, 6)$. Then $I^c = (1, 3, 4, 5)$, so $dx_I = dx_2\, dx_6$ and $dx_{I^c} = dx_1\, dx_3\, dx_4\, dx_5$. Therefore

$$\begin{aligned} dx_I\, dx_{I^c} &= dx_2\, dx_6\, dx_1\, dx_3\, dx_4\, dx_5 \\ &= dx_1\, dx_2\, dx_6\, dx_3\, dx_4\, dx_5 = -dx_1\, dx_2\, dx_3\, dx_4\, dx_5\, dx_6, \end{aligned}$$

which shows that $\varepsilon_I = -1$. Hence $*(dx_2\, dx_6) = -dx_1\, dx_3\, dx_4\, dx_5$.

2.12. EXAMPLE. On $\mathbf{R}^2$ we have $*dx = dy$ and $*dy = -dx$. On $\mathbf{R}^3$ we have

$$\begin{aligned} *dx &= dy\, dz, & *(dx\, dy) &= dz, \\ *dy &= -dx\, dz = dz\, dx, & *(dx\, dz) &= -dy, \\ *dz &= dx\, dy, & *(dy\, dz) &= dx. \end{aligned}$$

(This is the reason that 2-forms on $\mathbf{R}^3$ are sometimes written as $f\, dx\, dy + g\, dz\, dx + h\, dy\, dz$, in contravention of our usual rule to write the variables in increasing order. In higher dimensions it is better to stick to the rule.) On $\mathbf{R}^4$ we have

$$\begin{aligned} *dx_1 &= dx_2\, dx_3\, dx_4, & *dx_3 &= dx_1\, dx_2\, dx_4, \\ *dx_2 &= -dx_1\, dx_3\, dx_4, & *dx_4 &= -dx_1\, dx_2\, dx_3, \end{aligned}$$

and

$$\begin{aligned} *(dx_1\, dx_2) &= dx_3\, dx_4, & *(dx_2\, dx_3) &= dx_1\, dx_4, \\ *(dx_1\, dx_3) &= -dx_2\, dx_4, & *(dx_2\, dx_4) &= -dx_1\, dx_3, \\ *(dx_1\, dx_4) &= dx_2\, dx_3, & *(dx_3\, dx_4) &= dx_1\, dx_2. \end{aligned}$$

On $\mathbf{R}^n$ we have $*1 = dx_1\, dx_2 \cdots dx_n$, $*(dx_1\, dx_2 \cdots dx_n) = 1$, and

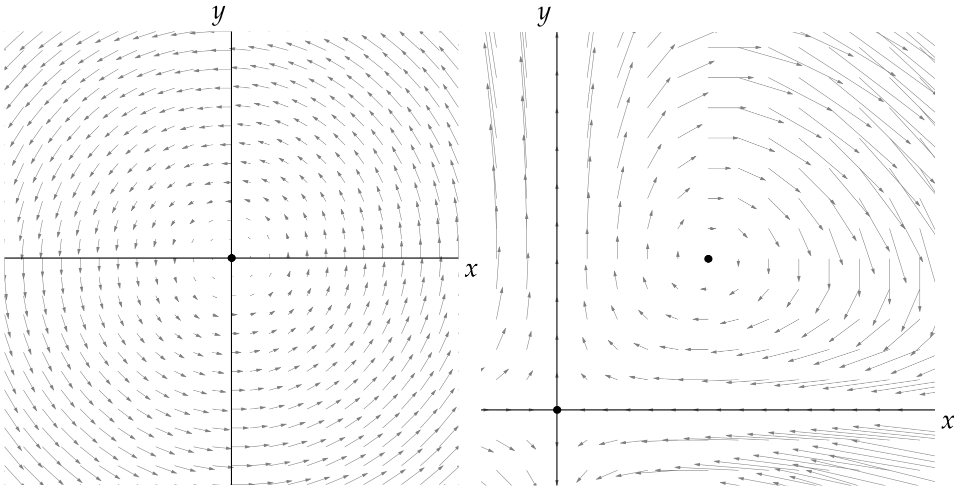
$$\begin{aligned} *dx_i &= (-1)^{i+1} dx_1\, dx_2 \cdots \widehat{dx_i} \cdots dx_n && \text{for } 1 \le i \le n, \\ *(dx_i\, dx_j) &= (-1)^{i+j+1} dx_1\, dx_2 \cdots \widehat{dx_i} \cdots \widehat{dx_j} \cdots dx_n && \text{for } 1 \le i < j \le n. \end{aligned}$$

2.5. div, grad and curl

A *vector field* on an open subset U of $\mathbf{R}^n$ is a smooth map $\mathbf{F}\colon U \to \mathbf{R}^n$. We can write $\mathbf{F}$ in components as

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \\ \vdots \\ F_n(\mathbf{x}) \end{pmatrix},$$

or alternatively as $\mathbf{F} = \sum_{i=1}^{n} F_i \mathbf{e}_i$, where $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ are the standard basis vectors of $\mathbf{R}^n$. Vector fields in the plane can be plotted by placing the vector $\mathbf{F}(\mathbf{x})$ with its tail at the point $\mathbf{x}$. The diagrams below represent the vector fields $-y\mathbf{e}_1 + x\mathbf{e}_2$ and $(-x + xy)\mathbf{e}_1 + (y - xy)\mathbf{e}_2$ (which you may recognize from Exercise 1.10). The arrows have been shortened so as not to clutter the pictures. The black dots are the *zeroes* of the vector fields (i.e. points $\mathbf{x}$ where $\mathbf{F}(\mathbf{x}) = \mathbf{0}$).



We can turn $\mathbf{F}$ into a 1-form α by using the F_i as coefficients: $\alpha = \sum_{i=1}^{n} F_i \, dx_i$. For instance, the 1-form $\alpha = -y\,dx + x\,dy$ corresponds to the vector field $\mathbf{F} = -y\mathbf{e}_1 + x\mathbf{e}_2$. Let us introduce the symbolic notation

$$d\mathbf{x} = \begin{pmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{pmatrix},$$

which we will think of as a vector-valued 1-form. Then we can write $\alpha = \mathbf{F} \cdot d\mathbf{x}$. Clearly, $\mathbf{F}$ is determined by α and vice versa. Thus vector fields and 1-forms are symbiotically associated to one another.

$$\boxed{\text{vector field } \mathbf{F} \longleftrightarrow \text{1-form } \alpha: \qquad \alpha = \mathbf{F} \cdot d\mathbf{x}.}$$

Intuitively, the vector-valued 1-form $d\mathbf{x}$ represents an infinitesimal displacement. If $\mathbf{F}$ represents a force field, such as gravity or an electric force acting on a particle, then $\alpha = \mathbf{F} \cdot d\mathbf{x}$ represents the *work* done by the force when the particle is displaced by an amount $d\mathbf{x}$. (If the particle travels along a path, the total work done by the force is found by *integrating* α along the path. We shall see how to do this in Section 4.1.)

The correspondence between vector fields and 1-forms behaves in an interesting way with respect to exterior differentiation and the Hodge star operator. For each function f the 1-form $df = \sum_{i=1}^n (\partial f/\partial x_i)\, dx_i$ is associated to the vector field

$$\operatorname{grad} f = \sum_{i=1}^n \frac{\partial f}{\partial x_i}\mathbf{e}_i = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

This vector field is called the *gradient* of f. (Equivalently, we can view $\operatorname{grad} f$ as the transpose of the Jacobi matrix of f.)

$$\boxed{\operatorname{grad} f \longleftrightarrow df: \qquad df = \operatorname{grad} f \cdot d\mathbf{x}.}$$

Starting with a vector field $\mathbf{F}$ and letting $\alpha = \mathbf{F} \cdot d\mathbf{x}$, we find

$$*\alpha = \sum_{i=1}^n F_i(*dx_i) = \sum_{i=1}^n F_i(-1)^{i+1}\, dx_1\, dx_2 \cdots \widehat{dx_i} \cdots dx_n,$$

Using the vector-valued $n-1$-form

$$*d\mathbf{x} = \begin{pmatrix} *dx_1 \\ *dx_2 \\ \vdots \\ *dx_n \end{pmatrix} = \begin{pmatrix} dx_2\, dx_3 \cdots dx_n \\ -dx_1\, dx_3 \cdots dx_n \\ \vdots \\ (-1)^{n+1} dx_1\, dx_2 \cdots dx_{n-1} \end{pmatrix}$$

we can also write $*\alpha = \mathbf{F} \cdot *d\mathbf{x}$. Intuitively, the vector-valued $n-1$-form $*d\mathbf{x}$ represents an infinitesimal $n-1$-dimensional hypersurface perpendicular to $d\mathbf{x}$. (This point of view will be justified in Section 8.3, after the proof of Theorem 8.14.) In fluid mechanics, the flow of a fluid or gas in $\mathbf{R}^n$ is represented by a vector field $\mathbf{F}$. The $n-1$-form $*\alpha$ then represents the *flux*, that is the amount of material passing through the hypersurface $*d\mathbf{x}$ per unit time. (The total amount of fluid passing through a hypersurface S is found by *integrating* α over S. We shall see how to do this in Section 5.1.) We have

$$\begin{aligned} d{*\alpha} = d(\mathbf{F} \cdot *d\mathbf{x}) &= \sum_{i=1}^n \frac{\partial F_i}{\partial x_i}(-1)^{i+1}\, dx_i\, dx_1\, dx_2 \cdots \widehat{dx_i} \cdots dx_n \\ &= \sum_{i=1}^n \frac{\partial F_i}{\partial x_i}\, dx_1\, dx_2 \cdots dx_i \cdots dx_n = \left(\sum_{i=1}^n \frac{\partial F_i}{\partial x_i}\right) dx_1\, dx_2 \cdots dx_n. \end{aligned}$$

The function $\operatorname{div} \mathbf{F} = \sum_{i=1}^n \partial F_i/\partial x_i$ is the *divergence* of $\mathbf{F}$. Thus if $\alpha = \mathbf{F} \cdot d\mathbf{x}$, then

$$d{*\alpha} = d(\mathbf{F} \cdot *d\mathbf{x}) = \operatorname{div} \mathbf{F}\, dx_1\, dx_2 \cdots dx_n.$$

An alternative way of writing this identity is obtained by applying $*$ to both sides, which gives

$$\boxed{\operatorname{div} \mathbf{F} = *d{*\alpha}.}$$

A very different identity is found by first applying d and then $*$ to α:

$$d\alpha = \sum_{i,j=1}^n \frac{\partial F_i}{\partial x_j}\, dx_j\, dx_i = \sum_{1 \le i < j \le n} \left(\frac{\partial F_j}{\partial x_i} - \frac{\partial F_i}{\partial x_j}\right) dx_i\, dx_j,$$

and hence

$$*d\alpha = \sum_{1\le i<j\le n} (-1)^{i+j+1}\Big(\frac{\partial F_j}{\partial x_i} - \frac{\partial F_i}{\partial x_j}\Big)\, dx_1\, dx_2 \cdots \widehat{dx_i} \cdots \widehat{dx_j} \cdots dx_n.$$

In three dimensions $*d\alpha$ is a 1-form and so is associated to a vector field, namely

$$\operatorname{curl}\mathbf{F} = \Big(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}\Big)\mathbf{e}_1 - \Big(\frac{\partial F_3}{\partial x_1} - \frac{\partial F_1}{\partial x_3}\Big)\mathbf{e}_2 + \Big(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\Big)\mathbf{e}_3,$$

the *curl* of $\mathbf{F}$. Thus, for $n = 3$, if $\alpha = \mathbf{F}\cdot d\mathbf{x}$, then

$$\boxed{\operatorname{curl}\mathbf{F}\cdot d\mathbf{x} = *d\alpha.}$$

You need not memorize every detail of this discussion. The point is rather to remember that exterior differentiation in combination with the Hodge star unifies and extends to arbitrary dimensions the classical differential operators of vector calculus.

Exercises

2.1. Compute the exterior derivative of the following forms. Recall that a hat indicates that a term has to be omitted.

(i) $e^{xyz}\, dx$.
(ii) $\sum_{i=1}^n x_i^2\, dx_1 \cdots \widehat{dx_i} \cdots dx_n$.
(iii) $\|\mathbf{x}\|^p \sum_{i=1}^n (-1)^{i+1} x_i\, dx_1 \cdots \widehat{dx_i} \cdots dx_n$, where p is a real constant. For what values of p is this form closed?

2.2. Consider the forms $\alpha = x\,dx - y\,dy$, $\beta = z\,dx\,dy + x\,dy\,dz$ and $\gamma = z\,dy$ on $\mathbf{R}^3$. Calculate

(i) $\alpha\beta$, $\alpha\beta\gamma$;
(ii) $d\alpha$, $d\beta$, $d\gamma$.

2.3. Write the coordinates on $\mathbf{R}^{2n}$ as $(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$. Let

$$\omega = dx_1\,dy_1 + dx_2\,dy_2 + \cdots + dx_n\,dy_n = \sum_{i=1}^n dx_i\,dy_i.$$

Compute $\omega^n = \omega\omega\cdots\omega$ (n-fold product). First work out the cases $n = 1, 2, 3$.

2.4. Write the coordinates on $\mathbf{R}^{2n+1}$ as $(x_1, y_1, x_2, y_2, \dots, x_n, y_n, z)$. Let

$$\alpha = dz + x_1\,dy_1 + x_2\,dy_2 + \cdots + x_n\,dy_n = dz + \sum_{i=1}^n x_i\,dy_i.$$

Compute $\alpha(d\alpha)^n = \alpha(d\alpha\,d\alpha\cdots d\alpha)$. First work out the cases $n = 1, 2, 3$.

2.5. Check that each of the following forms $\alpha \in \Omega^1(\mathbf{R}^3)$ is closed and find a function g such that $dg = \alpha$.

(i) $\alpha = (ye^{xy} - z\sin(xz))\,dx + (xe^{xy} + z^2)\,dy + (-x\sin(xz) + 2yz + 3z^2)\,dz$.
(ii) $\alpha = 2xy^3z^4\,dx + (3x^2y^2z^4 - ze^y\sin(ze^y))\,dy + (4x^2y^3z^3 - e^y\sin(ze^y) + e^z)\,dz$.

2.6. Let $\alpha = \sum_{i=1}^n f_i\,dx_i$ be a *closed* C^∞ 1-form on $\mathbf{R}^n$. Define a function g by

$$g(\mathbf{x}) = \int_0^{x_1} f_1(t, x_2, x_3, \dots, x_n)\,dt + \int_0^{x_2} f_2(0, t, x_3, x_4, \dots, x_n)\,dt + \int_0^{x_3} f_3(0, 0, t, x_4, x_5, \dots, x_n)\,dt + \cdots + \int_0^{x_n} f_n(0, 0, \dots, 0, t)\,dt.$$

Show that $dg = \alpha$. (Apply the fundamental theorem of calculus, formula (B.3), differentiate under the integral sign and don't forget to use $d\alpha = 0$.)

2.7. Let $\alpha = \sum_{i=1}^{n} f_i\, dx_i$ be a *closed* 1-form whose coefficients f_i are smooth functions defined on $\mathbf{R}^n - \{\mathbf{0}\}$ that are all homogeneous of the same degree $p \neq -1$. Let

$$g(\mathbf{x}) = \frac{1}{p+1} \sum_{i=1}^{n} x_i f_i(\mathbf{x}).$$

Show that $dg = \alpha$. (Use $d\alpha = 0$ and apply the identity proved in Exercise B.5 to each f_i.)

2.8. Let α and β be closed forms. Prove that $\alpha\beta$ is also closed.

2.9. Let α be closed and β exact. Prove that $\alpha\beta$ is exact.

2.10. Calculate $*\alpha$, $*\beta$, $*\gamma$, $*(\alpha\beta)$, where α, β and γ are as in Exercise 2.2.

2.11. Consider the form $\alpha = -x_2^2\, dx_1 + x_1^2\, dx_2$ on $\mathbf{R}^2$.

(i) Find $*\alpha$ and $*d{*}d\alpha$.
(ii) Repeat the calculation, regarding α as a form on $\mathbf{R}^3$.
(iii) Again repeat the calculation, now regarding α as a form on $\mathbf{R}^4$.

2.12. Prove that $**\alpha = (-1)^{kn+k}\alpha$ for every k-form α on $\mathbf{R}^n$.

2.13. Let $\alpha = \sum_I a_I\, dx_I$ and $\beta = \sum_I b_I\, dx_I$ be *constant* k-forms, i.e. with constant coefficients a_I and b_I. (We also assume, as usual, that the multi-indices I are increasing.) The *inner product* of α and β is the number defined by

$$(\alpha, \beta) = \sum_I a_I b_I.$$

Prove the following assertions.

(i) The dx_I form an orthonormal basis of the space of constant k-forms.
(ii) $(\alpha, \alpha) \geq 0$ for all α and $(\alpha, \alpha) = 0$ if and only if $\alpha = 0$.
(iii) $\alpha(*\beta) = (\alpha, \beta)\, dx_1\, dx_2 \cdots dx_n$.
(iv) $\alpha(*\beta) = \beta(*\alpha)$.
(v) The Hodge star operator is orthogonal, i.e. $(\alpha, \beta) = (*\alpha, *\beta)$.

2.14. The *Laplacian* of a smooth function on an open subset of $\mathbf{R}^n$ is defined by

$$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}.$$

Prove the following formulas.

(i) $\Delta f = *d{*}df$.
(ii) $\Delta(fg) = (\Delta f)g + f\Delta g + 2*(df(*dg))$. (Use Exercise 2.13(iv).)

2.15. Let $f\colon \mathbf{R}^n \to \mathbf{R}$ be a function and let $\alpha = f\, dx_i$.

(i) Calculate $d{*}d{*}\alpha$.
(ii) Calculate $*d{*}d\alpha$.
(iii) Show that $d{*}d{*}\alpha + (-1)^n{*}d{*}d\alpha = (\Delta f)\, dx_i$, where Δ is the Laplacian defined in Exercise 2.14.

2.16. (i) Let U be an open subset of $\mathbf{R}^n$ and let $f\colon U \to \mathbf{R}$ be a function satisfying $\operatorname{grad} f(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x}$ in U. On U define a vector field $\mathbf{n}$, an $n-1$-form ν and a 1-form α by

$$\begin{aligned} \mathbf{n}(\mathbf{x}) &= \|\operatorname{grad} f(\mathbf{x})\|^{-1} \operatorname{grad} f(\mathbf{x}), \\ \nu &= \mathbf{n} \cdot *d\mathbf{x}, \\ \alpha &= \|\operatorname{grad} f(\mathbf{x})\|^{-1} df. \end{aligned}$$

Prove that $dx_1\, dx_2 \cdots dx_n = \alpha\nu$ on U.

(ii) Let $r\colon \mathbf{R}^n \to \mathbf{R}$ be the function $r(\mathbf{x}) = \|\mathbf{x}\|$ (distance to the origin). Deduce from part (i) that $dx_1\, dx_2 \cdots dx_n = (dr)\nu$ on $\mathbf{R}^n - \{\mathbf{0}\}$, where $\nu = \|\mathbf{x}\|^{-1}\mathbf{x} \cdot *d\mathbf{x}$.

2.17. The *Minkowski* or *relativistic* inner product on $\mathbf{R}^{n+1}$ is given by

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i y_i - x_{n+1} y_{n+1}.$$

A vector $\mathbf{x} \in \mathbf{R}^{n+1}$ is *spacelike* if $(\mathbf{x}, \mathbf{x}) > 0$, *lightlike* if $(\mathbf{x}, \mathbf{x}) = 0$, and *timelike* if $(\mathbf{x}, \mathbf{x}) < 0$.

(i) Give examples of (nonzero) vectors of each type.
(ii) Show that for every $\mathbf{x} \neq \mathbf{0}$ there is a $\mathbf{y}$ such that $(\mathbf{x}, \mathbf{y}) \neq 0$.

A Hodge star operator corresponding to this inner product is defined as follows: if $\alpha = \sum_I f_I \, dx_I$, then

$$*\alpha = \sum_I f_I (*dx_I),$$

with

$$*dx_I = \begin{cases} \varepsilon_I dx_{I^c} & \text{if } I \text{ contains } n+1, \\ -\varepsilon_I dx_{I^c} & \text{if } I \text{ does not contain } n+1. \end{cases}$$

(Here ε_I and I^c are as in the definition of the ordinary Hodge star.)

(iii) Find $*1$, $*dx_i$ for $1 \leq i \leq n+1$, and $*(dx_1 \, dx_2 \cdots dx_n)$.
(iv) Compute the "relativistic Laplacian" (usually called the d'Alembertian or wave operator) $*d*df$ for any smooth function f on $\mathbf{R}^{n+1}$.
(v) For $n = 3$ (ordinary space-time) find $*(dx_i \, dx_j)$ for $1 \leq i < j \leq 4$.

2.18. One of the greatest advances in theoretical physics of the nineteenth century was Maxwell's formulation of the equations of electromagnetism:

$$\begin{aligned} \operatorname{curl} \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} && \text{(Faraday's Law)}, \\ \operatorname{curl} \mathbf{H} &= \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} && \text{(Ampère's Law)}, \\ \operatorname{div} \mathbf{D} &= 4\pi\rho && \text{(Gauß' Law)}, \\ \operatorname{div} \mathbf{B} &= 0 && \text{(no magnetic monopoles)}. \end{aligned}$$

Here c is the speed of light, $\mathbf{E}$ is the electric field, $\mathbf{H}$ is the magnetic field, $\mathbf{J}$ is the density of electric current, ρ is the density of electric charge, $\mathbf{B}$ is the magnetic induction and $\mathbf{D}$ is the dielectric displacement. $\mathbf{E}$, $\mathbf{H}$, $\mathbf{J}$, $\mathbf{B}$ and $\mathbf{D}$ are vector fields and ρ is a function on $\mathbf{R}^3$ and all depend on time t. The Maxwell equations look particularly simple in differential form notation, as we shall now see. In space-time $\mathbf{R}^4$ with coordinates (x_1, x_2, x_3, x_4), where $x_4 = ct$, introduce forms

$$\begin{aligned} \alpha &= (E_1 \, dx_1 + E_2 \, dx_2 + E_3 \, dx_3) \, dx_4 + B_1 \, dx_2 \, dx_3 + B_2 \, dx_3 \, dx_1 + B_3 \, dx_1 \, dx_2, \\ \beta &= -(H_1 \, dx_1 + H_2 \, dx_2 + H_3 \, dx_3) \, dx_4 + D_1 \, dx_2 \, dx_3 + D_2 \, dx_3 \, dx_1 + D_3 \, dx_1 \, dx_2, \\ \gamma &= \frac{1}{c} (J_1 \, dx_2 \, dx_3 + J_2 \, dx_3 \, dx_1 + J_3 \, dx_1 \, dx_2) \, dx_4 - \rho \, dx_1 \, dx_2 \, dx_3. \end{aligned}$$

(i) Show that Maxwell's equations are equivalent to

$$\begin{aligned} d\alpha &= 0, \\ d\beta + 4\pi\gamma &= 0. \end{aligned}$$

(ii) Conclude that γ is closed and that $\operatorname{div} \mathbf{J} + \partial\rho/\partial t = 0$.
(iii) In vacuum one has $\mathbf{E} = \mathbf{D}$ and $\mathbf{H} = \mathbf{B}$. Show that in vacuum $\beta = *\alpha$, the relativistic Hodge star of α defined in Exercise 2.17.
(iv) *Free space* is a vacuum without charges or currents. Show that the Maxwell equations in free space are equivalent to $d\alpha = d{*}\alpha = 0$.

(v) Let $f, g\colon \mathbf{R} \to \mathbf{R}$ be any smooth functions and define

$$\mathbf{E}(\mathbf{x}) = \begin{pmatrix} 0 \\ f(x_1 - x_4) \\ g(x_1 - x_4) \end{pmatrix}, \qquad \mathbf{B}(\mathbf{x}) = \begin{pmatrix} 0 \\ -g(x_1 - x_4) \\ f(x_1 - x_4) \end{pmatrix}.$$

Show that the corresponding 2-form α satisfies the free Maxwell equations $d\alpha = d{*}\alpha = 0$. Such solutions are called *electromagnetic waves*. Explain why. In what direction do these waves travel?

CHAPTER 3

Pulling back forms

3.1. Determinants

The determinant of a square matrix is the oriented volume of the block (parallelepiped) spanned by its column vectors. It is therefore not surprising that differential forms are closely related to determinants. This section is a review of some fundamental facts concerning determinants. Let

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{pmatrix}$$

be an $n \times n$-matrix with column vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. Its determinant is variously denoted by

$$\det A = \det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \det(a_{i,j})_{1 \le i,j \le n} = \begin{vmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{vmatrix}.$$

Expansion on the first column. You have probably seen the following definition of the determinant:

$$\det A = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A^{i,1}.$$

Here $A^{i,j}$ denotes the $(n-1) \times (n-1)$-matrix obtained from A by striking out the i-th row and the j-th column. This is a recursive definition, which reduces the calculation of any determinant to that of determinants of smaller size. (The recursion starts at $n = 1$; the determinant of a 1×1-matrix (a) is simply defined to be the number a.) It is a useful rule, but it has two serious flaws: first, it is extremely inefficient computationally (except for matrices containing lots of zeroes), and second, it obscures the relationship with volumes of parallelepipeds.

Axioms. A far better definition is available. The determinant can be completely characterized by three simple laws, which make good sense in view of its geometrical significance and which comprise an efficient algorithm for calculating any determinant.

3.1. DEFINITION. A *determinant* is a function det which assigns to every ordered n-tuple of vectors $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ a number $\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ subject to the following axioms:

(i) det is *multilinear* (i.e. linear in each column):

$$\begin{aligned}\det(\mathbf{a}_1, \mathbf{a}_2, \dots, c\mathbf{a}_i + c'\mathbf{a}_i', \dots, \mathbf{a}_n)\\ = c\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) + c'\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i', \dots, \mathbf{a}_n)\end{aligned}$$

for all scalars c, c' and all vectors $\mathbf{a}_1$, $\mathbf{a}_2, \dots$, $\mathbf{a}_i$, $\mathbf{a}_i', \dots$, $\mathbf{a}_n$;

(ii) det is *alternating* or *antisymmetric*:

$$\det(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n) = -\det(\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n)$$

for any $i \neq j$;

(iii) *normalization*: $\det(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 1$, where $\mathbf{e}_1$, $\mathbf{e}_2, \dots$, $\mathbf{e}_n$ are the standard basis vectors of $\mathbf{R}^n$.

We also write $\det A$ instead of $\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$, where A is the matrix whose columns are $\mathbf{a}_1$, $\mathbf{a}_2, \dots$, $\mathbf{a}_n$. Axiom (iii) lays down the value of $\det I$. Axioms (i) and (ii) govern the behaviour of oriented volumes under the elementary column operations on matrices. Recall that these operations come in three types: adding a multiple of any column of A to any other column (type I); multiplying a column by a nonzero constant (type II); and interchanging any two columns (type III). Type I does not affect the determinant, type II multiplies it by the corresponding constant, and type III causes a sign change. This can be restated as follows.

3.2. LEMMA. *If E is an elementary column operation, then* $\det(E(A)) = k\det A$, *where*

$$k = \begin{cases} 1 & \text{if } E \text{ is of type I,} \\ c & \text{if } E \text{ is of type II (multiplication of a column by } c\text{),} \\ -1 & \text{if } E \text{ is of type III.} \end{cases}$$

3.3. EXAMPLE. Identify the column operations applied at each step in the following calculation.

$$\begin{aligned}\begin{vmatrix} 1 & 1 & 1 \\ 4 & 10 & 9 \\ 1 & 5 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 4 & 6 & 5 \\ 1 & 4 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 4 & 1 & 5 \\ 1 & 1 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 3 & 1 & 2 \\ 0 & 1 & 0 \end{vmatrix}\\ = 2\begin{vmatrix} 1 & 0 & 0 \\ 3 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} = 2\begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} = -2\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -2.\end{aligned}$$

As this example suggests, the axioms (i)–(iii) suffice to calculate *any* $n \times n$-determinant. In other words, there is at most one function det which obeys these axioms. More precisely, we have the following result.

3.4. THEOREM (uniqueness of determinants). *Let* det *and* $\det'$ *be two functions satisfying Axioms* (i)–(iii). *Then* $\det A = \det' A$ *for all $n \times n$-matrices A.*

PROOF. Let $\mathbf{a}_1$, $\mathbf{a}_2, \dots$, $\mathbf{a}_n$ be the column vectors of A. Suppose first that A is *not* invertible. Then the columns of A are linearly dependent. For simplicity let us assume that the first column is a linear combination of the others: $\mathbf{a}_1 = c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$. Applying axioms (i) and (ii) we get

$$\det A = \sum_{i=2}^{n} c_i \det(\mathbf{a}_i, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) = 0,$$

and for the same reason $\det' A = 0$, so $\det A = \det' A$. Now assume that A is invertible. Then A is column equivalent to the identity matrix, i.e. it can be transformed to I by successive elementary column operations. Let $E_1, E_2, \dots, E_m$ be these elementary operations, so that $E_m E_{m-1} \cdots E_2 E_1(A) = I$. According to Lemma 3.2, each operation E_i has the effect of multiplying the determinant by a certain factor k_i, so axiom (iii) yields

$$1 = \det I = \det(E_m E_{m-1} \cdots E_2 E_1(A)) = k_m k_{m-1} \cdots k_2 k_1 \det A.$$

Applying the same reasoning to $\det' A$ we get $1 = k_m k_{m-1} \cdots k_2 k_1 \det' A$. Hence $\det A = 1/(k_1 k_2 \cdots k_m) = \det' A$. QED

3.5. REMARK (change of normalization). Suppose that $\det'$ is a function that satisfies the multilinearity axiom (i) and the antisymmetry axiom (ii) but is normalized differently: $\det' I = c$. Then the proof of Theorem 3.4 shows that $\det' A = c \det A$ for all $n \times n$-matrices A.

This result leaves an open question. We can calculate the determinant of any matrix by column reducing it to the identity matrix, but there are many different ways of performing this reduction. Do different column reductions lead to the same answer for the determinant? In other words, are the axioms (i)–(iii) consistent? We will answer this question by displaying an explicit formula for the determinant of any $n \times n$-matrix that does not involve any column reductions. Unlike Definition 3.1, this formula is not very practical for the purpose of calculating large determinants, but it has other uses, notably in the theory of differential forms.

3.6. THEOREM (existence of determinants). *Every $n \times n$-matrix A has a well-defined determinant. It is given by the formula*

$$\boxed{\det A = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}.}$$

This requires a little explanation. S_n stands for the collection of all *permutations* of the set $\{1, 2, \dots, n\}$. A permutation is a way of ordering the numbers $1, 2, \dots, n$. Permutations are usually written as row vectors containing each of these numbers exactly once. Thus for $n = 2$ there are only two permutations: $(1, 2)$ and $(2, 1)$. For $n = 3$ all possible permutations are

$$(1,2,3), \qquad (1,3,2), \qquad (2,1,3), \qquad (2,3,1), \qquad (3,1,2), \qquad (3,2,1).$$

For general n there are

$$n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1 = n!$$

permutations. An alternative way of thinking of a permutation is as a bijective (i.e. one-to-one and onto) map from the set $\{1, 2, \dots, n\}$ to itself. For example, for $n = 5$ a possible permutation is

$$(5, 3, 1, 2, 4),$$

and we think of this as a shorthand notation for the map σ given by $\sigma(1) = 5$, $\sigma(2) = 3$, $\sigma(3) = 1$, $\sigma(4) = 2$ and $\sigma(5) = 4$. The permutation $(1, 2, 3, \dots, n-1, n)$ then corresponds to the identity map on the set $\{1, 2, \dots, n\}$.

If σ is the identity permutation, then clearly $\sigma(i) < \sigma(j)$ whenever $i < j$. However, if σ is not the identity permutation, it cannot preserve the order in this way. An *inversion* in σ is any pair of numbers i and j such that $1 \le i < j \le n$ and

$\sigma(i) > \sigma(j)$. The *length* of σ, denoted by $l(\sigma)$, is the number of inversions in σ. A permutation is called *even* or *odd* according to whether its length is even, resp. odd. For instance, the permutation $(5, 3, 1, 2, 4)$ has length 6 and so is even. The *sign* of σ is

$$\operatorname{sign}(\sigma) = (-1)^{l(\sigma)} = \begin{cases} 1 & \text{if } \sigma \text{ is even,} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

Thus $\operatorname{sign}(5, 3, 1, 2, 4) = 1$. The permutations of $\{1, 2\}$ are $(1, 2)$, which has sign 1, and $(2, 1)$, which has sign -1, while for $n = 3$ we have the table below.

σ	$l(\sigma)$	$\operatorname{sign}(\sigma)$
$(1, 2, 3)$	0	1
$(1, 3, 2)$	1	-1
$(2, 1, 3)$	1	-1
$(2, 3, 1)$	2	1
$(3, 1, 2)$	2	1
$(3, 2, 1)$	3	-1

Thinking of permutations in S_n as bijective maps from $\{1, 2, \ldots, n\}$ to itself, we can form the composition $\sigma \circ \tau$ of any two permutations σ and τ in S_n. For permutations we usually write $\sigma\tau$ instead of $\sigma \circ \tau$ and call it the *product* of σ and τ. This is the permutation produced by *first* performing τ and *then* σ! For instance, if $\sigma = (5, 3, 1, 2, 4)$ and $\tau = (5, 4, 3, 2, 1)$, then

$$\tau\sigma = (1, 3, 5, 4, 2), \qquad \sigma\tau = (4, 2, 1, 3, 5).$$

A basic fact concerning signs, which we shall not prove here, is

$$\boxed{\operatorname{sign}(\sigma\tau) = \operatorname{sign}(\sigma)\operatorname{sign}(\tau).} \tag{3.1}$$

In particular, the product of two even permutations is even and the product of an even and an odd permutation is odd.

The determinant formula in Theorem 3.6 contains $n!$ terms, one for each permutation σ. Each term is a product which contains exactly one entry from each row and each column of A. For instance, for $n = 5$ the permutation $(5, 3, 1, 2, 4)$ contributes the term $a_{1,5}a_{2,3}a_{3,1}a_{4,2}a_{5,4}$. For 2×2- and 3×3-determinants Theorem 3.6 gives the well-known formulæ

$$\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1},$$

$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = a_{1,1}a_{2,2}a_{3,3} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} \\ + a_{1,3}a_{2,1}a_{3,2} - a_{1,3}a_{2,2}a_{3,1}.$$

PROOF OF THEOREM 3.6. We need to check that the right-hand side of the determinant formula in Theorem 3.6 obeys axioms (i)–(iii) of Definition 3.1. Let us for the moment denote the right-hand side by $f(A)$.

Axiom (i) is checked as follows: for every permutation σ the product

$$a_{1,\sigma(1)}a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

contains exactly one entry from each row and each column in A. So if we multiply the i-th row of A by c, each term in $f(A)$ is multiplied by c. Therefore

$$f(\mathbf{a}_1, \mathbf{a}_2, \dots, c\mathbf{a}_i, \dots, \mathbf{a}_n) = cf(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n).$$

Similarly,

$$f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i + \mathbf{a}_i', \dots, \mathbf{a}_n) = f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) + f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i', \dots, \mathbf{a}_n).$$

Axiom (ii) holds because if we interchange two columns in A, each term in $f(A)$ changes sign. To see this, let τ be the permutation in S_n that interchanges the two numbers i and j and leaves all others fixed. Then

$$\begin{aligned}
f(&\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) \\
&= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\tau\sigma(1)} a_{2,\tau\sigma(2)} \cdots a_{n,\tau\sigma(n)} \\
&= \sum_{\rho \in S_n} \operatorname{sign}(\tau\rho) a_{1,\rho(1)} a_{2,\rho(2)} \cdots a_{n,\rho(n)} && \text{substitute } \rho = \tau\sigma \\
&= \sum_{\rho \in S_n} \operatorname{sign}(\tau) \operatorname{sign}(\rho) a_{1,\rho(1)} a_{2,\rho(2)} \cdots a_{n,\rho(n)} && \text{by formula (3.1)} \\
&= -\sum_{\rho \in S_n} \operatorname{sign}(\rho) a_{1,\rho(1)} a_{2,\rho(2)} \cdots a_{n,\rho(n)} && \text{by Exercise 3.4} \\
&= -f(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n).
\end{aligned}$$

Finally, rule (iii) is correct because if $A = I$,

$$a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)} = \begin{cases} 1 & \text{if } \sigma = \text{identity}, \\ 0 & \text{otherwise}, \end{cases}$$

and therefore $f(I) = 1$. So f satisfies all three axioms for determinants. QED

Here are some further rules followed by determinants. Each can be deduced from Definition 3.1 or from Theorem 3.6. (Recall that the *transpose* of an $n \times n$-matrix $A = (a_{i,j})$ is the matrix A^T whose i, j-th entry is $a_{j,i}$.)

3.7. THEOREM. *Let A and B be $n \times n$-matrices.*

(i) $\det(AB) = \det A \det B$.
(ii) $\det A^T = \det A$.
(iii) *(Expansion on the j-th column)* $\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det A^{i,j}$ *for all $j = 1, 2, \dots, n$. Here $A^{i,j}$ denotes the $(n-1) \times (n-1)$-matrix obtained from A by striking out the i-th row and the j-th column.*
(iv) $\det A = a_{1,1} a_{2,2} \cdots a_{n,n}$ *if A is upper triangular (i.e. $a_{i,j} = 0$ for $i > j$).*

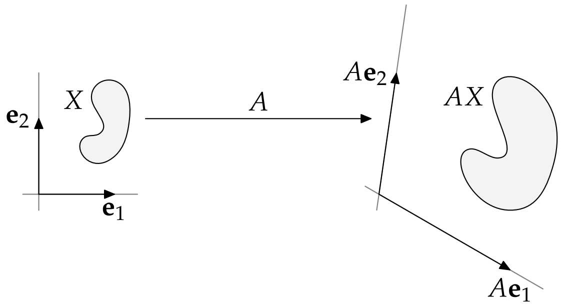
Volume change. We conclude this discussion with a slightly different geometric view of determinants. A square matrix A can be regarded as a linear map $A \colon \mathbf{R}^n \to \mathbf{R}^n$. The unit cube in $\mathbf{R}^n$,

$$[0, 1]^n = \{\, \mathbf{x} \in \mathbf{R}^n \mid 0 \le x_i \le 1 \quad \text{for } i = 1, 2, \dots, n \,\},$$

has n-dimensional volume 1. (For $n = 1$ it is usually called the *unit interval* and for $n = 2$ the *unit square*.) Its image $A([0,1]^n)$ under the map A is a parallelepiped with edges $A\mathbf{e}_1$, $A\mathbf{e}_2, \dots,$ $A\mathbf{e}_n$, the columns of A. Hence $A([0,1]^n)$

has n-dimensional volume $\operatorname{vol} A\big([0,1]^n\big) = |\det A| = |\det A| \operatorname{vol}[0,1]^n$. This rule generalizes as follows: if X is a measurable subset of $\mathbf{R}^n$, then

$$\operatorname{vol} A(X) = |\det A| \operatorname{vol} X.$$



So $|\det A|$ can be interpreted as a *volume change factor*. (A set is *measurable* if it has a well-defined, finite or infinite, n-dimensional volume. Explaining exactly what this means is rather hard, but it suffices for our purposes to know that all open and all closed subsets of $\mathbf{R}^n$ are measurable.)

3.2. Pulling back forms

By substituting new variables into a differential form we obtain a new form of the same degree but possibly in a different number of variables.

3.8. EXAMPLE. In Example 2.10 we defined the angle form on $\mathbf{R}^2 - \{\mathbf{0}\}$ to be

$$\alpha = \frac{-y\,dx + x\,dy}{x^2+y^2}.$$

By substituting $x = \cos t$ and $y = \sin t$ into the angle form we obtain the following 1-form on $\mathbf{R}$:

$$\frac{-\sin t\, d\cos t + \cos t\, d\sin t}{\cos^2 t + \sin^2 t} = \big((-\sin t)(-\sin t) + \cos^2 t\big)\,dt = dt.$$

We can take any k-form and substitute any number of variables into it to obtain a new k-form. This works as follows. Suppose α is a k-form defined on an open subset V of $\mathbf{R}^m$. Let us denote the coordinates on $\mathbf{R}^m$ by $y_1, y_2, \dots, y_m$ and let us write, as usual,

$$\alpha = \sum_I f_I\,dy_I,$$

where the functions f_I are defined on V. Suppose we want to substitute "new" variables $x_1, x_2, \dots, x_n$ and that the old variables are given in terms of the new by functions

$$\begin{aligned} y_1 &= \phi_1(x_1,\dots,x_n),\\ y_2 &= \phi_2(x_1,\dots,x_n),\\ &\;\;\vdots\\ y_m &= \phi_m(x_1,\dots,x_n). \end{aligned}$$

As usual we write $\mathbf{y} = \phi(\mathbf{x})$, where

$$\phi(\mathbf{x}) = \begin{pmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \\ \vdots \\ \phi_m(\mathbf{x}) \end{pmatrix}.$$

We assume that the functions ϕ_i are smooth and defined on a common domain U, which is an open subset of $\mathbf{R}^n$. We regard ϕ as a map from U to V. (In Example 3.8 we have $U = \mathbf{R}$, $V = \mathbf{R}^2 - \{\mathbf{0}\}$ and $\phi(t) = (\cos t, \sin t)$.) The *pullback* of α *along* ϕ is then the k-form $\phi^*\alpha$ on U obtained by substituting $y_i = \phi_i(x_1, \dots, x_n)$ for all i in the formula for α. That is to say, $\phi^*\alpha$ is defined by

$$\phi^*\alpha = \sum_I (\phi^* f_I)(\phi^* dy_I).$$

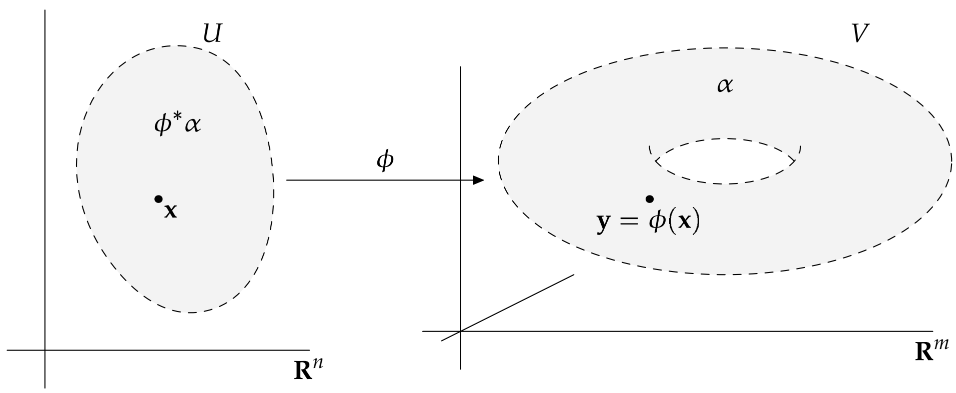
Here $\phi^* f_I$ is defined by

$$\phi^* f_I = f_I \circ \phi,$$

the composition of ϕ and f_I. This means $\phi^* f_I(\mathbf{x}) = f_I(\phi(\mathbf{x}))$; in other words, $\phi^* f_I$ is the function resulting from f_I by substituting $\mathbf{y} = \phi(\mathbf{x})$. The pullback $\phi^* dy_I$ is defined by replacing each y_i with ϕ_i. That is to say, if $I = (i_1, i_2, \dots, i_k)$ we put

$$\phi^* dy_I = \phi^*(dy_{i_1}\, dy_{i_2} \cdots dy_{i_k}) = d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k}.$$

The picture below is a schematic representation of the substitution process. The form $\alpha = \sum_I f_I\, dy_I$ is a k-form in $y_1, y_2, \dots, y_m$; its pullback $\phi^*\alpha = \sum_J g_J\, dx_J$ is a k-form in $x_1, x_2, \dots, x_n$. In Theorem 3.12 below we will give an explicit formula for the coefficients g_J in terms of f_I and ϕ.



3.9. EXAMPLE. The formula

$$\phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^3 x_2 \\ \ln(x_1 + x_2) \end{pmatrix}$$

defines a map $\phi\colon U \to \mathbf{R}^2$, where $U = \{\, \mathbf{x} \in \mathbf{R}^2 \mid x_1 + x_2 > 0 \,\}$. The components of ϕ are given by $\phi_1(x_1, x_2) = x_1^3 x_2$ and $\phi_2(x_1, x_2) = \ln(x_1 + x_2)$. Accordingly,

$$\begin{aligned}
\phi^* dy_1 &= d\phi_1 = d(x_1^3 x_2) = 3x_1^2 x_2\, dx_1 + x_1^3\, dx_2,\\
\phi^* dy_2 &= d\phi_2 = d\ln(x_1 + x_2) = (x_1 + x_2)^{-1}(dx_1 + dx_2),\\
\phi^*(dy_1\, dy_2) &= d\phi_1\, d\phi_2 = (3x_1^2 x_2\, dx_1 + x_1^3\, dx_2)(x_1 + x_2)^{-1}(dx_1 + dx_2)\\
&= \frac{3x_2^2 x_2 - x_1^3}{x_1 + x_2}\, dx_1\, dx_2.
\end{aligned}$$

Observe that the pullback operation turns k-forms on the target space V into k-forms on the source space U. Thus, while $\phi\colon U \to V$ is a map from U to V, ϕ^* is a map

$$\phi^*\colon \Omega^k(V) \to \Omega^k(U),$$

the opposite way from what you might naively expect. (Recall that $\Omega^k(U)$ stands for the collection of all k-forms on U.) The property that ϕ^* "turns the arrow around" is called *contravariance*. Pulling back forms is nicely compatible with the other operations that we learned about (except the Hodge star).

3.10. PROPOSITION. *Let $\phi\colon U \to V$ be a smooth map, where U is open in $\mathbf{R}^n$ and V is open in $\mathbf{R}^m$. The pullback operation is*

(i) *linear: $\phi^*(a\alpha + b\beta) = a\phi^*\alpha + b\phi^*\beta$;*
(ii) *multiplicative: $\phi^*(\alpha\beta) = (\phi^*\alpha)(\phi^*\beta)$;*
(iii) *natural: $\phi^*(\psi^*\alpha) = (\psi \circ \phi)^*\alpha$, where $\psi\colon V \to W$ is a second smooth map with W open in $\mathbf{R}^k$ and α a form on W.*

The term "natural" in property (iii) is a mathematical catchword meaning that a certain operation (in this case the pullback) is well-behaved with respect to composition of maps.

PROOF. If $\alpha = \sum_I f_I\, dy_I$ and $\beta = \sum_I g_I\, dy_I$ are two forms of the same degree, then $a\alpha + b\beta = \sum_I (af_I + bg_I)\, dy_I$, so

$$\phi^*(a\alpha + b\beta) = \sum_I \phi^*(af_I + bg_I)(\phi^* dy_I).$$

Now

$$\begin{aligned}
\phi^*(af_I + bg_I)(\mathbf{x}) = (af_I + bg_I)(\phi(\mathbf{x})) &= af_I(\phi(\mathbf{x})) + bg_I(\phi(\mathbf{x}))\\
&= a\phi^* f_I(\mathbf{x}) + b\phi^* g_I(\mathbf{x}),
\end{aligned}$$

so $\phi^*(a\alpha + b\beta) = \sum_I (a\phi^* f_I + b\phi^* g_I)(\phi^* dy_I) = a\phi^*\alpha + b\phi^*\beta$. This proves part (i). For the proof of part (ii) consider two forms $\alpha = \sum_I f_I\, dy_I$ and $\beta = \sum_J g_J\, dy_J$ (not necessarily of the same degree). Then $\alpha\beta = \sum_{I,J} f_I g_J\, dy_I\, dy_J$, so

$$\phi^*(\alpha\beta) = \sum_{I,J} \phi^*(f_I g_J)\phi^*(dy_I\, dy_J).$$

Now

$$\phi^*(f_I g_J)(\mathbf{x}) = f_I g_J(\phi(\mathbf{x})) = f_I(\phi(\mathbf{x}))g_J(\phi(\mathbf{x})) = (\phi^* f_I)(\phi^* g_J)(\mathbf{x}),$$

so $\phi^*(f_I g_J) = (\phi^* f_I)(\phi^* g_J)$. Furthermore,

$$\begin{aligned}\phi^*(dy_I\, dy_J) &= \phi^*(dy_{i_1}\phi^*(dy_{i_2}\cdots dy_{i_k}\, dy_{j_1}\cdots dy_{j_l})\\ &= d\phi_{i_1}\, d\phi_{i_2}\cdots d\phi_{i_k}\, d\phi_{j_1}\cdots d\phi_{j_l} = \phi^*(dy_I)(\phi^* dy_J),\end{aligned}$$

so

$$\begin{aligned}\phi^*(\alpha\beta) &= \sum_{I,J}(\phi^* f_I)(\phi^* g_J)\phi^*(dy_I)(\phi^* dy_J)\\ &= \Big(\sum_I(\phi^* f_I)\phi^*(dy_I)\Big)\Big(\sum_I(\phi^* g_J)(\phi^* dy_J)\Big) = (\phi^*\alpha)(\phi^*\beta),\end{aligned}$$

which establishes part (ii).

For the proof of property (iii) first consider a function f on W. Then

$$\begin{aligned}\phi^*(\psi^* f)(\mathbf{x}) &= (\psi^* f)(\phi(\mathbf{x})) = f(\psi(\phi(\mathbf{x}))) = f(\psi\circ\phi(\mathbf{x}))\\ &= (f\circ(\psi\circ\phi))(\mathbf{x}) = (\psi\circ\phi)^* f(\mathbf{x}),\end{aligned}$$

so $\phi^*(\psi^* f) = (\psi\circ\phi)^* f$. Next consider a 1-form $\alpha = dz_i$ on W, where $z_1, z_2, \ldots, z_k$ are the variables on $\mathbf{R}^k$. Then $\psi^*\alpha = d\psi_i = \sum_{j=1}^m \frac{\partial\psi_i}{\partial y_j}\, dy_j$, so

$$\begin{aligned}\phi^*(\psi^*\alpha) &= \sum_{j=1}^m \phi^*\Big(\frac{\partial\psi_i}{\partial y_j}\Big)\phi^* dy_j = \sum_{j=1}^m \phi^*\Big(\frac{\partial\psi_i}{\partial y_j}\Big)\, d\phi_j\\ &= \sum_{j=1}^m \phi^*\Big(\frac{\partial\psi_i}{\partial y_j}\Big)\sum_{l=1}^n \frac{\partial\phi_j}{\partial x_l}\, dx_l = \sum_{l=1}^n\Big(\sum_{j=1}^m \phi^*\Big(\frac{\partial\psi_i}{\partial y_j}\Big)\frac{\partial\phi_j}{\partial x_l}\Big)\, dx_l.\end{aligned}$$

By the chain rule, formula (B.6), the sum $\sum_{j=1}^m \phi^*(\partial\psi_i/\partial y_j)\partial\phi_j/\partial x_l$ is equal to $\partial(\phi^*\psi_i)/\partial x_l$. Therefore

$$\phi^*(\psi^*\alpha) = \sum_{l=1}^n \frac{\partial(\phi^*\psi_i)}{\partial x_l}\, dx_l = d(\phi^*\psi_i) = d((\psi\circ\phi)_i) = (\psi\circ\phi)^* dz_i = (\psi\circ\phi)^*\alpha.$$

Because every form on W is a sum of products of forms of type f and dz_i, property (iii) in general follows from the two special cases $\alpha = f$ and $\alpha = dz_i$. QED

Another application of the chain rule yields the following important result.

3.11. THEOREM. *Let $\phi\colon U \to V$ be a smooth map, where U is open in $\mathbf{R}^n$ and V is open in $\mathbf{R}^m$. Then $\phi^*(d\alpha) = d(\phi^*\alpha)$ for $\alpha \in \Omega^k(V)$. In short*

$$\boxed{\phi^* d = d\phi^*.}$$

PROOF. First let f be a function. Then

$$\begin{aligned}\phi^* df &= \phi^*\sum_{i=1}^m \frac{\partial f}{\partial y_i}\, dy_i = \sum_{i=1}^m \phi^*\Big(\frac{\partial f}{\partial y_i}\Big)\, d\phi_i = \sum_{i=1}^m \phi^*\Big(\frac{\partial f}{\partial y_i}\Big)\sum_{j=1}^n \frac{\partial\phi_i}{\partial x_j}\, dx_j\\ &= \sum_{j=1}^n\sum_{i=1}^m \phi^*\Big(\frac{\partial f}{\partial y_i}\Big)\frac{\partial\phi_i}{\partial x_j}\, dx_j.\end{aligned}$$

By the chain rule, formula (B.6), the quantity $\sum_{i=1}^m \phi^*(\partial f/\partial y_i)\partial\phi_i/\partial x_j$ is equal to $\partial(\phi^* f)/\partial x_j$. Hence

$$\phi^* df = \sum_{j=1}^n \frac{\partial(\phi^* f)}{\partial x_j}\, dx_j = d\phi^* f,$$

so the theorem is true for functions. Next let $\alpha = \sum_I f_I\, dy_I$. Then $d\alpha = \sum_I df_I\, dy_I$, so

$$\phi^* d\alpha = \sum_I \phi^*(df_I\, dy_I) = (\phi^* df_I)(\phi^* dy_I) = \sum_I d(\phi^* f_I)\, d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k},$$

because $\phi^* df_I = d\phi^* f_I$. On the other hand,

$$\begin{aligned} d\phi^*\alpha &= \sum_I d\big((\phi^* f_I)(\phi^* dy_I)\big) = \sum_I d\big((\phi^* f_I)\, d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k}\big) \\ &= \sum_I d(\phi^* f_I)\, d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k} + \sum_I (\phi^* f_I)\, d(d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k}) \\ &= \sum_I d(\phi^* f_I)\, d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k}. \end{aligned}$$

Here we have used the Leibniz rule for forms, Proposition 2.5(ii), plus the fact that the form $d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k}$ is always closed. (See Exercise 2.8.) Comparing the two equations above we see that $\phi^* d\alpha = d\phi^*\alpha$. QED

We finish this section by giving an explicit formula for the pullback $\phi^*\alpha$, which establishes a connection between forms and determinants. Let us do this first in degrees 1 and 2. The pullback of a 1-form $\alpha = \sum_{i=1}^m f_i\, dy_i$ is

$$\phi^*\alpha = \sum_{i=1}^m (\phi^* f_i)(\phi^* dy_i) = \sum_{i=1}^m (\phi^* f_i)\, d\phi_i.$$

Now $d\phi_i = \sum_{j=1}^n \frac{\partial\phi_i}{\partial x_j}\, dx_j$ and so

$$\phi^*\alpha = \sum_{i=1}^m \left((\phi^* f_i) \sum_{j=1}^n \frac{\partial\phi_i}{\partial x_j}\, dx_j\right) = \sum_{j=1}^n \sum_{i=1}^m (\phi^* f_i) \frac{\partial\phi_i}{\partial x_j}\, dx_j = \sum_{j=1}^n g_j\, dx_j,$$

with $g_j = \sum_{i=1}^m (\phi^* f_i) \frac{\partial\phi_i}{\partial x_j}$.

For a 2-form $\alpha = \sum_{1\le i<j\le m} f_{i,j}\, dy_i\, dy_j$ we get

$$\phi^*\alpha = \sum_{1\le i<j\le m} (\phi^* f_{i,j})\phi^*(dy_i\, dy_j) = \sum_{1\le i<j\le m} (\phi^* f_{i,j})\, d\phi_i\, d\phi_j.$$

Observe that

$$d\phi_i\, d\phi_j = \sum_{k,l=1}^n \frac{\partial\phi_i}{\partial x_k}\frac{\partial\phi_j}{\partial x_l}\, dx_k\, dx_l = \sum_{1\le k<l\le n} \left(\frac{\partial\phi_i}{\partial x_k}\frac{\partial\phi_j}{\partial x_l} - \frac{\partial\phi_i}{\partial x_l}\frac{\partial\phi_j}{\partial x_k}\right) dx_k\, dx_l,$$

where

$$\frac{\partial\phi_i}{\partial x_k}\frac{\partial\phi_j}{\partial x_l} - \frac{\partial\phi_i}{\partial x_l}\frac{\partial\phi_j}{\partial x_k} = \begin{vmatrix} \frac{\partial\phi_i}{\partial x_k} & \frac{\partial\phi_i}{\partial x_l} \\ \frac{\partial\phi_j}{\partial x_k} & \frac{\partial\phi_j}{\partial x_l} \end{vmatrix}$$

is the determinant of the 2×2-submatrix obtained from the Jacobi matrix $D\phi$ by extracting rows i and j and columns k and l. So we get

$$\begin{aligned}
\phi^*\alpha &= \sum_{1\le i<j\le m} \left((\phi^* f_{i,j}) \sum_{1\le k<l\le n} \begin{vmatrix} \frac{\partial\phi_i}{\partial x_k} & \frac{\partial\phi_i}{\partial x_k} \\ \frac{\partial\phi_j}{\partial x_k} & \frac{\partial\phi_j}{\partial x_l} \end{vmatrix} dx_k\, dx_l \right) \\
&= \sum_{1\le k<l\le n} \sum_{1\le i<j\le m} (\phi^* f_{i,j}) \begin{vmatrix} \frac{\partial\phi_i}{\partial x_k} & \frac{\partial\phi_i}{\partial x_k} \\ \frac{\partial\phi_j}{\partial x_k} & \frac{\partial\phi_j}{\partial x_l} \end{vmatrix} dx_k\, dx_l = \sum_{1\le k<l\le n} g_{k,l}\, dx_k\, dx_l
\end{aligned}$$

with

$$g_{k,l} = \sum_{1\le i<j\le m} (\phi^* f_{i,j}) \begin{vmatrix} \frac{\partial\phi_i}{\partial x_k} & \frac{\partial\phi_i}{\partial x_l} \\ \frac{\partial\phi_j}{\partial x_k} & \frac{\partial\phi_j}{\partial x_l} \end{vmatrix}.$$

For an arbitrary k-form $\alpha = \sum_I f_I\, dy_I$ we obtain

$$\phi^*\alpha = \sum_I (\phi^* f_I) \phi^*(dy_{i_1}\, dy_{i_2} \cdots dy_{i_k}) = \sum_I (\phi^* f_I)\, d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k}.$$

To write the product $d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k}$ in terms of the x-variables we use

$$d\phi_{i_l} = \sum_{m_l=1}^{n} \frac{\partial\phi_{i_l}}{\partial x_{m_l}}\, dx_{m_l}$$

for $l = 1, 2, \ldots, k$. This gives

$$\begin{aligned}
d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k} &= \sum_{m_1, m_2, \ldots, m_k = 1}^{n} \frac{\partial\phi_{i_1}}{\partial x_{m_1}} \frac{\partial\phi_{i_2}}{\partial x_{m_2}} \cdots \frac{\partial\phi_{i_k}}{\partial x_{m_k}}\, dx_{m_1}\, dx_{m_2} \cdots dx_{m_k} \\
&= \sum_M \frac{\partial\phi_{i_1}}{\partial x_{m_1}} \frac{\partial\phi_{i_2}}{\partial x_{m_2}} \cdots \frac{\partial\phi_{i_k}}{\partial x_{m_k}}\, dx_M,
\end{aligned}$$

in which the summation is over *all* n^k multi-indices $M = (m_1, m_2, \ldots, m_k)$. If a multi-index M has repeating entries, then $dx_M = 0$. If the entries of M are all distinct, we can rearrange them in increasing order by means of a permutation σ. In other words, we have $M = (m_1, m_2, \ldots, m_k) = (j_{\sigma(1)}, j_{\sigma(2)}, \ldots, j_{\sigma(k)})$, where $J = (j_1, j_2, \ldots, j_k)$ is an *increasing* multi-index and $\sigma \in S_k$ is a permutation. Thus we can rewrite the sum over all multi-indices M as a double sum over all increasing multi-indices J and all permutations σ:

$$\begin{aligned}
d\phi_{i_1}\, d\phi_{i_1} \cdots d\phi_{i_k} &= \sum_J \sum_{\sigma\in S_k} \frac{\partial\phi_{i_1}}{\partial x_{j_{\sigma(1)}}} \frac{\partial\phi_{i_2}}{\partial x_{j_{\sigma(2)}}} \cdots \frac{\partial\phi_{i_k}}{\partial x_{j_{\sigma(k)}}}\, dx_{j_{\sigma(1)}}\, dx_{j_{\sigma(2)}} \cdots dx_{j_{\sigma(k)}} \\
&= \sum_J \sum_{\sigma\in S_k} \operatorname{sign}(\sigma) \frac{\partial\phi_{i_1}}{\partial x_{j_{\sigma(1)}}} \frac{\partial\phi_{i_2}}{\partial x_{j_{\sigma(2)}}} \cdots \frac{\partial\phi_{i_k}}{\partial x_{j_{\sigma(k)}}}\, dx_J \qquad (3.2) \\
&= \sum_J \det D\phi_{I,J}\, dx_J. \qquad (3.3)
\end{aligned}$$

In (3.2) used the result of Exercise 3.7 and in (3.3) we applied Theorem 3.6. The notation $D\phi_{I,J}$ stands for the I, J-*submatrix* of $D\phi$, that is the $k \times k$-matrix obtained from the Jacobi matrix by extracting rows $i_1, i_2, \ldots, i_k$ and columns $j_1, j_2, \ldots, j_k$.

To sum up, we find

$$\phi^*\alpha = \sum_I \phi^* f_I \sum_J \det D\phi_{I,J}\, dx_J = \sum_J \left(\sum_I (\phi^* f_I) \det D\phi_{I,J} \right) dx_J.$$

This proves the following result.

3.12. THEOREM. *Let $\phi\colon U \to V$ be a smooth map, where U is open in $\mathbf{R}^n$ and V is open in $\mathbf{R}^m$. Let $\alpha = \sum_I f_I\, dy_I$ be a k-form on V. Then $\phi^*\alpha$ is the k-form on U given by $\phi^*\alpha = \sum_J g_J\, dx_J$ with*

$$\boxed{g_J = \sum_I (\phi^* f_I) \det D\phi_{I,J}.}$$

This formula is seldom used to calculate pullbacks in practice and you don't need to memorize the details of the proof. It is almost always easier to apply the definition of pullback directly. However, the formula has some important theoretical uses, one of which we record here.

Assume that $k = m = n$, that is to say, the number of new variables is equal to the number of old variables, and we are pulling back a form of top degree. Then

$$\alpha = f\, dy_1\, dy_2 \cdots dy_n, \qquad \phi^*\alpha = (\phi^* f)(\det D\phi)\, dx_1\, dx_2 \cdots dx_n.$$

If $f = 1$ (constant function) then $\phi^* f = 1$, so we see that $\det D\phi(\mathbf{x})$ can be interpreted as the ratio between the oriented volumes of two infinitesimal blocks positioned at $\mathbf{x}$: one with edges dx_1, $dx_2, \dots, dx_n$ and another with edges $d\phi_1$, $d\phi_2, \dots, d\phi_n$. Thus the Jacobi determinant is a measurement of how much the map ϕ changes oriented volume from point to point.

3.13. THEOREM. *Let $\phi\colon U \to V$ be a smooth map, where U and V are open in $\mathbf{R}^n$. Then the pullback of the volume form on V is equal to the Jacobi determinant times the volume form on U,*

$$\boxed{\phi^*(dy_1\, dy_2 \cdots dy_n) = (\det D\phi)\, dx_1\, dx_2 \cdots dx_n.}$$

Exercises

3.1. Deduce Theorem 3.7(iv) from Theorem 3.6.

3.2. Calculate the following determinants using column and/or row operations and Theorem 3.7(iv).

$$\begin{vmatrix} 1 & 3 & 1 & 1 \\ 2 & 1 & 5 & 2 \\ 1 & -1 & 2 & 3 \\ 4 & 1 & -3 & 7 \end{vmatrix}, \qquad \begin{vmatrix} 1 & 1 & -2 & 4 \\ 0 & 1 & 1 & 3 \\ 2 & -1 & 1 & 0 \\ 3 & 1 & 2 & 5 \end{vmatrix}.$$

3.3. Tabulate all permutations in S_4 with their lengths and signs.

3.4. Determine the length and the sign of the following permutations.

(i) A permutation of the form $(1, 2, \dots, i-1, j, \dots, j-1, i, \dots, n)$ where $1 \le i < j \le n$. (Such a permutation is called a *transposition*. It interchanges i and j and leaves all other numbers fixed.)

(ii) $(n, n-1, n-2, \dots, 3, 2, 1)$.

3.5. Find all permutations in S_n of length 1.

3.6. Calculate σ^{-1}, τ^{-1}, $\sigma\tau$ and $\tau\sigma$, where

(i) $\sigma = (3,6,1,2,5,4)$ and $\tau = (5,2,4,6,3,1)$;

(ii) $\sigma = (2,1,3,4,5,\dots,n-1,n)$ and $\tau = (n,2,3,\dots,n-2,n-1,1)$ (i.e. the transpositions interchanging 1 and 2, resp. 1 and n).

3.7. Show that

$$dx_{i_{\sigma(1)}}\,dx_{i_{\sigma(2)}}\cdots dx_{i_{\sigma(k)}} = \operatorname{sign}(\sigma)\,dx_{i_1}\,dx_{i_2}\cdots dx_{i_k}$$

for any multi-index $(i_1,i_2,\dots,i_k)$ and any permutation σ in S_k. (First show that the identity is true if σ is a transposition. Then show it is true for an arbitrary permutation σ by writing σ as a product $\sigma_1\sigma_2\cdots\sigma_l$ of transpositions and using formula (3.1) and Exercise 3.4(i).)

3.8. Show that for $n \geq 2$ the permutation group S_n has $n!/2$ even permutations and $n!/2$ odd permutations.

3.9. (i) Show that every permutation has the same length and sign as its inverse.

(ii) Deduce Theorem 3.7(ii) from Theorem 3.6.

3.10. The *i-th simple permutation* is defined by $\sigma_i = (1,2,\dots,i-1,i+1,i,i+2,\dots,n)$. So σ_i interchanges i and $i+1$ and leaves all other numbers fixed. S_n has $n-1$ simple permutations, namely $\sigma_1,\sigma_2,\dots,\sigma_{n-1}$. Prove the *Coxeter relations*

(i) $\sigma_i^2 = 1$ for $1 \leq i < n$,

(ii) $(\sigma_i\sigma_{i+1})^3 = 1$ for $1 \leq i < n-1$,

(iii) $(\sigma_i\sigma_j)^2 = 1$ for $1 \leq i,j < n$ and $i+1 < j$.

3.11. Let σ be a permutation of $\{1,2,\dots,n\}$. The *permutation matrix* corresponding to σ is the $n\times n$-matrix A_σ whose i-th column is the vector $\mathbf{e}_{\sigma(i)}$. In other words, $A_\sigma\mathbf{e}_i = \mathbf{e}_{\sigma(i)}$.

(i) Write down the permutation matrices for all permutations in S_3.

(ii) Show that $A_{\sigma\tau} = A_\sigma A_\tau$.

(iii) Show that $\det A_\sigma = \operatorname{sign}(\sigma)$.

3.12. (i) Suppose that A has the shape

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ 0 & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n,1} & \dots & a_{n,n} \end{pmatrix},$$

i.e. all entries below a_{11} are 0. Deduce from Theorem 3.6 that

$$\det A = a_{1,1}\begin{vmatrix} a_{2,2} & \dots & a_{2,n} \\ \vdots & & \vdots \\ a_{n,2} & \dots & a_{n,n} \end{vmatrix}.$$

(ii) Deduce from this the expansion rule, Theorem 3.7(iii).

3.13. Show that

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix} = \prod_{i<j}(x_j - x_i)$$

for any numbers $x_1, x_2, \dots, x_n$. (Starting at the bottom, from each row subtract x_1 times the row above it. This creates a new determinant whose first column is the standard basis vector $\mathbf{e}_1$. Expand on the first column and note that each column of the remaining determinant has a common factor.)

3.14. Let $\phi \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 x_2 \\ x_1 x_3 \\ x_2 x_3 \end{pmatrix}$. Find

(i) $\phi^* dy_1, \phi^* dy_2, \phi^* dy_3$;
(ii) $\phi^*(y_1 y_2 y_3), \phi^*(dy_1\, dy_2)$;
(iii) $\phi^*(dy_1\, dy_2\, dy_3)$.

3.15. Let $\phi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1^3 \\ x_1^2 x_2 \\ x_1 x_2^2 \\ x_2^3 \end{pmatrix}$. Find

(i) $\phi^*(y_1 + 3y_2 + 3y_3 + y_4)$;
(ii) $\phi^* dy_1, \phi^* dy_2, \phi^* dy_3, \phi^* dy_4$;
(iii) $\phi^*(dy_2\, dy_3)$.

3.16. Compute $\psi^*(x\,dy\,dz + y\,dz\,dx + z\,dx\,dy)$, where ψ is the map $\mathbf{R}^2 \to \mathbf{R}^3$ defined in Exercise B.7.

3.17. Let $P_3 \begin{pmatrix} r \\ \theta \\ \phi \end{pmatrix} = \begin{pmatrix} r\cos\phi\cos\theta \\ r\cos\phi\sin\theta \\ r\sin\phi \end{pmatrix}$ be spherical coordinates in $\mathbf{R}^3$.

(i) Calculate $P_3^* \alpha$ for the following forms α:

$$dx, \quad dy, \quad dz, \quad dx\,dy, \quad dx\,dz, \quad dy\,dz, \quad dx\,dy\,dz.$$

(ii) Find the inverse of the matrix DP_3.

3.18 (spherical coordinates in n dimensions). In this problem let us write a point in $\mathbf{R}^n$ as

$$\begin{pmatrix} r \\ \theta_1 \\ \vdots \\ \theta_{n-1} \end{pmatrix}.$$

Let P_1 be the function $P_1(r) = r$. For each $n \geq 1$ define a map $P_{n+1} \colon \mathbf{R}^{n+1} \to \mathbf{R}^{n+1}$ by

$$P_{n+1} \begin{pmatrix} r \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix} = \begin{pmatrix} (\cos\theta_n) P_n \begin{pmatrix} r \\ \theta_1 \\ \vdots \\ \theta_{n-1} \end{pmatrix} \\ r\sin\theta_n \end{pmatrix}.$$

(This is an example of a recursive definition. If you know P_1, you can compute P_2, and then P_3, etc.)

(i) Show that P_2 and P_3 are the usual polar, resp. spherical coordinates on $\mathbf{R}^2$, resp. $\mathbf{R}^3$.
(ii) Give an explicit formula for P_4.
(iii) Let $\mathbf{p}$ be the first column vector of the Jacobi matrix of P_n. Show that $P_n = r\mathbf{p}$.
(iv) Show that the Jacobi matrix of P_{n+1} is a $(n+1) \times (n+1)$-matrix of the form

$$DP_{n+1} = \begin{pmatrix} A & \mathbf{u} \\ \mathbf{v} & w \end{pmatrix},$$

where A is an $n \times n$-matrix, $\mathbf{u}$ is a column vector, $\mathbf{v}$ is a row vector and w is a function given respectively by

$$A = \cos\theta_n\, DP_n, \qquad \mathbf{u} = -(\sin\theta_n) P_n,$$
$$\mathbf{v} = (\sin\theta_n, 0, 0, \ldots, 0), \qquad w = r\cos\theta_n.$$

(v) Show that $\det DP_{n+1} = r\cos^{n-1}\theta_n \det DP_n$ for $n \geq 1$. (Expand $\det DP_{n+1}$ with respect to the last row, using the formula in part (iv), and apply the result of part (iii).)
(vi) Using the formula in part (v) calculate $\det DP_n$ for $n = 1, 2, 3, 4$.
(vii) Find an explicit formula for $\det DP_n$ for general n.
(viii) Show that $\det DP_n \neq 0$ for $r \neq 0$.

CHAPTER 4

Integration of 1-forms

Like functions, forms can be integrated as well as differentiated. Differentiation and integration are related via a multivariable version of the fundamental theorem of calculus, known as Stokes' theorem. In this chapter we investigate the case of 1-forms.

4.1. Definition and elementary properties of the integral

Let U be an open subset of $\mathbf{R}^n$. A *parametrized curve* in U is a smooth mapping $c\colon I \to U$ from an interval I into U. We want to integrate over I. To avoid problems with improper integrals we assume I to be closed and bounded, $I = [a, b]$. (Strictly speaking we have not defined what we mean by a smooth map $c\colon [a, b] \to U$. The easiest definition is that c should be the restriction of a smooth map $\tilde{c}\colon (a-\varepsilon, b+\varepsilon) \to U$ defined on a slightly larger *open* interval.) Let α be a 1-form on U. The pullback $c^*\alpha$ is a 1-form on $[a, b]$, and can therefore be written as $c^*\alpha = g\,dt$ (where t is the coordinate on $\mathbf{R}$). The *integral* of α over c is now defined by

$$\int_c \alpha = \int_{[a,b]} c^*\alpha = \int_a^b g(t)\,dt.$$

More explicitly, writing α in components, $\alpha = \sum_{i=1}^n f_i\,dx_i$, we have

$$c^*\alpha = \sum_{i=1}^n (c^* f_i) dc_i = \sum_{i=1}^n (c^* f_i) \frac{dc_i}{dt}\,dt, \tag{4.1}$$

so

$$\int_c \alpha = \sum_{i=1}^n \int_a^b f_i(c(t)) \frac{dc_i}{dt}(t)\,dt.$$

4.1. EXAMPLE. Let U be the punctured plane $\mathbf{R}^2 - \{\mathbf{0}\}$. Let $c\colon [0, 2\pi] \to U$ be the usual parametrization of the circle, $c(t) = (\cos t, \sin t)$, and let α be the angle form,

$$\alpha = \frac{-y\,dx + x\,dy}{x^2 + y^2}.$$

Then $c^*\alpha = dt$ (see Example 3.8), so $\int_c \alpha = \int_0^{2\pi} dt = 2\pi$.

A curve $c\colon [a, b] \to U$ can be *reparametrized* by substituting a new variable, $t = p(s)$, where s ranges over another interval $[\bar{a}, \bar{b}]$. We shall assume p to be a one-to-one mapping from $[\bar{a}, \bar{b}]$ onto $[a, b]$ satisfying $p'(s) \neq 0$ for $\bar{a} \leq s \leq \bar{b}$. Such a p is called a *reparametrization*. The parametrized curve

$$c \circ p\colon [\bar{a}, \bar{b}] \to U$$

has the same image as the original curve c, but it is traversed at a different rate. Since $p'(s) \neq 0$ for all $s \in [\bar{a}, \bar{b}]$ we have either $p'(s) > 0$ for all s (in which case p is increasing) or $p'(s) < 0$ for all s (in which case p is decreasing). If p is increasing, we say that it *preserves* the orientation of the curve (or that the curves c and $c \circ p$ have the *same orientation*); if p is decreasing, we say that it *reverses* the orientation (or that c and $c \circ p$ have *opposite orientations*). In the orientation-reversing case, $c \circ p$ traverses the curve in the opposite direction to c.

4.2. EXAMPLE. The curve $c\colon [0, 2\pi] \to \mathbf{R}^2$ defined by $c(t) = (\cos t, \sin t)$ represents the unit circle in the plane, traversed at a constant rate (angular velocity) of 1 radian per second. Let $p(s) = 2s$. Then p maps $[0, \pi]$ to $[0, 2\pi]$ and $c \circ p$, regarded as a map $[0, \pi] \to \mathbf{R}^2$, represents the same circle, but traversed at 2 radians per second. (It is important to restrict the domain of p to the interval $[0, \pi]$. If we allowed s to range over $[0, 2\pi]$, then $(\cos 2s, \sin 2s)$ would traverse the circle twice. This is not considered a reparametrization of the original curve c.) Now let $p(s) = -s$. Then $c \circ p\colon [0, 2\pi] \to \mathbf{R}^2$ traverses the unit circle in the clockwise direction. This reparametrization reverses the orientation; the angular velocity is now -1 radian per second. Finally let $p(s) = 2\pi s^2$. Then p maps $[0, 1]$ to $[0, 2\pi]$ and $c \circ p\colon [0, 1] \to \mathbf{R}^2$ runs once counterclockwise through the unit circle, but at a variable rate. What is the angular velocity as a function of s?

It turns out that the integral of a form along a curve is almost completely independent of the parametrization.

4.3. THEOREM. *Let α be a 1-form on U and $c\colon [a, b] \to U$ a curve in U. Let $p\colon [\bar{a}, \bar{b}] \to [a, b]$ be a reparametrization. Then*

$$\int_{c\circ p} \alpha = \begin{cases} \int_c \alpha & \text{if } p \text{ preserves the orientation,} \\ -\int_c \alpha & \text{if } p \text{ reverses the orientation.} \end{cases}$$

PROOF. It follows from the definition of the integral and from the naturality of pullbacks (Proposition 3.10(iii)) that

$$\int_{c\circ p} \alpha = \int_{[\bar{a},\bar{b}]} (c \circ p)^*\alpha = \int_{[\bar{a},\bar{b}]} p^*(c^*\alpha).$$

Now let us write $c^*\alpha = g\,dt$ and $t = p(s)$. Then $p^*(c^*\alpha) = p^*(g\,dt) = (p^*g)dp = (p^*g)(dp/ds)\,ds$, so

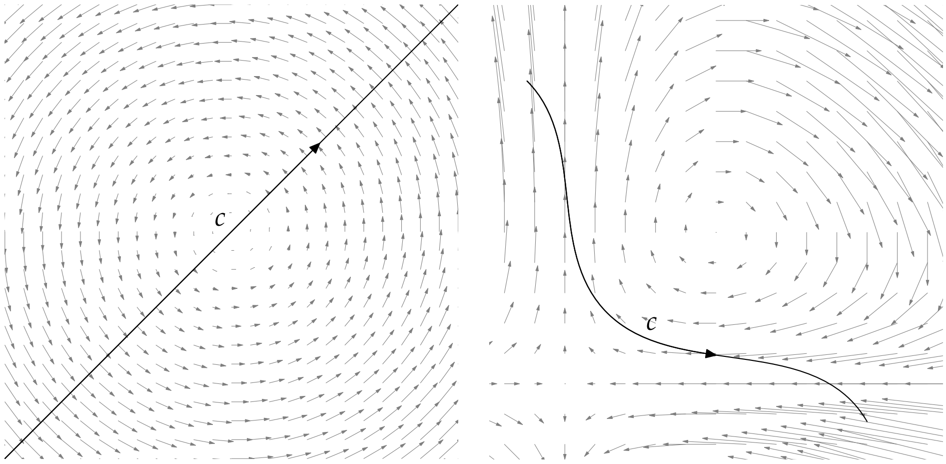
$$\int_{c\circ p} \alpha = \int_{[\bar{a},\bar{b}]} (p^*g)\frac{dp}{ds}\,ds = \int_{\bar{a}}^{\bar{b}} g(p(s))p'(s)\,ds.$$

On the other hand, $\int_c \alpha = \int_a^b g(t)\,dt$, so by the substitution formula, Theorem B.7, we have $\int_{c\circ p} \alpha = \pm \int_c \alpha$, where the $+$ occurs if $p' > 0$ and the $-$ if $p' < 0$. QED

Interpretation of the integral. Integrals of 1-forms play an important role in physics and engineering. A curve $c\colon [a, b] \to U$ models a particle travelling through the region U. Recall from Section 2.5 that to a 1-form $\alpha = \sum_{i=1}^n F_i\,dx_i$ corresponds a vector field $\mathbf{F} = \sum_{i=1}^n F_i\mathbf{e}_i$, which can be thought of as a force field acting on the particle. Symbolically we write $\alpha = \mathbf{F} \cdot d\mathbf{x}$, where we think of $d\mathbf{x}$ as an infinitesimal vector tangent to the curve. Thus α represents the *work* done by the force field along an infinitesimal vector $d\mathbf{x}$. From (4.1) we see that $c^*\alpha = \mathbf{F}(c(t)) \cdot c'(t)\,dt$. Accordingly, the *total work* done by the force $\mathbf{F}$ on the particle during its trip along c is the integral

$$\int_c \alpha = \int_c \mathbf{F} \cdot d\mathbf{x} = \int_a^b \mathbf{F}(c(t)) \cdot c'(t)\,dt.$$

In particular, the work and the total work are nil if the force is perpendicular to the path, as in the picture on the left. The work done by the force in the picture on the right is negative.



Theorem 4.3 can be translated into this language as follows: the work done by the force does not depend on the rate at which the particle speeds along its path, but only on the path itself and on the direction of travel.

The field $\mathbf{F}$ is *conservative* if it can be written as the gradient of a function, $\mathbf{F} = \operatorname{grad} g$. The function $-g$ is called a *potential* for the field and is interpreted as the potential energy of the particle. In terms of forms this means that $\alpha = dg$, i.e. α is exact.

4.2. Integration of exact 1-forms

Integrating an exact 1-form $\alpha = dg$ is easy once the function g is known.

4.4. THEOREM (fundamental theorem of calculus in $\mathbf{R}^n$). *Let $\alpha = dg$ be an exact 1-form on an open subset U of $\mathbf{R}^n$. Let $c\colon [a, b] \to U$ be a parametrized curve. Then*

$$\boxed{\int_c \alpha = g(c(b)) - g(c(a)).}$$

PROOF. By Theorem 3.11 we have $c^*\alpha = c^*dg = dc^*g$. Writing $h(t) = c^*g(t) = g(c(t))$ we have $c^*\alpha = dh$, so

$$\int_c \alpha = \int_{[a,b]} c^*\alpha = \int_a^b dh = h(b) - h(a),$$

where we used the (ordinary) fundamental theorem of calculus, formula (B.1). Hence $\int_c \alpha = g(c(b)) - g(c(a))$. QED

The physical interpretation of this result is that when a particle moves in a conservative force field, its potential energy decreases by the amount of work done by the field. This clarifies what it means for a field to be conservative: it means that the work done is entirely converted into mechanical energy and that none is dissipated by friction into heat, radiation, etc. Thus the fundamental theorem of calculus "explains" the law of conservation of energy.

It also yields a necessary and sufficient criterion for a 1-form on U to be exact. A curve $c\colon [a,b] \to U$ is called *closed* if $c(a) = c(b)$.

4.5. THEOREM. *Let α be a 1-form on an open subset U of $\mathbf{R}^n$. Then the following statements are equivalent.*

(i) *α is exact.*
(ii) *$\int_c \alpha = 0$ for all closed curves c.*
(iii) *$\int_c \alpha$ depends only on the endpoints of c for every curve c in U.*

PROOF. (i) $\implies$ (ii): if $\alpha = dg$ and c is closed, then $\int_c \alpha = g(c(b)) - g(c(a)) = 0$ by the fundamental theorem of calculus, Theorem 4.4.

(ii) $\implies$ (iii): assume $\int_c \alpha = 0$ for all closed curves c. Let

$$c_1\colon [a_1, b_1] \to U \qquad \text{and} \qquad c_2\colon [a_2, b_2] \to U$$

be two curves with the same endpoints, i.e. $c_1(a_1) = c_2(a_2)$ and $c_1(b_1) = c_2(b_2)$. We need to show that $\int_{c_1} \alpha = \int_{c_2} \alpha$. After reparametrizing c_1 and c_2 we may assume that $a_1 = a_2 = 0$ and $b_1 = b_2 = 1$. Define a new curve c by

$$c(t) = \begin{cases} c_1(t) & \text{for } 0 \le t \le 1, \\ c_2(2-t) & \text{for } 1 \le t \le 2. \end{cases}$$

(First traverse c_1, then traverse c_2 backwards.) Then c is closed, so $\int_c \alpha = 0$. But Theorem 4.3 implies $\int_c \alpha = \int_{c_1} \alpha - \int_{c_2} \alpha$, so $\int_{c_1} \alpha = \int_{c_2} \alpha$.

(iii) $\implies$ (i): assume that, for all c, $\int_c \alpha$ depends only on the endpoints of c. We must define a function g such that $\alpha = dg$. Fix a point $\mathbf{x}_0$ in U. For each point $\mathbf{x}$ in U choose a curve $c_{\mathbf{x}}\colon [0,1] \to U$ which joins $\mathbf{x}_0$ to $\mathbf{x}$. Define

$$g(\mathbf{x}) = \int_{c_{\mathbf{x}}} \alpha.$$

We assert that dg is well-defined and equal to α. Write $\alpha = \sum_{i=1}^n f_i\, dx_i$. We must show that $\partial g/\partial x_i = f_i$. From the definition of partial differentiation,

$$\frac{\partial g}{\partial x_i}(\mathbf{x}) = \lim_{h\to 0} \frac{g(\mathbf{x} + h\mathbf{e}_i) - g(\mathbf{x})}{h} = \lim_{h\to 0} \frac{1}{h}\left(\int_{c_{\mathbf{x}+h\mathbf{e}_i}} \alpha - \int_{c_{\mathbf{x}}} \alpha\right).$$

Now consider a curve $\tilde{c}$ composed of two pieces: for $0 \le t \le 1$ travel from $\mathbf{x}_0$ to $\mathbf{x}$ along the curve $c_{\mathbf{x}}$ and then for $1 \le t \le 2$ travel from $\mathbf{x}$ to $\mathbf{x} + h\mathbf{e}_i$ along the straight line given by $l(t) = \mathbf{x} + (t-1)h\mathbf{e}_i$. Then $\tilde{c}$ has the same endpoints as $c_{\mathbf{x}+h\mathbf{e}_i}$. Therefore $\int_{c_{\mathbf{x}+h\mathbf{e}_i}} \alpha = \int_{\tilde{c}} \alpha$, and hence

$$\begin{aligned} \frac{\partial g}{\partial x_i}(\mathbf{x}) = \lim_{h\to 0} \frac{1}{h}\left(\int_{\tilde{c}} \alpha - \int_{c_{\mathbf{x}}} \alpha\right) &= \lim_{h\to 0} \frac{1}{h}\left(\int_{c_{\mathbf{x}}} \alpha + \int_l \alpha - \int_{c_{\mathbf{x}}} \alpha\right) \\ &= \lim_{h\to 0} \frac{1}{h} \int_l \alpha = \lim_{h\to 0} \frac{1}{h} \int_{[1,2]} l^*\alpha. \end{aligned} \tag{4.2}$$

Let $\delta_{i,j}$ be the *Kronecker delta*, which is defined by $\delta_{i,i} = 1$ and $\delta_{i,j} = 0$ if $i \neq j$. Then we can write $l_j(t) = x_j + \delta_{i,j}(t-1)h$, and hence $l_j'(t) = \delta_{i,j}h$. This shows that

$$\begin{aligned} l^*\alpha = \sum_{j=1}^n f_j(\mathbf{x} + (t-1)h\mathbf{e}_i)dl_j &= \sum_{j=1}^n f_j(\mathbf{x} + (t-1)h\mathbf{e}_i)l_j'(t)\,dt \\ &= \sum_{j=1}^n f_j(\mathbf{x} + (t-1)h\mathbf{e}_i)\delta_{i,j}h\,dt = hf_i(\mathbf{x} + (t-1)h\mathbf{e}_i)\,dt. \end{aligned} \tag{4.3}$$

Taking equations (4.2) and (4.3) together we find

$$\begin{aligned} \frac{\partial g}{\partial x_i}(\mathbf{x}) &= \lim_{h\to 0}\frac{1}{h}\int_1^2 hf_i(\mathbf{x} + (t-1)h\mathbf{e}_i)\,dt = \lim_{h\to 0}\int_0^1 f_i(\mathbf{x} + sh\mathbf{e}_i)\,ds \\ &= \int_0^1 \lim_{h\to 0} f_i(\mathbf{x} + sh\mathbf{e}_i)\,ds = \int_0^1 f_i(\mathbf{x})\,ds = f_i(\mathbf{x}). \end{aligned}$$

This proves that g is smooth and that $dg = \alpha$. QED

This theorem, and its proof, can be used in many different ways. For example, it tells us that once we know a 1-form α to be exact we can find an "antiderivative" $g(\mathbf{x})$ by integrating α along an arbitrary path running from a fixed point $\mathbf{x}_0$ to $\mathbf{x}$. (See Exercises 4.3–4.5 for an application.) On the other hand, the theorem also enables us to detect closed 1-forms that are not exact.

4.6. EXAMPLE. The angle form $\alpha \in \Omega^1(\mathbf{R}^2 - \{\mathbf{0}\})$ of Example 2.10 is closed, but not exact. Indeed, its integral around the circle is $2\pi \neq 0$. Mark the contrast with closed 1-forms on $\mathbf{R}^n$, which are always exact! (See Exercise 2.6.) This phenomenon underlines the importance of being careful about the domain of definition of a form.

4.3. The global angle function and the winding number

In this section we will have a closer look at the angle form and see that it carries interesting information of a "topological" nature. Throughout this section U will be the punctured plane $\mathbf{R}^2 - \{\mathbf{0}\}$, α will denote the angle form,

$$\alpha = \frac{-y\,dx + x\,dy}{x^2 + y^2},$$

and ξ and η will denote the functions

$$\xi = \frac{x}{\sqrt{x^2 + y^2}}, \qquad \eta = \frac{y}{\sqrt{x^2 + y^2}}.$$

Then α is a closed 1-form and ξ and η are smooth functions on U. In fact, ξ and η are just the components of $\mathbf{x}/\|\mathbf{x}\|$, the unit vector pointing in the direction of $\mathbf{x}$. These functions satisfy

$$\alpha = \xi d\eta - \eta d\xi. \tag{4.4}$$

(You will be asked to check this formula in Exercise 4.6.) Now let $\theta\colon U \to [0, 2\pi)$ be the angle between a point and the positive x-axis, chosen to lie in the interval $[0, 2\pi)$. Then $\xi = \cos\theta$ and $\eta = \sin\theta$, so by equation (4.4)

$$\alpha = \cos\theta\,d\sin\theta - \sin\theta\,d\cos\theta = \cos^2\theta\,d\theta + (-\sin\theta)^2\,d\theta = d\theta.$$

This equation is not valid on all of U (it cannot be because we saw in Example 4.6 that α is not exact), but only where θ is differentiable, i.e. on the complement of the positive x-axis. Hence the nonexactness of α is closely related to the impossibility of defining a global differentiable angle function on U. (The precise meaning of this assertion will become clear in Exercise 4.6.)

However, along a curve $c\colon [a,b] \to U$ we can define a continuous angle function, and the fact that $\alpha = d\theta$ almost everywhere suggests how: by integrating α along c! For simplicity assume that $a = 0$ and $b = 1$. Start by fixing any ϑ_0 such that $\cos\vartheta_0 = \xi(c(0))$ and $\sin\vartheta_0 = \eta(c(0))$ and then define

$$\vartheta(t) = \vartheta_0 + \int_{[0,t]} c^*\alpha.$$

The following result says that $\vartheta(t)$ measures the angle between $c(t)$ and the positive x-axis (up to an integer multiple of 2π) and that the function $\vartheta\colon [0,1] \to \mathbf{R}$ is smooth. In this sense ϑ is a "differentiable choice of angle" along the curve c.

4.7. THEOREM. *The function ϑ is smooth and satisfies*

$$\vartheta(0) = \vartheta_0, \quad \cos\vartheta(t) = \xi(c(t)) \quad \textit{and} \quad \sin\vartheta(t) = \eta(c(t)).$$

PROOF. To see that $\vartheta(0) = \vartheta_0$, plug $t = 0$ into the definition of ϑ. To prove the other assertions we rescale the curve $c(t)$ to a new curve $c(t)/\|c(t)\|$ moving on the unit circle. Let $f(t)$ and $g(t)$ be the x- and y-components of this new curve. Then $f(t) = \xi(c(t))$ and $g(t) = \eta(c(t))$ and

$$f(t)^2 + g(t)^2 = 1$$

for all t. In other words $f = c^*\xi$ and $g = c^*\eta$, so it follows from formula (4.4) that $c^*\alpha = f\,dg - g\,df$. Therefore

$$\vartheta(t) = \vartheta_0 + \int_0^t \big(f(s)g'(s) - g(s)f'(s)\big)\,ds.$$

By the fundamental theorem of calculus, formula (B.2), ϑ is differentiable and

$$\vartheta' = fg' - gf'. \tag{4.5}$$

Since the right-hand side is smooth, ϑ is smooth as well. To prove that $\cos\vartheta(t) = f(t)$ and $\sin\vartheta(t) = g(t)$ for all t it is enough to show that the difference vector

$$\begin{pmatrix} f(t) \\ g(t) \end{pmatrix} - \begin{pmatrix} \cos\vartheta(t) \\ \sin\vartheta(t) \end{pmatrix}$$

has length 0. Its length is equal to

$$\begin{aligned}(f - \cos\vartheta)^2 + (g - \sin\vartheta)^2 &= f^2 + g^2 - 2(f\cos\vartheta + g\sin\vartheta) + \cos^2\vartheta + \sin^2\vartheta \\ &= 2 - 2(f\cos\vartheta + g\sin\vartheta).\end{aligned}$$

Hence we need to show that the function $u = f\cos\vartheta + g\sin\vartheta$ is a constant equal to 1. For $t = 0$ we have

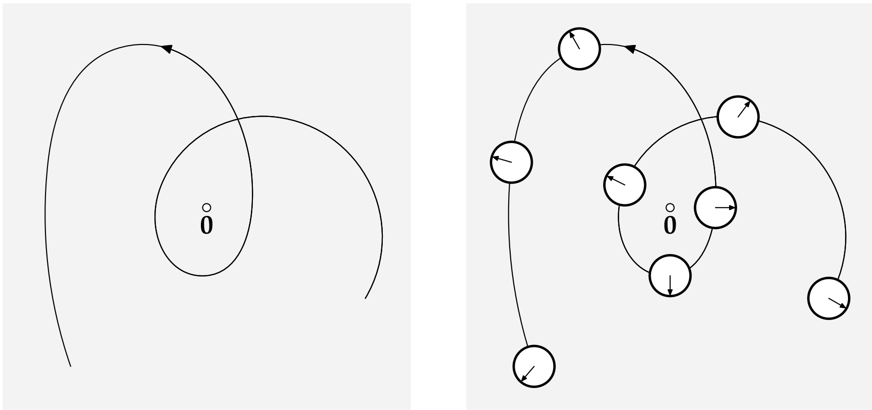
$$u(0) = f(0)\cos\vartheta_0 + g(0)\sin\vartheta_0 = \cos^2\vartheta_0 + \sin^2\vartheta_0 = 1.$$

Furthermore, the derivative of u is

$$\begin{aligned} u' &= f'\cos\vartheta - f\vartheta'\sin\vartheta + g'\sin\vartheta + g\vartheta'\cos\vartheta \\ &= (f' - g^2f' + fgg')\cos\vartheta + (g' - f^2g' + fgf')\sin\vartheta && \text{by formula (4.5)} \\ &= (f^2f' + fgg')\cos\vartheta + (g^2g' + fgf')\sin\vartheta && \text{since } f^2 + g^2 = 1 \\ &= f(ff' + gg')\cos\vartheta + g(gg' + ff')\sin\vartheta. \end{aligned}$$

Now $f^2 + g^2 = 1$ implies $ff' + gg' = 0$, so $u'(t) = 0$ for all t. Hence u is a constant function, so $u(t) = 1$ for all t. QED

It is useful to think of the vector $(f(t), g(t))^T = c(t)/\|c(t)\|$ as a dial that points in the same direction as the vector $c(t)$.



As t increases from 0 to 1, the dial starts at the angle $\vartheta(0) = \vartheta_0$, it moves around the meter, and ends up at the final angle $\vartheta(1)$. The difference $\vartheta(1) - \vartheta(0)$ measures the total angle swept out by the dial.

4.8. COROLLARY. *If $c\colon [0,1] \to U$ is a closed curve, then $\vartheta(1) - \vartheta(0) = 2\pi k$, where k is an integer.*

PROOF. By Theorem 4.7, $c(0) = c(1)$ implies

$$\begin{aligned}\big(\cos\vartheta(0), \sin\vartheta(0)\big) = \big(\xi(c(0)), \eta(c(0))\big) = \big(\xi(c(1)), \eta(c(1))\big)\\ = \big(\cos\vartheta(1), \sin\vartheta(1)\big).\end{aligned}$$

In other words $\cos\vartheta(0) = \cos\vartheta(1)$ and $\sin\vartheta(0) = \sin\vartheta(1)$, so $\vartheta(0)$ and $\vartheta(1)$ differ by an integer multiple of 2π. QED

The integer $k = (2\pi)^{-1}\int_c \alpha$ is called the *winding number* of the closed curve c about the origin. It measures how many times the curve loops around the origin.

$$\boxed{\text{winding number of a closed curve about origin} = \frac{1}{2\pi}\int_c \alpha.} \tag{4.6}$$

4.9. EXAMPLE. By Example 4.1, the winding number of the circle $c(t) = (\cos t, \sin t)^T$ $(0 \leq t \leq 2\pi)$ is equal to 1.

Exercises

4.1. Consider the curve $c\colon [0, \pi/2] \to \mathbf{R}^2$ defined by $c(t) = (a\cos t, b\sin t)^T$, where a and b are positive constants. Let $\alpha = xy\,dx + x^2y\,dy$.

(i) Sketch the curve c for $a = 2$ and $b = 1$.
(ii) Find $\int_c \alpha$ (for arbitrary a and b).

4.2. Restate Theorem 4.5 in terms of force fields, potentials and energy. Explain why the result is plausible on physical grounds.

4.3. Consider the 1-form $\alpha = \|\mathbf{x}\|^a \sum_{i=1}^n x_i\,dx_i$ on $\mathbf{R}^n - \{\mathbf{0}\}$, where a is a real constant. For every $\mathbf{x} \neq \mathbf{0}$ let $c_{\mathbf{x}}$ be the line segment starting at the origin and ending at $\mathbf{x}$.

(i) Show that α is closed for any value of a.
(ii) Determine for which values of a the function $g(\mathbf{x}) = \int_{c_{\mathbf{x}}} \alpha$ is well-defined and compute it.
(iii) For the values of a you found in part (ii) check that $dg = \alpha$.

4.4. Let $\alpha \in \Omega^1(\mathbf{R}^n - \{\mathbf{0}\})$ be the 1-form of Exercise 4.3. Now let $c_{\mathbf{x}}$ be the halfline pointing from $\mathbf{x}$ radially outward to infinity. Parametrize $c_{\mathbf{x}}$ by travelling from infinity inward to $\mathbf{x}$. (You can do this by using an infinite time interval $(-\infty, 0]$ in such a way that $c_{\mathbf{x}}(0) = \mathbf{x}$.)

(i) Determine for which values of a the function $g(\mathbf{x}) = \int_{c_{\mathbf{x}}} \alpha$ is well-defined and compute it.
(ii) For the values of a you found in part (i) check that $dg = \alpha$.
(iii) Show how to recover from this computation the potential energy for Newton's gravitational force. (See Exercise B.4.)

4.5. Let $\alpha \in \Omega^1(\mathbf{R}^n - \{\mathbf{0}\})$ be as in Exercise 4.3. There is one value of a which is not covered by Exercises 4.3 and 4.4. For this value of a find a smooth function g on $\mathbf{R}^n - \{\mathbf{0}\}$ such that $dg = \alpha$.

4.6. (i) Verify equation (4.4).
(ii) Let $U = \mathbf{R}^2 - \{\mathbf{0}\}$. Prove that there does not exist a smooth function $\theta\colon U \to \mathbf{R}$ satisfying $\cos\theta(x,y) = x/\sqrt{x^2+y^2}$ and $\sin\theta(x,y) = y/\sqrt{x^2+y^2}$ for all $(x,y) \in U$. (Argue by contradiction, by letting $\alpha = (-y\,dx + x\,dy)/(x^2+y^2)$ and showing that $\alpha = d\theta$ if θ was such a function.)

4.7. Calculate directly from the definition the winding number about the origin of the curve $c\colon [0, 2\pi] \to \mathbf{R}^2$ given by $c(t) = (\cos kt, \sin kt)^T$.

4.8. Let $\mathbf{x}_0$ be a point in $\mathbf{R}^2$ and c a closed curve which does not pass through $\mathbf{x}_0$. How would you define the winding number of c around $\mathbf{x}_0$? Try to formulate two different definitions: a "geometric" definition and a definition in terms of an integral over c of a certain 1-form analogous to formula (4.6).

4.9. Let $c\colon [0,1] \to \mathbf{R}^2 - \{\mathbf{0}\}$ be a closed curve with winding number k. Determine the winding numbers of the following curves $\tilde{c}\colon [0,1] \to \mathbf{R}^2 - \{\mathbf{0}\}$ by using the formula, and then explain the answer by appealing to geometric intuition.

(i) $\tilde{c}(t) = c(1-t)$;
(ii) $\tilde{c}(t) = \rho(t)c(t)$, where $\rho\colon [0,1] \to (0,\infty)$ is a function satisfying $\rho(0) = \rho(1)$;
(iii) $\tilde{c}(t) = \|c(t)\|^{-1}c(t)$;
(iv) $\tilde{c}(t) = \phi(c(t))$, where $\phi(x,y) = (y,x)^T$;
(v) $\tilde{c}(t) = \phi(c(t))$, where $\phi(x,y) = \dfrac{1}{x^2+y^2}(x,-y)^T$.

4.10. For each of the following closed curves $c\colon [0,2\pi] \to \mathbf{R}^2 - \{\mathbf{0}\}$ set up the integral defining the winding number about the origin. Evaluate the integral if you can (but don't give up too soon). If not, sketch the curve (the use of software is allowed) and obtain the answer geometrically.

(i) $c(t) = (a\cos t, b\sin t)^T$, where $a > 0$ and $b > 0$;
(ii) $c(t) = (\cos t - 2, \sin t)^T$;
(iii) $c(t) = (\cos^3 t, \sin^3 t)^T$;
(iv) $c(t) = \big((a\cos t + b)\cos t + (b-a)/2, (a\cos t + b)\sin t\big)^T$, where $0 < b < a$.

4.11. Let $b > 0$ and $a \neq 0$ be constants with $|a| \neq b$. Define a planar curve $c\colon [0, 2\pi] \to \mathbf{R}^2 - \{\mathbf{0}\}$ by

$$c(t) = \left((a+b)\cos t + a\cos\frac{a+b}{a}t, (a+b)\sin t + a\sin\frac{a+b}{a}t\right)^T.$$

(i) Sketch the curve c for $a = \pm b/3$.
(ii) For what values of a and b is the curve closed?
(iii) Assume c is closed. Set up the integral defining the winding number of c around the origin and evaluate it. If you get stuck, find the answer geometrically.

4.12. Let U be an open subset of $\mathbf{R}^2$ and let $\mathbf{F} = F_1\mathbf{e}_1 + F_2\mathbf{e}_2\colon U \to \mathbf{R}^2$ be a smooth vector field. The differential form

$$\beta = \frac{F_1 dF_2 - F_2 dF_1}{F_1^2 + F_2^2}$$

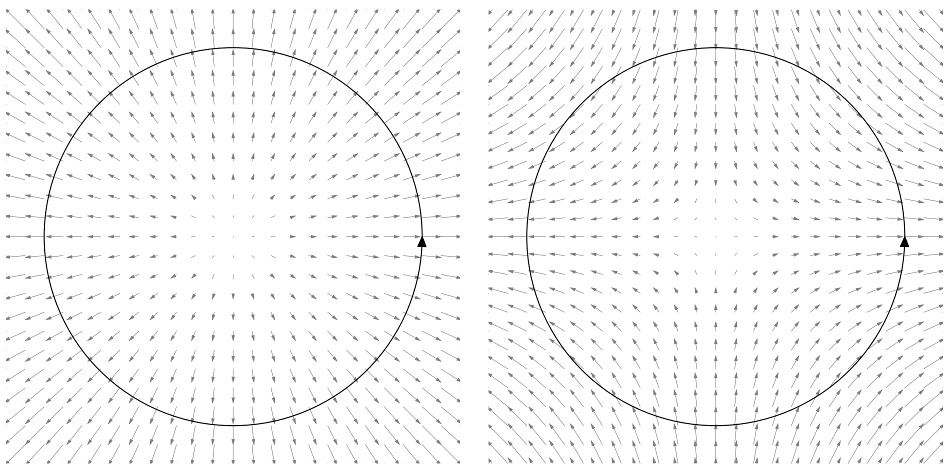
is well-defined at all points $\mathbf{x}$ of U where $\mathbf{F}(\mathbf{x}) \neq 0$. Let c be a parametrized circle contained in U, traversed once in the counterclockwise direction. Assume that $\mathbf{F}(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in c$. The *index* of $\mathbf{F}$ relative to c is

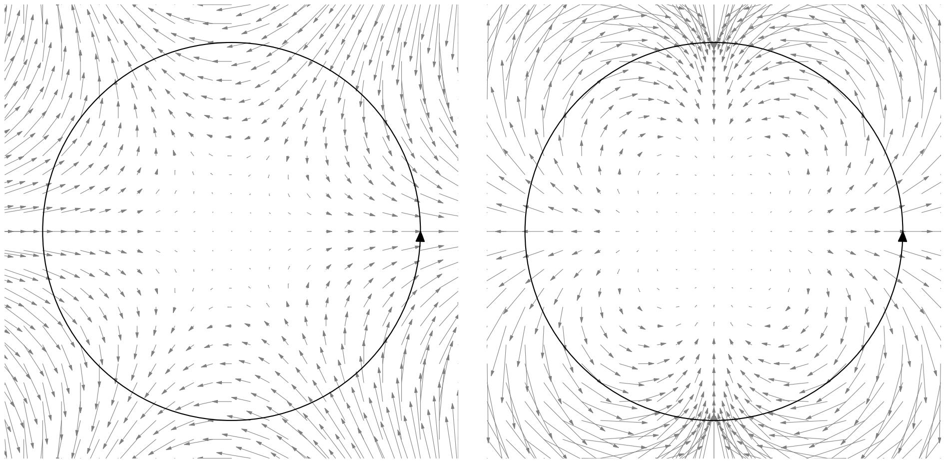
$$\operatorname{index}(\mathbf{F}, c) = \frac{1}{2\pi}\int_c \beta.$$

Prove the following assertions.

(i) $\beta = \mathbf{F}^*\alpha$, where α is the angle form $(-y\,dx + x\,dy)/(x^2+y^2)$;
(ii) β is closed;
(iii) $\operatorname{index}(\mathbf{F}, c)$ is the winding number of the curve $\mathbf{F} \circ c$ about the origin;
(iv) $\operatorname{index}(\mathbf{F}, c)$ is an integer.

4.13. (i) Find the indices of the following vector fields around the indicated circles.





(ii) Draw diagrams of three vector fields in the plane with respective indices 0, 2 and 4 around suitable circles.

CHAPTER 5

Integration and Stokes' theorem

5.1. Integration of forms over chains

In this chapter we generalize the theory of Chapter 4 to higher dimensions. In the same way that 1-forms are integrated over parametrized curves, k-forms can be integrated over k-dimensional parametrized regions. Let U be an open subset of $\mathbf{R}^n$ and let α be a k-form on U. The simplest k-dimensional analogue of an interval is a *rectangular block* in $\mathbf{R}^k$ whose edges are parallel to the coordinate axes. This is a set of the form

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_k, b_k] = \{\, \mathbf{t} \in \mathbf{R}^k \mid a_i \le t_i \le b_i \quad \text{for } 1 \le i \le k \,\},$$

where $a_i < b_i$. The k-dimensional analogue of a parametrized path is a smooth map $c\colon R \to U$. Although the image $c(R)$ may look very different from the block R, we think of the map c as a parametrization of the subset $c(R)$ of U: each choice of a point $\mathbf{t}$ in R gives rise to a point $c(\mathbf{t})$ in $c(R)$. The pullback $c^*\alpha$ is a k-form on R and therefore looks like $g(\mathbf{t})\, dt_1\, dt_2 \cdots dt_k$ for some function $g\colon R \to \mathbf{R}$. The *integral* of α over c is defined as

$$\int_c \alpha = \int_R c^*\alpha = \int_{a_k}^{b_k} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} g(\mathbf{t})\, dt_1\, dt_2 \cdots dt_k.$$

For $k = 1$ this reproduces the definition given in Chapter 4. (The definition makes sense if we replace the rectangular block R by more general shapes in $\mathbf{R}^k$, such as skew blocks, k-dimensional balls, cylinders, etc. In fact any compact subset of $\mathbf{R}^k$ will do.)

The case $k = 0$ is also worth examining. A zero-dimensional "block" R in $\mathbf{R}^0 = \{0\}$ is just the point 0. We can therefore think of a map $c\colon R \to U$ as a collection $\{\mathbf{x}\}$ consisting of a single point $\mathbf{x} = c(0)$ in U. The integral of a 0-form (function) f over c is by definition the value of f at $\mathbf{x}$,

$$\int_c f = f(\mathbf{x}).$$

As in the one-dimensional case, integrals of k-forms are almost wholly unaffected by a change of variables. Let

$$\bar{R} = [\bar{a}_1, \bar{b}_1] \times [\bar{a}_2, \bar{b}_2] \times \cdots \times [\bar{a}_k, \bar{b}_k]$$

be a second rectangular block. A *reparametrization* is a map $p\colon \bar{R} \to R$ satisfying the following conditions: p is bijective (i.e. one-to-one and onto) and the $k \times k$-matrix $Dp(\mathbf{s})$ is invertible for all $\mathbf{s} \in \bar{R}$. Then $\det Dp(\mathbf{s}) \neq 0$ for all $\mathbf{s} \in \bar{R}$, so either $\det Dp(\mathbf{s}) > 0$ for all $\mathbf{s}$ or $\det Dp(\mathbf{s}) < 0$ for all $\mathbf{s}$. In these cases we say that the reparametrizion *preserves*, respectively *reverses* the orientation of c.

5.1. THEOREM. *Let α be a k-form on U and $c\colon R \to U$ a smooth map. Let $p\colon \bar{R} \to R$ be a reparametrization. Then*

$$\int_{c\circ p} \alpha = \begin{cases} \int_c \alpha & \text{if } p \text{ preserves the orientation,} \\ -\int_c \alpha & \text{if } p \text{ reverses the orientation.} \end{cases}$$

PROOF. Almost verbatim the same proof as for $k = 1$ (Theorem 4.3). It follows from the definition of the integral and from the naturality of pullbacks, Proposition 3.10(iii), that

$$\int_{c\circ p} \alpha = \int_{\bar{R}} (c \circ p)^* \alpha = \int_{\bar{R}} p^*(c^*\alpha).$$

Now let us write $c^*\alpha = g\, dt_1\, dt_2 \cdots dt_k$ and $\mathbf{t} = p(\mathbf{s})$. Then

$$p^*(c^*\alpha) = p^*(g\, dt_1\, dt_2 \cdots dt_k) = (p^*g) \det Dp\, ds_1\, ds_2 \cdots ds_k$$

by Theorem 3.13, so

$$\int_{c\circ p} \alpha = \int_{\bar{R}} g(p(\mathbf{s})) \det Dp(\mathbf{s})\, ds_1\, ds_2 \cdots ds_k.$$

On the other hand, $\int_c \alpha = \int_R g(t)\, dt_1\, dt_2 \cdots dt_k$, so by the substitution formula, Theorem B.7, we have $\int_{c\circ p} \alpha = \pm \int_c \alpha$, where the $+$ occurs if $\det Dp > 0$ and the $-$ if $\det Dp < 0$. QED

5.2. EXAMPLE. The *unit interval* is the interval $[0, 1]$ in the real line. Any curve $c\colon [a, b] \to U$ can be reparametrized to a curve $c \circ p\colon [0, 1] \to U$ by means of the reparametrization $p(s) = (b - a)s + a$. Similarly, the *unit cube* in $\mathbf{R}^k$ is the rectangular block

$$[0, 1]^k = \{\, \mathbf{t} \in \mathbf{R}^k \mid t_i \in [0, 1] \quad \text{for } 1 \le i \le k \,\}.$$

Let R be any other block, given by $a_i \le t_i \le b_i$. Define $p\colon [0, 1]^k \to R$ by $p(\mathbf{s}) = A\mathbf{s} + \mathbf{a}$, where

$$A = \begin{pmatrix} b_1 - a_1 & 0 & \dots & 0 \\ 0 & b_2 - a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_k - a_k \end{pmatrix} \qquad \text{and} \qquad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}.$$

("Squeeze the unit cube until it has the same edgelengths as R and then move it to the position of R.") Then p is one-to-one and onto and $Dp(\mathbf{s}) = A$, so $\det Dp(\mathbf{s}) = \det A = \operatorname{vol} R > 0$ for all $\mathbf{s}$, so p is an orientation-preserving reparametrization. Hence $\int_{c\circ p} \alpha = \int_c \alpha$ for any k-form α on U.

5.3. REMARK. A useful fact you learned in calculus is that one may interchange the order of integration in a multiple integral, as in the formula

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} f(t_1, t_2)\, dt_1\, dt_2 = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(t_1, t_2)\, dt_2\, dt_1. \tag{5.1}$$

(This follows for instance from the substitution formula, Theorem B.7.) On the other hand, we have also learned that $f(t_1, t_2)\, dt_2\, dt_1 = -f(t_1, t_2)\, dt_1\, dt_2$. How can this be squared with formula (5.1)? The explanation is as follows. Let $\alpha = f(t_1, t_2)\, dt_1\, dt_2$. Then the left-hand side of formula (5.1) is the integral of α over

$c\colon [a_1, b_1] \times [a_2, b_2] \to \mathbf{R}^2$, the parametrization of the rectangle given by $c(t_1, t_2) = (t_1, t_2)$. The right-hand side is the integral of $-\alpha$ not over c, but over $c \circ p$, where

$$p\colon [a_2, b_2] \times [a_1, b_1] \to [a_1, b_1] \times [a_2, b_2]$$

is the reparametrization $p(s_1, s_2) = (s_2, s_1)$. Since p reverses the orientation, Theorem 5.1 says that $\int_{c\circ p} \alpha = -\int_c \alpha$; in other words $\int_c \alpha = \int_{c\circ p}(-\alpha)$, which is exactly formula (5.1). Analogously we have

$$\int_{[0,1]^k} f(t_1, t_2, \dots, t_k)\, dt_1\, dt_2 \cdots dt_k = \int_{[0,1]^k} f(t_1, t_2, \dots, t_k)\, dt_i\, dt_1\, dt_2 \cdots \widehat{dt_i} \cdots dt_k$$

for any i.

We see from Example 5.2 that an integral over any rectangular block can be written as an integral over the unit cube. For this reason, from now on we shall usually take R to be the unit cube. A smooth map $c\colon [0,1]^k \to U$ is called a *k-cube* in U (or sometimes a *singular* k-cube, the word singular meaning that the map c is not assumed to be one-to-one, so that the image can have self-intersections.)

It is often necessary to integrate over regions that are made up of several pieces. A *k-chain* in U is a formal linear combination of k-cubes,

$$c = a_1 c_1 + a_2 c_2 + \cdots + a_p c_p,$$

where $a_1, a_2, \dots, a_p$ are real coefficients and $c_1, c_2, \dots, c_p$ are k-cubes. For any k-form α we then define

$$\int_c \alpha = \sum_{i=1}^p a_i \int_{c_i} \alpha.$$

(In the language of linear algebra, the k-chains form an abstract vector space with a basis consisting of the k-cubes. Integration, which is a priori only defined on cubes, is extended to chains in such a way as to be linear.)

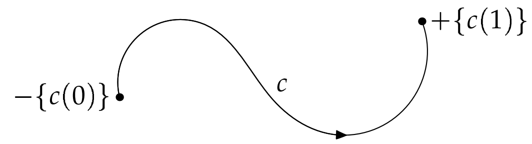
Recall that a 0-cube is nothing but a singleton $\{\mathbf{x}\}$ consisting of a single point $\mathbf{x}$ in U. Thus a 0-chain is a formal linear combination of points, $c = \sum_{i=1}^p a_i\{\mathbf{x}_i\}$. A good way to think of c is as a collection of p point charges, with an electric charge a_i placed at the point $\mathbf{x}_i$. (You must carefully distinguish between the formal linear combination $\sum_{i=1}^p a_i\{\mathbf{x}_i\}$, which represents a distribution of point charges, and the linear combination of vectors $\sum_{i=1}^p a_i\mathbf{x}_i$, which represents a vector in $\mathbf{R}^n$.) The integral of a function f over the 0-chain is by definition

$$\int_c f = \sum_{i=1}^p a_i \int_{\{\mathbf{x}_i\}} f = \sum_{i=1}^p a_i f(\mathbf{x}_i).$$

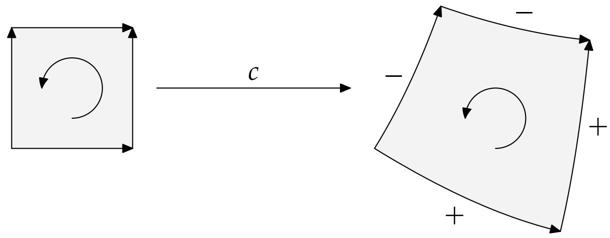
Likewise, a k-chain $\sum_{i=1}^p a_i c_i$ can be pictured as a charge distribution, with an electric charge a_i spread along the k-dimensional "patch" c_i.

5.2. The boundary of a chain

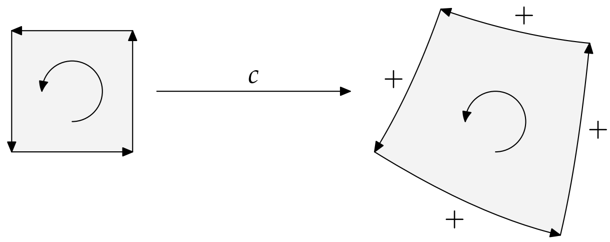
Consider a curve ("1-cube") $c\colon [0,1] \to U$. Its *boundary* is by definition the 0-chain defined by $\partial c = \{c(1)\} - \{c(0)\}$.



The boundary of a 2-cube $c\colon [0,1]^2 \to U$ consists of four pieces corresponding to the edges of the unit square: $c_1(t) = c(t,0)$, $c_2(t) = c(1,t)$, $c_3(t) = c(t,1)$ and $c_4(t) = c(0,t)$. The picture below suggests that we should define $\partial c = c_1 + c_2 - c_3 - c_4$.



(Alternatively we could define $\partial c = c_1 + c_2 + \bar{c}_3 + \bar{c}_4$, with $\bar{c}_3(t) = c(1-t,1)$ and $\bar{c}_4(t) = c(0,1-t)$, which corresponds to the following picture:



This would work equally well, but is technically less convenient.)

A k-cube $c\colon [0,1]^k \to U$ has $2k$ faces of dimension $k-1$, which are described as follows. Let $\mathbf{t} = (t_1, t_2, \dots, t_{k-1}) \in [0,1]^{k-1}$ and for $i = 1, 2, \dots, k$ put

$$\begin{aligned} c_{i,0}(\mathbf{t}) &= c(t_1, t_2, \dots, t_{i-1}, 0, t_i, \dots, t_{k-1}), \\ c_{i,1}(\mathbf{t}) &= c(t_1, t_2, \dots, t_{i-1}, 1, t_i, \dots, t_{k-1}). \end{aligned}$$

("Insert 0, resp. 1 in the i-th slot".) Now define

$$\partial c = \sum_{i=1}^{k} (-1)^i (c_{i,0} - c_{i,1}) = \sum_{i=1}^{k} \sum_{\rho=0,1} (-1)^{i+\rho} c_{i,\rho}.$$

For an arbitrary k-chain $c = \sum_i a_i c_i$ we put $\partial c = \sum_i a_i \partial c_i$. Then ∂ is a linear map from k-chains to $k-1$-chains. You should check that for $k = 0$ and $k = 1$ this definition is consistent with the one- and two-dimensional cases considered above.

There are a number of curious similarities between the boundary operator ∂ and the exterior derivative d, the most important of which is the following. (There are also many differences, such as the fact that d *raises* the degree of a form by 1, whereas ∂ *lowers* the dimension of a chain by 1.)

5.4. PROPOSITION. $\partial(\partial c) = 0$ *for every k-chain c in U. In short,*

$$\boxed{\partial^2 = 0.}$$

PROOF. By linearity of ∂ it suffices to prove this for k-cubes $c\colon [0,1]^k \to U$. Let $\mathbf{t} = (t_1, t_2, \dots, t_{k-2}) \in [0,1]^{k-2}$ and let ρ and σ be 0 or 1. Then for $1 \le i \le j \le k-1$

we have

$$\begin{aligned}(c_{i,\rho})_{j,\sigma}(t_1,t_2,\dots,t_{k-2}) &= c_{i,\rho}(t_1,t_2,\dots,t_{j-1},\sigma,t_j,\dots,t_{k-2})\\ &= c(t_1,t_2,\dots,t_{i-1},\rho,t_i,\dots,t_{j-1},\sigma,t_j,\dots,t_{k-2}).\end{aligned}$$

On the other hand,

$$\begin{aligned}(c_{j+1,\sigma})_{i,\rho}(t_1,t_2,\dots,t_{k-2}) &= c_{j+1,\sigma}(t_1,t_2,\dots,t_{i-1},\rho,t_i,\dots,t_{k-2})\\ &= c(t_1,t_2,\dots,t_{i-1},\rho,t_i,\dots,t_{j-1},\sigma,t_j,\dots,t_{k-2}),\end{aligned}$$

because in the vector $(t_1,t_2,\dots,t_{i-1},\rho,t_i,\dots,t_{k-2})$ the entry t_j occupies the $j+1$st slot! We conclude that $(c_{i,\rho})_{j,\sigma} = (c_{j+1,\sigma})_{i,\rho}$ for $1 \le i \le j \le k-1$. Therefore the $k-2$-chain $\partial(\partial c)$ is given by

$$\begin{aligned}\partial(\partial c) &= \partial \sum_{i=1}^{k}\sum_{\rho=0,1}(-1)^{i+\rho}c_{i,\rho} = \sum_{i=1}^{k}\sum_{\rho=0,1}(-1)^{i+\rho}\partial c_{i,\rho}\\ &= \sum_{i=1}^{k}\sum_{j=1}^{k-1}\sum_{\rho,\sigma=0,1}(-1)^{i+j+\rho+\sigma}(c_{i,\rho})_{j,\sigma}.\end{aligned}$$

The double sum over i and j can be rearranged in a sum over $i \le j$ and a sum over $i > j$ to give

$$\begin{aligned}\partial(\partial c) = &\sum_{1\le i\le j\le k-1}\sum_{\rho,\sigma=0,1}(-1)^{i+j+\rho+\sigma}(c_{i,\rho})_{j,\sigma}\\ &+\sum_{1\le j<i\le k}\sum_{\rho,\sigma=0,1}(-1)^{i+j+\rho+\sigma}(c_{i,\rho})_{j,\sigma}.\end{aligned} \tag{5.2}$$

In the first term on the right in (5.2) we substitute $(c_{i,\rho})_{j,\sigma} = (c_{j+1,\sigma})_{i,\rho}$ and then $r = j+1$, $s = i$, $\mu = \sigma$, $\nu = \rho$ to get

$$\begin{aligned}\sum_{1\le i\le j\le k-1}\sum_{\rho,\sigma=0,1}(-1)^{i+j+\rho+\sigma}(c_{i,\rho})_{j,\sigma} &= \sum_{1\le i\le j\le k-1}\sum_{\rho,\sigma=0,1}(-1)^{i+j+\rho+\sigma}(c_{j+1,\sigma})_{i,\rho}\\ &= \sum_{1\le s<r\le k}\sum_{\mu,\nu=0,1}(-1)^{s+r-1+\nu+\mu}(c_{r,\mu})_{s,\nu}\\ &= -\sum_{1\le s<r\le k}\sum_{\mu,\nu=0,1}(-1)^{s+r+\nu+\mu}(c_{r,\mu})_{s,\nu}.\end{aligned}$$

Thus the two terms on the right in (5.2) cancel out. QED

5.3. Cycles and boundaries

A k-cube c is *degenerate* if $c(t_1,\dots,t_k)$ is independent of t_i for some i. A k-chain c is *degenerate* if it is a linear combination of degenerate cubes. In particular, a degenerate 1-cube is a constant curve. The work done by a force field on a motionless particle is 0. More generally we have the following.

5.5. LEMMA. *Let α be a k-form and c a degenerate k-chain. Then $\int_c \alpha = 0$.*

PROOF. By linearity we may assume that c is a degenerate cube. Suppose c is constant as a function of t_i. Then

$$c(t_1,\dots,t_i,\dots,t_k) = c(t_1,\dots,0,\dots,t_k) = g\big(f(t_1,\dots,t_i,\dots,t_k)\big),$$

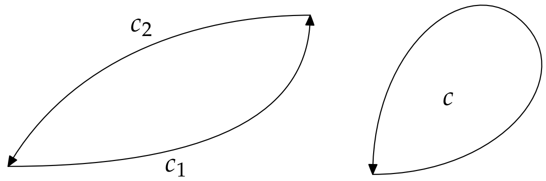
where $f\colon [0,1]^k \to [0,1]^{k-1}$ and $g\colon [0,1]^{k-1} \to U$ are given respectively by

$$f(t_1, \dots, t_i, \dots, t_k) = (t_1, \dots, \hat{t}_i, \dots, t_k),$$
$$g(s_1, \dots, s_{k-1}) = c(s_1, \dots, s_{i-1}, 0, s_{i+1}, \dots, t_{k-1}).$$

Now $g^*\alpha$ is a k-form on $[0,1]^{k-1}$ and hence equal to 0, and so $c^*\alpha = f^*(g^*\alpha) = 0$. We conclude $\int_c \alpha = \int_{[0,1]^k} c^*\alpha = 0$. QED

So degenerate chains are irrelevant where integration is concerned. This motivates the following definition. A k-chain c is *closed*, or a *cycle*, if ∂c is a degenerate $k-1$-chain. A k-chain c is a *boundary* if $c = \partial b + c'$ for some $k+1$-chain b and some degenerate k-chain c'.

5.6. EXAMPLE. If c_1 and c_2 are curves arranged head to tail as in the picture below, then $c_1 + c_2$ is a 1-cycle. Likewise, the closed curve c is a 1-cycle.



5.7. LEMMA. *The boundary of a degenerate k-chain is a degenerate $k-1$-chain.*

PROOF. By linearity it suffices to consider the case of a degenerate k-cube c. Suppose c is constant as a function of t_i. Then $c_{i,0} = c_{i,1}$, so

$$\partial c = \sum_{j \neq i} (-1)^j (c_{j,0} - c_{j,1}).$$

Let $\mathbf{t} = (t_1, t_2, \dots, t_{k-1})$. For $j > i$ the cubes $c_{j,0}(\mathbf{t})$ and $c_{j,1}(\mathbf{t})$ are independent of t_i and for $j < i$ they are independent of t_{i-1}. So ∂c is a combination of degenerate $k-1$-cubes and hence is degenerate. QED

5.8. COROLLARY. *Every boundary is a cycle.*

PROOF. Suppose $c = \partial b + c'$ with c' degenerate. Then by Lemma 5.5 $\partial c = \partial(\partial b) + \partial c' = \partial c'$, where we used Proposition 5.4. Lemma 5.7 says that $\partial c'$ is degenerate, and therefore so is ∂c. QED

5.9. EXAMPLE. Consider the unit circle in the plane $c(t) = (\cos 2\pi t, \sin 2\pi t)$ with $0 \leq t \leq 1$. This is a closed 1-cube. The circle is the boundary of the disc of radius 1 and therefore it is reasonable to expect that c is a boundary of a 2-cube. This is indeed true in the sense defined above, that $c = \partial b + c'$ where c' is a constant 1-chain. (It is actually not possible to find a b such that $c = \partial b$; see Exercise 5.2.) The 2-cube b is defined by "shrinking c to a point", $b(t_1, t_2) = (1 - t_2)c(t_1)$ for (t_1, t_2) in the unit square. Then

$$b(t_1, 0) = c(t_1), \qquad b(0, t_2) = b(1, t_2) = (1 - t_2, 0), \qquad b(t_1, 1) = (0, 0),$$

so that $\partial b = c - c'$, where c' is the constant curve located at the origin. Therefore $c = \partial b + c'$, a boundary plus a degenerate 1-cube.

In the same way that a closed form is not necessarily exact, it may happen that a 1-cycle is not a boundary. See Example 5.11.

5.4. Stokes' theorem

In the language of chains and boundaries we can rewrite the fundamental theorem of calculus, Theorem 4.4, as follows:

$$\int_c dg = g(c(1)) - g(c(0)) = \int_{\{c(1)\}} g - \int_{\{c(0)\}} g = \int_{\{c(1)\}-\{c(0)\}} g = \int_{\partial c} g,$$

i.e. $\int_c dg = \int_{\partial c} g$. This is the form in which the fundamental theorem of calculus generalizes to higher dimensions. This generalization is perhaps the neatest relationship between the exterior derivative and the boundary operator. It contains as special cases the classical integration formulas of vector calculus (Green, Gauß and Stokes) and for that reason has Stokes' name attached to it, although it would perhaps be better to call it the "fundamental theorem of multivariable calculus".

5.10. THEOREM (Stokes' theorem). *Let α be a $k-1$-form on an open subset U of $\mathbf{R}^n$ and let c be a k-chain in U. Then*

$$\boxed{\int_c d\alpha = \int_{\partial c} \alpha.}$$

PROOF. By the definition of the integral and by Theorem 3.11 we have

$$\int_c d\alpha = \int_{[0,1]^k} c^* d\alpha = \int_{[0,1]^k} dc^*\alpha.$$

Since $c^*\alpha$ is a $k-1$-form on $[0,1]^k$, it can be written as

$$c^*\alpha = \sum_{i=1}^k g_i \, dt_1 \, dt_2 \cdots \widehat{dt_i} \cdots dt_k$$

for certain functions $g_1, g_2, \dots, g_k$ defined on $[0,1]^k$. Therefore

$$\int_c d\alpha = \sum_{i=1}^k \int_{[0,1]^k} d\big(g_i \, dt_1 \, dt_2 \cdots \widehat{dt_i} \cdots dt_k\big) = \sum_{i=1}^k (-1)^{i+1} \int_{[0,1]^k} \frac{\partial g_i}{\partial t_i} \, dt_1 \, dt_2 \cdots dt_k.$$

Changing the order of integration (cf. Remark 5.3) and subsequently applying the fundamental theorem of calculus in one variable, formula (B.1), gives

$$\begin{aligned}\int_{[0,1]^k} \frac{\partial g_i}{\partial t_i} \, dt_1 \, dt_2 \cdots dt_k &= \int_{[0,1]^k} \frac{\partial g_i}{\partial t_i} \, dt_i \, dt_1 \, dt_2 \cdots \widehat{dt_i} \cdots dt_k \\ &= \int_{[0,1]^{k-1}} \big(g_i(t_1, \dots, t_{i-1}, 1, t_{i+1}, \dots, t_k) \\ &\qquad - g_i(t_1, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_k)\big) \, dt_1 \, dt_2 \cdots \widehat{dt_i} \cdots dt_k.\end{aligned}$$

The forms

$$g_i(t_1, \dots, t_{i-1}, 1, t_{i+1}, \dots, t_k) \, dt_1 \, dt_2 \cdots \widehat{dt_i} \cdots dt_k \quad \text{and}$$
$$g_i(t_1, \dots, t_{i-1}, 0, t_{i+1}, \dots, t_k) \, dt_1 \, dt_2 \cdots \widehat{dt_i} \cdots dt_k$$

are nothing but $c_{i,1}^*\alpha$, resp. $c_{i,0}^*\alpha$. Accordingly,

$$\begin{aligned}\int_c d\alpha &= \sum_{i=1}^k (-1)^{i+1} \int_{[0,1]^k} \frac{\partial g_i}{\partial t_i}\, dt_i\, dt_1\, dt_2 \cdots \widehat{dt_i} \cdots dt_k \\ &= \sum_{i=1}^k (-1)^{i+1} \int_{[0,1]^{k-1}} (c_{i,1}^*\alpha - c_{i,0}^*\alpha) \\ &= \sum_{i=1}^k \sum_{\rho=0,1} (-1)^{i+\rho} \int_{[0,1]^{k-1}} c_{i,\rho}^*\alpha \\ &= \sum_{i=1}^k \sum_{\rho=0,1} (-1)^{i+\rho} \int_{c_{i,\rho}} \alpha = \int_{\partial c} \alpha,\end{aligned}$$

which proves the result. QED

5.11. EXAMPLE. The unit circle $c(t) = (\cos 2\pi t, \sin 2\pi t)$ is a 1-cycle in the punctured plane $U = \mathbf{R}^2 - \{\mathbf{0}\}$. Considered as a chain in $\mathbf{R}^2$ it is also a boundary, as we saw in Example 5.9. However, we claim that it is not a boundary in U in the sense that there exist no 2-chain b and no degenerate 1-chain c' *both contained in* U such that $c = \partial b + c'$. Indeed, suppose that $c = \partial b + c'$. Let $\alpha = (-y\,dx + x\,dy)/(x^2 + y^2)$ be the angle form. Then $\int_c \alpha = 2\pi$ by Example 4.1.On the other hand,

$$\int_c \alpha = \int_{\partial b + c'} \alpha = \int_b d\alpha = 0,$$

where we have used Stokes' theorem, Lemma 5.5 and the fact that α is closed. This is a contradiction. The moral of this example is that the presence of the puncture in U is responsible both for the existence of the non-exact closed 1-form α (see Example 4.6) and for the closed 1-chain c which is not a boundary. We detected both phenomena by using Stokes' theorem.

Exercises

5.1. Let U be an open subset of $\mathbf{R}^n$, V an open subset of $\mathbf{R}^m$ and $\phi\colon U \to V$ a smooth map. Let c be a k-cube in U and α a k-form on V. Prove that $\int_c \phi^*\alpha = \int_{\phi\circ c} \alpha$.

5.2. Let U be an open subset of $\mathbf{R}^n$. Its boundary is a linear combination of $k-1$-cubes, $\partial c = \sum_i a_i c_i$.

(i) Let b be a $k+1$-chain in U. Its boundary is a linear combination of k-cubes, $\partial b = \sum_i a_i c_i$. Prove that $\sum_i a_i = 0$.
(ii) Let c be a k-cube in U. Conclude that there exists no $k+1$-chain b in U satisfying $\partial b = c$.

5.3. Define a 2-cube $c\colon [0,1]^2 \to \mathbf{R}^3$ by $c(t_1, t_2) = \left(t_1^2, t_1 t_2, t_2^2\right)$, and let $\alpha = x_1\,dx_2 + x_1\,dx_3 + x_2\,dx_3$.

(i) Sketch the image of c.
(ii) Calculate both $\int_c d\alpha$ and $\int_{\partial c} \alpha$ and check that they are equal.

5.4. Define a 3-cube $c\colon [0,1]^3 \to \mathbf{R}^3$ by $c(t_1, t_2, t_3) = (t_2 t_3, t_1 t_3, t_1 t_2)$, and let $\alpha = x_1\,dx_2\,dx_3$. Calculate both $\int_c d\alpha$ and $\int_{\partial c} \alpha$ and check that they are equal.

5.5. Using polar coordinates in n dimensions (cf. Exercise 3.18) write the $n-1$-dimensional unit sphere S^{n-1} in $\mathbf{R}^n$ as the image of an $n-1$-cube c.For $n = 2$, 3, 4, calculate the boundary ∂c of this cube. (The domain of c will not be the unit cube in $\mathbf{R}^{n-1}$, but a rectangular block R dictated by the formula in Exercise 3.18. Choose R in such a way as to cover the sphere as economically as possible.)

5.6. Deduce the following classical integration formulas from the generalized version of Stokes' theorem. All functions, vector fields, chains etc. are smooth and are defined in an open subset U of $\mathbf{R}^n$. (Some formulas hold only for special values of n, as indicated.)

(i) $\displaystyle\int_c \operatorname{grad} g \cdot d\mathbf{x} = g(c(1)) - g(c(0))$ for any function g and any curve c.

(ii) Green's formula: $\displaystyle\int_c \Big(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\Big)\, dx\, dy = \int_{\partial c} (f\, dx + g\, dy)$ for any functions f, g and any 2-chain c. (Here $n = 2$.)

(iii) Gauß' formula: $\displaystyle\int_c \operatorname{div} \mathbf{F}\, dx_1\, dx_2 \cdots dx_n = \int_{\partial c} \mathbf{F} \cdot {*d\mathbf{x}}$ for any vector field $\mathbf{F}$ and any n-chain c.

(iv) Stokes' formula: $\displaystyle\int_c \operatorname{curl} \mathbf{F} \cdot {*d\mathbf{x}} = \int_{\partial c} \mathbf{F} \cdot d\mathbf{x}$ for any vector field $\mathbf{F}$ and any 2-chain c. (Here $n = 3$.)

In parts (iii) and (iv) we use the notations $d\mathbf{x}$ and $*d\mathbf{x}$ explained in Section 2.5. We shall give a geometric interpretation of the entity $*d\mathbf{x}$ in terms of volume forms later on. (See Corollary 8.15.)

CHAPTER 6

Manifolds

6.1. The definition

Intuitively, an n-dimensional manifold in the Euclidean space $\mathbf{R}^N$ is a subset that in the neighbourhood of every point "looks like" $\mathbf{R}^n$ up to "smooth distortions". The formal definition is given below and is unfortunately a bit long. It will help to consider first the basic example of the surface of the earth, which is a two-dimensional sphere placed in three-dimensional space. A useful way to represent the earth is by means of a world atlas, which is a collection of maps. Each map depicts a portion of the world, such as a country or an ocean. The correspondence between points on a map and points on the earth's surface is not entirely faithful, because charting a curved surface on a flat piece of paper inevitably distorts the distances between points. But the distortions are continuous, indeed differentiable (in most traditional cartographic projections). Maps of neighbouring areas overlap near their edges and the totality of all maps in a world atlas covers the whole world.

An arbitrary manifold is defined similarly, as an n-dimensional "world" represented by an "atlas" consisting of "maps". These maps are a special kind of parametrizations known as embeddings.

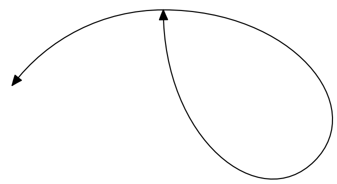
6.1. DEFINITION. Let U be an open subset of $\mathbf{R}^n$. An *embedding* of U into $\mathbf{R}^N$ is a C^∞ map $\psi\colon U \to \mathbf{R}^N$ satisfying the following conditions:

(i) ψ is one-to-one (i.e. if $\psi(\mathbf{t}_1) = \psi(\mathbf{t}_2)$, then $\mathbf{t}_1 = \mathbf{t}_2$);
(ii) $D\psi(\mathbf{t})$ is one-to-one for all $\mathbf{t} \in U$;
(iii) the inverse of ψ, which is a map $\psi^{-1}\colon \psi(U) \to U$, is continuous.

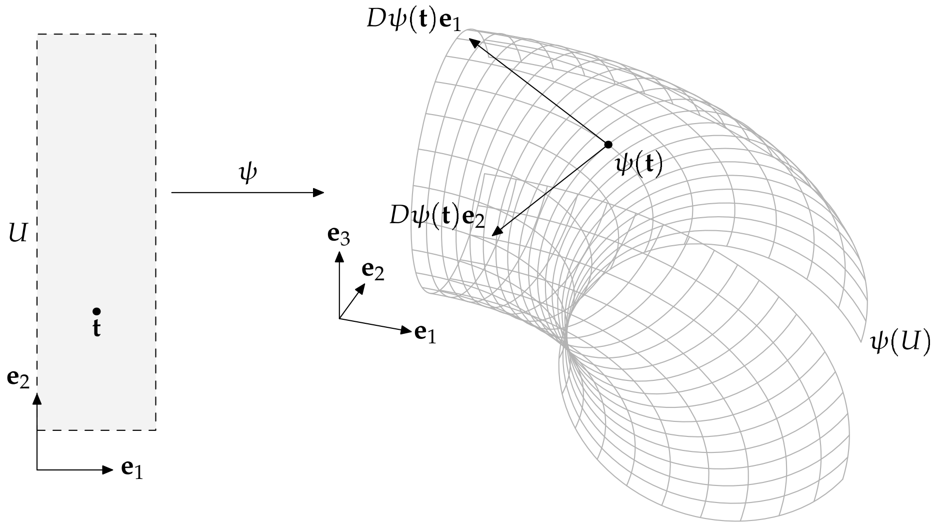
The *image* of the embedding is the set $\psi(U) = \{\, \psi(\mathbf{t}) \mid \mathbf{t} \in U \,\}$ consisting of all points of the form $\psi(\mathbf{t})$ with $\mathbf{t} \in U$. The inverse map ψ^{-1} is called a *chart* or *coordinate map*. You should think of $\psi(U)$ as an n-dimensional "patch" in $\mathbf{R}^N$ parametrized by the map ψ. Condition (i) means that to distinct values of the "parameter" $\mathbf{t}$ must correspond distinct points $\psi(\mathbf{t})$ in the patch $\psi(U)$. Thus the patch $\psi(U)$ has no self-intersections. Condition (ii) means that for each $\mathbf{t}$ in U all n columns of the Jacobi matrix $D\psi(\mathbf{t})$ must be independent. This is imposed to prevent the occurrence of cusps and other singularities in the image $\psi(U)$. Since $D\psi(\mathbf{t})$ has N rows, this condition also implies that $N \geq n$: the target space $\mathbf{R}^N$ must have dimension greater than or equal to that of the source space U, or else ψ cannot be an embedding. The column space of $D\psi(\mathbf{t})$ is called the *tangent space* to the patch at the point $\mathbf{x} = \psi(\mathbf{t})$ and is denoted by $T_{\mathbf{x}}\psi(U)$,

$$\boxed{T_{\mathbf{x}}\psi(U) = D\psi(\mathbf{t})(\mathbf{R}^n).}$$

The tangent space at each point is an n-dimensional subspace of $\mathbf{R}^N$ because $D\psi(\mathbf{t})$ has n independent columns. Condition (iii) can be restated as the requirement that if $\mathbf{t}_i$ is any sequence of points in U such that $\lim_{i\to\infty} \psi(\mathbf{t}_i)$ exists and is equal to $\psi(\mathbf{t})$ for some $\mathbf{t} \in U$, then $\lim_{i\to\infty} \mathbf{t}_i = \mathbf{t}$. This is intended to avoid situations where the image $\psi(U)$ doubles back on itself "at infinity". (See Exercise 6.4 for an example.)



6.2. EXAMPLE. The picture below shows an embedding of an open rectangle in the plane into three-space, the image of which is a portion of a torus. Try to write a formula for such an embedding! (If we chose U too big, the image would self-intersect and the map would not be an embedding.) For one particular value of $\mathbf{t}$ the column vectors of the Jacobi matrix are also shown. As you can see, they span the tangent plane at the image point.



6.3. EXAMPLE. Let U be an open subset of $\mathbf{R}^n$ and let $f\colon U \to \mathbf{R}^m$ be a smooth map. The *graph* of f is the collection

$$\operatorname{graph} f = \left\{ \begin{pmatrix} \mathbf{t} \\ f(\mathbf{t}) \end{pmatrix} \,\middle|\, \mathbf{t} \in U \right\}.$$

Since $\mathbf{t}$ is an n-vector and $f(\mathbf{t})$ an m-vector, the graph is a subset of $\mathbf{R}^N$ with $N = n + m$. We claim that the graph is the image of an embedding $\psi\colon U \to \mathbf{R}^N$. Define

$$\psi(\mathbf{t}) = \begin{pmatrix} \mathbf{t} \\ f(\mathbf{t}) \end{pmatrix}.$$

Then by definition $\operatorname{graph} f = \psi(U)$. Furthermore ψ is an embedding. Indeed, $\psi(\mathbf{t}_1) = \psi(\mathbf{t}_2)$ implies $\mathbf{t}_1 = \mathbf{t}_2$, so ψ is one-to-one. Also,

$$D\psi(\mathbf{t}) = \begin{pmatrix} I_n \\ Df(\mathbf{t}) \end{pmatrix},$$

so $D\psi(\mathbf{t})$ has n independent columns. Finally the inverse of ψ is given by

$$\psi^{-1}\begin{pmatrix}\mathbf{t}\\ f(\mathbf{t})\end{pmatrix} = \mathbf{t},$$

which is continuous. Hence ψ is an embedding.

A manifold is an object patched together out of the images of several embeddings. More precisely,

6.4. DEFINITION. An *n-dimensional manifold*[1] (or *n-manifold* for short) in $\mathbf{R}^N$ is a subset M of $\mathbf{R}^N$ such that for all $\mathbf{x} \in M$ there exist

- an open subset $V \subseteq \mathbf{R}^N$ containing $\mathbf{x}$,
- an open subset $U \subseteq \mathbf{R}^n$,
- and an embedding $\psi\colon U \to \mathbf{R}^N$ satisfying $\psi(U) = V \cap M$.

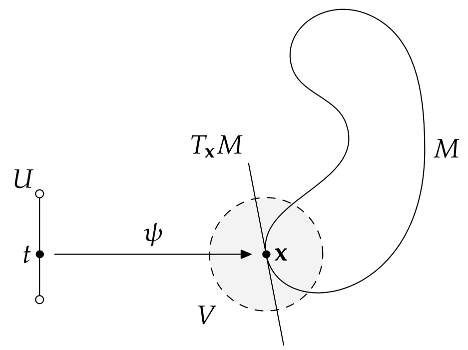
The *codimension* of M in $\mathbf{R}^N$ is $N - n$. Choose $\mathbf{t} \in U$ such that $\psi(\mathbf{t}) = \mathbf{x}$. Then the *tangent space* to M at $\mathbf{x}$ is the column space of $D\psi(\mathbf{t})$,

$$\boxed{T_{\mathbf{x}}M = D\psi(\mathbf{t})(\mathbf{R}^n).}$$

(Using the chain rule one can show that $T_{\mathbf{x}}M$ is independent of the choice of the embedding ψ.) The elements of $T_{\mathbf{x}}M$ are *tangent vectors* to M at $\mathbf{x}$. A collection of embeddings $\psi_i\colon U_i \to \mathbf{R}^N$ with U_i open in $\mathbf{R}^n$ and such that M is the union of all the sets $\psi_i(U_i)$ is an *atlas* for M.

One-dimensional manifolds are called *(smooth) curves*, two-dimensional manifolds *(smooth) surfaces*, and n-manifolds in $\mathbf{R}^{n+1}$ *(smooth) hypersurfaces*. In these cases the tangent spaces are usually called *tangent lines*, *tangent planes*, and *tangent hyperplanes*, respectively.

The following picture illustrates the definition. Here M is a curve in the plane, so we have $N = 2$ and $n = 1$. U is an open interval in $\mathbf{R}$ and V is an open disc in $\mathbf{R}^2$. The map ψ sends t to $\mathbf{x}$ and parametrizes the portion of the curve inside V. Since $n = 1$, the Jacobi matrix $D\psi(t)$ consists of a single column vector, which is tangent to the curve at $\mathbf{x} = \psi(t)$. The tangent line $T_{\mathbf{x}}M$ is the line spanned by this vector.



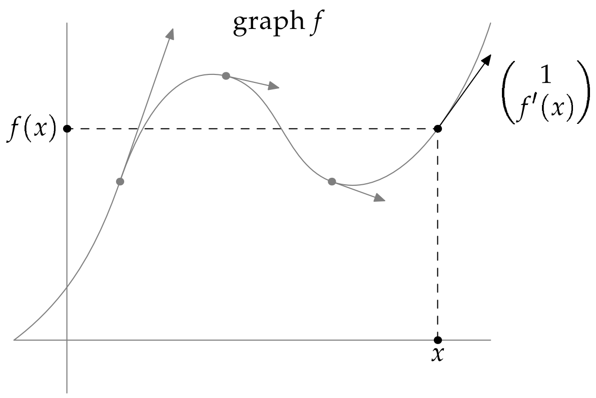
Sometimes a manifold has an atlas consisting of one single chart. In that event we can take $V = \mathbf{R}^N$, and choose one open $U \subseteq \mathbf{R}^n$ and an embedding $\psi\colon U \to \mathbf{R}^N$ such that $M = \psi(U)$. However, usually one needs more than one chart to cover a manifold. (For instance, one chart is not enough for the curve M in the picture above.)

[1]In the literature this is usually called a *submanifold* of Euclidean space. It is possible to define manifolds more abstractly, without reference to a surrounding vector space. However, it turns out that practically all abstract manifolds can be embedded into a vector space of sufficiently high dimension. Hence the abstract notion of a manifold is not substantially more general than the notion of a submanifold of a vector space.

6.5. EXAMPLE. An open subset U of $\mathbf{R}^n$ can be regarded as a manifold of dimension n (hence of codimension 0). Indeed, U is the image of the map $\psi\colon U \to \mathbf{R}^n$ given by $\psi(\mathbf{x}) = \mathbf{x}$, the identity map. The tangent space to U at any point is $\mathbf{R}^n$ itself.

6.6. EXAMPLE. Let $N \geq n$ and define $\psi\colon \mathbf{R}^n \to \mathbf{R}^N$ by $\psi(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n, 0, 0, \dots, 0)$. It is easy to check that ψ is an embedding. Hence the image $\psi(\mathbf{R}^n)$ is an n-manifold in $\mathbf{R}^N$. (Note that $\psi(\mathbf{R}^n)$ is just a linear subspace isomorphic to $\mathbf{R}^n$; e.g. if $N = 3$ and $n = 2$ it is just the xy-plane. We shall usually identify $\mathbf{R}^n$ with its image in $\mathbf{R}^N$.) Combining this example with the previous one, we see that if U is any open subset of $\mathbf{R}^n$, then $\psi(U)$ is a manifold in $\mathbf{R}^N$ of codimension $N - n$. Its tangent space at any point is $\mathbf{R}^n$.

6.7. EXAMPLE. Let $M = \operatorname{graph} f$, where $f\colon U \to \mathbf{R}^m$ is a smooth map. As shown in Example 6.3, M is the image of a single embedding $\psi\colon U \to \mathbf{R}^{n+m}$, so M is an n-dimensional manifold in $\mathbf{R}^{n+m}$, covered by a single chart. At a point $(\mathbf{x}, f(\mathbf{x}))$ in the graph the tangent space is spanned by the columns of $D\psi$. For instance, if $n = m = 1$, M is one-dimensional and the tangent line to M at $(x, f(x))$ is spanned by the vector $(1, f'(x))$. This is equivalent to the well-known fact that the slope of the tangent line to the graph at x is $f'(x)$.

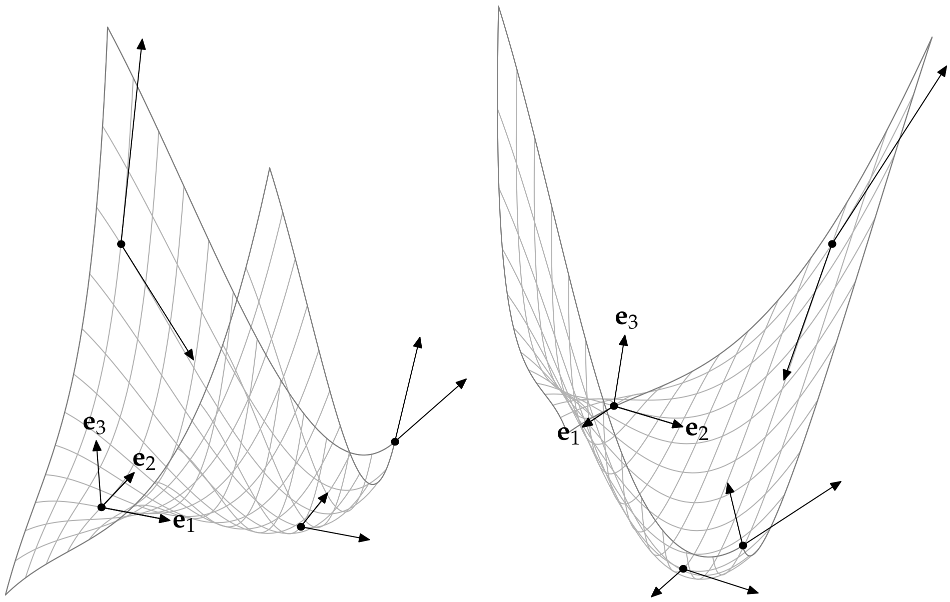


For $n = 2$ and $m = 1$, M is a surface in $\mathbf{R}^3$. The tangent plane to M at a point $(x, y, f(x, y))$ is spanned by the columns of $D\psi(x, y)$, namely

$$\begin{pmatrix} 1 \\ \frac{\partial f}{\partial x}(x, y) \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial y}(x, y) \end{pmatrix}.$$

The diagram below shows the graph of $f(x, y) = x^3 + y^3 - 3xy$ from two different angles, together with a few points and tangent vectors. (To improve the scale the

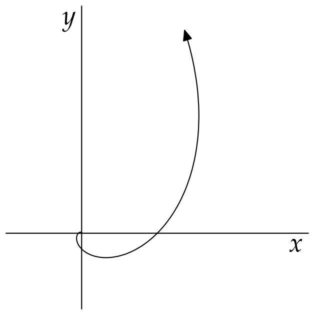
z-coordinate and the tangent vectors have been compressed by a factor of 2.)



6.8. EXAMPLE. Consider the path $\psi\colon \mathbf{R} \to \mathbf{R}^2$ given by $\psi(t) = e^t(\cos t, \sin t)$. Let us check that ψ is an embedding. Observe first that $\|\psi(t)\| = e^t$. Therefore $\psi(t_1) = \psi(t_2)$ implies $e^{t_1} = e^{t_2}$. The exponential function is one-to-one, so $t_1 = t_2$. This shows that ψ is one-to-one. The velocity vector is

$$\psi'(t) = e^t \begin{pmatrix} \cos t - \sin t \\ \cos t + \sin t \end{pmatrix}.$$

Therefore $\psi'(t) = \mathbf{0}$ if and only if $\cos t = \sin t = 0$, which is impossible because $\cos^2 t + \sin^2 t = 1$. So $\psi'(t) \neq \mathbf{0}$ for all t. Moreover we have $t = \ln e^t = \ln\|\psi(t)\|$. Hence the inverse of ψ is given by $\psi^{-1}(\mathbf{x}) = \ln\|\mathbf{x}\|$ for $\mathbf{x} \in \psi(\mathbf{R})$ and so is continuous. Therefore ψ is an embedding and $\psi(\mathbf{R})$ is a 1-manifold. The image $\psi(\mathbf{R})$ is a spiral, which for $t \to -\infty$ converges to the origin. It winds infinitely many times around the origin, although that is hard to see in the picture.



Even though $\psi(\mathbf{R})$ is a manifold, the set $\psi(\mathbf{R}) \cup \{\mathbf{0}\}$ is not: it has a very nasty singularity at the origin!

6.9. EXAMPLE. An example of a manifold which cannot be covered by a single chart is the unit sphere $M = S^{n-1}$ in $\mathbf{R}^n$. Let $U = \mathbf{R}^{n-1}$ and let $\psi\colon U \to \mathbf{R}^n$ be the map

$$\psi(\mathbf{t}) = \frac{1}{\|\mathbf{t}\|^2 + 1}\bigl(2\mathbf{t} + (\|\mathbf{t}\|^2 - 1)\mathbf{e}_n\bigr)$$

given in Exercise B.7. As we saw in that exercise, the image of ψ is the punctured sphere $M - \{\mathbf{e}_n\}$, so if we let V be the open set $\mathbf{R}^n - \{\mathbf{e}_n\}$, then $\psi(U) = M \cap V$. Also we saw that ψ has a two-sided inverse $\phi\colon \psi(U) \to U$, the stereographic projection from the north pole. Therefore ψ is one-to-one and its inverse is continuous (indeed, differentiable). Moreover, $\phi \circ \psi(\mathbf{t}) = \mathbf{t}$ implies $D\phi(\psi(\mathbf{t}))D\psi(\mathbf{t})\mathbf{v} = \mathbf{v}$ for all $\mathbf{v}$ in $\mathbf{R}^{n-1}$ by the chain rule. Therefore, if $\mathbf{v}$ is in the nullspace of $D\psi(\mathbf{t})$,

$$\mathbf{v} = D\phi(\psi(\mathbf{t}))D\psi(\mathbf{t})\mathbf{v} = D\phi(\psi(\mathbf{t}))\mathbf{0} = \mathbf{0}.$$

Thus we see that ψ is an embedding. To cover all of M we need a second map, for example the inverse of the stereographic projection from the south pole. This is also an embedding and its image is $M - \{-\mathbf{e}_n\} = M \cap V$, where $V = \mathbf{R}^n - \{e_n\}$. This finishes the proof that M is an $n-1$-manifold in $\mathbf{R}^n$.

As this example shows, the definition of a manifold can be a little awkward to work with in practice, even for a very simple manifold. Aside from the above examples, in practice it can be rather hard to decide whether a given subset is a manifold using the definition alone. Fortunately there exists a more manageable criterion for a set to be a manifold.

6.2. The regular value theorem

Definition 6.4 is based on the notion of an embedding, which can be regarded as an "explicit" way of describing a manifold. However, embeddings can be hard find in practice. Instead, manifolds are often given "implicitly", by a system of m equations in N unknowns,

$$\begin{aligned}\phi_1(x_1, \dots, x_N) &= c_1,\\ \phi_2(x_1, \dots, x_N) &= c_2,\\ &\;\vdots\\ \phi_m(x_1, \dots, x_N) &= c_m.\end{aligned}$$

Here the ϕ_i's are smooth functions presumed to be defined on some common open subset U of $\mathbf{R}^N$. Writing in the usual way

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}, \quad \phi(\mathbf{x}) = \begin{pmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \\ \vdots \\ \phi_m(\mathbf{x}) \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix},$$

we can abbreviate this system to a single equation

$$\phi(\mathbf{x}) = \mathbf{c}.$$

For a fixed vector $\mathbf{c} \in \mathbf{R}^m$ we denote the solution set by

$$\phi^{-1}(\mathbf{c}) = \{\, \mathbf{x} \in U \mid \phi(\mathbf{x}) = \mathbf{c} \,\}$$

and call it the *level set* or the *fibre* of ϕ at $\mathbf{c}$. (The notation $\phi^{-1}(\mathbf{c})$ for the solution set is standard, but a bit unfortunate because it suggests falsely that ϕ is invertible, which it is usually not.) If ϕ is a linear map, the system of equations is inhomogeneous linear and by linear algebra the solution set is an affine subspace of $\mathbf{R}^N$. The dimension of this affine subspace is $N - m$, provided that ϕ has rank m (i.e. has m independent columns). We can generalize this idea to nonlinear equations as follows. We say that $\mathbf{c} \in \mathbf{R}^m$ is a *regular value* of ϕ if the Jacobi matrix $D\phi(\mathbf{x})\colon \mathbf{R}^N \to \mathbf{R}^m$ has rank m for all $\mathbf{x} \in \phi^{-1}(\mathbf{c})$. A vector that is not a regular value is called a *singular* value. (As an extreme, though slightly silly, special case, if $\phi^{-1}(\mathbf{c})$ is empty, then $\mathbf{c}$ is automatically a regular value.)

The following result is the most useful criterion for a set to be a manifold. (Don't get carried away though, because it does not apply to every possible manifold. In other words, it is a sufficient but not a necessary criterion.) The proof uses the following important fact from linear algebra,

$$\boxed{\operatorname{nullity} A + \operatorname{rank} A = l,}$$

valid for any $k \times l$-matrix A. Here the rank is the number of independent columns of A (in other words the dimension of the column space $A(\mathbf{R}^l)$) and the nullity is the number of independent solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$ (in other words the dimension of the nullspace $\ker A$).

6.10. THEOREM (regular value theorem). *Let U be open in $\mathbf{R}^N$ and let $\phi\colon U \to \mathbf{R}^m$ be a smooth map. Suppose that $\mathbf{c}$ is a regular value of ϕ and that $M = \phi^{-1}(\mathbf{c})$ is nonempty. Then M is a manifold in $\mathbf{R}^N$ of codimension m. Its tangent space at $\mathbf{x}$ is the nullspace of $D\phi(\mathbf{x})$,*

$$\boxed{T_{\mathbf{x}}M = \ker D\phi(\mathbf{x}).}$$

PROOF. Let $\mathbf{x} \in M$. Then $D\phi(\mathbf{x})$ has rank m and so has m independent columns. After relabelling the coordinates on $\mathbf{R}^N$ we may assume the last m columns are independent and therefore constitute an invertible $m \times m$-submatrix A of $D\phi(\mathbf{x})$. Let us put $n = N - m$. Identify $\mathbf{R}^N$ with $\mathbf{R}^n \times \mathbf{R}^m$ and correspondingly write an N-vector as a pair $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u}$ a n-vector and $\mathbf{v}$ an m-vector. Also write $\mathbf{x} = (\mathbf{u}_0, \mathbf{v}_0)$. Now refer to Appendix B.4 and observe that the submatrix A is nothing but the "partial" Jacobian $D_{\mathbf{v}}\phi(\mathbf{u}_0, \mathbf{v}_0)$. This matrix being invertible, by the implicit function theorem, Theorem B.4, there exist open neighbourhoods U of $\mathbf{u}_0$ in $\mathbf{R}^n$ and V of $\mathbf{v}_0$ in $\mathbf{R}^m$ such that for each $\mathbf{u} \in U$ there exists a unique $\mathbf{v} = f(\mathbf{u}) \in V$ satisfying $\phi(\mathbf{u}, f(\mathbf{u})) = \mathbf{c}$. The map $f\colon U \to V$ is C^∞. In other words $M \cap (U \times V) = \operatorname{graph} f$ is the graph of a smooth map. We conclude from Example 6.7 that $M \cap (U \times V)$ is an n-manifold, namely the image of the embedding $\psi\colon U \to \mathbf{R}^N$ given by $\psi(\mathbf{u}) = (\mathbf{u}, f(\mathbf{u}))$. Since $U \times V$ is open in $\mathbf{R}^N$ and the above argument is valid for every $\mathbf{x} \in M$, we see that M is an n-manifold. To compute $T_{\mathbf{x}}M$ note that $\phi(\psi(\mathbf{u})) = \mathbf{c}$, a constant, for all $\mathbf{u} \in U$. Hence $D\phi(\psi(\mathbf{u}))D\psi(\mathbf{u}) = 0$ by the chain rule. Plugging in $\mathbf{u} = \mathbf{u}_0$ gives

$$D\phi(\mathbf{x})D\psi(\mathbf{u}_0) = 0.$$

The tangent space $T_{\mathbf{x}}M$ is by definition the column space of $D\psi(\mathbf{u}_0)$, so every tangent vector $\mathbf{v}$ to M at $\mathbf{x}$ is of the form $\mathbf{v} = D\psi(\mathbf{u}_0)\mathbf{a}$ for some $\mathbf{a} \in \mathbf{R}^n$. Therefore $D\phi(\mathbf{x})\mathbf{v} = D\phi(\mathbf{x})D\psi(\mathbf{u}_0)\mathbf{a} = \mathbf{0}$, i.e. $T_{\mathbf{x}}M \subseteq \ker D\phi(\mathbf{x})$. The tangent space $T_{\mathbf{x}}M$

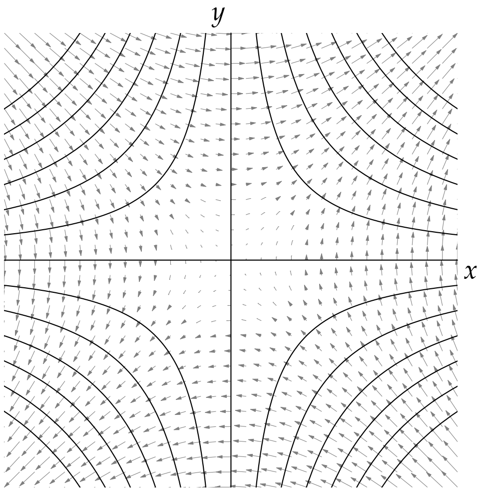
is n-dimensional (because the n columns of $D\psi(\mathbf{u}_0)$ are independent) and so is the nullspace of $D\phi(\mathbf{x})$ (because nullity $D\phi(\mathbf{x}) = N - m = n$). Hence $T_{\mathbf{x}}M = \ker D\phi(\mathbf{x})$. QED

The case of one single equation ($m = 1$) is especially important. Then $D\phi$ is a single row vector and its transpose is the gradient of ϕ: $D\phi^T = \operatorname{grad}\phi$. It has rank 1 at $\mathbf{x}$ if and only if it is nonzero, i.e. at least one of the partials of ϕ does not vanish at $\mathbf{x}$. The solution set of a scalar equation $\phi(\mathbf{x}) = c$ is known as a *level hypersurface*. Level hypersurfaces, especially level curves, occur frequently in all kinds of applications. For example, isotherms in weathercharts and contour lines in topographical maps are types of level curves.

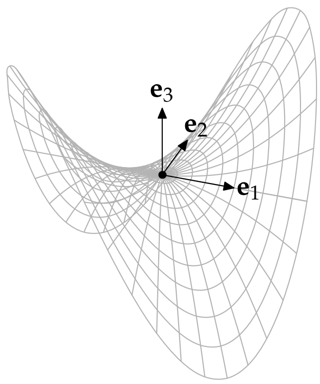
6.11. COROLLARY (level hypersurfaces). *Let U be open in $\mathbf{R}^N$ and let $\phi\colon U \to \mathbf{R}$ be a smooth function. Suppose that $M = \phi^{-1}(c)$ is nonempty and that* $\operatorname{grad}\phi(\mathbf{x}) \neq \mathbf{0}$ *for all $\mathbf{x}$ in M. Then M is a manifold in $\mathbf{R}^N$ of codimension 1. Its tangent space at $\mathbf{x}$ is the orthogonal complement of* $\operatorname{grad}\phi(\mathbf{x})$,

$$\boxed{T_{\mathbf{x}}M = (\operatorname{grad}\phi(\mathbf{x}))^{\perp}.}$$

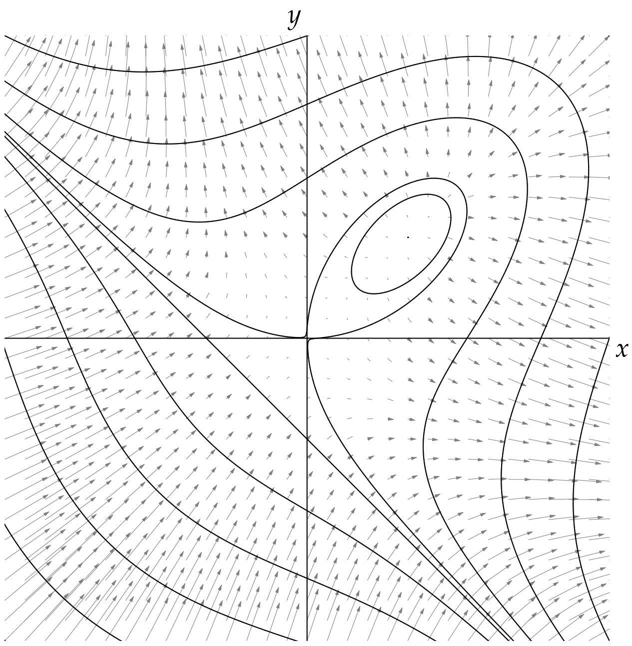
6.12. EXAMPLE. Let $U = \mathbf{R}^2$ and $\phi(x, y) = xy$. The level curves of ϕ are hyperbolas in the plane and the gradient is $\operatorname{grad}\phi(\mathbf{x}) = (y, x)^T$. The diagram below shows a few level curves as well as the gradient vector field, which as you can see is perpendicular to the level curves.



The gradient vanishes only at the origin, so $\phi(\mathbf{0}) = 0$ is the only singular value of ϕ. By Corollary 6.11 this means that $\phi^{-1}(c)$ is a 1-manifold for $c \neq 0$. The fibre $\phi^{-1}(0)$ is the union of the two coordinate axes, which has a self-intersection and so is not a manifold. However, the set $\phi^{-1}(0) - \{\mathbf{0}\}$ is a 1-manifold since the gradient is nonzero outside the origin. Think of this diagram as a topographical map representing the surface $z = \phi(x, y)$ shown below. The level curves of ϕ are the contour lines of the surface, obtained by intersecting the surface with horizontal planes at different heights. As explained in Appendix B.2, the gradient points in the direction of steepest ascent. Where the contour lines self-intersect the surface has a "mountain pass" or saddle point.



6.13. EXAMPLE. A more interesting example of an equation in two variables is $\phi(x, y) = x^3 + y^3 - 3xy = c$. Here $\operatorname{grad}\phi(\mathbf{x}) = 3(x^2 - y, y^2 - x)^T$, so $\operatorname{grad}\phi$ vanishes at the origin and at $(1, 1)^T$. The corresponding values of ϕ are 0, resp. -1, which are the singular values of ϕ.



The level "curve" $\phi^{-1}(-1)$ is not a curve at all, but consists of the single point $(1, 1)^T$. Here ϕ has a minimum and the surface $z = \phi(x, y)$ has a "valley". The level curve $\phi^{-1}(0)$ has a self-intersection at the origin, which corresponds to a saddle point on the surface. These features are also clearly visible in the surface itself, which is shown in Example 6.7.

6.14. EXAMPLE. Let $U = \mathbf{R}^N$ and $\phi(\mathbf{x}) = \|x\|^2$. Then $\operatorname{grad}\phi(\mathbf{x}) = 2\mathbf{x}$, so as in Example 6.12 $\operatorname{grad}\phi$ vanishes only at the origin $\mathbf{0}$, which is contained in $\phi^{-1}(0)$. So again any $c \neq 0$ is a regular value of ϕ. Clearly, $\phi^{-1}(c)$ is empty for $c < 0$. For $c > 0$, $\phi^{-1}(c)$ is an $N - 1$-manifold, the sphere of radius $\sqrt{c}$ in $\mathbf{R}^N$. The tangent space to the sphere at $\mathbf{x}$ is the set of all vectors perpendicular to grad $\phi(\mathbf{x}) = 2\mathbf{x}$. In other words,

$$T_{\mathbf{x}}M = \mathbf{x}^{\perp} = \{\, \mathbf{y} \in \mathbf{R}^N \mid \mathbf{y} \cdot \mathbf{x} = 0 \,\}.$$

Finally, 0 is a singular value (the absolute minimum) of ϕ and $\phi^{-1}(0) = \{\mathbf{0}\}$ is not an $N-1$-manifold. (It happens to be a 0-manifold, though, just like the singular fibre $\phi^{-1}(-1)$ in Example 6.13. So if $\mathbf{c}$ is a singular value, you cannot be certain that $\phi^{-1}(\mathbf{c})$ is *not* a manifold. However, even if a singular fibre happens to be a manifold, it is often of the "wrong" dimension.)

Here is an example of a manifold given by two equations ($m = 2$).

6.15. EXAMPLE. Define $\phi\colon \mathbf{R}^4 \to \mathbf{R}^2$ by

$$\phi(\mathbf{x}) = \begin{pmatrix} x_1^2 + x_2^2 \\ x_1x_3 + x_2x_4 \end{pmatrix}.$$

Then

$$D\phi(\mathbf{x}) = \begin{pmatrix} 2x_1 & 2x_2 & 0 & 0 \\ x_3 & x_4 & x_1 & x_2 \end{pmatrix}.$$

If $x_1 \neq 0$ the first and third columns of $D\phi(\mathbf{x})$ are independent, and if $x_2 \neq 0$ the second and fourth columns are independent. On the other hand, if $x_1 = x_2 = 0$, $D\phi(\mathbf{x})$ has rank 1 and $\phi(\mathbf{x}) = \mathbf{0}$. This shows that the origin $\mathbf{0}$ in $\mathbf{R}^2$ is the only singular value of ϕ. Therefore, by the regular value theorem, for every nonzero vector $\mathbf{c}$ the set $\phi^{-1}(\mathbf{c})$ is a two-manifold in $\mathbf{R}^4$. For instance, $M = \phi^{-1}\binom{1}{0}$ is a two-manifold. Note that M contains the point $\mathbf{x} = (1, 0, 0, 0)^T$. Let us find a basis of the tangent space $T_{\mathbf{x}}M$. Again by the regular value theorem, this tangent space is equal to the nullspace of

$$D\phi(\mathbf{x}) = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

which is equal the set of all vectors $\mathbf{y}$ satisfying $y_1 = y_3 = 0$. A basis of $T_{\mathbf{x}}M$ is therefore given by the standard basis vectors $\mathbf{e}_2$ and $\mathbf{e}_4$.

We now come to a more sophisticated example of a manifold determined by a large system of equations.

6.16. EXAMPLE. Recall that an $n \times n$-matrix A is *orthogonal* if $A^TA = I$. This means that the columns (and also the rows) of A are perpendicular to one another and have length 1. (In other words, they form an ortho*normal* basis of $\mathbf{R}^n$—note the regrettable inconsistency in the terminology.) The collection of orthogonal matrices form a group under matrix multiplication, which is usually called the *orthogonal group* and denoted by $\mathbf{O}(n)$. Let us prove using the regular value theorem that $\mathbf{O}(n)$ is a manifold. First observe that that $(A^TA)^T = A^TA$, so A^TA is a symmetric matrix. In other words, if $V = \mathbf{R}^{n\times n}$ is the vector space of all $n \times n$-matrices and $W = \{\, C \in V \mid C = C^T \,\}$ the linear subspace of all symmetric matrices, then

$$\phi(A) = A^TA$$

defines a map $\phi\colon V \to W$. Clearly $\mathbf{O}(n) = \phi^{-1}(I_n)$, so to prove that $\mathbf{O}(n)$ is a manifold it suffices to show that I is a regular value of ϕ. The derivative of ϕ can be computed by using the formula derived in Exercise B.3:

$$\begin{aligned} D\phi(A)B &= \lim_{h\to 0}\frac{1}{h}(\phi(A+hB)-\phi(A)) \\ &= \lim_{h\to 0}\frac{1}{h}(A^TA + hA^TB + hB^TA + h^2B^TB - A^TA) \\ &= BA^T + AB^T. \end{aligned}$$

We need to show that for $A \in \mathbf{O}(n)$ the linear map $D\phi(A)\colon V \to W$ has rank equal to the dimension of W. By linear algebra this amounts to showing that the equation

$$BA^T + AB^T = C \tag{6.1}$$

is solvable for B, given any orthogonal A and any symmetric C. Here is a way of guessing a solution: observe that $C = \frac{1}{2}(C + C^T)$ and first try to solve $BA^T = \frac{1}{2}C$. Left multiplying both sides by A and using $A^TA = I$ gives $B = \frac{1}{2}CA$. It is now easy to check that $B = \frac{1}{2}CA$ is a solution of equation (6.1).

Exercises

6.1. This is a continuation of Exercise 1.1. Define $\psi\colon \mathbf{R} \to \mathbf{R}^2$ by $\psi(t) = (t - \sin t, 1 - \cos t)^T$. Show that ψ is one-to-one. Determine all t for which $\psi'(t) = \mathbf{0}$. Prove that $\psi(\mathbf{R})$ is not a manifold at these points.

6.2. Let $a \in (0,1)$ be a constant. Prove that the map $\psi\colon \mathbf{R} \to \mathbf{R}^2$ given by $\psi(t) = (t - a\sin t, 1 - a\cos t)^T$ is an embedding. (This becomes easier if you first show that $t - a\sin t$ is an increasing function of t.) Graph the curve defined by ψ.

6.3. Prove that the map $\psi\colon \mathbf{R} \to \mathbf{R}^2$ given by $\psi(t) = \frac{1}{2}(e^t + e^{-t}, e^t - e^{-t})^T$ is an embedding. Conclude that $M = \psi(\mathbf{R})$ is a 1-manifold. Graph the curve M. Compute the tangent line to M at $(1,0)$ and try to find an equation for M.

6.4. Let I be the open interval $(-1,\infty)$ and let $\psi\colon I \to \mathbf{R}^2$ be the map $\psi(t) = (3at/(1+t^3), 3at^2/(1+t^3))^T$, where a is a nonzero constant. Show that ψ is one-to-one and that $\psi'(t) \neq \mathbf{0}$ for all $t \in I$. Is ψ an embedding and is $\psi(I)$ a manifold? (Observe that $\psi(I)$ is a portion of the curve studied in Exercise 1.2.)

6.5. Define $\psi\colon \mathbf{R} \to \mathbf{R}^2$ by

$$\psi(t) = \begin{cases} \big(-f(t), f(t)\big)^T & \text{if } t \leq 0, \\ \big(f(t), f(t)\big)^T & \text{if } t \geq 0, \end{cases}$$

where f is the function given in Exercise B.6. Show that ψ is smooth, one-to-one and that its inverse $\psi^{-1}\colon \psi(\mathbf{R}) \to \mathbf{R}$ is continuous. Sketch the image of ψ. Is $\psi(\mathbf{R})$ a manifold?

6.6. Define a map $\psi\colon \mathbf{R}^2 \to \mathbf{R}^4$ by

$$\psi\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} t_1^3 \\ t_1^2 t_2 \\ t_1 t_2^2 \\ t_2^3 \end{pmatrix}.$$

(i) Show that ψ is one-to-one.
(ii) Show that $D\psi(\mathbf{t})$ is one-to-one for all $\mathbf{t} \neq \mathbf{0}$.
(iii) Let U be the punctured plane $\mathbf{R}^2 - \{\mathbf{0}\}$. Show that $\psi\colon U \to \mathbf{R}^4$ is an embedding. Conclude that $\psi(U)$ is a two-manifold in $\mathbf{R}^4$.

(iv) Find a basis of the tangent plane to $\psi(U)$ at the point $\psi(1,1)$.

6.7. Let $\phi\colon \mathbf{R}^n - \{\mathbf{0}\} \to \mathbf{R}$ be a homogeneous function of degree p as defined in Exercise B.5. Assume that ϕ is smooth and that $p \neq 0$. Show that 0 is the only possible singular value of ϕ. (Use the result of Exercise B.5.) Conclude that, if nonempty, $\phi^{-1}(c)$ is an $n-1$-manifold for $c \neq 0$.

6.8. Let $\phi(\mathbf{x}) = a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2$, where the a_i are nonzero constants. Determine the regular and singular values of ϕ. For $n = 3$ sketch the level surface $\phi^{-1}(c)$ for a regular value c. (You have to distinguish between a few different cases.)

6.9. Show that the trajectories of the Lotka-Volterra system of Exercise 1.10 are one-dimensional manifolds.

6.10. Compute the dimension of the orthogonal group $\mathbf{O}(n)$ and show that its tangent space at the identity matrix I is the set of all antisymmetric $n \times n$-matrices.

6.11. Let V be the vector space of $n \times n$-matrices and define $\phi\colon V \to \mathbf{R}$ by $\phi(A) = \det A$.

(i) Show that

$$D\phi(A)B = \sum_{i=1}^{n} \det(\mathbf{a}_1, \mathbf{a}_2 \dots, \mathbf{a}_{i-1}, \mathbf{b}_i, \mathbf{a}_{i+1}, \dots, \mathbf{a}_n),$$

where $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ and $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ denote the column vectors of A, resp. B. (Apply the formula of Exercise B.3 for the derivative and use the multilinearity of the determinant.)

(ii) The *special linear group* is the subset of V defined by

$$\mathbf{SL}(n) = \{\, A \in V \mid \det A = 1 \,\}.$$

Show that $\mathbf{SL}(n)$ is a manifold. What is its dimension?

(iii) Show that for $A = I$, the identity matrix, we have $D\phi(A)B = \sum_{i=1}^{n} b_{i,i} = \operatorname{tr} B$, the *trace* of B. Conclude that the tangent space to $\mathbf{SL}(n)$ at I is the set of *traceless* matrices, i.e. matrices A satisfying $\operatorname{tr} A = 0$.

6.12. (i) Let W be punctured 4-space $\mathbf{R}^4 - \{\mathbf{0}\}$ and define $\phi\colon W \to \mathbf{R}$ by

$$\phi(\mathbf{x}) = x_1x_4 - x_2x_3.$$

Show that 0 is a regular value of ϕ.

(ii) Let A be a real 2×2-matrix. Show that $\operatorname{rank} A = 1$ if and only if $\det A = 0$ and $A \neq 0$.

(iii) Let M be the set of 2×2-matrices of rank 1. Show that M is a three-dimensional manifold.

(iv) Compute T_AM, where $A = \left(\begin{smallmatrix}1 & 1\\ 0 & 0\end{smallmatrix}\right)$.

6.13. Define $\phi\colon \mathbf{R}^4 \to \mathbf{R}^2$ by

$$\phi(\mathbf{x}) = \begin{pmatrix} x_1 + x_2 + x_3x_4 \\ x_1x_2x_3 + x_4 \end{pmatrix}.$$

(i) Show that $D\phi(\mathbf{x})$ has rank 2 unless $\mathbf{x}$ is of the form $(t^{-2}, t^{-2}, t, t^{-3})$ for some $t \neq 0$. (Compute all 2×2-subdeterminants of $D\phi$ and set them equal to 0.)

(ii) Show that $M = \phi^{-1}(\mathbf{0})$ is a 2-manifold (where $\mathbf{0}$ is the origin in $\mathbf{R}^2$).

(iii) Find a basis of the tangent space $T_{\mathbf{x}}M$ for all $\mathbf{x} \in M$ with $x_3 = 0$. (The answer depends on $\mathbf{x}$.)

6.14. Let U be an open subset of $\mathbf{R}^n$ and let $\phi\colon U \to \mathbf{R}^m$ be a smooth map. Let M be the manifold $\phi^{-1}(\mathbf{c})$, where $\mathbf{c}$ is a regular value of ϕ. Let $f\colon U \to \mathbf{R}$ be a smooth function. A point $\mathbf{x} \in M$ is called a *critical point* for the *restricted function* $f|M$ if $Df(\mathbf{x})\mathbf{v} = 0$ for all tangent vectors $\mathbf{v} \in T_{\mathbf{x}}M$. Prove that $\mathbf{x} \in M$ is critical for $f|M$ if and only if there exist numbers $\lambda_1, \lambda_2, \dots, \lambda_m$ such that

$$\operatorname{grad} f(\mathbf{x}) = \lambda_1 \operatorname{grad} \phi_1(\mathbf{x}) + \lambda_2 \operatorname{grad} \phi_2(\mathbf{x}) + \cdots + \lambda_m \operatorname{grad} \phi_m(\mathbf{x}).$$

(Use the characterization of $T_{\mathbf{x}}M$ given by the regular value theorem.)

6.15. Find the critical points of the function $f(x, y, z) = -x + 2y + 3z$ over the circle C given by

$$x^2 + y^2 + z^2 = 1,$$
$$x + z = 0.$$

Where are the maxima and minima of $f|C$?

6.16 (eigenvectors via calculus). Let $A = A^T$ be a symmetric $n \times n$-matrix and define $f\colon \mathbf{R}^n \to \mathbf{R}$ by $f(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x}$. Let M be the unit sphere $\{\, \mathbf{x} \in \mathbf{R}^n \mid \|\mathbf{x}\| = 1 \,\}$.

(i) Calculate $\operatorname{grad} f(\mathbf{x})$.
(ii) Show that $\mathbf{x} \in M$ is a critical point of $f|M$ if and only if $\mathbf{x}$ is an eigenvector for A of length 1.
(iii) Given an eigenvector $\mathbf{x}$ of length 1, show that $f(\mathbf{x})$ is the corresponding eigenvalue of $\mathbf{x}$.

CHAPTER 7

Differential forms on manifolds

7.1. First definition

There are several different ways to define differential forms on manifolds. In this section we present a practical, workaday definition. A more theoretical approach is taken in Section 7.2.

Let M be an n-manifold in $\mathbf{R}^N$ and let us first consider what we might mean by a 0-form or smooth function on M. A function $f\colon M \to \mathbf{R}$ is simply an assignment of a unique number $f(\mathbf{x})$ to each point $\mathbf{x}$ in M. For instance, M could be the surface of the earth and f could represent temperature at a given time, or height above sea level. But how would we define such a function to be differentiable? The difficulty here is that if $\mathbf{x}$ is in M and $\mathbf{e}_j$ is one of the standard basis vectors, the straight line $\mathbf{x} + h\mathbf{e}_j$ may not be contained in M, so we cannot form the limit $\partial f/\partial x_j = \lim_{h\to 0}(f(\mathbf{x}+h\mathbf{e}_j) - f(\mathbf{x}))/h$.

Here is one way out of this difficulty. Because M is a manifold there exist open sets U_i in $\mathbf{R}^n$ and embeddings $\psi_i\colon U_i \to \mathbf{R}^N$ such that the images $\psi_i(U_i)$ cover M: $M = \bigcup_i \psi_i(U_i)$. (Here i ranges over some unspecified, possibly infinite, index set.) For each i we define a function $f_i\colon U_i \to \mathbf{R}$ by $f_i(\mathbf{t}) = f(\psi_i(\mathbf{t}))$, i.e. $f_i = \psi_i^* f$. We call f_i the *local representative* of f relative to the embedding ψ_i. (For instance, if M is the earth's surface, f is temperature, and ψ_i is a map of New York State, then f_i represents a temperature chart of NY.) Since f_i is defined on the open subset U_i of $\mathbf{R}^n$, it makes sense to ask whether its partial derivatives exist. We say that f is C^k if each of the local representatives f_i is C^k. Now suppose that $\mathbf{x}$ is in the overlap of two charts. Then we have two indices i and j and vectors $\mathbf{t} \in U_i$ and $\mathbf{u} \in U_j$ such that $\mathbf{x} = \psi_i(\mathbf{t}) = \psi_j(\mathbf{u})$. Then we must have $f(\mathbf{x}) = f(\psi_i(\mathbf{t})) = f(\psi_j(\mathbf{u}))$, so $f_i(\mathbf{t}) = f_j(\mathbf{u})$. Also $\psi_i(\mathbf{t}) = \psi_j(\mathbf{u})$ implies $\mathbf{t} = \psi_i^{-1} \circ \psi_j(\mathbf{u})$ and therefore $f_j(\mathbf{u}) = f_i\big(\psi_i^{-1} \circ \psi_j(\mathbf{u})\big)$. This identity must hold for all $\mathbf{u} \in U_j$ such that $\psi_j(\mathbf{u}) \in \psi_i(U_i)$, i.e. for all $\mathbf{u}$ in $\psi_j^{-1}(\psi_i(U_i))$. We can abbreviate this by saying that

$$f_j = (\psi_i^{-1} \circ \psi_j)^* f_i$$

on $\psi_j^{-1}(\psi_i(U_i))$. This is a consistency condition on the functions f_i imposed by the fact that they are pullbacks of a single function f defined everywhere on M. The map $\psi_i^{-1} \circ \psi_j$ is often called a *change of coordinates* and the consistency condition is also known as the *transformation law* for the local representatives f_i. (Pursuing the weather chart analogy, it expresses nothing but the obvious fact that where the maps of New York and Pennsylvania overlap, the corresponding two temperature charts must show the same temperatures.) Conversely, the collection of all local representatives f_i determines f, because we have $f(\mathbf{x}) = f_i(\psi_i^{-1}(\mathbf{x}))$ if $\mathbf{x} \in \psi_i(U_i)$.

(That is to say, if we have a complete set of weather charts for the whole world, we know the temperature everywhere.)

Following this cue we formulate the following definition.

7.1. DEFINITION. A *differential form of degree* k, or simply a k*-form*, α on M is a collection of k-forms α_i on U_i satisfying the transformation law

$$\boxed{\alpha_j = (\psi_i^{-1} \circ \psi_j)^* \alpha_i} \tag{7.1}$$

on $\psi_j^{-1}(\psi_i(U_i))$. We call α_i the *local representative* of α relative to the embedding ψ_i and denote it by $\alpha_i = \psi_i^* \alpha$. The collection of all k-forms on M is denoted by $\Omega^k(M)$.

This definition is rather indirect, but it works really well if a specific atlas for the manifold M is known. Definition 7.1 is particularly tractible if M is the image of a single embedding $\psi\colon U \to \mathbf{R}^N$. In that case the compatibility relation (7.1) is vacuous and a k-form α on M is determined by one single representative, a k-form $\psi^* \alpha$ on U.

Sometimes it is useful to write the transformation law (7.1) in components. We can do this by appealing to Theorem 3.12. If

$$\alpha_i = \sum_I f_I \, dt_I \qquad \text{and} \qquad \alpha_j = \sum_J g_J \, dt_J$$

are two local representatives for α, then

$$g_J = \sum_I (\psi_i^{-1} \circ \psi_j)^* f_I \, \det D(\psi_i^{-1} \circ \psi_j)_{I,J}.$$

on $\psi_j^{-1}(\psi_i(U_i))$.

Just like forms on $\mathbf{R}^n$, forms on a manifold can be added, multiplied, differentiated and integrated. For example, suppose α is a k-form and β an l-form on M. Suppose α_i, resp. β_i, is the local representative of α, resp. β, relative to an embedding $\psi_i\colon U_i \to M$. Then we define the product $\gamma = \alpha\beta$ by setting $\gamma_i = \alpha_i \beta_i$. To see that this definition makes sense, we check that the forms γ_i satisfy the transformation law (7.1):

$$\gamma_j = \alpha_j \beta_j = (\psi_i^{-1} \circ \psi_j)^* \alpha_i (\psi_i^{-1} \circ \psi_j)^* \beta_i = (\psi_i^{-1} \circ \psi_j)^* (\alpha_i \beta_i) = (\psi_i^{-1} \circ \psi_j)^* \gamma_i.$$

Here we have used the multiplicative property of pullbacks, Proposition 3.10(ii). Similarly, the exterior derivative of α is defined by setting $(d\alpha)_i = d\alpha_i$. As before, let us check that the forms $(d\alpha)_i$ satisfy the transformation law (7.1):

$$(d\alpha)_j = d\alpha_j = d(\psi_i^{-1} \circ \psi_j)^* \alpha_i = (\psi_i^{-1} \circ \psi_j)^* d\alpha_i = (\psi_i^{-1} \circ \psi_j)^* (d\alpha)_i,$$

where we used Theorem 3.11.

7.2. Second definition

This section presents some of the algebraic underpinnings of the theory of differential forms. This branch of algebra, now called *exterior* or *alternating algebra* was invented by Graßmann in the mid-nineteenth century and is a prerequisite for much of the more advanced literature on the subject.

Covectors. Before giving a rigorous definition of differential forms on manifolds we need to be more precise about the definition of a differential form on $\mathbf{R}^n$. Recall that $\mathbf{R}^n$ is the collection of all column vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Let U be an open subset of $\mathbf{R}^n$. The definition of a 0-form on U requires no further clarification: it is simply a function on U. Formally, a 1-form α on U can be defined as a *row vector*

$$\alpha = (f_1, f_2, \dots, f_n)$$

whose entries are functions on U. The form is called *constant* if the entries $f_1, \dots,$ f_n are constant. The set of constant row vectors is denoted by $(\mathbf{R}^n)^*$ and is called the *dual* of $\mathbf{R}^n$. Constant 1-forms are also known as *covariant vectors* or *covectors* and arbitrary 1-forms as *covariant vector fields* or *covector fields*. By definition dx_i is the constant 1-form

$$dx_i = \mathbf{e}_i^T = (0, \dots, 0, 1, 0, \dots, 0),$$

the transpose of $\mathbf{e}_i$, the i-th standard basis vector of $\mathbf{R}^n$. Every 1-form can thus be written as

$$\alpha = (f_1, f_2, \dots, f_n) = \sum_{i=1}^n f_i \, dx_i.$$

Using this formalism we can write for any smooth function g on U

$$dg = \sum_{i=1}^n \frac{\partial g}{\partial x_i} \, dx_i = \Big(\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}, \dots, \frac{\partial g}{\partial x_n} \Big),$$

so dg is simply the Jacobi matrix Dg of g! (This is the reason that many authors use the notation dg for the Jacobi matrix.)

We would like to extend the notions of covectors and 1-forms to vector spaces other than $\mathbf{R}^n$. To see how, let us start by observing that a row vector $\mathbf{y}$ is nothing but a $1 \times n$-matrix. We can multiply it by a column vector $\mathbf{x}$ to obtain a number,

$$\mathbf{y}\mathbf{x} = (y_1, y_2, \dots, y_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^n y_i x_i.$$

Obviously we have $\mathbf{y}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\mathbf{y}\mathbf{x}_1 + c_2\mathbf{y}\mathbf{x}_2$. Thus a row vector can be viewed as a linear map which sends column vectors in $\mathbf{R}^n$ to one-dimensional vectors (scalars) in $\mathbf{R}^1 = \mathbf{R}$.

This motivates the following definition. If V is any vector space over the real numbers (for example $\mathbf{R}^n$ or a subspace of $\mathbf{R}^n$), then V^*, the *dual* of V, is the set of linear maps from V to $\mathbf{R}$. Elements of V^* are called *dual vectors* or *covectors* or *linear functionals*. The dual is a vector space in its own right: if μ_1 and μ_2 are in V^* we define $\mu_1 + \mu_2$ and $c\mu_1$ by setting $(\mu_1 + \mu_2)(v) = \mu_1(v) + \mu_2(v)$ for all $v \in V$ and $(c\mu_1)(v) = c\mu_1(v)$.

7.2. EXAMPLE. Let $V = C^0([a,b], \mathbf{R})$, the collection of all continuous real-valued functions on a closed and bounded interval $[a,b]$. A linear combination of continuous functions is continuous, so V is a vector space. Define $\mu(f) = \int_a^b f(x)\,dx$. Then $\mu(c_1 f_1 + c_2 f_2) = c_1\mu(f_1) + c_2\mu(f_2)$, so μ is a linear functional on V.

7.3. EXAMPLE. Let $V = \mathbf{R}^n$ and fix $\mathbf{v} \in V$. Define $\mu(\mathbf{x}) = \mathbf{v} \cdot \mathbf{x}$, where "$\cdot$" is the standard inner product on $\mathbf{R}^n$. Then μ is a linear functional on V.

Now suppose that V is a vector space of finite dimension n and choose a basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ of V. Then every vector $\mathbf{v} \in V$ can be written in a unique way as a linear combination $\sum_j c_j \mathbf{v}_j$. Define a covector $\lambda_i \in V^*$ by $\lambda_i(\mathbf{v}) = c_i$. In other words, λ_i is determined by the rule

$$\lambda_i(\mathbf{v}_j) = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

We call λ_i the *i-th coordinate function*.

7.4. LEMMA. *The coordinate functions $\lambda_1, \lambda_2, \dots, \lambda_n$ constitute a basis of V^*. Hence* $\dim V^* = n = \dim V$.

PROOF. Let $\lambda \in V^*$. We need to write λ as a linear combination $\lambda = \sum_{i=1}^n c_i \lambda_i$. Assuming for the moment that this is possible, we can apply both sides to the vector $\mathbf{v}_j$ to obtain

$$\lambda(\mathbf{v}_j) = \sum_{i=1}^n c_i \lambda_i(\mathbf{v}_j) = \sum_{i=1}^n c_i \delta_{i,j} = c_j. \tag{7.2}$$

So $c_j = \lambda(\mathbf{v}_j)$ is the only possible choice for the coefficient c_j. To show that this choice of coefficients works, let us define $\lambda' = \sum_{i=1}^n \lambda(\mathbf{v}_i)\lambda_i$. Then by equation (7.2), $\lambda'(\mathbf{v}_j) = \lambda(\mathbf{v}_j)$ for all j, so $\lambda' = \lambda$, i.e. $\lambda = \sum_{i=1}^n \lambda(\mathbf{v}_i)\lambda_i$. We have proved that every $\lambda \in V^*$ can be written uniquely as a linear combination of the λ_i. QED

The basis $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ of V^* is said to be *dual* to the basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of V.

7.5. EXAMPLE. Consider $\mathbf{R}^n$ with standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Then $dx_i(\mathbf{e}_j) = \mathbf{e}_i^T \mathbf{e}_j = \delta_{i,j}$, so the dual basis of $(\mathbf{R}^n)^*$ is $\{dx_1, dx_2, \dots, dx_n\}$.

Dual bases come in handy when writing the matrix of a linear map. Let $L\colon V \to W$ be a linear map between abstract vector spaces V and W. To write the matrix of L we need to start by picking a basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ of V and a basis $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ of W. Then for each $j = 1, 2, \dots, n$ the vector $L\mathbf{v}_j$ can be expanded uniquely in terms of the $\mathbf{w}$'s: $L\mathbf{v}_j = \sum_{i=1}^m l_{i,j}\mathbf{w}_i$. The $m \times n$ numbers $l_{i,j}$ make up the matrix of L relative to the two bases of V and W.

7.6. LEMMA. *Let $\lambda_1, \lambda_2, \dots, \lambda_n \in V^*$ be the dual basis of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ and $\mu_1, \mu_2, \dots, \mu_n \in W^*$ the dual basis of $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$. Then the i, j-th matrix element of a linear map $L\colon V \to W$ is equal to $l_{i,j} = \mu_i(L\mathbf{v}_j)$.*

PROOF. We have $L\mathbf{v}_j = \sum_{k=1}^m l_{k,j}\mathbf{w}_k$, so

$$\mu_i(L\mathbf{v}_j) = \sum_{k=1}^m l_{k,j}\mu_i(\mathbf{w}_k) = \sum_{k=1}^m l_{k,j}\delta_{ik} = l_{i,j},$$

that is $l_{i,j} = \mu_i(L\mathbf{v}_j)$. QED

Multilinear algebra. Let V be a vector space and let V^k denote the Cartesian product $V \times \cdots \times V$ (k times). Thus an element of V^k is an ordered k-tuple $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ of vectors in V. A *k-multilinear* function on V is a function $\lambda \colon V^k \to \mathbf{R}$ which is linear in each vector, i.e.

$$\lambda(\mathbf{v}_1, \mathbf{v}_2, \dots, c\mathbf{v}_i + c'\mathbf{v}_i', \dots, \mathbf{v}_k) \\ = c\lambda(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) + c'\lambda(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i', \dots, \mathbf{v}_k)$$

for all scalars c, c' and all vectors $\mathbf{v}_1$, $\mathbf{v}_2, \dots$, $\mathbf{v}_i$, $\mathbf{v}_i'$, $\dots$, $\mathbf{v}_k$.

7.7. EXAMPLE. Let $V = \mathbf{R}^n$ and let $\lambda(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$, the inner product of $\mathbf{x}$ and $\mathbf{y}$. Then λ is bilinear (i.e. 2-multilinear).

7.8. EXAMPLE. Let $V = \mathbf{R}^4$, $k = 2$. The function $\lambda(\mathbf{v}, \mathbf{w}) = v_1 w_2 - v_2 w_1 + v_3 w_4 - v_4 w_3$ is bilinear on $\mathbf{R}^4$.

7.9. EXAMPLE. Let $V = \mathbf{R}^n$, $k = n$. The determinant $\det(\mathbf{v}_1 \mathbf{v}_2, , \dots, \mathbf{v}_n)$ is an n-multilinear function on $\mathbf{R}^n$.

A k-multilinear function is *alternating* or *antisymmetric* if it has the alternating property,

$$\lambda(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_k) = -\lambda(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_k).$$

More generally, if λ is alternating, then for any permutation $\sigma \in S_k$ we have

$$\lambda(\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(k)}) = \operatorname{sign}(\sigma)\lambda(\mathbf{v}_1, \dots, \mathbf{v}_k).$$

7.10. EXAMPLE. The inner product of Example 7.7 is bilinear, but it is not alternating. Indeed it is *symmetric*: $\mathbf{y} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{y}$. The bilinear function of Example 7.8 is alternating, and so is the determinant function of Example 7.9.

Here is a useful trick to generate alternating k-multilinear functions starting from k covectors λ_1, $\lambda_2, \dots$, $\lambda_k \in V^*$. The *(wedge) product* is the function

$$\lambda_1 \lambda_2 \cdots \lambda_k \colon V^k \to \mathbf{R}$$

defined by

$$\boxed{\lambda_1 \lambda_2 \cdots \lambda_k(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \det\bigl(\lambda_i(\mathbf{v}_j)\bigr)_{1 \le i,j \le k}.}$$

(The determinant on the right is a $k \times k$-determinant.) It follows from the multilinearity and the alternating property of the determinant that $\lambda_1 \lambda_2 \cdots \lambda_k$ is an alternating k-multilinear function. The wedge product is often denoted by $\lambda_1 \wedge \lambda_2 \wedge \cdots \wedge \lambda_k$ to distinguish it from other products, such as the tensor product defined in Exercise 7.6.

The collection of all alternating k-multilinear functions is denoted by $A^k V$.

For $k = 1$ the alternating property is vacuous, so an alternating 1-multilinear function is nothing but a linear function. Thus $A^1 V = V^*$.

For $k = 0$ a k-multilinear function is defined to be a single number. Thus $A^0 V = \mathbf{R}$.

For any k, k-multilinear functions can be added and scalar-multiplied just like ordinary linear functions, so the set $A^k V$ forms a vector space.

There is a nice way to construct a basis of the vector space $A^k V$ starting from a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V. The idea is to take wedge products of dual basis vectors.

Let $\{\lambda_1, \dots, \lambda_n\}$ be the corresponding dual basis of V^*. Let $I = (i_1, i_2, \dots, i_k)$ be an increasing multi-index, i.e. $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. Write

$$\lambda_I = \lambda_{i_1} \lambda_{i_2} \cdots \lambda_{i_k} \in A^k V,$$
$$\mathbf{v}_I = (\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \dots, \mathbf{v}_{i_k}) \in V^k.$$

7.11. EXAMPLE. Let $V = \mathbf{R}^3$ with standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. The dual basis of $(\mathbf{R}^3)^*$ is $\{dx_1, dx_2, dx_3\}$. Let $k = 2$ and $I = (1,2)$, $J = (2,3)$. Then

$$dx_I(\mathbf{e}_I) = \begin{vmatrix} dx_1(\mathbf{e}_1) & dx_1(\mathbf{e}_2) \\ dx_2(\mathbf{e}_1) & dx_2(\mathbf{e}_2) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1,$$

$$dx_I(\mathbf{e}_J) = \begin{vmatrix} dx_1(\mathbf{e}_2) & dx_1(\mathbf{e}_3) \\ dx_2(\mathbf{e}_2) & dx_2(\mathbf{e}_3) \end{vmatrix} = \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix} = 0,$$

$$dx_J(\mathbf{e}_I) = \begin{vmatrix} dx_2(\mathbf{e}_1) & dx_2(\mathbf{e}_2) \\ dx_3(\mathbf{e}_1) & dx_3(\mathbf{e}_2) \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix} = 0,$$

$$dx_J(\mathbf{e}_J) = \begin{vmatrix} dx_2(\mathbf{e}_2) & dx_2(\mathbf{e}_3) \\ dx_3(\mathbf{e}_2) & dx_3(\mathbf{e}_3) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

This example generalizes as follows.

7.12. LEMMA. *Let I and J be increasing multi-indices of degree k. Then*

$$\lambda_I(\mathbf{v}_J) = \delta_{I,J} = \begin{cases} 1 & \text{if } I = J, \\ 0 & \text{if } I \neq J. \end{cases}$$

PROOF. Let $I = (i_1, \dots, i_k)$ and $J = (j_1, \dots, j_k)$. Then

$$\lambda_I(\mathbf{v}_J) = \det\big(\lambda_{i_r}(\mathbf{v}_{j_s})\big)_{1 \leq r,s \leq k} = \begin{vmatrix} \delta_{i_1,j_1} & \dots & \delta_{i_1,j_k} \\ \vdots & & \vdots \\ \delta_{i_l,j_1} & \dots & \delta_{i_l,j_k} \\ \vdots & & \vdots \\ \delta_{i_k,j_1} & \dots & \delta_{i_k,j_k} \end{vmatrix} = 1 \quad \text{if } I = J.$$

If $I \neq J$, then $i_l \neq j_l$ for some l. Choose l as small as possible, so that $i_m = j_m$ for $m < l$. There are two cases: $i_l < j_l$ and $i_l > j_l$. If $i_l < j_l$, then $i_l < j_l < j_{l+1} < \cdots < j_k$ because J is increasing, so all entries δ_{i_l,j_m} in the determinant with $m \geq l$ are 0. For $m < l$ we have $j_m = i_m < i_l$ because I is increasing, so $\delta_{i_l,j_m} = 0$ for $m < l$. In other words the l-th row in the determinant is 0 and hence $\lambda_I(\mathbf{v}_J) = 0$. If $i_l > j_l$ we find that the l-th column in the determinant is 0 and therefore again $\lambda_I(\mathbf{v}_J) = 0$. QED

We need one further technical result before showing that the functions λ_I are a basis of $A^k V$.

7.13. LEMMA. *Let $\lambda \in A^k V$. Suppose $\lambda(\mathbf{v}_I) = 0$ for all increasing multi-indices I of degree k. Then $\lambda = 0$.*

PROOF. The assumption implies

$$\lambda(\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k}) = 0 \tag{7.3}$$

for *all* multi-indices $(i_1, \dots, i_k)$, because of the alternating property. We need to show that $\lambda(\mathbf{w}_1, \dots, \mathbf{w}_k) = 0$ for arbitrary vectors $\mathbf{w}_1, \dots, \mathbf{w}_k$. We can expand the $\mathbf{w}_i$ using the basis:

$$\begin{aligned}\mathbf{w}_1 &= a_{11}\mathbf{v}_1 + \cdots + a_{1k}\mathbf{v}_k,\\ &\vdots\\ \mathbf{w}_k &= a_{k1}\mathbf{v}_1 + \cdots + a_{kk}\mathbf{v}_k.\end{aligned}$$

Therefore by multilinearity

$$\lambda(\mathbf{w}_1, \dots, \mathbf{w}_k) = \sum_{i_1=1}^{k} \cdots \sum_{i_k=1}^{k} a_{1i_1} \cdots a_{ki_k} \lambda(\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_k}).$$

Each term in the right-hand side is 0 by equation (7.3). QED

7.14. THEOREM. *Let V be an n-dimensional vector space with basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. Let $\{\lambda_1, \dots, \lambda_n\}$ be the corresponding dual basis of V^*. Then the alternating k-multilinear functions $\lambda_I = \lambda_{i_1} \cdots \lambda_{i_k}$, where I ranges over the set of all increasing multi-indices of degree k, form a basis of $A^k V$. Hence* $\dim A^k V = \binom{n}{k}$.

PROOF. The proof is closely analogous to that of Lemma 7.4. Let $\lambda \in A^k V$. We need to write λ as a linear combination $\lambda = \sum_I c_I \lambda_I$. Assuming for the moment that this is possible, we can apply both sides to the k-tuple of vectors $\mathbf{v}_J$. Using Lemma 7.12 we obtain

$$\lambda(\mathbf{v}_J) = \sum_I c_I \lambda_I(\mathbf{v}_J) = \sum_I c_I \delta_{I,J} = c_J.$$

So $c_J = \lambda(\mathbf{v}_J)$ is the only possible choice for the coefficient c_J. To show that this choice of coefficients works, let us define $\lambda' = \sum_I \lambda(\mathbf{v}_I)\lambda_I$. Then for all increasing multi-indices I we have $\lambda(\mathbf{v}_I) - \lambda'(\mathbf{v}_I) = \lambda(\mathbf{v}_I) - \lambda(\mathbf{v}_I) = 0$. Applying Lemma 7.13 to $\lambda - \lambda'$ we find $\lambda - \lambda' = 0$. In other words, $\lambda = \sum_I \lambda(\mathbf{v}_I)\lambda_I$. We have proved that every $\lambda \in V^*$ can be written uniquely as a linear combination of the λ_i. QED

7.15. EXAMPLE. Let $V = \mathbf{R}^n$ with standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. The dual basis of $(\mathbf{R}^n)^*$ is $\{dx_1, \dots, dx_n\}$. Therefore $A^k V$ has a basis consisting of all k-multilinear functions of the form

$$dx_I = dx_{i_1}\, dx_{i_2} \cdots dx_{i_k},$$

with $1 \le i_1 < \cdots < i_k \le n$. Hence a general alternating k-multilinear function λ on $\mathbf{R}^n$ looks like

$$\lambda = \sum_I a_I\, dx_I,$$

with a_I constant. By Lemma 7.12, $\lambda(\mathbf{e}_J) = \sum_I a_I\, dx_I(\mathbf{e}_J) = \sum_I a_I \delta_{I,J} = a_J$, so a_I is equal to $\lambda(\mathbf{e}_I)$.

An arbitary k-form α on a region U in $\mathbf{R}^n$ is now defined as a choice of an alternating k-multilinear function $\alpha_{\mathbf{x}}$ for each $\mathbf{x} \in U$; hence it looks like $\alpha_{\mathbf{x}} = \sum_I f_I(\mathbf{x})\, dx_I$, where the coefficients f_I are *functions* on U. We shall abbreviate this to

$$\alpha = \sum_I f_I\, dx_I,$$

and we shall always assume the coefficients f_I to be smooth functions. By Example 7.15 we can express the coefficients as $f_I = \alpha(\mathbf{e}_I)$ (which is to be interpreted as $f_I(\mathbf{x}) = \alpha_{\mathbf{x}}(\mathbf{e}_I)$ for all $\mathbf{x}$).

Pullbacks re-examined. In the light of this new definition we can give a fresh interpretation of a pullback. This will be useful in our study of forms on manifolds. Let U and V be open subsets of $\mathbf{R}^n$, resp. $\mathbf{R}^m$, and $\phi\colon U \to V$ a smooth map. For a k-form $\alpha \in \Omega^k(V)$ define the pullback $\phi^*\alpha \in \Omega^k(U)$ by

$$\boxed{(\phi^*\alpha)_{\mathbf{x}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \alpha_{\phi(\mathbf{x})}(D\phi(\mathbf{x})\mathbf{v}_1, D\phi(\mathbf{x})\mathbf{v}_2, \dots, D\phi(\mathbf{x})\mathbf{v}_k).}$$

Let us check that this formula agrees with the old definition. Suppose $\alpha = \sum_I f_I\, dy_I$ and $\phi^*\alpha = \sum_J g_J\, dx_J$. What is the relationship between g_J and f_I? We use $g_J = \phi^*\alpha(\mathbf{e}_J)$, our new definition of pullback and the definition of the wedge product to get

$$\begin{aligned} g_J(\mathbf{x}) = (\phi^*\alpha)_{\mathbf{x}}(\mathbf{e}_J) &= \alpha_{\phi(\mathbf{x})}(D\phi(\mathbf{x})\mathbf{e}_{j_1}, D\phi(\mathbf{x})\mathbf{e}_{j_2}, \dots, D\phi(\mathbf{x})\mathbf{e}_{j_k}) \\ &= \sum_I f_I(\phi(\mathbf{x}))\, dy_I(D\phi(\mathbf{x})\mathbf{e}_{j_1}, D\phi(\mathbf{x})\mathbf{e}_{j_2}, \dots, D\phi(\mathbf{x})\mathbf{e}_{j_k}) \\ &= \sum_I \phi^* f_I(\mathbf{x}) \det\big(dy_{i_r}(D\phi(\mathbf{x})\mathbf{e}_{j_s})\big)_{1\le r,s\le k}. \end{aligned}$$

By Lemma 7.6 the number $dy_{i_r}(D\phi(\mathbf{x})\mathbf{e}_{j_s})$ is the $i_r j_s$-matrix entry of the Jacobi matrix $D\phi(\mathbf{x})$ (with respect to the standard basis $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ of $\mathbf{R}^n$ and the standard basis $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ of $\mathbf{R}^m$). In other words, $g_J(\mathbf{x}) = \sum_I \phi^* f_I(\mathbf{x}) \det D\phi_{I,J}(\mathbf{x})$. This formula is identical to the one in Theorem 3.12 and therefore our new definition agrees with the old!

Forms on manifolds. Let M be an n-dimensional manifold in $\mathbf{R}^N$. For each point $\mathbf{x}$ in M the tangent space $T_{\mathbf{x}}M$ is an n-dimensional linear subspace of $\mathbf{R}^N$. A *differential form of degree k* or a *k-form* α on M is a choice of an alternating k-multilinear map $\alpha_{\mathbf{x}}$ on the vector space $T_{\mathbf{x}}M$, one for each $\mathbf{x} \in M$. This alternating map $\alpha_{\mathbf{x}}$ is required to depend smoothly on $\mathbf{x}$ in the following sense. According to the definition of a manifold, for each $\mathbf{x} \in M$ there exists an embedding $\psi\colon U \to \mathbf{R}^N$ such that $\psi(U) = M \cap V$ for some open set V in $\mathbf{R}^N$ containing $\mathbf{x}$. The tangent space at $\mathbf{x}$ is then $T_{\mathbf{x}}M = D\psi(\mathbf{t})(\mathbf{R}^n)$, where $\mathbf{t} \in U$ is chosen such that $\psi(\mathbf{t}) = \mathbf{x}$. The *pullback* of α under the local parametrization ψ is defined by

$$(\psi^*\alpha)_{\mathbf{t}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \alpha_{\psi(\mathbf{t})}(D\psi(\mathbf{t})\mathbf{v}_1, D\psi(\mathbf{t})\mathbf{v}_2, \dots, D\psi(\mathbf{t})\mathbf{v}_k).$$

Then $\psi^*\alpha$ is a k-form on U, an open subset of $\mathbf{R}^n$, so $\psi^*\alpha = \sum_I f_I\, dt_I$ for certain functions f_I defined on U. We will require the functions f_I to be smooth. (The form $\psi^*\alpha = \sum_I f_I\, dt_I$ is the *local representative* of α relative to the embedding ψ, as introduced in Section 7.1.) To recapitulate:

7.16. DEFINITION. A *k-form* α on M is a choice, for each $\mathbf{x} \in M$, of an alternating k-multilinear map $\alpha_{\mathbf{x}}$ on $T_{\mathbf{x}}M$, which depends smoothly on $\mathbf{x}$.

The book **[BT82]** describes a k-form as an "animal" that inhabits a "world" M, eats ordered k-tuples of tangent vectors, and spits out numbers.

7.17. EXAMPLE. Let M be a one-dimensional manifold in $\mathbf{R}^N$. Let us choose an orientation ("direction") on M. A tangent vector to M is *positive* if it points in the same direction as the orientation and *negative* if it points in the opposite direction. Define a 1-form α on M as follows. For $\mathbf{x} \in M$ and a tangent vector $\mathbf{v} \in T_{\mathbf{x}}M$ put

$$\alpha_{\mathbf{x}}(\mathbf{v}) = \begin{cases} \|\mathbf{v}\| & \text{if } \mathbf{v} \text{ is positive,} \\ -\|\mathbf{v}\| & \text{if } \mathbf{v} \text{ is negative.} \end{cases}$$

This form is the *element of arc length* of M. We shall see in Chapter 8 how to generalize it to higher-dimensional manifolds and in Chapter 9 how to use it to calculate arc lengths and volumes.

We can calculate the local representative $\psi^*\alpha$ of a k-form α for any embedding $\psi\colon U \to \mathbf{R}^N$ parametrizing a portion of M. Suppose we had two different such embeddings $\psi_i\colon U_i \to M$ and $\psi_j\colon U_j \to M$, such that $\mathbf{x}$ is contained in both $W_i = \psi_i(U_i)$ and $W_j = \psi_j(U_j)$. How do the local expressions $\alpha_i = \psi_i^*\alpha$ and $\alpha_j = \psi_j^*\alpha$ for α compare? To answer this question, consider the coordinate change map $\psi_j^{-1} \circ \psi_i$, which maps $\psi_i^{-1}(W_i \cap W_j)$ to $\psi_j^{-1}(W_i \cap W_j)$. From $\alpha_i = \psi_i^*\alpha$ and $\alpha_j = \psi_j^*\alpha$ we recover the transformation law (7.1)

$$\alpha_j = (\psi_i^{-1} \circ \psi_j)^*\alpha_i.$$

This shows that Definitions 7.1 and 7.16 of differential forms on a manifold are equivalent.

Exercises

7.1. The vectors $\mathbf{e}_1 + \mathbf{e}_2$ and $\mathbf{e}_1 - \mathbf{e}_2$ form a basis of $\mathbf{R}^2$. What is the dual basis of $(\mathbf{R}^2)^*$?

7.2. Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be a basis of $\mathbf{R}^n$ and let $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ be the dual basis of $(\mathbf{R}^n)^*$. Let A be an invertible $n \times n$-matrix. Then by elementary linear algebra the set of vectors $\{A\mathbf{v}_1, \dots, A\mathbf{v}_n\}$ is also a basis of $\mathbf{R}^n$. Show that the corresponding dual basis is the set of row vectors $\{\lambda_1 A^{-1}, \lambda_2 A^{-1}, \dots, \lambda_n A^{-1}\}$.

7.3. Suppose that μ is a bilinear function on a vector space V satisfying $\mu(\mathbf{v}, \mathbf{v}) = 0$ for all vectors $\mathbf{v} \in V$. Prove that μ is alternating. Generalize this observation to k-multilinear functions.

7.4. Show that the bilinear function μ of Example 7.8 is equal to $dx_1\, dx_2 + dx_3\, dx_4$.

7.5. The wedge product is a generalization of the cross product to arbitrary dimensions in the sense that

$$\mathbf{x} \times \mathbf{y} = \big(*(\mathbf{x}^T \wedge \mathbf{y}^T)\big)^T$$

for all $\mathbf{x}, \mathbf{y} \in \mathbf{R}^3$. Prove this formula. (Interpretation: $\mathbf{x}$ and $\mathbf{y}$ are column vectors, $\mathbf{x}^T$ and $\mathbf{y}^T$ are row vectors, $\mathbf{x}^T \wedge \mathbf{y}^T$ is a 2-form on $\mathbf{R}^3$, $*(\mathbf{x}^T \wedge \mathbf{y}^T)$ is a 1-form, i.e. a row vector. So both sides of the formula represent column vectors.)

7.6. Let V be a vector space and let $\mu_1, \mu_2, \dots, \mu_k \in V^*$ be covectors. Their *tensor product* is the function

$$\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_k \colon V^k \to \mathbf{R}$$

defined by

$$\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_k(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \mu_1(\mathbf{v}_1)\mu_2(\mathbf{v}_2) \cdots \mu_k(\mathbf{v}_k).$$

Show that $\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_k$ is a k-multilinear function.

7.7. Let $\mu\colon V^k \to \mathbf{R}$ be a k-multilinear function. Define a new function $\operatorname{Alt}\mu\colon V^k \to \mathbf{R}$ by

$$\operatorname{Alt}\mu(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sign}(\sigma)\mu(\mathbf{v}_{\sigma(1)}, \mathbf{v}_{\sigma(2)}, \dots, \mathbf{v}_{\sigma(k)}).$$

Prove the following.

(i) $\operatorname{Alt}\mu$ is an *alternating* k-multilinear function.
(ii) $\operatorname{Alt}\mu = \mu$ if μ is alternating.
(iii) $\operatorname{Alt}\operatorname{Alt}\mu = \operatorname{Alt}\mu$ for all k-multilinear μ.
(iv) Let $\mu_1, \mu_2, \dots, \mu_k \in V^*$. Then

$$\mu_1\mu_2\cdots\mu_k = \frac{1}{k!}\operatorname{Alt}(\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_k).$$

7.8. Show that $\det(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = dx_1\,dx_2\cdots dx_n(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ for all vectors $\mathbf{v}_1$, $\mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbf{R}^n$. In short,

$$\det = dx_1\,dx_2\cdots dx_n.$$

7.9. Let V and W be vector spaces and $L\colon V \to W$ a linear map. Show that $L^*(\lambda\mu) = (L^*\lambda)(L^*\mu)$ for all covectors λ, $\mu \in W^*$.

CHAPTER 8

Volume forms

8.1. n-Dimensional volume in $\mathbf{R}^N$

Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ be vectors in $\mathbf{R}^N$. The *block* or *parallelepiped* spanned by these vectors is the set of all vectors of the form $\sum_{i=1}^n c_i\mathbf{a}_i$, where the coefficients c_i range over the unit interval $[0,1]$. For $n = 1$ this is also called a *line segment* and for $n = 2$ a *parallelogram*. We will need a formula for the volume of a block. If $n < N$ there is no coherent way of defining an orientation on all n-blocks in $\mathbf{R}^N$, so this volume will be not an oriented but an absolute volume. We approach this problem in a similar way as the problem of defining the determinant, namely by imposing a few reasonable axioms.

8.1. Definition. An *absolute n-dimensional Euclidean volume function* is a function

$$\mathrm{vol}_n \colon \underbrace{\mathbf{R}^N \times \mathbf{R}^N \times \cdots \times \mathbf{R}^N}_{n \text{ times}} \to \mathbf{R}$$

with the following properties:

(i) homogeneity:

$$\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \ldots, c\mathbf{a}_i, \ldots, \mathbf{a}_n) = |c| \,\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)$$

for all scalars c and all vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$;

(ii) invariance under shear transformations:

$$\mathrm{vol}_n(\mathbf{a}_1, \ldots, \mathbf{a}_i + c\mathbf{a}_j, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) = \mathrm{vol}_n(\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n)$$

for all scalars c and any $i \neq j$;

(iii) invariance under Euclidean motions:

$$\mathrm{vol}_n(Q\mathbf{a}_1, \ldots, Q\mathbf{a}_n) = \mathrm{vol}_n(\mathbf{a}_1, \ldots, \mathbf{a}_n)$$

for all orthogonal matrices Q;

(iv) normalization: $\mathrm{vol}_n(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = 1$.

We shall shortly see that these axioms uniquely determine the n-dimensional volume function.

8.2. Lemma. (i) *$\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \|\mathbf{a}_1\| \, \|\mathbf{a}_2\| \cdots \|\mathbf{a}_n\|$ if $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are orthogonal vectors.*

(ii) *$\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = 0$ if the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are dependent.*

Proof. Suppose $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are orthogonal. First assume they are nonzero. Then we can define $\mathbf{q}_i = \|\mathbf{a}_i\|^{-1}\mathbf{a}_i$. The vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ are ortho*normal*. Complete them to an orthonormal basis $\mathbf{q}_1, \mathbf{q}_2 \ldots, \mathbf{q}_n, \mathbf{q}_{n+1}, \ldots, \mathbf{q}_N$ of $\mathbf{R}^N$. Let

Q be the matrix whose i-th column is $\mathbf{q}_i$. Then Q is orthogonal and $Q\mathbf{e}_i = \mathbf{q}_i$. Therefore

$$\begin{aligned}
\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) &= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\|\,\mathrm{vol}_n(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n) && \text{by Axiom (i)}\\
&= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\|\,\mathrm{vol}_n(Q\mathbf{e}_1, Q\mathbf{e}_2, \dots, Q\mathbf{e}_n)\\
&= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\|\,\mathrm{vol}_n(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) && \text{by Axiom (iii)}\\
&= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\| && \text{by Axiom (iv),}
\end{aligned}$$

which proves part (i) if all $\mathbf{a}_i$ are nonzero. If one of the $\mathbf{a}_i$ is $\mathbf{0}$, the vectors $\mathbf{a}_1, \mathbf{a}_2, \dots,$ $\mathbf{a}_n$ are dependent, so the statement follows from part (ii), which we prove next.

Assume $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are dependent. For simplicity suppose $\mathbf{a}_1$ is a linear combination of the other vectors, $\mathbf{a}_1 = \sum_{i=2}^n c_i\mathbf{a}_i$. By repeatedly applying Axiom (ii) we get

$$\begin{aligned}
\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) &= \mathrm{vol}_n\left(\sum_{i=2}^n c_i\mathbf{a}_i, \mathbf{a}_2, \dots, \mathbf{a}_n\right)\\
&= \mathrm{vol}_n\left(\sum_{i=3}^n c_i\mathbf{a}_i, \mathbf{a}_2, \dots, \mathbf{a}_n\right) = \cdots = \mathrm{vol}_n(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n).
\end{aligned}$$

Now by Axiom (i),

$$\mathrm{vol}_n(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n) = \mathrm{vol}_n(0\,\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n) = 0\,\mathrm{vol}_n(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n) = 0,$$

which proves property (ii). QED

This brings us to the volume formula. We can form a matrix A out of the column vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. It does not make sense to take $\det A$ because A is not square, unless $n = N$. However, the product A^TA is square and we can take *its* determinant.

8.3. THEOREM. *There exists a unique n-dimensional volume function on* $\mathbf{R}^N$. *Let* $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbf{R}^N$ *and let* A *be the* $N \times n$*-matrix whose i-th column is* $\mathbf{a}_i$. *Then*

$$\boxed{\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \sqrt{\det(A^TA)}.}$$

PROOF. We leave it to the reader to check that the function $\sqrt{\det(A^TA)}$ satisfies the axioms for a n-dimensional volume function on $\mathbf{R}^N$. (See Exercise 8.2.) Here we prove only the uniqueness part of the theorem.

Case 1. First assume that $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are orthogonal. Then A^TA is a diagonal matrix. Its i-th diagonal entry is $\|\mathbf{a}_i\|^2$, so $\sqrt{\det(A^TA)} = \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\|$, which is equal to $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ by Lemma 8.2(i).

Case 2. Next assume that $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are dependent. Then the matrix A has a nontrivial nullspace, i.e. there exists a nonzero n-vector $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{0}$. But then $A^TA\mathbf{v} = \mathbf{0}$, so the columns of A^TA are dependent as well. Since A^TA is square, this implies $\det A^TA = 0$, so $\sqrt{\det(A^TA)} = 0$, which is equal to $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ by Lemma 8.2(ii).

Case 3. Finally consider an arbitrary sequence of independent vectors $\mathbf{a}_1,$ $\mathbf{a}_2, \dots, \mathbf{a}_n$. This sequence can be transformed into an orthogonal sequence $\mathbf{v}_1,$ $\mathbf{v}_2, \dots, \mathbf{v}_n$ by the Gram-Schmidt process. This works as follows: let $\mathbf{b}_1 = \mathbf{0}$ and for $i > 1$ let $\mathbf{b}_i$ be the orthogonal projection of $\mathbf{a}_i$ onto the span of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{i-1}$; then

$\mathbf{v}_i = \mathbf{a}_i - \mathbf{b}_i$. (See illustration below.) Let V be the $N \times n$-matrix whose i-th column is $\mathbf{v}_i$. Then by repeated applications of Axiom (ii),

$$\begin{aligned}\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) &= \mathrm{vol}_n(\mathbf{v}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \mathrm{vol}_n(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{a}_n) = \cdots \\ &= \mathrm{vol}_n(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \sqrt{\det(V^T V)}, \quad (8.1)\end{aligned}$$

where the last equality follows from Case 1. Since $\mathbf{v}_i = \mathbf{a}_i - \mathbf{b}_i$, where $\mathbf{b}_i$ is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{i-1}$, we have $V = AU$, where U is a $n \times n$-matrix of the form

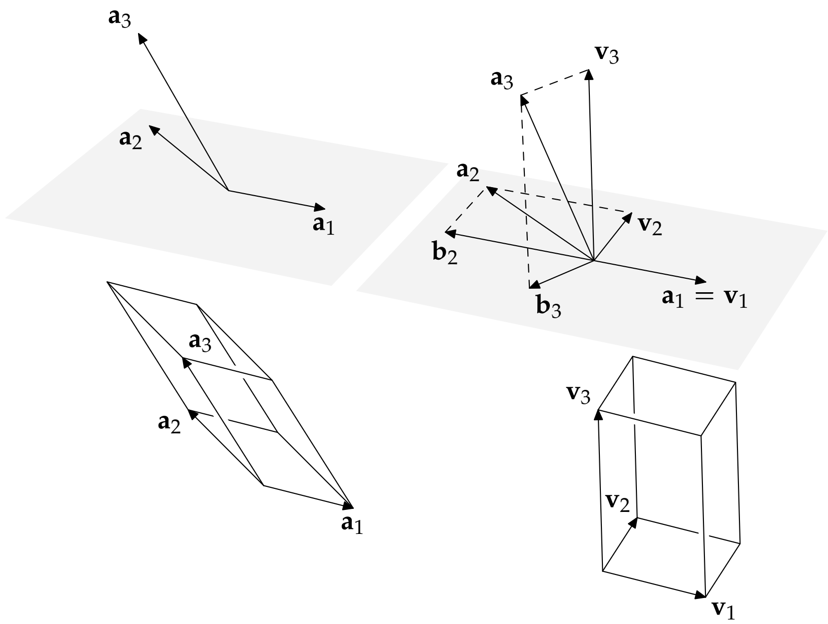
$$U = \begin{pmatrix} 1 & * & * & \cdots & * \\ 0 & 1 & * & \cdots & * \\ 0 & 0 & 1 & \cdots & * \\ \vdots & \vdots & & \ddots & * \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Note that U has determinant 1. This implies that $V^T V = U^T A^T A U$ and

$$\det(A^T A) = \det U^T \det(A^T A) \det U = \det(U^T A^T A U) = \det(V^T V).$$

Using formula (8.1) we get $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \sqrt{\det(A^T A)}$. QED

The Gram-Schmidt process transforms a sequence of n independent vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ into an orthogonal sequence $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. (The horizontal "floor" represents the plane spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$.) The block spanned by the $\mathbf{a}$'s has the same volume as the rectangular block spanned by the $\mathbf{v}$'s.



For $n = N$ Theorem 8.3 gives the following result.

8.4. COROLLARY. *Let $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ be vectors in $\mathbf{R}^n$ and let A be the $n \times n$-matrix whose i-th column is $\mathbf{a}_i$. Then* $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = |\det A|$.

PROOF. A is square, so $\det(A^T A) = \det A^T \det A = (\det A)^2$ by Theorem 3.7(ii) and therefore $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \sqrt{(\det A)^2} = |\det A|$ by Theorem 8.3. QED

8.2. Orientations

Oriented vector spaces. You are probably familiar with orientations on vector spaces of dimension ≤ 3. An orientation of a line is an assignment of a direction. An orientation of a plane is a choice of a direction of rotation, clockwise versus counterclockwise. An orientation of a three-dimensional space is a choice of "handedness", i.e. a choice of a right-hand rule versus a left-hand rule.

These notions can be generalized as follows. Let V be an n-dimensional vector space over the real numbers. Suppose that $\mathscr{B} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ and $\mathscr{B}' = (\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n)$ are two ordered bases of V. Then we can write $\mathbf{v}'_i = \sum_j a_{i,j} \mathbf{v}_j$ and $\mathbf{v}_i = \sum_j b_{i,j} \mathbf{v}'_j$ for suitable coefficients $a_{i,j}$ and $b_{i,j}$. The $n \times n$-matrices $A = (a_{i,j})$ and $B = (b_{i,j})$ satisfy $AB = BA = I$ and are therefore invertible. We say that the bases $\mathscr{B}$ and $\mathscr{B}'$ define the *same orientation* of V if $\det A > 0$. If $\det A < 0$, the two bases define *opposite orientations*.

For instance, if $(\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n) = (\mathbf{v}_2, \mathbf{v}_1, \dots, \mathbf{v}_n)$, then

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

so $\det A = -1$. Hence the ordered bases $(\mathbf{v}_2, \mathbf{v}_1, \dots, \mathbf{v}_n)$ and $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ define opposite orientations.

We know now what it means for two bases to have the same orientation, but how do we define the concept of an orientation itself? In typical mathematician's fashion we define the orientation of V determined by the basis $\mathscr{B}$ to be the collection of all ordered bases that have the same orientation as $\mathscr{B}$. (There is an analogous definition of the number 1, namely as the collection of all sets that contain one element.) The orientation determined by $\mathscr{B} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ is denoted by $[\mathscr{B}]$ or $[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$. So if $\mathscr{B}$ and $\mathscr{B}'$ define the same orientation then $[\mathscr{B}] = [\mathscr{B}']$. If they define opposite orientations we write $[\mathscr{B}] = -[\mathscr{B}']$. Because the determinant of an invertible matrix is either positive or negative, there are two possible orientations of V. An *oriented vector space* is a vector space together with a choice of an orientation. This preferred orientation is then called *positive*.

For $n = 0$ we need to make a special definition, because a zero-dimensional space has an empty basis. In this case we define an orientation of V to be a choice of sign, $+$ or $-$.

8.5. EXAMPLE. The *standard orientation* on $\mathbf{R}^n$ is the orientation $[\mathbf{e}_1, \dots, \mathbf{e}_n]$ defined by the standard ordered basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$. We shall always use this orientation on $\mathbf{R}^n$.

Maps and orientations. Let V and W be oriented vector spaces of the same dimension and let $L\colon V \to W$ be an invertible linear map. Choose a positively oriented basis $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ of V. Because L is invertible, the ordered n-tuple

$(L\mathbf{v}_1, L\mathbf{v}_2, \dots, L\mathbf{v}_n)$ is an ordered basis of W. If this basis is positively, resp. negatively, oriented we say that L is *orientation-preserving*, resp. *orientation-reversing*. This definition does not depend on the choice of the basis, for if $(\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_n)$ is another positively oriented basis of V, then $\mathbf{v}'_i = \sum_j a_{i,j}\mathbf{v}_j$ with $\det(a_{i,j}) > 0$. Therefore $L\mathbf{v}'_i = L\left(\sum_j a_{i,j}\mathbf{v}_j\right) = \sum_j a_{i,j}L\mathbf{v}_j$, and hence the two bases $(L\mathbf{v}_1, L\mathbf{v}_2, \dots, L\mathbf{v}_n)$ and $(L\mathbf{v}'_1, L\mathbf{v}'_2, \dots, L\mathbf{v}'_n)$ of W determine the same orientation.

Oriented manifolds. Now let M be a manifold. We define an *orientation* of M to be a choice of an orientation for each tangent space $T_\mathbf{x}M$ which varies continuously over M. "Continuous" means that for every $\mathbf{x} \in M$ there exists a local parametrization $\psi\colon W \to M$, with W open in $\mathbf{R}^n$ and $\mathbf{x} \in \psi(W)$, such that $D\psi_\mathbf{y}\colon \mathbf{R}^n \to T_\mathbf{y}M$ preserves the orientation for all $\mathbf{y} \in W$. (Here $\mathbf{R}^n$ is equipped with its standard orientation.) A manifold is *orientable* if it possesses an orientation; it is *oriented* if a specific orientation has been chosen.

Hypersurfaces. The case of a hypersurface, a manifold of codimension 1, is particularly instructive. A *unit normal vector field* on a manifold M in $\mathbf{R}^n$ is a smooth function $\mathbf{n}\colon M \to \mathbf{R}^n$ such that $\mathbf{n}(\mathbf{x}) \perp T_\mathbf{x}M$ and $\|\mathbf{n}(\mathbf{x})\| = 1$ for all $\mathbf{x} \in M$.

8.6. PROPOSITION. *A hypersurface in $\mathbf{R}^n$ is orientable if and only if it possesses a unit normal vector field.*

PROOF. Let M be a hypersurface in $\mathbf{R}^n$. Suppose M possesses a unit normal vector field. Let $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ be an ordered basis of $T_\mathbf{x}M$ for some $\mathbf{x} \in M$. Then $(\mathbf{n}(\mathbf{x}), \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ is a basis of $\mathbf{R}^n$, because $\mathbf{n}(\mathbf{x}) \perp \mathbf{v}_i$ for all i. We say that $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ is positively oriented if $(\mathbf{n}(\mathbf{x}), \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ is a positively oriented basis of $\mathbf{R}^n$. This defines an orientation on M, called the orientation *induced* by the normal vector field $\mathbf{n}$.

Conversely, let us suppose that M is an oriented hypersurface in $\mathbf{R}^n$. For each $\mathbf{x} \in M$ the tangent space $T_\mathbf{x}M$ is $n-1$-dimensional, so its orthogonal complement $(T_\mathbf{x}M)^\perp$ is a line. There are therefore precisely two vectors of length 1 which are perpendicular to $T_\mathbf{x}M$. We can pick a preferred unit normal vector as follows. Let $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ be a positively oriented basis of $T_\mathbf{x}M$. The *positive* unit normal vector is that unit normal vector $\mathbf{n}(\mathbf{x})$ that makes $(\mathbf{n}(\mathbf{x}), \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ a positively oriented basis of $\mathbf{R}^n$. In Exercise 8.8 you will be asked to check that $\mathbf{n}(\mathbf{x})$ depends smoothly on $\mathbf{x}$. In this way we have produced a unit normal vector field on M. QED

8.7. EXAMPLE. Let us regard $\mathbf{R}^{n-1}$ as the subspace of $\mathbf{R}^n$ spanned by the first $n-1$ standard basis vectors $\mathbf{e}_1$, $\mathbf{e}_2, \dots$, $\mathbf{e}_{n-1}$. The standard orientation on $\mathbf{R}^n$ is $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n]$, and the standard orientation on $\mathbf{R}^{n-1}$ is $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}]$. Since

$$[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n] = (-1)^{n+1}[\mathbf{e}_n, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}]$$

by Exercise 8.5, the positive unit normal to $\mathbf{R}^{n-1}$ in $\mathbf{R}^n$ is $(-1)^{n+1}\mathbf{e}_n$.

The positive unit normal on an oriented hypersurface M in $\mathbf{R}^n$ can be regarded as a map $\mathbf{n}$ from M into the unit sphere S^{n-1}, which is often called the *Gauß map* of M. The unit normal enables one to distinguish between two sides of M: the direction of $\mathbf{n}$ is "out" or "up"; the opposite direction is "in" or "down". For this reason orientable hypersurfaces are often called *two-sided*, whereas the nonorientable ones are called *one-sided*. Let us show that a hypersurface given by a single equation is always orientable.

8.8. PROPOSITION. *Let U be open in $\mathbf{R}^n$ and let $\phi\colon U \to \mathbf{R}$ be a smooth function. Let c be a regular value of ϕ. Then the manifold $\phi^{-1}(c)$ has a unit normal vector field given by $\mathbf{n}(\mathbf{x}) = \operatorname{grad}\phi(\mathbf{x})/\|\operatorname{grad}\phi(\mathbf{x})\|$ and is therefore orientable.*

PROOF. The regular value theorem tells us that $M = \phi^{-1}(c)$ is a hypersurface in $\mathbf{R}^n$ (if nonempty), and also that $T_{\mathbf{x}}M = \ker D\phi_{\mathbf{x}} = \big(\operatorname{grad}\phi(\mathbf{x})\big)^{\perp}$. The function $\mathbf{n}(\mathbf{x}) = \operatorname{grad}\phi(\mathbf{x})/\|\operatorname{grad}\phi(\mathbf{x})\|$ therefore defines a unit normal vector field on M. Appealing to Proposition 8.6 we conclude that M is orientable. QED

8.9. EXAMPLE. Taking $\phi(\mathbf{x}) = \|\mathbf{x}\|^2$ and $c = r^2$ we obtain that the $(n-1)$-sphere of radius r about the origin is orientable. The unit normal is

$$\mathbf{n}(\mathbf{x}) = \operatorname{grad}\phi(\mathbf{x})/\|\operatorname{grad}\phi(\mathbf{x})\| = \mathbf{x}/\|\mathbf{x}\|.$$

8.3. Volume forms

Now let M be an *oriented* n-manifold in $\mathbf{R}^N$. Choose a collection of embeddings $\psi_i\colon U_i \to \mathbf{R}^N$ with U_i open in $\mathbf{R}^n$ such that $M = \bigcup_i \psi_i(U_i)$ and such that $D\psi_i(\mathbf{t})\colon \mathbf{R}^n \to T_{\mathbf{x}}M$ is orientation-preserving for all $\mathbf{t} \in U_i$. The *volume form* μ_M, also denoted by μ, is the n-form on M whose local representative relative to the embedding ψ_i is defined by

$$\mu_i = \psi_i^*\mu = \sqrt{\det(D\psi_i(\mathbf{t})^T D\psi_i(\mathbf{t}))}\, dt_1\, dt_2 \cdots dt_n.$$

By Theorem 8.3 the square-root factor measures the volume of the n-dimensional block in the tangent space $T_{\mathbf{x}}M$ spanned by the columns of $D\psi_i(\mathbf{t})$, the Jacobi matrix of ψ_i at $\mathbf{t}$. Hence you should think of μ as measuring the volume of infinitesimal blocks inside M.

8.10. THEOREM. *For any oriented n-manifold M in $\mathbf{R}^N$ the volume form μ_M is a well-defined n-form.*

PROOF. To show that μ is well-defined we need to check that its local representatives satisfy the transformation law (7.1). So let us put $\phi = \psi_i^{-1} \circ \psi_j$ and substitute $\mathbf{t} = \phi(\mathbf{u})$ into μ_i. Since each of the embeddings ψ_i is orientation-preserving, we have $\det D\phi > 0$. Hence by Theorem 3.13 we have

$$\phi^*(dt_1\, dt_2 \cdots dt_n) = \det D\phi(\mathbf{u})\, du_1\, du_2 \cdots du_n = |\det D\phi(\mathbf{u})|\, du_1\, du_2 \cdots du_n.$$

Therefore

$$\begin{aligned}
\phi^*\mu_i &= \sqrt{\det\big(D\psi_i(\phi(\mathbf{u}))^T D\psi_i(\phi(\mathbf{u}))\big)}\, |\det D\phi(\mathbf{u})|\, du_1\, du_2 \cdots du_n \\
&= \sqrt{\det D\phi(\mathbf{u})^T \det\big(D\psi_i(\phi(\mathbf{u}))^T D\psi_i(\phi(\mathbf{u}))\big) \det D\phi(\mathbf{u})}\, du_1\, du_2 \cdots du_n \\
&= \sqrt{\det\big((D\psi_i(\phi(\mathbf{u}))D\phi(\mathbf{u}))^T D\psi_i(\phi(\mathbf{u}))D\phi(\mathbf{u})\big)}\, du_1\, du_2 \cdots du_n \\
&= \sqrt{\det((D\psi_j(\mathbf{u}))^T D\psi_j(\mathbf{u}))}\, du_1\, du_2 \cdots du_n = \mu_j,
\end{aligned}$$

where in the second to last identity we applied the chain rule. QED

For $n = 1$ the volume form is usually called the *element of arc length*, for $n = 2$, the *element of surface area*, and for $n = 3$, the *volume element*. Traditionally these are denoted by ds, dA, and dV, respectively. Don't be misled by this old-fashioned notation: volume forms are seldom exact! The volume form μ_M is highly dependent on the embedding of M into $\mathbf{R}^N$. It changes if we dilate or shrink or otherwise deform M.

8.11. EXAMPLE. Let U be an open subset of $\mathbf{R}^n$. Recall from Example 6.5 that U is a manifold covered by a single embedding, namely the identity map $\psi\colon U \to U$, $\psi(\mathbf{x}) = \mathbf{x}$. Then $\det(D\psi^T D\psi) = 1$, so the volume form on U is simply $dt_1\, dt_2 \cdots dt_n$, the ordinary volume form on $\mathbf{R}^n$.

8.12. EXAMPLE. Let I be an interval in the real line and $f\colon I \to \mathbf{R}$ a smooth function. Let $M \subseteq \mathbf{R}^2$ be the graph of f. By Example 6.7 M is a 1-manifold in $\mathbf{R}^2$. Indeed, M is the image of the embedding $\psi\colon I \to \mathbf{R}^2$ given by $\psi(t) = (t, f(t))$. Let us give M the orientation induced by the embedding ψ, i.e. "from left to right". What is the element of arc length of M? Let us compute the pullback $\psi^*\mu$, a 1-form on I. We have

$$D\psi(t) = \begin{pmatrix} 1 \\ f'(t) \end{pmatrix}, \qquad D\psi(t)^T D\psi(t) = \begin{pmatrix} 1 & f'(t) \end{pmatrix} \begin{pmatrix} 1 \\ f'(t) \end{pmatrix} = 1 + f'(t)^2,$$

so $\psi^*\mu = \sqrt{\det(D\psi(t)^T D\psi(t))}\, dt = \sqrt{1 + f'(t)^2}\, dt$.

The next result can be regarded as an alternative definition of μ_M. It is perhaps more intuitive, but it requires familiarity with Section 7.2.

8.13. PROPOSITION. *Let M be an oriented n-manifold in $\mathbf{R}^N$. Let $\mathbf{x} \in M$ and $\mathbf{v}_1$, $\mathbf{v}_2, \dots, \mathbf{v}_n \in T_{\mathbf{x}}M$. Then the volume form of M is given by*

$$\mu_{M,\mathbf{x}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \begin{cases} \operatorname{vol}_n(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) & \text{if } \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \text{ are positively oriented,} \\ -\operatorname{vol}_n(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) & \text{if } \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \text{ are negatively oriented,} \\ 0 & \text{if } \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \text{ are linearly dependent,} \end{cases}$$

i.e. $\mu_{M,\mathbf{x}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ is the oriented volume of the n-dimensional parallelepiped in $T_{\mathbf{x}}M$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

PROOF. For each $\mathbf{x}$ in M and n-tuple of tangent vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ at $\mathbf{x}$ let $\omega_{\mathbf{x}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ be the oriented volume of the block spanned by these n vectors. This defines an n-form ω on M and we must show that $\omega = \mu_M$. Let U be an open subset of $\mathbf{R}^n$ and $\psi\colon U \to \mathbf{R}^N$ an orientation-preserving embedding with $\psi(U) \subseteq M$ and $\psi(\mathbf{t}) = \mathbf{x}$ for some $\mathbf{t}$ in U. Let us calculate the n-form $\psi^*\omega$ on U. We have $\psi^*\omega = g\, dt_1\, dt_2 \cdots dt_n$ for some function g. By Lemma 7.12 this function is given by

$$g(\mathbf{t}) = \psi^*\omega_{\mathbf{x}}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = \omega_{\mathbf{x}}(D\psi(\mathbf{t})\mathbf{e}_1, D\psi(\mathbf{t})\mathbf{e}_2, \dots, D\psi(\mathbf{t})\mathbf{e}_n),$$

where in the second equality we used the definition of pullback. The vectors $D\psi(\mathbf{t})\mathbf{e}_1$, $D\psi(\mathbf{t})\mathbf{e}_2, \dots, D\psi(\mathbf{t})\mathbf{e}_n$ are a positively oriented basis of $T_{\mathbf{x}}M$ and, moreover, are the columns of the matrix $D\psi(\mathbf{t})$, so by Theorem 8.3 they span a positive volume of magnitude $\sqrt{\det(D\psi(\mathbf{t})^T D\psi(\mathbf{t}))}$. This shows that $g = \sqrt{\det(D\psi^T D\psi)}$ and therefore

$$\psi^*\omega = \sqrt{\det(D\psi^T D\psi)}\, dt_1\, dt_2 \cdots dt_n.$$

Thus $\psi^*\omega$ is equal to the local representative of μ_M with respect to the embedding ψ. Since this holds for all embeddings ψ, we have $\omega = \mu_M$. QED

Volume form of a hypersurface. For oriented hypersurfaces M in $\mathbf{R}^n$ there is a more convenient expression for the volume form μ_M. Recall the vector-valued forms

$$d\mathbf{x} = \begin{pmatrix} dx_1 \\ \vdots \\ dx_n \end{pmatrix} \qquad \text{and} \qquad *d\mathbf{x} = \begin{pmatrix} *dx_1 \\ \vdots \\ *dx_n \end{pmatrix}$$

introduced in Section 2.5. Let $\mathbf{n}$ be the positive unit normal vector field on M and let $\mathbf{F}$ be any vector field on M, i.e. a smooth map $\mathbf{F}\colon M \to \mathbf{R}^n$. Then the inner product $\mathbf{F} \cdot \mathbf{n}$ is a function defined on M. It measures the component of $\mathbf{F}$ orthogonal to M. The product $(\mathbf{F} \cdot \mathbf{n})\mu_M$ is an $n-1$-form on M. On the other hand we have the $n-1$-form $*(\mathbf{F} \cdot d\mathbf{x}) = \mathbf{F} \cdot *d\mathbf{x}$.

8.14. THEOREM. *On the hypersurface M we have*

$$\boxed{\mathbf{F} \cdot *d\mathbf{x} = (\mathbf{F} \cdot \mathbf{n})\mu_M.}$$

FIRST PROOF. This proof is short but requires familiarity with the material in Section 7.2. Let $\mathbf{x} \in M$. Let us change the coordinates on $\mathbf{R}^n$ in such a way that the first $n-1$ standard basis vectors $(\mathbf{e}_1, \mathbf{e}_2 \dots, \mathbf{e}_{n-1})$ form a positively oriented basis of $T_\mathbf{x}M$. Then, according to Example 8.7, the positive unit normal at $\mathbf{x}$ is given by $\mathbf{n}(\mathbf{x}) = (-1)^{n+1}\mathbf{e}_n$ and the volume form satisfies $\mu_{M,\mathbf{x}}(\mathbf{e}_1, \dots, \mathbf{e}_{n-1}) = 1$. Writing $\mathbf{F} = \sum_{i=1}^n F_i\mathbf{e}_i$, we have $\mathbf{F}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = (-1)^{n+1}F_n(\mathbf{x})$. On the other hand

$$\mathbf{F} \cdot *d\mathbf{x} = \sum_i (-1)^{i+1} F_i\, dx_1 \cdots \widehat{dx_i} \cdots dx_n,$$

and therefore $(\mathbf{F} \cdot *d\mathbf{x})(\mathbf{e}_1, \dots, \mathbf{e}_{n-1}) = (-1)^{n+1}F_n$. This proves that

$$(\mathbf{F} \cdot *d\mathbf{x})_\mathbf{x}(\mathbf{e}_1, \dots, \mathbf{e}_{n-1}) = (\mathbf{F}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}))\mu_M(\mathbf{e}_1, \dots, \mathbf{e}_{n-1}),$$

which implies $(\mathbf{F} \cdot *d\mathbf{x})_\mathbf{x} = (\mathbf{F}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}))\mu_M$. Since this equality holds for every $\mathbf{x} \in M$, we find $\mathbf{F} \cdot *d\mathbf{x} = (\mathbf{F} \cdot \mathbf{n})\mu_M$. QED

SECOND PROOF. Choose an embedding $\psi\colon U \to \mathbf{R}^n$, where U is open in $\mathbf{R}^{n-1}$, such that $\psi(U) \subseteq M$, $\mathbf{x} \in \psi(U)$. Let $\mathbf{t} \in U$ be the point satisfying $\psi(\mathbf{t}) = \mathbf{x}$. As a preliminary step in the proof we are going to replace the embedding ψ with a new one enjoying a particularly nice property. Let us change the coordinates on $\mathbf{R}^n$ in such a way that the first $n-1$ standard basis vectors $(\mathbf{e}_1, \mathbf{e}_2 \dots, \mathbf{e}_{n-1})$ form a positively oriented basis of $T_\mathbf{x}M$. Then at $\mathbf{x}$ the positive unit normal is given by $\mathbf{n}(\mathbf{x}) = (-1)^{n+1}\mathbf{e}_n$. Since the columns of the Jacobi matrix $D\psi(\mathbf{t})$ are independent, there exist unique vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{n-1}$ in $\mathbf{R}^{n-1}$ such that $D\psi(\mathbf{t})\mathbf{a}_i = \mathbf{e}_i$ for $i = 1$, $2, \dots, n-1$. These vectors $\mathbf{a}_i$ are independent, because the $\mathbf{e}_i$ are independent. Therefore the $(n-1) \times (n-1)$-matrix A with i-th column vector equal to $\mathbf{a}_i$ is invertible. Put $\tilde{U} = A^{-1}(U)$, $\tilde{\mathbf{t}} = A^{-1}\mathbf{t}$ and $\tilde{\psi} = \psi \circ A$. Then $\tilde{U}$ is open in $\mathbf{R}^{n-1}$, $\tilde{\psi}(\tilde{\mathbf{t}}) = \mathbf{x}$, $\tilde{\psi}\colon \tilde{U} \to \mathbf{R}^n$ is an embedding with $\tilde{\psi}(\tilde{U}) = \psi(U)$, and

$$D\tilde{\psi}(\tilde{\mathbf{t}}) = D\psi(\mathbf{t}) \circ DA(\tilde{\mathbf{t}} = D\psi(\mathbf{t}) \circ A$$

by the chain rule. Therefore the i-th column vector of $D\tilde{\psi}(\tilde{\mathbf{t}})$ is

$$D\tilde{\psi}(\tilde{\mathbf{t}})\mathbf{e}_i = D\psi(\mathbf{t})A\mathbf{e}_i = D\psi(\mathbf{t})\mathbf{a}_i = \mathbf{e}_i \tag{8.2}$$

for $i = 1, 2, \ldots, n-1$. (On the left $\mathbf{e}_i$ denotes the i-th standard basis vector in $\mathbf{R}^{n-1}$, on the right it denotes the i-th standard basis vector in $\mathbf{R}^n$.) In other words, the Jacobi matrix of $\tilde{\psi}$ at $\tilde{\mathbf{t}}$ is the $(n-1) \times n$-matrix

$$D\tilde{\psi}(\tilde{\mathbf{t}}) = \begin{pmatrix} I_{n-1} \\ \mathbf{0} \end{pmatrix},$$

where I_{n-1} is the $(n-1) \times (n-1)$ identity matrix and $\mathbf{0}$ denotes a row consisting of $n-1$ zeros.

Let us now calculate $\tilde{\psi}^*\big((\mathbf{F} \cdot \mathbf{n})\mu_M\big)$ and $\tilde{\psi}^*(\mathbf{F} \cdot *d\mathbf{x})$ at the point $\tilde{\mathbf{t}}$. Writing $\mathbf{F} \cdot \mathbf{n} = \sum_{i=1}^n F_i n_i$ and using the definition of μ_M we get

$$\tilde{\psi}^*\big((\mathbf{F} \cdot \mathbf{n})\mu_M\big) = \left(\sum_{i=1}^n \tilde{\psi}^*(F_i n_i)\right) \sqrt{\det(D\tilde{\psi}^T D\tilde{\psi})}\, d\tilde{t}_1\, d\tilde{t}_2 \cdots d\tilde{t}_{n-1}.$$

From formula (8.2) we have $\det\big(D\tilde{\psi}(\tilde{\mathbf{t}})^T D\tilde{\psi}(\tilde{\mathbf{t}})\big) = 1$. So evaluating this expression at the point $\tilde{\mathbf{t}}$ and using $\mathbf{n}(\mathbf{x}) = (-1)^{n+1}\mathbf{e}_n$ we get

$$\big(\tilde{\psi}^*(\mathbf{F} \cdot \mathbf{n})\mu_M\big)_{\tilde{\mathbf{t}}} = (-1)^{n+1} F_n(\mathbf{x})\, d\tilde{t}_1\, d\tilde{t}_2 \cdots d\tilde{t}_{n-1}.$$

From $\mathbf{F} \cdot *d\mathbf{x} = \sum_{i=1}^n (-1)^{i+1} F_i\, dx_1\, dx_2 \cdots \widehat{dx_i} \cdots dx_n$ we get

$$\tilde{\psi}^*(\mathbf{F} \cdot *d\mathbf{x}) = \sum_{i=1}^n (-1)^{i+1} \tilde{\psi}^* F_i\, d\tilde{\psi}_1\, d\tilde{\psi}_2 \cdots \widehat{d\tilde{\psi}_i} \cdots d\tilde{\psi}_n.$$

From formula (8.2) we see $\partial\tilde{\psi}_i(\tilde{\mathbf{t}})/\partial\tilde{t}_j = \delta_{i,j}$ for $1 \le i, j \le n-1$ and $\partial\tilde{\psi}_n(\tilde{\mathbf{t}})/\partial\tilde{t}_j = 0$ for $1 \le j \le n-1$. Therefore

$$\big(\tilde{\psi}^*(\mathbf{F} \cdot *d\mathbf{x})\big)_{\tilde{\mathbf{t}}} = (-1)^{n+1} F_n(\mathbf{x})\, d\tilde{t}_1\, d\tilde{t}_2 \cdots d\tilde{t}_{n-1}.$$

We conclude that $\big(\tilde{\psi}^*(\mathbf{F} \cdot \mathbf{n})\mu_M\big)_{\tilde{\mathbf{t}}} = \big(\tilde{\psi}^*(\mathbf{F} \cdot *d\mathbf{x})\big)_{\tilde{\mathbf{t}}}$, in other words $\big((\mathbf{F} \cdot \mathbf{n})\mu_M\big)_{\mathbf{x}} = (\mathbf{F} \cdot *d\mathbf{x})_{\mathbf{x}}$. Since this holds for all $\mathbf{x} \in M$ we have $\mathbf{F} \cdot *d\mathbf{x} = (\mathbf{F} \cdot \mathbf{n})\mu_M$. QED

This theorem gives insight into the physical interpretation of $n-1$-forms. Think of the vector field $\mathbf{F}$ as representing the flow of a fluid or gas. The direction of the vector $\mathbf{F}$ indicates the direction of the flow and its magnitude measures the strength of the flow. Then Theorem 8.14 says that the $n-1$-form $\mathbf{F} \cdot *d\mathbf{x}$ measures, for any unit vector $\mathbf{n}$ in $\mathbf{R}^n$, the amount of fluid per unit of time passing through a hyperplane of unit volume perpendicular to $\mathbf{n}$. We call $\mathbf{F} \cdot *d\mathbf{x}$ the *flux* of the vector field $\mathbf{F}$.

Another application of the theorem is the following formula for the volume form on a hypersurface. The formula provides a heuristic interpretation of the vector-valued form $*d\mathbf{x}$: if $\mathbf{n}$ is a unit vector in $\mathbf{R}^n$, then the scalar-valued $n-1$-form $\mathbf{n} \cdot *d\mathbf{x}$ measures the volume of an infinitesimal $n-1$-dimensional parallelepiped perpendicular to $\mathbf{n}$.

8.15. COROLLARY. *Let $\mathbf{n}$ be the unit normal vector field and μ_M the volume form of an oriented hypersurface M in $\mathbf{R}^n$. Then*

$$\boxed{\mu_M = \mathbf{n} \cdot *d\mathbf{x}.}$$

PROOF. Set $\mathbf{F} = \mathbf{n}$ in Proposition 8.14. Then $\mathbf{F} \cdot \mathbf{n} = 1$ because $\|\mathbf{n}\| = 1$. QED

8.16. EXAMPLE. Suppose the hypersurface M is given by an equation $\phi(\mathbf{x}) = c$, where c is a regular value of a function $\phi\colon U \to \mathbf{R}$, with U open in $\mathbf{R}^n$. Then by Proposition 8.8 M has a unit normal $\mathbf{n} = \operatorname{grad}\phi/\|\operatorname{grad}\phi\|$. The volume form is therefore $\mu = \|\operatorname{grad}\phi\|^{-1}\operatorname{grad}\phi \cdot {*d\mathbf{x}}$. In particular, if M is the sphere of radius R about the origin in $\mathbf{R}^n$, then $\mathbf{n}(\mathbf{x}) = \mathbf{x}/R$, so $\mu_M = R^{-1}\mathbf{x}\cdot {*d\mathbf{x}}$.

Exercises

8.1. Deduce from Theorem 8.3 that the area of the parallelogram spanned by a pair of vectors $\mathbf{a}$, $\mathbf{b}$ in $\mathbf{R}^n$ is given by $\|\mathbf{a}\|\,\|\mathbf{b}\|\,\sin\phi$, where ϕ is the angle between $\mathbf{a}$ and $\mathbf{b}$ (which is taken to lie between 0 and π). Show that $\|\mathbf{a}\|\,\|\mathbf{b}\|\,\sin\phi = \|\mathbf{a}\times\mathbf{b}\|$ in $\mathbf{R}^3$.

8.2. Check that the function $\operatorname{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \sqrt{\det(A^TA)}$ satisfies the axioms of Definition 8.1.

8.3. Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l$ be vectors in $\mathbf{R}^N$ satisfying $\mathbf{u}_i\cdot\mathbf{v}_j = 0$ for $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, l$. ("The $\mathbf{u}$'s are perpendicular to the $\mathbf{v}$'s.") Prove that

$$\operatorname{vol}_{k+l}(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l) = \operatorname{vol}_k(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)\operatorname{vol}_l(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l).$$

8.4. Let $a_1, a_2, \dots, a_n$ be real numbers, let $c = \sqrt{1 + \sum_{i=1}^n a_i^2}$ and let

$$\mathbf{u}_1 = \begin{pmatrix}1\\0\\\vdots\\0\\a_1\end{pmatrix},\quad \mathbf{u}_2 = \begin{pmatrix}0\\1\\\vdots\\0\\a_2\end{pmatrix},\quad \dots,\quad \mathbf{u}_n = \begin{pmatrix}0\\0\\\vdots\\1\\a_n\end{pmatrix},\quad \mathbf{u}_{n+1} = \frac{1}{c}\begin{pmatrix}-a_1\\-a_2\\\vdots\\-a_n\\1\end{pmatrix}$$

be vectors in $\mathbf{R}^{n+1}$.

(i) Deduce from Exercise 8.3 that

$$\operatorname{vol}_n(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = \operatorname{vol}_{n+1}(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n, \mathbf{u}_{n+1}).$$

(ii) Prove that

$$\begin{vmatrix} 1+a_1^2 & a_1a_2 & a_1a_3 & \dots & a_1a_n\\ a_2a_1 & 1+a_2^2 & a_2a_3 & \dots & a_2a_n\\ a_3a_1 & a_3a_2 & 1+a_3^2 & \dots & a_3a_n\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ a_na_1 & a_na_2 & a_na_3 & \dots & 1+a_n^2 \end{vmatrix} = 1 + \sum_{i=1}^n a_i^2.$$

8.5. Justify the following identities concerning orientations of a vector space V. Here the $\mathbf{v}$'s form a basis of V (which in part (i) is n-dimensional and in parts (ii)–(iii) two-dimensional).

(i) If $\sigma \in S_n$ is any permutation, then

$$[\mathbf{v}_{\sigma(1)}, \mathbf{v}_{\sigma(2)}, \dots, \mathbf{v}_{\sigma(n)}] = \operatorname{sign}(\sigma)[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n].$$

(ii) $[-\mathbf{v}_1, \mathbf{v}_2] = -[\mathbf{v}_1, \mathbf{v}_2]$.
(iii) $[3\mathbf{v}_1, 5\mathbf{v}_2] = [\mathbf{v}_1, \mathbf{v}_2]$.

8.6. Let U be open in $\mathbf{R}^n$ and let $f\colon U \to \mathbf{R}$ be a smooth function. Let $\psi\colon U \to \mathbf{R}^{n+1}$ be the embedding $\psi(\mathbf{x}) = (\mathbf{x}, f(\mathbf{x}))$ and let $M = \psi(U)$, the graph of f. Define an orientation on M by requiring ψ to be orientation-preserving. Deduce from Exercise 8.4 that the volume form of M is given by $\psi^*\mu_M = \sqrt{1 + \|\operatorname{grad} f(\mathbf{x})\|^2}\,dx_1\,dx_2\cdots dx_n$.

8.7. Let $M = \operatorname{graph} f$ be the oriented hypersurface of Exercise 8.6.

(i) Show that the positive unit normal vector field on M is given by

$$\mathbf{n} = \frac{(-1)^{n+1}}{\sqrt{1+\|\operatorname{grad} f(\mathbf{x})\|^2}} \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \\ 1 \end{pmatrix}.$$

(ii) Derive the formula $\psi^* \mu_M = \sqrt{1+\|\operatorname{grad} f(\mathbf{x})\|^2}\, dx_1\, dx_2 \cdots dx_n$ from Corollary 8.15 by substituting $x_{n+1} = f(x_1, x_2, \dots, x_n)$. (Caution: for consistency you must replace n with $n+1$ in Corollary 8.15.)

8.8. Show that the unit normal vector field $\mathbf{n}\colon M \to \mathbf{R}^n$ defined in the proof of Proposition 8.6 is smooth. (Compute $\mathbf{n}$ in terms of an orientation-preserving parametrization $\psi\colon U \to M$ of an open subset of M.)

8.9. Let $\psi\colon (a,b) \to \mathbf{R}^n$ be an embedding. Let μ be the element of arc length on the embedded curve $M = \psi(a,b)$. Show that $\psi^*\mu$ is the 1-form on (a,b) given by $\|\psi'(t)\|\, dt = \sqrt{\psi_1'(t)^2 + \psi_2'(t)^2 + \cdots + \psi_n'(t)^2}\, dt$.

CHAPTER 9

Integration and Stokes' theorem on manifolds

In this chapter we will see how to integrate an n-form over an oriented n-manifold. In particular, by integrating the volume form we find the volume of the manifold. We will also discuss a version of Stokes' theorem for manifolds. This requires the slightly more general notion of a manifold with boundary.

9.1. Manifolds with boundary

The notion of a spherical earth developed in classical Greece around the time of Plato and Aristotle. Older cultures (and also Western culture until the rediscovery of Greek astronomy in the late Middle Ages) visualized the earth as a flat disc surrounded by an ocean or a void. A closed disc is not a manifold, because no neighbourhood of a point on the edge is the image of an open subset of $\mathbf{R}^2$ under an embedding. Rather, it is a manifold with boundary, a notion which can be defined as follows. The *n-dimensional halfspace* is

$$\mathbf{H}^n = \{\, x \in \mathbf{R}^n \mid x_n \geq 0 \,\}.$$

The *boundary* of $\mathbf{H}^n$ is $\partial\mathbf{H}^n = \{\, x \in \mathbf{R}^n \mid x_n = 0 \,\} = \mathbf{R}^{n-1}$ and its *interior* is $\operatorname{int}\mathbf{H}^n = \{\, x \in \mathbf{R}^n \mid x_n > 0 \,\}$.

9.1. DEFINITION. An *n-dimensional manifold with boundary* (or *n-manifold with boundary*) in $\mathbf{R}^N$ is a subset M of $\mathbf{R}^N$ such that for all $\mathbf{x} \in M$ there exist

- an open subset $V \subseteq \mathbf{R}^N$ containing $\mathbf{x}$,
- an open subset $U \subseteq \mathbf{R}^n$,
- and an embedding $\psi\colon U \to \mathbf{R}^N$ satisfying $\psi(U \cap \mathbf{H}^n) = V \cap M$.

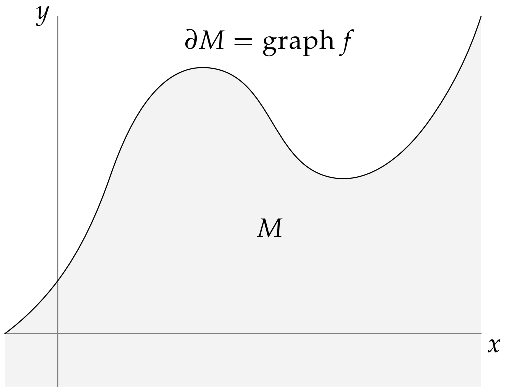
You should compare this definition carefully with Definition 6.4 of a manifold. If $\mathbf{x} = \psi(\mathbf{t})$ with $\mathbf{t} \in \partial\mathbf{H}^n$, then $\mathbf{x}$ is a *boundary point* of M. The *boundary* of M is the set of all boundary points and is denoted by ∂M. Its complement $M - \partial M$ is the *interior* of M and is denoted by $\operatorname{int} M$.

Somewhat confusingly, the boundary of a manifold with boundary is allowed to be empty. If nonempty, the boundary ∂M is an $n-1$-dimensional manifold. Likewise the interior $\operatorname{int} M$ is an n-manifold.

The most obvious example of an n-manifold with boundary is the halfspace $\mathbf{H}^n$ itself, which has boundary $\partial\mathbf{H}^n = \mathbf{R}^{n-1}$ and interior the open halfspace $\{\, \mathbf{x} \in \mathbf{R}^n \mid x_n > 0 \,\}$. Here is a more interesting type of example, which generalizes the graph of a function.

9.2. EXAMPLE. Let U' be an open subset of $\mathbf{R}^{n-1}$ and let $f\colon U' \to \mathbf{R}$ be a smooth function. Put $U = U' \times \mathbf{R}$ and write elements of U as $\binom{\mathbf{x}}{y}$ with $\mathbf{x}$ in U' and

y in $\mathbf{R}$. The *region below the graph of f* is the set consisting of all $\binom{\mathbf{x}}{y}$ in U such that $y \le f(\mathbf{x})$.



We assert that the region below the graph is an n-manifold whose boundary is exactly the graph of f. We will prove this by describing it as the image of a single embedding. Define $\psi\colon U \to \mathbf{R}^n$ by

$$\psi\begin{pmatrix}\mathbf{t}\\ u\end{pmatrix} = \begin{pmatrix}\mathbf{t}\\ f(\mathbf{t}) - u\end{pmatrix}.$$

As in Example 6.3 one verifies that ψ is an embedding, using the fact that

$$D\psi\begin{pmatrix}\mathbf{t}\\ u\end{pmatrix} = \begin{pmatrix}I_{n-1} & \mathbf{0}\\ Df(\mathbf{t}) & -1\end{pmatrix},$$

where $\mathbf{0}$ is the origin in $\mathbf{R}^{n-1}$. By definition therefore, the set $M = \psi(U \cap \mathbf{H}^n)$ is an n-manifold in $\mathbf{R}^n$ with boundary $\partial M = \psi(U \cap \partial\mathbf{H}^n)$. What are M and ∂M? A point $\binom{\mathbf{x}}{y}$ is in M if and only if it is of the form

$$\begin{pmatrix}\mathbf{x}\\ y\end{pmatrix} = \psi\begin{pmatrix}\mathbf{t}\\ u\end{pmatrix} = \begin{pmatrix}\mathbf{t}\\ f(\mathbf{t}) - u\end{pmatrix}$$

for some $\binom{\mathbf{t}}{u}$ in $U \cap \mathbf{H}^n$. Since $\mathbf{H}^n$ is given by $u \ge 0$, this is equivalent to $\mathbf{x} \in U'$ and $y \le f(\mathbf{x})$. Thus M is exactly the region below the graph. On $\partial\mathbf{H}^n$ we have $u = 0$, so ∂M is given by the equality $y = f(\mathbf{x})$, i.e. ∂M is the graph.

9.3. EXAMPLE. If $f\colon U' \to \mathbf{R}^m$ is a vector-valued map one cannot speak about the region "below" the graph, but one can do the following. Again put $U = U' \times \mathbf{R}$. Let $N = n + m - 1$ and think of $\mathbf{R}^N$ as the set of vectors $\binom{\mathbf{x}}{\mathbf{y}}$ with $\mathbf{x}$ in $\mathbf{R}^{n-1}$ and $\mathbf{y}$ in $\mathbf{R}^m$. Define $\psi\colon U \to \mathbf{R}^N$ by

$$\psi\begin{pmatrix}\mathbf{t}\\ u\end{pmatrix} = \begin{pmatrix}\mathbf{t}\\ f(\mathbf{t}) - u\mathbf{e}_m\end{pmatrix}$$

This time we have

$$D\psi(\mathbf{t}) = \begin{pmatrix}I_{n-1} & \mathbf{0}\\ Df(\mathbf{t}) & -\mathbf{e}_m\end{pmatrix}$$

and again ψ is an embedding. Therefore $M = \psi(U \cap \mathbf{H}^n)$ is an n-manifold in $\mathbf{R}^N$ with boundary $\partial M = \psi(U \cap \partial\mathbf{H}^n)$. This time M is the set of points $\binom{\mathbf{x}}{\mathbf{y}}$ of the form

$$\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \begin{pmatrix}\mathbf{t}\\ f(\mathbf{t}) - u\mathbf{e}_m\end{pmatrix}$$

with $\mathbf{t} \in U'$ and $u \geq 0$. Hence M is the set of points $\binom{\mathbf{x}}{\mathbf{y}}$ where $\mathbf{x}$ is in U' and where $\mathbf{y}$ satisfies $m-1$ equalities and one inequality:

$$y_1 = f_1(\mathbf{x}), \quad y_2 = f_2(\mathbf{x}), \quad \dots, \quad y_{m-1} = f_{m-1}(\mathbf{x}), \quad y_m \leq f_m(\mathbf{x}).$$

Again ∂M is given by $\mathbf{y} = f(\mathbf{x})$, so ∂M is the graph of f.

Here is an extension of the regular value theorem, Theorem 6.10, to manifolds with boundary. The proof, which we will not spell out, is similar to that of Theorem 6.10.

9.4. THEOREM (regular value theorem for manifolds with boundary). *Let U be open in $\mathbf{R}^N$ and let $\phi\colon U \to \mathbf{R}^m$ be a smooth map. Let M be the set of $\mathbf{x}$ in $\mathbf{R}^N$ satisfying*

$$\phi_1(\mathbf{x}) = c_1, \quad \phi_2(\mathbf{x}) = c_2, \quad \dots, \quad \phi_{m-1}(\mathbf{x}) = c_{m-1}, \quad \phi_m(\mathbf{x}) \leq c_m.$$

Suppose that $\mathbf{c} = (c_1, c_2, \dots, c_m)$ is a regular value of ϕ and that M is nonempty. Then M is a manifold in $\mathbf{R}^N$ of codimension $m-1$ and with boundary $\partial M = \phi^{-1}(\mathbf{c})$.

9.5. EXAMPLE. Let $U = \mathbf{R}^n$, $m = 1$ and $\phi(\mathbf{x}) = \|\mathbf{x}\|^2$. The set given by the inequality $\phi(\mathbf{x}) \leq 1$ is then the closed unit ball $\{\, \mathbf{x} \in \mathbf{R}^n \mid \|\mathbf{x}\| \leq 1 \,\}$. Since $\operatorname{grad} \phi(\mathbf{x}) = 2\mathbf{x}$, any nonzero value is a regular value of ϕ. Hence the ball is an n-manifold in $\mathbf{R}^n$, whose boundary is $\phi^{-1}(1)$, the unit sphere S^{n-1}.

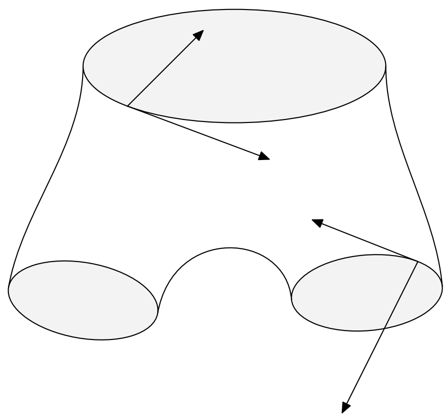
If more than one inequality is involved, singularities often arise. A simple example is the closed quadrant in $\mathbf{R}^2$ given by the pair of inequalities $x \geq 0$ and $y \geq 0$. This is not a manifold with boundary because its edge has a sharp angle at the origin. Similarly, a closed square is not a manifold with boundary.

However, one can show that a set given by a pair of inequalities of the form $a \leq f(\mathbf{x}) \leq b$, where a and b are both regular values of a function f, is a manifold with boundary. For instance, the spherical shell

$$\{\, x \in \mathbf{R}^n \mid R_1 \leq \|x\| \leq R_2 \,\}$$

is an n-manifold whose boundary is a union of two concentric spheres.

Other examples of manifolds with boundary are the *pair of pants*, a 2-manifold whose boundary consists of three closed curves,



and the Möbius band shown in Chapter 1. The Möbius band is a nonorientable manifold with boundary. We will not give a proof of this fact, but you can convince yourself that it is true by trying to paint the two sides of a Möbius band in different colours.

An n-manifold with boundary contained in $\mathbf{R}^n$ (i.e. of codimension 0) is often called a *domain*. For instance, a closed ball is a domain in $\mathbf{R}^n$.

To define the *tangent space* to a manifold with boundary M at a point $\mathbf{x}$ choose U and ψ as in the definition and put

$$T_{\mathbf{x}}M = D\psi(\mathbf{t})(\mathbf{R}^n).$$

As in the case of a manifold, this does not depend on the choice of the embedding ψ. Now suppose $\mathbf{x}$ is a boundary point of M and let $\mathbf{v} \in T_{\mathbf{x}}M$ be a tangent vector. Then $\mathbf{v} = D\psi(\mathbf{t})\mathbf{u}$ for some $\mathbf{u} \in \mathbf{R}^n$. We say that $\mathbf{v}$ points *inwards* if $u_n > 0$ and *outwards* if $u_n < 0$. If $u_n = 0$, then $\mathbf{v}$ is tangent to the boundary. In other words,

$$T_{\mathbf{x}}\partial M = D\phi(\mathbf{t})(\mathbf{R}^{n-1}).$$

The above picture of the pair of pants shows some tangent vectors at boundary points that are tangent to the boundary or outward-pointing.

Orienting the boundary. Let M be an oriented manifold with boundary. The orientation on M induces an orientation on ∂M by a method very similar to the proof of Proposition 8.6. Namely, for $\mathbf{x} \in \partial M$ define $\mathbf{n}(\mathbf{x}) \in T_{\mathbf{x}}M$ to be the unique outward-pointing tangent vector of length 1 which is orthogonal to $T_{\mathbf{x}}\partial M$. This defines the *unit outward-pointing normal vector field* on ∂M. A basis $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ of $T_{\mathbf{x}}\partial M$ is called positively oriented if $(\mathbf{n}(\mathbf{x}), \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n-1})$ is a positively oriented basis of $T_{\mathbf{x}}M$. This defines an orientation of ∂M, called the *induced orientation*. For instance, let $M = \mathbf{H}^n$ with the standard orientation $[\mathbf{e}_1, \dots, \mathbf{e}_n]$. At each point of $\partial M = \mathbf{R}^{n-1}$ the outward pointing normal is $-\mathbf{e}_n$. This implies that induced orientation on ∂M is $(-1)^n[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}]$, because

$$[-\mathbf{e}_n, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}] = (-1)^n[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}, \mathbf{e}_n].$$

9.2. Integration over orientable manifolds

As we saw in Chapter 5, a form of degree n can be integrated over a chain of dimension n. The integral does not change if we reparametrize the chain in an orientation-preserving manner. This opens up the possibility of integrating an n-form over an oriented n-manifold.

Let M be an n-dimensional oriented manifold (possibly with boundary) in $\mathbf{R}^N$ and let α be an n-form on M. To define the integral of α over M let us assume that M is compact. (A subset of $\mathbf{R}^N$ is called *compact* if it is closed and bounded; see Appendix A.2. This assumption is made to ensure that the integral is a proper integral and therefore converges.) For a start, let us also make the assumption that there exists a smooth map $c\colon [0,1]^n \to \mathbf{R}^N$ such that

- $c\big([0,1]^n\big) = M$ and
- the restriction of c to $(0,1)^n$ is an orientation-preserving embedding.

For instance, this assumption is satisfied for the n-sphere S^n (see Exercise 5.5) and the torus $S^1 \times S^1$ (see Exercise 9.4). The pullback $c^*\alpha$ is then an n-form on the cube $[0,1]^n$. We define

$$\int_M \alpha = \int_{[0,1]^n} c^*\alpha.$$

Suppose $\bar{c}\colon [0,1]^n \to \mathbf{R}^N$ is a smooth map with the same properties as c. To ensure that $\int_M \alpha$ is well-defined we need to check the following equality.

9.6. LEMMA. $\displaystyle\int_{[0,1]^n} c^*\alpha = \int_{[0,1]^n} \bar{c}^*\alpha.$

SKETCH OF PROOF. Let us denote the closed cube $[0,1]^n$ by R and let U be the open cube $(0,1)^n$. Let $V = U \cap c^{-1}(\bar{c}(U))$ and $\bar{V} = U \cap \bar{c}^{-1}(c(U))$. The complement of V and of $\bar{V}$ in R are negligeable in the sense that

$$\int_R c^*\alpha = \int_V c^*\alpha \qquad \text{and} \qquad \int_R \bar{c}^*\alpha = \int_{\bar{V}} \bar{c}^*\alpha. \tag{9.1}$$

By assumption the restriction of c to U is an embedding. This implies that $c\colon V \to M$ is a bijection onto its image, and so we see that $c^{-1} \circ \bar{c}$ is a bijection from $\bar{V}$ onto V. It is orientation-preserving, because c and $\bar{c}$ are orientation-preserving. Therefore, by Theorem 5.1,

$$\int_V c^*\alpha = \int_{\bar{V}} (c^{-1} \circ \bar{c})^*(c^*\alpha) = \int_{\bar{V}} (c \circ c^{-1} \circ \bar{c})^*\alpha = \int_{\bar{V}} \bar{c}^*\alpha.$$

Combining this with the equalities (9.1) we get the result. QED

Not every manifold can be covered with one single n-cube. However, it can be shown that there always exists a finite collection of n-cubes $c_i\colon [0,1]^n \to M$ for $i = 1, 2, \dots, k$, such that

(i) $\bigcup_{i=1}^k c_i\big([0,1]^n\big) = M$,
(ii) $c_i\big((0,1)^n\big) \cap c_j\big((0,1)^n\big)$ is empty for $i \neq j$,
(iii) for each i the restriction of c_i to $(0,1)^n$ is an orientation-preserving embedding.

We can then define

$$\int_M \alpha = \sum_{i=1}^k \int_{[0,1]^n} c_i^*\alpha,$$

and check as in Lemma 9.6 that the result does not depend on the maps c_i. (The condition (ii) on the maps is imposed to avoid "double counting" in the integral.)

9.7. DEFINITION. Let M a compact oriented manifold in $\mathbf{R}^N$. The *volume* of M is $\operatorname{vol} M = \int_M \mu$, where μ is the volume form on M. (If $\dim M = 1$, resp. 2, we speak of the *arc length*, resp. *surface area* of M.) The *integral* of a function f on M is defined as $\int_M f\mu$. The *mean* or *average* of f is the number $\bar{f} = (\operatorname{vol} M)^{-1} \int_M f\mu$. The *centroid* or *barycentre* of M is the point $\bar{\mathbf{x}}$ in R^n whose i-th coordinate is the mean value of x_i over M, i.e.

$$\bar{x}_i = \frac{1}{\operatorname{vol} M} \int_M x_i \mu.$$

The volume form depends on the embedding of M into $\mathbf{R}^N$, so the notions defined above depend on the embedding as well.

The most important property of the integral is the following version of Stokes' theorem, which can be viewed as a parametrization-independent version of Theorem 5.10 and is proved in a similar way.

9.8. THEOREM (Stokes' theorem for manifolds). *Let α be an $n-1$-form on a compact oriented n-manifold with boundary M. Give the boundary ∂M the induced orientation. Then*

$$\boxed{\int_M d\alpha = \int_{\partial M} \alpha.}$$

9.3. Gauß and Stokes

Stokes' theorem, Theorem 9.8, contains as special cases the integral theorems of vector calculus. These classical results involve a vector field $\mathbf{F} = \sum_{i=1}^n F_i \mathbf{e}_i$ defined on an open subset U of $\mathbf{R}^n$. As discussed in Section 2.5, to this vector field corresponds a 1-form $\alpha = \mathbf{F} \cdot d\mathbf{x} = \sum_{i=1}^n F_i\, dx_i$, which we can think of as the work done by the force $\mathbf{F}$ along an infinitesimal line segment $d\mathbf{x}$. We will now derive the classical integral theorems by applying Theorem 9.8 to one-dimensional, resp. n-dimensional, resp. two-dimensional manifolds M contained in U.

Fundamental theorem of calculus. If $\mathbf{F}$ is conservative, $\mathbf{F} = \operatorname{grad} g$ for a function g, then $\alpha = \operatorname{grad} g \cdot d\mathbf{x} = dg$. If M is a compact oriented 1-manifold with boundary in $\mathbf{R}^n$, then $\int_M dg = \int_{\partial M} g$ by Theorem 9.8. The boundary consists of two points $\mathbf{a}$ and $\mathbf{b}$ (if M is connected). If the orientation of M is "from $\mathbf{a}$ to $\mathbf{b}$", then $\mathbf{a}$ acquires a minus and $\mathbf{b}$ a plus. Stokes' theorem therefore gives the fundamental theorem of calculus in $\mathbf{R}^n$,

$$\boxed{\int_M \mathbf{F} \cdot d\mathbf{x} = g(\mathbf{b}) - g(\mathbf{a}).}$$

If we interpret $\mathbf{F}$ as a force acting on a particle travelling along M, then $-g$ stands for the potential energy of the particle in the force field. Thus the potential energy of the particle decreases by the amount of work done.

Gauß' divergence theorem. We have

$$*\alpha = \mathbf{F} \cdot *d\mathbf{x} \qquad \text{and} \qquad d{*\alpha} = \operatorname{div} \mathbf{F}\, dx_1\, dx_2 \cdots dx_n.$$

If N is a oriented hypersurface in $\mathbf{R}^n$ with positive unit normal $\mathbf{n}$, then $*\alpha = (\mathbf{F} \cdot \mathbf{n})\mu_N$ on N by Theorem 8.14. In this situation it is best to think of $\mathbf{F}$ as the flow vector field of a fluid, where the direction of $\mathbf{F}(\mathbf{x})$ gives the direction of the flow at a point $\mathbf{x}$ and the magnitude $\|\mathbf{F}(\mathbf{x})\|$ gives the mass of the amount of fluid passing per unit time through a hypersurface of unit area placed at $\mathbf{x}$ perpendicular to the vector $\mathbf{F}(\mathbf{x})$. Then $*\alpha$ describes the amount of fluid passing per unit time and per unit area through the hypersurface N. For this reason the $n-1$-form $*\alpha$ is also called the *flux* of $\mathbf{F}$, and its integral over N the *total flux* through N.

Applying Stokes' theorem to a compact domain M in $\mathbf{R}^n$ we get $\int_M d{*\alpha} = \int_{\partial M} \alpha$. Written in terms of the vector field $\mathbf{F}$ this is Gauß' divergence theorem,

$$\boxed{\int_M \operatorname{div} \mathbf{F}\, dx_1\, dx_2 \cdots dx_n = \int_{\partial M} (\mathbf{F} \cdot \mathbf{n})\mu_{\partial M}.}$$

Thus the total flux out of the hypersurface ∂M is the integral of $\operatorname{div} \mathbf{F}$ over M. If the fluid is incompressible (e.g. most liquids) then this formula leads to the interpretation of the divergence of $\mathbf{F}$ (or equivalently $d{*\alpha}$) as a measure of the sources or sinks of the flow. Thus $\operatorname{div} \mathbf{F} = 0$ for an incompressible fluid without sources or sinks. If the fluid is a gas and if there are no sources or sinks then $\operatorname{div} \mathbf{F}(\mathbf{x}) > 0$ (resp. < 0) indicates that the gas is expanding (resp. being compressed) at $\mathbf{x}$.

Classical version of Stokes' theorem. Now let M be a compact two-dimensional oriented surface with boundary and let us rewrite Stokes' theorem $\int_M d\alpha = \int_{\partial M} \alpha$ in terms of the vector field $\mathbf{F}$. The right-hand side represents the work of $\mathbf{F}$ done around the boundary curve(s) of M, which is not necessarily 0 if $\mathbf{F}$ is not conservative. The left-hand side has a nice interpretation if $n = 3$. Then $*d\alpha = \operatorname{curl} \mathbf{F} \cdot d\mathbf{x}$, so $d\alpha = \operatorname{curl} \mathbf{F} \cdot *d\mathbf{x}$. Hence if $\mathbf{n}$ is the positive unit normal of the surface M in $\mathbf{R}^3$, then $d\alpha = (\operatorname{curl} \mathbf{F} \cdot \mathbf{n})\mu_M$ on M. In this way we get the classical formula of Stokes,

$$\boxed{\int_M (\operatorname{curl} \mathbf{F} \cdot \mathbf{n})\mu_M = \int_{\partial M} \mathbf{F} \cdot d\mathbf{x}.}$$

In other words, the total flux of $\operatorname{curl} \mathbf{F}$ through the surface M is equal to the work done by $\mathbf{F}$ around the boundary curves of M. This formula shows that $\operatorname{curl} \mathbf{F}$, or equivalently $*d\alpha$, can be regarded as a measure of the vorticity of the vector field.

Exercises

9.1. Let U be an open subset of $\mathbf{R}^n$ and let f, $g\colon U \to \mathbf{R}$ be two smooth functions satisfying $f(\mathbf{x}) < g(\mathbf{x})$ for all $\mathbf{x}$ in U. Let M be the set of all pairs $(\mathbf{x}, y)$ such that $\mathbf{x}$ in U and $f(\mathbf{x}) \le y \le g(\mathbf{x})$.

(i) Draw a picture of M if U is the open unit disc given by $x^2 + y^2 < 1$ and $f(x, y) = -\sqrt{1 - x^2 - y^2}$ and $g(x, y) = 2 - x^2 - y^2$.
(ii) Show directly from the definition that M is a manifold with boundary. (Use two embeddings to cover M.) What is the dimension of M and what are the boundary and the interior?
(iii) Give an example showing that M is not necessarily a manifold with boundary if the condition $f(\mathbf{x}) \le y \le g(\mathbf{x})$ fails.

9.2. (i) Let $\alpha = x\,dy - y\,dx$ and let M be a compact domain in the plane $\mathbf{R}^2$. Show that $\int_{\partial M} \alpha$ is twice the surface area of M.
(ii) Apply the observation of part (i) to find the area enclosed by the astroid $x = \cos^3 t$, $y = \sin^3 t$.
(iii) Let $\alpha = \mathbf{x} \cdot *d\mathbf{x}$ and let M be a compact domain in $\mathbf{R}^n$. Show that $\int_{\partial M} \alpha$ is a constant times the volume of M. What is the value of the constant?

9.3. Write the divergence theorem for the vector field $\mathbf{F} = -cx_n\mathbf{e}_n$ on $\mathbf{R}^n$, where c is a positive constant. Deduce Archimedes' Law: the buoyant force exerted on a submerged body is equal to the weight of the displaced fluid. $E\acute{\upsilon}\rho\eta\kappa\alpha$!

9.4. Let $R_1 \ge R_2 \ge 0$ be constants. Define a 3-cube $c\colon [0, R_2] \times [0, 2\pi] \times [0, 2\pi] \to \mathbf{R}^3$ by

$$c\begin{pmatrix} r \\ \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} (R_1 + r\cos\theta_2)\cos\theta_1 \\ (R_1 + r\cos\theta_2)\sin\theta_1 \\ r\sin\theta_2 \end{pmatrix}.$$

(i) Sketch the image of c.
(ii) Let x_1, x_2, x_3 be the standard coordinates on $\mathbf{R}^3$. Compute c^*dx_1, c^*dx_2, c^*dx_3 and $c^*(dx_1\,dx_2\,dx_3)$.
(iii) Find the volume of the solid parametrized by c.
(iv) Find the surface area of the boundary of this solid.

9.5. Let M be a compact domain in $\mathbf{R}^n$. Let f and g be smooth functions on M. The *Dirichlet integral* of f and g is $D(f,g) = \int_M (\operatorname{grad} f \cdot \operatorname{grad} g)\mu$, where $\mu = dx_1\,dx_2 \cdots dx_n$ is the volume form on M.

(i) Show that $df(*dg) = (\operatorname{grad} f \cdot \operatorname{grad} g)\mu$.

(ii) Show that $d{*}dg = (\Delta g)\mu$, where $\Delta g = \sum_{i=1}^n \partial^2 g/\partial x_i^2$.

(iii) Deduce from parts (i)–(ii) that $d(f(*dg)) = (\operatorname{grad} f \cdot \operatorname{grad} g + f\Delta g)\mu$.

(iv) Let $\mathbf{n}$ be the outward-pointing unit normal vector field on ∂M. Write $\partial g/\partial \mathbf{n}$ for the directional derivative $(Dg)\mathbf{n} = \operatorname{grad} g \cdot \mathbf{n}$. Show that

$$\int_{\partial M} f(*dg) = \int_{\partial M} f \frac{\partial g}{\partial \mathbf{n}} \mu_{\partial M}.$$

(v) Deduce from parts (iii) and (iv) Green's formula,

$$\int_{\partial M} f \frac{\partial g}{\partial \mathbf{n}} \mu_{\partial M} = D(f,g) + \int_M (f\Delta g)\mu.$$

(vi) Deduce Green's symmetric formula,

$$\int_{\partial M} \Big(f \frac{\partial g}{\partial \mathbf{n}} - g \frac{\partial f}{\partial \mathbf{n}} \Big) \mu_{\partial M} = \int_M (f\Delta g - g\Delta f)\mu.$$

9.6. In this problem we will calculate the volume of a ball and a sphere in Euclidean space. Let $B(R)$ be the closed ball of radius R about the origin in $\mathbf{R}^n$. Then its boundary $S(R) = \partial B(R)$ is the sphere of radius R. Put $V_n(R) = \operatorname{vol}_n B(R)$ and $A_n(R) = \operatorname{vol}_{n-1} S(R)$. Also put $V_n = V_n(1)$ and $A_n = A_n(1)$.

(i) Deduce from Corollary 8.15 that the volume form on $S(R)$ is the restriction of ν to $S(R)$, where ν is as in Exercise 2.16. Conclude that $A_n(R) = \int_{S(R)} \nu$.

(ii) Show that $V_n(R) = R^n V_n$ and $A_n(R) = R^{n-1} A_n$. (Substitute $\mathbf{y} = R\mathbf{x}$ in the volume forms of $B(R)$ and $S(R)$.)

(iii) Let $f\colon [0,\infty) \to \mathbf{R}$ be a continuous function. Define $g\colon \mathbf{R}^n \to \mathbf{R}$ by $g(\mathbf{x}) = f(\|\mathbf{x}\|)$. Use Exercise 2.16(ii) to prove that

$$\int_{B(R)} g\,dx_1\,dx_2 \cdots dx_n = \int_0^R f(r) A_n(r)dr = A_n \int_0^R f(r) r^{n-1} dr.$$

(iv) Show that

$$\Big(\int_{-\infty}^{\infty} e^{-r^2} dr \Big)^n = A_n \int_0^\infty e^{-r^2} r^{n-1} dr.$$

(Take $f(r) = e^{-r^2}$ in part (iii) and let $R \to \infty$.)

(v) Using Exercises B.10 and B.11 conclude that

$$A_n = \frac{2\pi^{\frac{n}{2}}}{\Gamma\big(\frac{n}{2}\big)}, \quad \text{whence} \quad A_{2m} = \frac{2\pi^m}{(m-1)!} \quad \text{and} \quad A_{2m+1} = \frac{2^{m+1}\pi^m}{1 \cdot 3 \cdot 5 \cdots (2m-1)}.$$

(vi) By taking $f(r) = 1$ in part (iii) show that $A_n = nV_n$ and $A_n(R) = \partial V_n(R)/\partial R$.

(vii) Deduce that

$$V_n = \frac{\pi^{\frac{n}{2}}}{\Gamma\big(\frac{n}{2}+1\big)}, \quad \text{whence} \quad V_{2m} = \frac{\pi^m}{m!} \quad \text{and} \quad V_{2m+1} = \frac{2^{m+1}\pi^m}{1 \cdot 3 \cdot 5 \cdots (2m+1)}.$$

(viii) Complete the following table. (Conventions: a space of negative dimension is empty; the volume of a zero-dimensional manifold is its number of points.)

n	0	1	2	3	4	5
$V_n(R)$			πR^2	$\frac{4}{3}\pi R^3$		
$A_n(R)$			$2\pi R$			

(ix) Find $\lim_{n\to\infty} A_n$, $\lim_{n\to\infty} V_n$ and $\lim_{n\to\infty}(A_{n+1}/A_n)$. Use Stirling's formula,

$$\lim_{x\to\infty} \frac{\Gamma(x+1)e^x}{x^{x+\frac{1}{2}}} = \sqrt{2\pi}.$$

CHAPTER 10

Applications to topology

10.1. Brouwer's fixed point theorem

Let M be a manifold, possibly with boundary. A *retraction* of M onto a subset A is a smooth map $\phi\colon M \to A$ such that $\phi(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x}$ in A. For instance, let M be the punctured unit ball in n-space,

$$M = \{\, \mathbf{x} \in \mathbf{R}^n \mid 0 < \|\mathbf{x}\| \leq 1 \,\}.$$

Then the normalization map $\phi(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ is a retraction of M onto its boundary $A = \partial M$, the unit sphere. The following theorem says that a retraction onto the boundary is impossible if M is compact and orientable.

10.1. THEOREM. *Let M be a compact orientable manifold with nonempty boundary. Then there does not exist a retraction from M onto ∂M.*

PROOF. Suppose $\phi\colon M \to \partial M$ was a retraction. Let us choose an orientation of M and equip ∂M with the induced orientation. Let $\beta = \mu_{\partial M}$ be the volume form on the boundary (relative to some embedding of M into $\mathbf{R}^N$). Let $\alpha = \phi^*\beta$ be its pullback to M. Let n denote the dimension of M. Note that β is an $n-1$-form on the $n-1$-manifold ∂M, so $d\beta = 0$. Therefore $d\alpha = d\phi^*\beta = \phi^* d\beta = 0$ and hence by Stokes' theorem $0 = \int_M d\alpha = \int_{\partial M} \alpha$. But ϕ is a retraction onto ∂M, so the restriction of ϕ to ∂M is the identity map and therefore $\alpha = \beta$ on ∂M. Thus

$$0 = \int_{\partial M} \alpha = \int_{\partial M} \beta = \operatorname{vol} \partial M \neq 0,$$

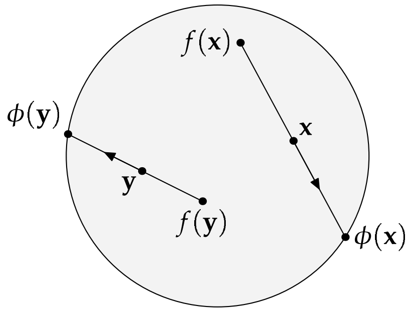
which is a contradiction. Therefore ϕ does not exist. QED

This brings us to one of the oldest results in topology. Suppose f is a map from a set X into itself. An element x of X is a *fixed point* of f if $f(x) = x$.

10.2. THEOREM (Brouwer's fixed point theorem). *Every smooth map from the closed unit ball into itself has at least one fixed point.*

PROOF. Let $M = \{\, \mathbf{x} \in \mathbf{R}^n \mid \|\mathbf{x}\| \leq 1 \,\}$ be the closed unit ball. Suppose $f\colon M \to M$ was a smooth map without fixed points. Then $f(\mathbf{x}) \neq \mathbf{x}$ for all $\mathbf{x}$. For each $\mathbf{x}$ in the ball consider the halfline starting at $f(\mathbf{x})$ and pointing in the direction of $\mathbf{x}$. This halfline intersects the unit sphere ∂M in a unique point that we shall call

$\phi(\mathbf{x})$, as in the following picture.



This defines a smooth map $\phi\colon M \to \partial M$. If $\mathbf{x}$ is in the unit sphere, then $\phi(\mathbf{x}) = \mathbf{x}$, so ϕ is a retraction of the ball onto its boundary, which contradicts Theorem 10.1. Therefore f must have a fixed point. QED

This theorem can be stated imprecisely as saying that after you stir a cup of coffee, at least one molecule must return to its original position. Brouwer originally stated his result for arbitrary continuous maps. This more general statement can be derived from Theorem 10.2 by an argument from analysis which shows that every continuous map is homotopic to a smooth map. (See Section 10.2 for the definition of homotopy.) The theorem also remains valid if the closed ball is replaced by a closed cube or a similar shape.

10.2. Homotopy

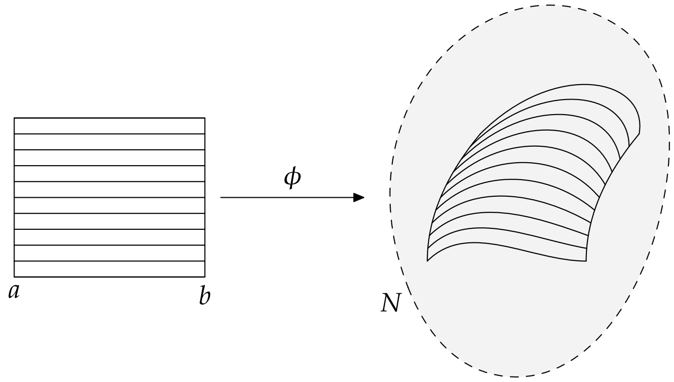
Definition and first examples. Suppose that ϕ_0 and ϕ_1 are two maps from a manifold M to a manifold N and that α is a form on N. What is the relationship between the pullbacks $\phi_0^*\alpha$ and $\phi_1^*\alpha$? There is a reasonable answer to this question if ϕ_0 can be smoothly deformed into ϕ_1. More formally, we say that ϕ_0 and ϕ_1 are *homotopic* if there exists a smooth map $\phi\colon M \times [0,1] \to N$ such that $\phi(\mathbf{x},0) = \phi_0(\mathbf{x})$ and $\phi(\mathbf{x},1) = \phi_1(\mathbf{x})$ for all $\mathbf{x}$ in M. The map ϕ is called a *homotopy*. Instead of $\phi(\mathbf{x},t)$ we often write $\phi_t(\mathbf{x})$. Then each ϕ_t is a map from M to N and we can think of ϕ_t as a family of maps parametrized by t in the unit interval that interpolates between ϕ_0 and ϕ_1, or as a one-second "movie" that at time 0 starts at ϕ_0 and at time 1 ends up at ϕ_1.

10.3. EXAMPLE. Let $M = N = \mathbf{R}^n$ and $\phi_0(\mathbf{x}) = \mathbf{x}$ (identity map) and $\phi_1(\mathbf{x}) = \mathbf{0}$ (constant map). Then ϕ_0 and ϕ_1 are homotopic. A homotopy is given by $\phi(\mathbf{x},t) = (1-t)\mathbf{x}$. This homotopy collapses Euclidean space onto the origin by moving each point radially inward. There are other ways to accomplish this. For instance $(1-t)^2\mathbf{x}$ and $(1-t^2)\mathbf{x}$ are two other homotopies between the same maps. We can also interchange ϕ_0 and ϕ_1: if $\phi_0(\mathbf{x}) = \mathbf{0}$ and $\phi_1(\mathbf{x}) = \mathbf{x}$, then we find a homotopy by reversing time (playing the movie backwards), $\phi(\mathbf{x},t) = t\mathbf{x}$.

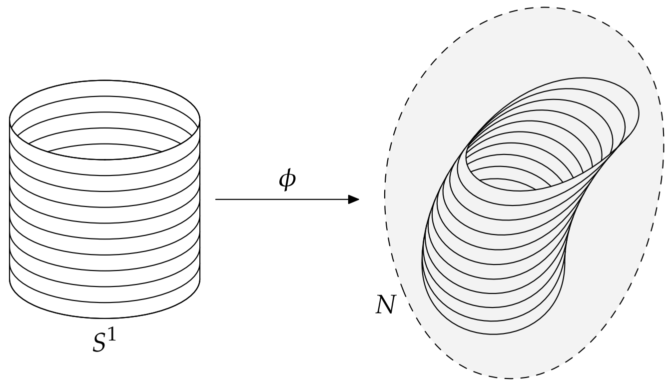
10.4. EXAMPLE. Let $M = N$ be the punctured Euclidean space $\mathbf{R}^n - \{\mathbf{0}\}$ and let $\phi_0(\mathbf{x}) = \mathbf{x}$ (identity map) and $\phi_1(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$ (normalization map). Then ϕ_0 and ϕ_1 are homotopic. A homotopy is given for instance by $\phi(\mathbf{x},t) = \mathbf{x}/\|\mathbf{x}\|^t$ or by $\phi(\mathbf{x},t) = (1-t)\mathbf{x} + t\mathbf{x}/\|\mathbf{x}\|$. Either of these homotopies collapses punctured Euclidean space onto the unit sphere about the origin by smoothly stretching or shrinking each vector until it has length 1.

10.5. EXAMPLE. A manifold M is said to be *contractible* if there exists a point $\mathbf{x}_0$ in M such that the constant map $\phi_0(\mathbf{x}) = \mathbf{x}_0$ is homotopic to the identity map $\phi(\mathbf{x}) = \mathbf{x}$. A specific homotopy $\phi\colon M \times [0,1] \to M$ from ϕ_0 to ϕ_1 is a *contraction* of M onto $\mathbf{x}_0$. (Perhaps "expansion" would be a more accurate term, a "contraction" being the result of replacing t with $1-t$.) Example 10.3 shows that $\mathbf{R}^n$ is contractible onto the origin. (In fact it is contractible onto any point $\mathbf{x}_0$. Can you write a contraction of $\mathbf{R}^n$ onto $\mathbf{x}_0$?) The same formula shows that an open or closed ball around the origin is contractible. We shall see in Theorem 10.19 that punctured n-space $\mathbf{R}^n - \{\mathbf{0}\}$ is *not* contractible.

Homotopy of curves. If M is an interval $[a,b]$ and N any manifold, then maps from M to N are nothing but parametrized curves in N. A homotopy of curves can be visualized as a piece of string moving through the manifold N.

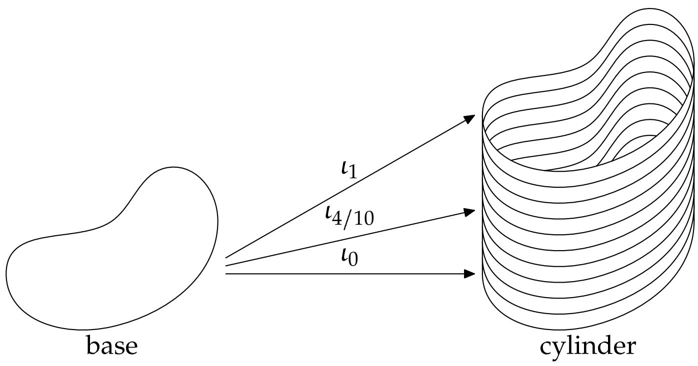


Homotopy of loops. A *loop* in a manifold N is a smooth map from the unit circle S^1 into N. This can be visualized as a thin rubber band sitting in N. A homotopy of loops $\phi\colon S^1 \times [0,1] \to N$ can be pictured as a rubber band floating through N from time 0 until time 1.



10.6. EXAMPLE. Consider the two loops $\phi_0, \phi_1\colon S^1 \to \mathbf{R}^2$ in the plane given by $\phi_0(\mathbf{x}) = \mathbf{x}$ and $\phi_1(\mathbf{x}) = \mathbf{x} + \binom{2}{0}$. A homotopy of loops is given by shifting ϕ_0 to the right, $\phi_t(\mathbf{x}) = \mathbf{x} + \binom{2t}{0}$. What if we regard ϕ_0 and ϕ_1 as loops in the punctured plane $\mathbf{R}^2 - \{\mathbf{0}\}$? Clearly the homotopy ϕ does not work, because it moves the loop through the forbidden point $\mathbf{0}$. (E.g. $\phi_t(\mathbf{x}) = \mathbf{0}$ for $\mathbf{x} = \binom{-1}{0}$ and $t = 1/2$.) In fact, however you try to move ϕ_0 to ϕ_1 you get stuck at the origin, so it seems intuitively clear that there exists no homotopy of loops from ϕ_0 to ϕ_1 in the punctured plane. This is indeed the case, as we shall see in Example 10.13.

The homotopy formula. The product $M \times [0,1]$ is often called the *cylinder* with *base* M. The two maps defined by $\iota_0(\mathbf{x}) = (\mathbf{x}, 0)$ and $\iota_1(\mathbf{x}) = (\mathbf{x}, 1)$ send M to the bottom, resp. the top of the cylinder. A homotopy $\iota\colon M \times [0,1] \to M \times [0,1]$ between these maps is given by the identity map $\iota(\mathbf{x}, t) = (\mathbf{x}, t)$. ("Slide the bottom to the top at speed 1.")



If M is an open subset of $\mathbf{R}^n$, a $k+1$-form on the cylinder can be written as

$$\gamma = \sum_I f_I(x,t)\, dx_I + \sum_J g_J(x,t)\, dt\, dx_J,$$

with I running over multi-indices of degree $k+1$ and J over multi-indices of degree k. (Here we write the dt in front of the dx's because that is more convenient in what follows.) The *cylinder operator* turns forms on the cylinder into forms on the base lowering the degree by 1,

$$\kappa\colon \Omega^{k+1}(M \times [0,1]) \to \Omega^k(M),$$

by taking the piece of γ involving dt and integrating it over the unit interval,

$$\kappa\gamma = \sum_J \left(\int_0^1 g_J(x,t)\, dt \right) dx_J.$$

(In particular $\kappa\gamma = 0$ for any γ that does not involve dt.) For a general manifold M we can write a $k+1$-form on the cylinder as $\gamma = \beta + dt\,\gamma$, where β and γ are forms on $M \times [0,1]$ (of degree $k+1$ and k respectively) that do not involve dt. We then define $\kappa\gamma = \int_0^1 \gamma\, dt$.

The following result will enable us to compare pullbacks of forms under homotopic maps. It can be regarded as an application of Stokes' theorem, but we shall give a direct proof.

10.7. LEMMA (cylinder formula). *Let M be a manifold. Then $\iota_1^*\gamma - \iota_0^*\gamma = \kappa\, d\gamma + d\kappa\gamma$ for all $k+1$-forms γ on $M \times [0,1]$. In short,*

$$\boxed{\iota_1^* - \iota_0^* = \kappa d + d\kappa.}$$

PROOF. We write out the proof for an open subset of $\mathbf{R}^n$. The proof for arbitrary manifolds is similar. It suffices to consider two cases: $\gamma = f\, dx_I$ and $\gamma = g\, dt\, dx_J$.

Case 1. If $\gamma = f\,dx_I$, then $\kappa\gamma = 0$ and $d\kappa\gamma = 0$. Also

$$d\gamma = \frac{\partial f}{\partial t}\,dt\,dx_I + \sum_i \frac{\partial f}{\partial x_i}\,dx_i\,dx_I = \frac{\partial f}{\partial t}\,dt\,dx_I + \text{terms not involving } dt,$$

so

$$\begin{aligned} d\kappa\gamma + \kappa\,d\gamma = \kappa\,d\gamma &= \left(\int_0^1 \frac{\partial f}{\partial t}(x,t)\,dt\right)dx_I \\ &= \big(f(x,1) - f(x,0)\big)\,dx_I = \iota_1^*\gamma - \iota_0^*\gamma. \end{aligned}$$

Case 2. If $\gamma = g\,dt\,dx_J$, then $\iota_0^*\gamma = \iota_1^*\gamma = 0$ and

$$d\gamma = \sum_i \frac{\partial g}{\partial x_i}\,dx_i\,dt\,dx_J = -\sum_i \frac{\partial g}{\partial x_i}\,dt\,dx_i\,dx_J,$$

so

$$\kappa\,d\gamma = -\sum_{i=1}^n \left(\int_0^1 \frac{\partial g}{\partial x_i}(x,t)\,dt\right)dx_i\,dx_J.$$

Also $\kappa\gamma = \big(\int_0^1 g(x,t)\,dt\big)\,dx_J$, so

$$d\kappa\gamma = \sum_{i=1}^n \frac{\partial}{\partial x_i}\left(\int_0^1 g(x,t)\,dt\right)dx_i\,dx_J = \sum_{i=1}^n \left(\int_0^1 \frac{\partial g}{\partial x_i}(x,t)\,dt\right)dx_i\,dx_J.$$

Hence $d\kappa\gamma + \kappa\,d\gamma = 0 = \iota_1^*\gamma - \iota_0^*\gamma$. QED

Now suppose we have a pair of maps ϕ_0 and ϕ_1 going from a manifold M to a manifold N and that $\phi\colon M \times [0,1] \to N$ is a homotopy between ϕ_0 and ϕ_1. For $\mathbf{x}$ in M we have $\phi \circ \iota_0(\mathbf{x}) = \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x})$, in other words $\phi_0 = \phi \circ \iota_0$. Similarly $\phi_1 = \phi \circ \iota_1$. Hence for any $k+1$-form α on N we have $\iota_0^*\phi^*\alpha = \phi_0^*\alpha$ and $\iota_1^*\phi^*\alpha = \phi_1^*\alpha$. Applying the cylinder formula to $\gamma = \phi^*\alpha$ we see that the pullbacks $\phi_0^*\alpha$ and $\phi_1^*\alpha$ are related in the following manner.

10.8. THEOREM (homotopy formula). *Let $\phi_0, \phi_1\colon M \to N$ be smooth maps from a manifold M to a manifold N and let $\phi\colon M \times [0,1] \to N$ be a homotopy from ϕ_0 to ϕ_1. Then $\phi_1^*\alpha - \phi_0^*\alpha = \kappa\phi^*d\alpha + d\kappa\phi^*\alpha$ for all $k+1$-forms α on N. In short,*

$$\boxed{\phi_1^* - \phi_0^* = \kappa\phi^*d + d\kappa\phi^*.}$$

In particular, if $d\alpha = 0$ we get $\phi_1^*\alpha = \phi_0^*\alpha + d\kappa\phi^*\alpha$.

10.9. COROLLARY. *If $\phi_0, \phi_1\colon M \to N$ are homotopic maps between manifolds and α is a closed form on N, then $\phi_0^*\alpha$ and $\phi_1^*\alpha$ differ by an exact form.*

This implies that if the degree of α is equal to the dimension of M, $\phi_0^*\alpha$ and $\phi_1^*\alpha$ have the same integral.

10.10. THEOREM. *Let M and N be manifolds and let α be a closed n-form on N, where $n = \dim M$. Suppose M is compact and oriented and has no boundary. Let ϕ_0 and ϕ_1 be homotopic maps from M to N. Then*

$$\int_M \phi_0^*\alpha = \int_M \phi_1^*\alpha.$$

PROOF. By Corollary 10.9, $\phi_1^*\alpha - \phi_0^*\alpha = d\beta$ for an $n-1$-form β on M. Hence by Stokes' theorem

$$\int_M (\phi_1^*\alpha - \phi_0^*\alpha) = \int_M d\beta = \int_{\partial M} \beta = 0,$$

because ∂M is empty. QED

ALTERNATIVE PROOF. Here is a proof based on Stokes' theorem for the manifold with boundary $M \times [0,1]$. The boundary of $M \times [0,1]$ consists of two copies of M, namely $M \times \{1\}$ and $M \times \{0\}$, the first of which is counted with a plus sign and the second with a minus. Therefore, if $\phi\colon M \times [0,1] \to N$ is a homotopy between ϕ_0 and ϕ_1,

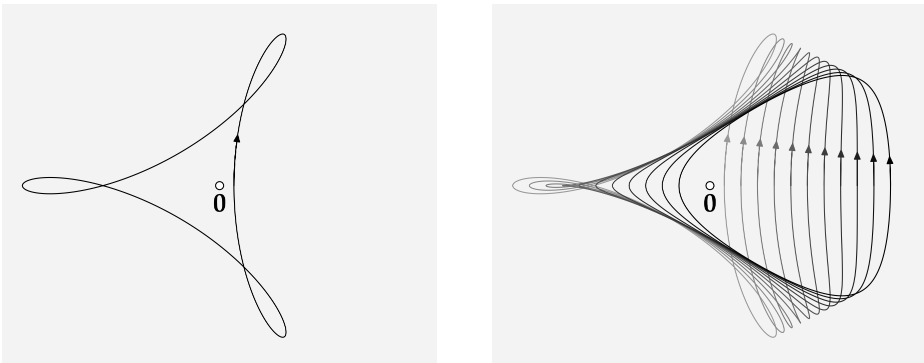
$$0 = \int_{M\times[0,1]} \phi^* d\alpha = \int_{M\times[0,1]} d\phi^*\alpha = \int_{\partial(M\times[0,1])} \phi^*\alpha = \int_M \phi_1^*\alpha - \int_M \phi_0^*\alpha.$$

QED

If M is the circle S^1, N the punctured plane $\mathbf{R}^2 - \{\mathbf{0}\}$ and α the angle form $(-y\,dx + x\,dy)/(x^2+y^2)$ of Example 3.8, then a map from M to N is a loop in N and the integral of α is 2π times the winding number of the loop. Thus Theorem 10.10 gives the following result.

10.11. COROLLARY. *Homotopic loops in $\mathbf{R}^2 - \{\mathbf{0}\}$ have the same winding number about the origin.*

10.12. EXAMPLE. Unfolding the three self-intersections in the curve pictured below does not affect its winding number.



10.13. EXAMPLE. The two circles ϕ_0 and ϕ_1 of Example 10.6 have winding number 1, resp. 0 and therefore are not homotopic (as loops in the punctured plane).

10.3. Closed and exact forms re-examined

The homotopy formula throws light on our old problem of when a closed form is exact, which we looked into in Section 2.3. The answer turns out to depend on the "shape" of the manifold on which the forms are defined. On some manifolds all closed forms (of positive degree) are exact, on others this is true only in certain degrees. Failure of exactness is typically detected by integrating over a submanifold of the correct dimension and finding a nonzero answer. In a certain sense all obstructions to exactness are of this nature. We shall not attempt to say the last word on this problem, but study a few representative special cases. The matter is explored in [**Fla89**] and at a more advanced level in [**BT82**].

0**-forms.** A closed 0-form on a manifold is a smooth function f satisfying $df = 0$. This means that f is constant (on each connected component of M). If this constant is nonzero, then f is not exact (because forms of degree -1 are by definition 0). So a closed 0-form is never exact (unless it is 0) for a rather uninteresting reason.

1**-forms and simple connectivity.** Let us now consider 1-forms on a manifold M. Theorem 4.5 says that the integral of an exact 1-form along a loop is 0. With a stronger assumption on the loop the same is true for arbitrary *closed* 1-forms. A loop $c\colon S^1 \to M$ is *null-homotopic* if it is homotopic to a constant loop. The integral of a 1-form along a constant loop is 0, so from Theorem 10.10 (where we set the M of the theorem equal to S^1) we get the following.

10.14. PROPOSITION. *Let c be a null-homotopic loop in M. Then $\int_c \alpha = 0$ for all closed forms α on M.*

A manifold is *simply connected* if every loop in it is null-homotopic.

10.15. THEOREM. *All closed* 1*-forms on a simply connected manifold are exact.*

PROOF. Let α be a closed 1-form and c a loop in M. Then c is null-homotopic, so $\int_c \alpha = 0$ by Proposition 10.14. The result now follows from Theorem 4.5. QED

10.16. EXAMPLE. The punctured plane $\mathbf{R}^2 - \{\mathbf{0}\}$ is not simply connected, because it possesses a nonexact closed 1-form. (See Example 4.6.) In contrast it can be proved that for $n \geq 3$ the sphere S^{n-1} and punctured n-space $\mathbf{R}^n - \{\mathbf{0}\}$ are simply connected. Intuitively, the reason is that in two dimensions a loop that encloses the puncture at the origin cannot be crumpled up to a point without getting stuck at the puncture, whereas in higher dimensions there is enough room to slide any loop away from the puncture and then squeeze it to a point.

The Poincaré lemma. On a contractible manifold *all* closed forms of positive degree are exact.

10.17. THEOREM (Poincaré lemma). *All closed k-forms on a contractible manifold are exact for $k \geq 1$.*

PROOF. Let M be a manifold and let $\phi\colon M \times [0,1] \to M$ be a contraction onto a point $\mathbf{x}_0$ in M, i.e. a smooth map satisfying $\phi(\mathbf{x}, 0) = \mathbf{x}_0$ and $\phi(\mathbf{x}, 1) = \mathbf{x}$ for all $\mathbf{x}$. Let α be a closed k-form on M with $k \geq 1$. Then $\phi_1^*\alpha = \alpha$ and $\phi_0^*\alpha = 0$, so putting $\beta = \kappa\phi^*\alpha$ we get

$$d\beta = d\kappa\phi^*\alpha = \phi_1^*\alpha - \phi_0^*\alpha - \kappa\, d\phi^*\alpha = \alpha.$$

Here we used the homotopy formula, Theorem 10.8, and the assumption that $d\alpha = 0$. Hence $d\beta = \alpha$. QED

The proof provides us with a formula for the "antiderivative", namely $\beta = \kappa\phi^*\alpha$, which can be made quite explicit in certain cases.

10.18. EXAMPLE. Let M be $\mathbf{R}^n$ and let $\phi(\mathbf{x}, t) = t\mathbf{x}$ be the radial contraction. Let $\alpha = \sum_i g_i \, dx_i$ be a 1-form. Then

$$\phi^*\alpha = \sum_i g_i(t\mathbf{x}) d(tx_i) = \sum_i g_i(t\mathbf{x})(x_i \, dt + t \, dx_i),$$

so

$$\beta = \kappa\phi^*\alpha = \sum_i x_i \int_0^1 g_i(t\mathbf{x}) \, dt.$$

According to the proof of the Poincaré lemma, the function β satisfies $d\beta = \alpha$ provided that $d\alpha = 0$. It is instructive to compare β with the function f constructed in the proof of Theorem 4.5. (See Exercise 10.5.)

Another typical application of the Poincaré lemma is showing that a manifold is not contractible by exhibiting a closed form that is not exact. For example, the punctured plane $\mathbf{R}^2 - \{\mathbf{0}\}$ is not contractible because it possesses a nonexact closed 1-form, namely the angle form. (See Example 4.6.) The angle form generalizes to an $n-1$-form on punctured n-space $\mathbf{R}^n - \{\mathbf{0}\}$,

$$\alpha = \frac{\mathbf{x} \cdot *d\mathbf{x}}{\|\mathbf{x}\|^n}.$$

10.19. THEOREM. *α is a closed but non-exact $n-1$-form on punctured n-space. Hence punctured n-space is not contractible.*

PROOF. $d\alpha = 0$ follows from Exercise 2.1(ii). The $n-1$-sphere $M = S^{n-1}$ has unit normal vector field $\mathbf{x}$, so by Corollary 8.15 on M we have $\alpha = \mu$, the volume form. Hence $\int_M \alpha = \operatorname{vol} M \neq 0$. On the other hand, suppose α was exact, $\alpha = d\beta$ for an $n-1$-form β. Then

$$\int_M \alpha = \int_M d\beta = \int_{\partial M} \alpha = 0$$

by Stokes' theorem, Theorem 9.8. This is a contradiction, so α is not exact. It now follows from the Poincaré lemma, Theorem 10.17, that $\mathbf{R}^n - \{\mathbf{0}\}$ is not contractible. QED

Using the same form α, but restricting it to the unit sphere S^{n-1}, we see that S^{n-1} is not contractible. But how about forms of degree not equal to $n-1$? Without proof we state the following fact.

10.20. THEOREM. *On $\mathbf{R}^n - \{\mathbf{0}\}$ and on S^{n-1} every closed form of degree $k \neq 1$, $n-1$ is exact.*

For a compact oriented hypersurface without boundary M contained in in $\mathbf{R}^n - \{\mathbf{0}\}$ the integral

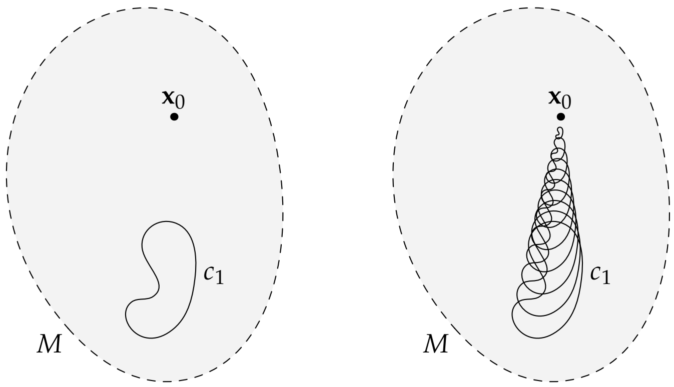
$$\boxed{\frac{1}{\operatorname{vol}_{n-1} S^{n-1}} \int_M \frac{\mathbf{x} \cdot *d\mathbf{x}}{\|\mathbf{x}\|^n}}$$

is the *winding number* of M about the origin. It generalizes the winding number of a closed curve in $\mathbf{R}^2 - \{\mathbf{0}\}$ around the origin. It can be shown that the winding number in any dimension is always an integer. It provides a measure of how many times the hypersurface wraps around the origin. For instance, the proof of Theorem 10.19 shows that the winding number of the $n-1$-sphere about the origin is 1.

Contractibility versus simple connectivity. Theorems 10.15 and 10.17 suggest that the notions of contractibility and simple connectivity are not independent.

10.21. PROPOSITION. *A contractible manifold is simply connected.*

PROOF. Use a contraction to collapse any loop onto a point.



Formally, let $c_1\colon S^1 \to M$ be a loop, $\phi\colon M \times [0,1] \to M$ a contraction of M onto $\mathbf{x}_0$. Put $c(s,t) = \phi(c_1(s),t)$. Then c is a homotopy between c_1 and the constant loop $c_0(t) = \phi(c_1(s),0) = \mathbf{x}_0$ positioned at $\mathbf{x}_0$. QED

As mentioned in Example 10.16, the sphere S^{n-1} and punctured n-space $\mathbf{R}^n - \{\mathbf{0}\}$ are simply connected for $n \geq 3$, although it follows from Theorem 10.19 that they are not contractible. Thus simple connectivity is weaker than contractibility.

The Poincaré conjecture. Not long after inventing the fundamental group Poincaré posed the following question. Let M be a compact three-dimensional manifold without boundary. Suppose M is simply connected. Is M homeomorphic to the three-dimensional sphere? (This means: does there exist a bijective map $M \to S^3$ which is continuous and has a continuous inverse?) This question became (inaccurately) known as the *Poincaré conjecture*. It is famously difficult and was the force that drove many of the developments in twentieth-century topology. It has an n-dimensional analogue, called the *generalized* Poincaré conjecture, which asks whether every compact n-dimensional manifold without boundary which is homotopy equivalent to S^n is homeomorphic to S^n. We cannot here go into this fascinating problem in any serious way, other than to report that it has now been completely solved. Strangely, the case $n \geq 5$ of the generalized Poincaré conjecture conjecture was the easiest and was confirmed by S. Smale in 1960. The case $n = 4$ was done by M. Freedman in 1982. The case $n = 3$, the original version of the conjecture, turned out to be the hardest, but was finally confirmed by G. Perelman in 2002-03. For a discussion and references, see the paper *Towards the Poincaré conjecture and the classification of* 3*-manifolds* by J. Milnor, which appeared in the November 2003 issue of the Notices of the American Mathematical Society and can be read online at `http://www.ams.org/notices/`.

Exercises

10.1. Write a formula for the map ϕ figuring in the proof of Brouwer's fixed point theorem and prove that it is smooth.

10.2. Let $\mathbf{x}_0$ be any point in $\mathbf{R}^n$. By analogy with the radial contraction onto the origin, write a formula for radial contraction onto the point $\mathbf{x}_0$. Deduce that any open or closed ball centred at $\mathbf{x}_0$ is contractible.

10.3. A subset M of $\mathbf{R}^n$ is *star-shaped* relative to a point $\mathbf{x}_0 \in M$ if for all $\mathbf{x} \in M$ the straight line segment joining $\mathbf{x}_0$ to $\mathbf{x}$ is entirely contained in M. Show that if M is star-shaped relative to $\mathbf{x}_0$, then it is contractible onto $\mathbf{x}_0$. Give an example of a contractible set that is not star-shaped.

10.4. A subset M of $\mathbf{R}^n$ is *convex* if for all $\mathbf{x}$ and $\mathbf{y}$ in M the straight line segment joining $\mathbf{x}$ to $\mathbf{y}$ is entirely contained in M. Prove the following assertions.

(i) M is convex if and only if it is star-shaped relative to each of its points. Give an example of a star-shaped set that is not convex.
(ii) The closed ball $B(\varepsilon, \mathbf{x})$ of radius ε centred at $\mathbf{x}$ is convex.
(iii) Same for the open ball $B^\circ(\varepsilon, \mathbf{x})$.

10.5. Let $\mathbf{x} \in \mathbf{R}^n$ and let $c_\mathbf{x}$ be the straight line connecting the origin to $\mathbf{x}$. Let α be a 1-form on $\mathbf{R}^n$ and let β be the function defined in Example 10.18. Show that $\beta(\mathbf{x}) = \int_{c_\mathbf{x}} \alpha$.

10.6. Let α be the k-form $f\,dx_I = f\,dx_{i_1}\,dx_{i_2}\cdots dx_{i_k}$ on $\mathbf{R}^n$ and let $\phi\colon \mathbf{R}^n \times [0,1] \to \mathbf{R}^n$ be the radial contraction $\phi(\mathbf{x}, t) = t\mathbf{x}$. Verify that

$$\kappa\phi^*\alpha = \sum_{m=1}^{k} (-1)^{m+1} \biggl(\int_0^1 f(t\mathbf{x}) t^{k-1}\,dt\biggr) x_{i_m}\,dx_{i_1}\,dx_{i_2}\cdots \widehat{dx_{i_m}} \cdots dx_{i_k},$$

and check directly that $d\kappa\phi^*\alpha + \kappa\,d\phi^*\alpha = \alpha$ for $k \geq 1$.

10.7. Let $\alpha = f\,dx\,dy + g\,dz\,dx + h\,dy\,dz$ be a 2-form on $\mathbf{R}^3$ and let $\phi(x, y, z, t) = t(x, y, z)$ be the radial contraction of $\mathbf{R}^3$ onto the origin. Verify that

$$\kappa\phi^*\alpha = \biggl(\int_0^1 f(tx, ty, tz) t\,dt\biggr)(x\,dy - y\,dx) + \biggl(\int_0^1 g(tx, ty, tz) t\,dt\biggr)(z\,dx - x\,dz) + \biggl(\int_0^1 h(tx, ty, tz) t\,dt\biggr)(y\,dz - z\,dy).$$

10.8. Let $\alpha = \sum_I f_I\,dx_I$ be a *closed* k-form whose coefficients f_I are smooth functions defined on $\mathbf{R}^n - \{\mathbf{0}\}$ that are all homogeneous of the same degree $p \neq -k$. Let

$$\beta = \frac{1}{p+k} \sum_I \sum_{l=1}^{k} (-1)^{l+1} x_{i_l} f_I\,dx_{i_1}\,dx_{i_2}\cdots \widehat{dx_{i_l}} \cdots dx_{i_k}.$$

Show that $d\beta = \alpha$. (Use $d\alpha = 0$ and apply the identity proved in Exercise B.5 to each f_I; see also Exercise 2.7.)

10.9. Let M and N be manifolds and ϕ_0, $\phi_1\colon M \to N$ homotopic maps. Show that $\int_c \phi_0^*\alpha = \int_c \phi_1^*\alpha$ for all closed k-chains c in M and all closed k-forms α on N.

10.10. Prove that any two maps ϕ_0 and ϕ_1 from M to N are homotopic if M or N is contractible. (First show that every map $M \to N$ is homotopic to a constant map $\phi(\mathbf{x}) = \mathbf{y}_0$.)

10.11. Let $\mathbf{x}_0 = (2, 0)$ and let M be the twice-punctured plane $\mathbf{R}^2 - \{\mathbf{0}, \mathbf{x}_0\}$. Let c_1, c_2, $c_3\colon [0, 2\pi] \to M$ be the loops defined by $c_1(t) = (\cos t, \sin t)$, $c_2(t) = (2 + \cos t, \sin t)$ and $c_3(t) = (1 + 2\cos t, 2\sin t)$. Show that c_1, c_2 and c_3 are not homotopic. (Construct a 1-form α on M such that the integrals $\int_{c_1} \alpha$, $\int_{c_2} \alpha$ and $\int_{c_3} \alpha$ are distinct.)

10.12. A function $g\colon \mathbf{R} \to \mathbf{R}$ is 2π*-periodic* if $g(x + 2\pi) = g(x)$ for all x.

(i) Let $g\colon \mathbf{R} \to \mathbf{R}$ be a smooth 2π-periodic function and let $\beta = g\,dt$, where t is the coordinate on $\mathbf{R}$. Prove that there is a unique number k such that $\beta - k\,dt = dh$ for some smooth 2π-periodic function h. (To find k, integrate the equation $\beta - k\,dt = dh$ over $[0, 2\pi]$. Then check that this value of k works.)

(ii) Let α be any 1-form on the unit circle S^1 and let μ be the element of arc length of S^1. (You can think of μ as the restriction to S^1 of the angle form.) Prove that there is a unique number k such that $\alpha - k\mu$ is exact. (Use the parametrization $c(t) = (\cos t, \sin t)$ and apply the result of part (i).)

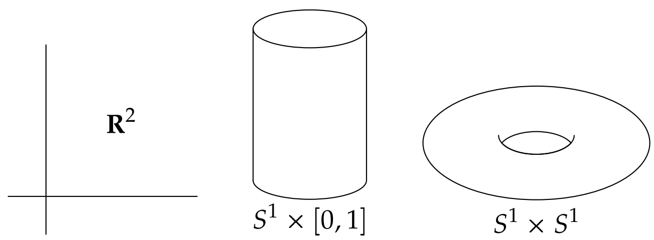
APPENDIX A

Sets and functions

A.1. Glossary

We start with a list of set-theoretical notations that are frequently used in the text. Let X and Y be sets.

$x \in X$: x is an element of X.
$\{a, b, c\}$: the set containing the elements a, b and c.
$X \subseteq Y$: X is a subset of Y, i.e. every element of X is an element of Y.
$X \cap Y$: the intersection of X and Y. This is defined as the set of all x such that $x \in X$ *and* $x \in Y$.
$X \cup Y$: the union of X and Y. This is defined as the set of all x such that $x \in X$ *or* $x \in Y$.
$X - Y$: the complement of Y in X. This is defined as the set of x in X such that x is not in Y.
$X \times Y$: the Cartesian product of X and Y. This is by definition the set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$. Examples: $\mathbf{R} \times \mathbf{R}$ is the Euclidean plane, usually written $\mathbf{R}^2$; $S^1 \times [0,1]$ is a cylinder wall of height 1; $S^1 \times S^1$ is a torus.



$\{\, x \in X \mid P(x) \,\}$: the set of all $x \in X$ which have the property $P(x)$. Examples:

$$\begin{aligned}
\{\, x \in \mathbf{R} \mid 1 \leq x < 3 \,\} &\text{ is the interval } [1,3),\\
\{\, x \mid x \in X \text{ and } x \in Y \,\} &\text{ is the intersection } X \cap Y,\\
\{\, x \mid x \in X \text{ or } x \in Y \,\} &\text{ is the union } X \cup Y,\\
\{\, x \in X \mid x \notin Y \,\} &\text{ is the complement } X - Y.
\end{aligned}$$

$f\colon X \to Y$: f is a function (also called a map) from X to Y. This means that f assigns to each $x \in X$ a unique element $f(x) \in Y$. The set X is called the *domain* or *source* of f, and Y is called the *codomain* or *target* of f.

$f(A)$: the image of a A under the map f. If A is a subset of X, then its image under f is by definition the set

$$f(A) = \{\, y \in Y \mid y = f(x) \text{ for some } x \in A \,\}.$$

$f^{-1}(B)$: the preimage of B under the map f. If B is a subset of Y, this is by definition the set

$$f^{-1}(B) = \{\, x \in X \mid f(x) \in B \,\}.$$

(This is a somewhat confusing notation. It is *not* meant to imply that f is required to have an inverse.)

$f^{-1}(c)$: an abbreviation for $f^{-1}(\{c\})$, i.e. the set $\{\, x \in X \mid f(x) = c \,\}$. This is often called the *fibre* or *level set* of f at c.

$f|A$: the restriction of f to A. If A is a subset of X, $f|A$ is the function defined by

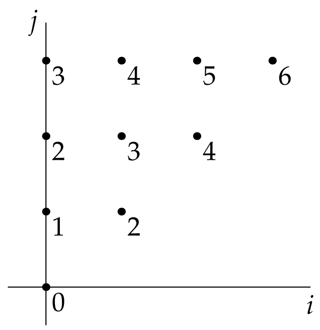
$$(f|A)(x) = \begin{cases} f(x) & \text{if } x \in A, \\ \text{not defined} & \text{if } x \notin A. \end{cases}$$

In other words, $f|A$ is equal to f on A, but "forgets" the values of f at points outside A.

$g \circ f$: the composition of f and g. If $f\colon X \to Y$ and $g\colon Y \to Z$ are functions, then $g \circ f\colon X \to Z$ is defined by $(g \circ f(x) = g(f(x))$. We often say that the function $g \circ f$ is obtained by "substituting $y = f(x)$ into $g(y)$".

A function $f\colon X \to Y$ is *injective* or *one-to-one* if $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$. (Equivalently, f is injective if $f(x_1) = f(x_2)$ implies $x_1 = x_2$.) It is called *surjective* or *onto* if $f(X) = Y$, i.e. if $y \in Y$ then $y = f(x)$ for some $x \in X$. It is called *bijective* if it is both injective and surjective. The function f is bijective if and only if it has a two-sided inverse $f^{-1}\colon Y \to X$ satisfying $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$.

If X is a finite set and $f\colon X \to \mathbf{R}$ a real-valued function, the sum of all the numbers $f(x)$, where x ranges through X, is denoted by $\sum_{x \in X} f(x)$. The set X is called the *index* set for the sum. This notation is often abbreviated or abused in various ways. For instance, if X is the collection $\{1, 2, \ldots, n\}$, one uses the familiar notation $\sum_{i=1}^{n} f(i)$. In these notes we will often deal with indices which are pairs or k-tuples of integers, also known as *multi-indices*. As a simple example, let n be a fixed nonnegative integer, let X be the set of all pairs of integers (i, j) satisfying $0 \leq i \leq j \leq n$, and let $f(i, j) = i + j$. For $n = 3$ we can display X and f in a tableau as follows.



The sum $\sum_{x \in X} f(x)$ of all these numbers is written as

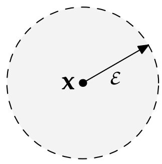
$$\sum_{0 \leq i \leq j \leq n} (i + j).$$

You will be asked to evaluate it explicitly in Exercise A.2.

A.2. General topology of Euclidean space

Let $\mathbf{x}$ be a point in Euclidean space $\mathbf{R}^n$. The *open ball* of radius ε about a point $\mathbf{x}$ is the collection of all points $\mathbf{y}$ whose distance to $\mathbf{x}$ is less than ε,

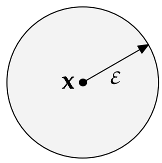
$$B^\circ(\varepsilon, \mathbf{x}) = \{\, \mathbf{y} \in \mathbf{R}^n \mid \|\mathbf{y} - \mathbf{x}\| < \varepsilon \,\}.$$



A subset O of $\mathbf{R}^n$ is *open* if for every $\mathbf{x} \in O$ there exists an $\varepsilon > 0$ such that $B^\circ(\varepsilon, \mathbf{x})$ is contained in O. Intuitively this means that at every point in O there is a little bit of room inside O to move around in any direction you like. An *open neighbourhood* of $\mathbf{x}$ is any open set containing $\mathbf{x}$.

A subset C of $\mathbf{R}^n$ is *closed* if its complement $\mathbf{R}^n - C$ is open. This definition is equivalent to the following: C is closed if and only if for every sequence of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \ldots$ that converges to a point $\mathbf{x}$ in $\mathbf{R}^n$, the limit $\mathbf{x}$ is contained in C. Loosely speaking, closed means "closed under taking limits". An example of a closed set is the *closed ball* of radius ε about a point $\mathbf{x}$, which is defined as the collection of all points $\mathbf{y}$ whose distance to $\mathbf{x}$ is less than or equal to ε,

$$B(\varepsilon, \mathbf{x}) = \{\, \mathbf{y} \in \mathbf{R}^n \mid \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon \,\}.$$



Closed is not the opposite of open! There exist lots of subsets of $\mathbf{R}^n$ that are neither open nor closed, for example the interval $[0, 1)$ in $\mathbf{R}$. (On the other hand, there are not so many subsets that are both open and closed, namely just the empty set and $\mathbf{R}^n$ itself.)

A subset A of $\mathbf{R}^n$ is *bounded* if there exists some $R > 0$ such that $\|x\| \leq R$ for all $\mathbf{x}$ in A. (That is, A is contained in the ball $B(R, \mathbf{0})$ for some value of R.) A *compact* subset of $\mathbf{R}^n$ is one that is both closed and bounded. The importance of the notion of compactness, as far as these notes are concerned, is that the integral of a continuous function over a compact subset of $\mathbf{R}^n$ is always a well-defined, finite number.

Exercises

A.1. Parts (iii) and (iv) of this problem require the use of an atlas (or the Web; see e.g. `http://nationalatlas.gov`). Let X be the surface of the earth, let Y be the real line and let

$f\colon X \to Y$ be the function which assigns to each $x \in X$ its geographical latitude measured in degrees.

(i) Find $f(X)$.
(ii) Find $f^{-1}(0)$, $f^{-1}(90)$, $f^{-1}(-90)$.
(iii) Let A be the contiguous United States. Find $f(A)$. Round the numbers to whole degrees.
(iv) Let $B = f(A)$, where A is as in part (iii). Find (a) a country other than A that is contained in $f^{-1}(B)$; (b) a country that intersects $f^{-1}(B)$ but is not contained in $f^{-1}(B)$; and (c) a country in the northern hemisphere that does not intersect $f^{-1}(B)$.

A.2. Let $S(n) = \sum_{0 \le i \le j \le n}(i + j)$. Prove the following assertions.

(i) $S(0) = 0$ and $S(n+1) = S(n) + \frac{3}{2}(n+1)(n+2)$.
(ii) $S(n) = \frac{1}{2}n(n+1)(n+2)$. (Use induction on n.)

A.3. Prove that the open ball $B^\circ(\varepsilon, \mathbf{x})$ is open. (This is not a tautology! State your reasons as precisely as you can, using the definition of openness stated in the text. You will need the triangle inequality $\|\mathbf{y} - \mathbf{x}\| \le \|\mathbf{y} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{x}\|$.)

A.4. Prove that the closed ball is $B(\varepsilon, \mathbf{x})$ is closed. (Same comments as for Exercise A.3.)

A.5. Show that the two definitions of closedness given in the text are equivalent.

A.6. Complete the following table. Here S^{n-1} denotes the unit sphere about the origin in $\mathbf{R}^n$, that is the set of vectors of length 1.

	closed?	bounded?	compact?
$[-3, 5]$	yes	yes	yes
$[-3, 5)$			
$[-3, \infty)$			
$(-3, \infty)$			
$B(\varepsilon, \mathbf{x})$			
$B^\circ(\varepsilon, \mathbf{x})$			
S^{n-1}			
xy-plane in $\mathbf{R}^3$			
unit cube $[0, 1]^n$			

APPENDIX B

Calculus review

This appendix is a brief review of some single- and multi-variable calculus needed in the study of manifolds. References for this material are **[Edw94]**, **[HH02]** and **[MT03]**.

B.1. The fundamental theorem of calculus

Suppose that F is a differentiable function of a single variable x and that the derivative $f = F'$ is continuous. Let $[a, b]$ be an interval contained in the domain of F. The fundamental theorem of calculus says that

$$\boxed{\int_a^b f(t)\,dt = F(b) - F(a).} \tag{B.1}$$

There are two useful alternative ways of writing this theorem. Replacing b with x and differentiating with respect to x we find

$$\boxed{\frac{d}{dx}\int_a^x f(t)\,dt = f(x).} \tag{B.2}$$

Writing g instead of F and g' instead of f and adding $g(a)$ to both sides in formula (B.1) we get

$$\boxed{g(x) = g(a) + \int_a^x g'(t)\,dt.} \tag{B.3}$$

Formulas (B.1)–(B.3) are equivalent, but they emphasize different aspects of the fundamental theorem of calculus. Formula (B.1) is a formula for a definite integral: it tells you how to find the (signed) surface area between the graph of the function f and the x-axis. Formula (B.2) says that the integral of a continuous function is a differentiable function of the upper limit; and the derivative is the integrand. Formula (B.3) is an "integral formula", which expresses the function g in terms of the value $g(a)$ and the derivative g'. (See Exercise B.1 for an application.)

B.2. Derivatives

Let $\phi_1, \phi_2, \dots, \phi_m$ be functions of n variables $x_1, x_2, \dots, x_n$. As usual we write

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \phi(\mathbf{x}) = \begin{pmatrix} \phi_1(\mathbf{x}) \\ \phi_2(\mathbf{x}) \\ \vdots \\ \phi_m(\mathbf{x}) \end{pmatrix},$$

and view $\phi(\mathbf{x})$ as a single map from $\mathbf{R}^n$ to $\mathbf{R}^m$. (In calculus the word "map" is often used for vector-valued functions, while the word "function" is generally reserved for real-valued functions.) We say that ϕ is *continuously differentiable* if the partial derivatives

$$\frac{\partial \phi_i}{\partial x_j}(\mathbf{x}) = \lim_{h \to 0} \frac{\phi_i(\mathbf{x} + h\mathbf{e}_j) - \phi_i(\mathbf{x})}{h} \tag{B.4}$$

are well-defined and continuous functions of $\mathbf{x}$ for all $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Here

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ldots, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

are the standard basis vectors of $\mathbf{R}^n$. The *(total) derivative* or *Jacobi matrix* of ϕ at $\mathbf{x}$ is then the $m \times n$-matrix

$$D\phi(\mathbf{x}) = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(\mathbf{x}) & \ldots & \frac{\partial \phi_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial \phi_m}{\partial x_1}(\mathbf{x}) & \ldots & \frac{\partial \phi_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

If $\mathbf{v}$ is any vector in $\mathbf{R}^n$, the *directional derivative* of ϕ along $\mathbf{v}$ is defined to be the vector $D\phi(\mathbf{x})\mathbf{v}$ in $\mathbf{R}^m$, obtained by multiplying the matrix $D\phi(\mathbf{x})$ by the vector $\mathbf{v}$.

For $n = 1$ ϕ is a vector-valued function of one variable x, often called a *path* or *(parametrized) curve* in $\mathbf{R}^m$. In this case the matrix $D\phi(x)$ consists of a single column vector, called the *velocity vector*, and is usually denoted simply by $\phi'(x)$.

For $m = 1$ ϕ is a scalar-valued function of n variables and $D\phi(\mathbf{x})$ is a single row vector. The transpose matrix of $D\phi(\mathbf{x})$ is therefore a column vector, usually called the *gradient* of ϕ:

$$D\phi(\mathbf{x})^T = \operatorname{grad} \phi(\mathbf{x}).$$

The directional derivative of ϕ along $\mathbf{v}$ can then be written as an inner product, $D\phi(\mathbf{x})\mathbf{v} = \operatorname{grad} \phi(\mathbf{x}) \cdot \mathbf{v}$. There is an important characterization of the gradient, which is based on the identity $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos\theta$. Here $0 \leq \theta \leq \pi$ is the angle subtended by $\mathbf{a}$ and $\mathbf{b}$. If $\mathbf{v}$ is a unit vector ($\|\mathbf{v}\| = 1$), then

$$D\phi(\mathbf{x})\mathbf{v} = \operatorname{grad} \phi(\mathbf{x}) \cdot \mathbf{v} = \|\operatorname{grad} \phi(\mathbf{x})\| \cos\theta,$$

where θ is the angle between $\operatorname{grad} \phi(\mathbf{x})$ and $\mathbf{v}$. So $D\phi(\mathbf{x})\mathbf{v}$ takes on its maximal value if $\cos\theta = 1$, i.e. $\theta = 0$. This means that $\mathbf{v}$ points in the same direction as $\operatorname{grad} \phi(\mathbf{x})$. Thus the *direction* of the vector $\operatorname{grad} \phi(\mathbf{x})$ is the direction of *steepest ascent*, i.e. in which ϕ increases fastest, and the *magnitude* of $\operatorname{grad} \phi(\mathbf{x})$ is equal to the directional derivative $D\phi(\mathbf{x})\mathbf{v}$, where $\mathbf{v}$ is the unit vector pointing along $\operatorname{grad} \phi(\mathbf{x})$.

Frequently a function is not defined on all of $\mathbf{R}^n$, but only on a subset U. We must be a little careful in defining the derivative of such a function. Let us assume that U is an *open* set. Let $\phi\colon U \to \mathbf{R}^m$ be a function defined on U and let $\mathbf{x} \in U$. Because U is open, there exists $\varepsilon > 0$ such that the points $\mathbf{x} + t\mathbf{e}_j$ are contained in U for $-\varepsilon < t < \varepsilon$. Therefore $\phi_i(\mathbf{x} + t\mathbf{e}_j)$ is well-defined for $-\varepsilon < t < \varepsilon$ and thus it makes sense to ask whether the partial derivatives (B.4) exist. If they do, for all $\mathbf{x} \in U$ and all i and j, and if they are continuous, the function ϕ is called *continuously differentiable* or C^1.

If the second partial derivatives

$$\frac{\partial^2 \phi_i}{\partial x_j \partial x_k}(\mathbf{x})$$

exist and are continuous for all $\mathbf{x} \in U$ and for all $i = 1, 2, \dots, n$ and $j, k = 1, 2, \dots, m$, then ϕ is called *twice continuously differentiable* or C^2. Likewise, if all r-fold partial derivatives

$$\frac{\partial^r \phi_i}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_r}}(\mathbf{x})$$

exist and are continuous, then ϕ is *r times continuously differentiable* or C^r. If ϕ is C^r for all $r \geq 1$, then we say that ϕ is *infinitely many times differentiable*, C^∞, or *smooth*. This means that ϕ can be differentiated arbitrarily many times with respect to any of the variables.

Let us now review some of the most important facts concerning derivatives.

B.3. The chain rule

Recall that if A, B and C are sets and $\phi\colon A \to B$ and $\psi\colon B \to C$ are functions, we can apply ψ after ϕ to obtain the composite function $(\psi \circ \phi)(x) = \psi(\phi(x))$.

B.1. THEOREM (chain rule). *Let $U \subseteq \mathbf{R}^n$ and $V \subseteq \mathbf{R}^m$ be open and let $\phi\colon U \to V$ and $\psi\colon V \to \mathbf{R}^k$ be C^r. Then $\psi \circ \phi$ is C^r and*

$$\boxed{D(\psi \circ \phi)(\mathbf{x}) = D\psi(\phi(\mathbf{x}))D\phi(\mathbf{x})}$$

for all $\mathbf{x} \in U$.

Here $D\psi(\phi(\mathbf{x}))D\phi(\mathbf{x})$ denotes the composition or the product of the two matrices $D\psi(\phi(\mathbf{x}))$ and $D\phi(\mathbf{x})$.

B.2. EXAMPLE. In the one-variable case $n = m = k = 1$ the derivatives $D\phi$ and $D\psi$ are 1×1-matrices $(\phi'(x))$ and $(\psi'(y))$, and matrix multiplication is ordinary multiplication, so we get the usual chain rule

$$(\psi \circ \phi)'(x) = \psi'(\phi(x))\phi'(x).$$

B.3. EXAMPLE. If $n = k = 1$, then $\psi \circ \phi$ is a real-valued function of one variable x, so $D(\psi \circ \phi)$ is a 1×1-matrix containing the single entry $(\psi \circ \phi)'$. Moreover,

$$D\phi(x) = \begin{pmatrix} \frac{d\phi_1}{dx}(x) \\ \vdots \\ \frac{d\phi_m}{dx}(x) \end{pmatrix} \qquad \text{and} \qquad D\psi(\mathbf{y}) = \begin{pmatrix} \frac{\partial\psi}{\partial y_1}(\mathbf{y}) & \dots & \frac{\partial\psi}{\partial y_m}(\mathbf{y}) \end{pmatrix},$$

so by the chain rule

$$\frac{d(\psi \circ \phi)}{dx}(x) = D\psi(\phi(x))D\phi(x) = \sum_{i=1}^m \frac{\partial\psi}{\partial y_i}(\phi(x))\frac{d\phi_i}{dx}(x). \tag{B.5}$$

This is perhaps the most important special case of the chain rule. Sometimes we are sloppy and abbreviate this identity to

$$\frac{d(\psi \circ \phi)}{dx} = \sum_{i=1}^m \frac{\partial\psi}{\partial y_i}\frac{d\phi_i}{dx}.$$

An even sloppier, but nevertheless quite common, notation is

$$\frac{d\psi}{dx} = \sum_{i=1}^{m} \frac{\partial \psi}{\partial y_i} \frac{d\phi_i}{dx}.$$

In these notes we frequently use the so-called "pullback" notation. Instead of $\psi \circ \phi$ we often write $\phi^*\psi$, so that $\phi^*\psi(x)$ stands for $\psi(\phi(x))$. Similarly, $\phi^*(\partial\psi/\partial y_i)(x)$ stands for $\partial\psi/\partial y_i(\phi(x))$. In this notation we have

$$\frac{d\phi^*\psi}{dx} = \sum_{i=1}^{m} \phi^*\left(\frac{\partial \psi}{\partial y_i}\right) \frac{d\phi_i}{dx}. \tag{B.6}$$

B.4. The implicit function theorem

Let $\phi\colon W \to \mathbf{R}^m$ be a continuously differentiable function defined on an open subset W of $\mathbf{R}^{n+m}$. Let us think of a vector in $\mathbf{R}^{n+m}$ as an ordered pair of vectors $(\mathbf{u}, \mathbf{v})$ with $\mathbf{u} \in \mathbf{R}^n$ and $\mathbf{v} \in \mathbf{R}^m$. Consider the equation

$$\phi(\mathbf{u}, \mathbf{v}) = \mathbf{0}.$$

Under what circumstances is it possible to solve for $\mathbf{v}$ as a function of $\mathbf{u}$? The answer is given by the implicit function theorem. We form the Jacobi matrices of ϕ with respect to the $\mathbf{u}$- and $\mathbf{v}$-variables separately,

$$D_{\mathbf{u}}\phi = \begin{pmatrix} \frac{\partial \phi_1}{\partial u_1} & \cdots & \frac{\partial \phi_1}{\partial u_n} \\ \vdots & & \vdots \\ \frac{\partial \phi_m}{\partial u_1} & \cdots & \frac{\partial \phi_m}{\partial u_n} \end{pmatrix}, \qquad D_{\mathbf{v}}\phi = \begin{pmatrix} \frac{\partial \phi_1}{\partial v_1} & \cdots & \frac{\partial \phi_1}{\partial v_m} \\ \vdots & & \vdots \\ \frac{\partial \phi_m}{\partial v_1} & \cdots & \frac{\partial \phi_m}{\partial v_m} \end{pmatrix}.$$

Observe that the matrix $D_{\mathbf{v}}\phi$ is square. We are in business if we have a point $(\mathbf{u}_0, \mathbf{v}_0)$ at which ϕ is $\mathbf{0}$ and $D_{\mathbf{v}}\phi$ is invertible.

B.4. THEOREM (implicit function theorem). *Let $\phi\colon W \to \mathbf{R}^m$ be C^r, where W is open in $\mathbf{R}^{n+m}$. Suppose that $(\mathbf{u}_0, \mathbf{v}_0) \in W$ is a point such that $\phi(\mathbf{u}_0, \mathbf{v}_0) = \mathbf{0}$ and $D_{\mathbf{v}}\phi(\mathbf{u}_0, \mathbf{v}_0)$ is invertible. Then there are open neighbourhoods $U \subseteq \mathbf{R}^n$ of $\mathbf{u}_0$ and $V \subseteq \mathbf{R}^m$ of $\mathbf{v}_0$ such that for each $\mathbf{u} \in U$ there exists a unique $\mathbf{v} = f(\mathbf{u}) \in V$ satisfying $\phi(\mathbf{u}, f(\mathbf{u})) = \mathbf{0}$. The function $f\colon U \to V$ is C^r with derivative given by implicit differentiation:*

$$\boxed{Df(\mathbf{u}) = -D_{\mathbf{v}}\phi(\mathbf{u}, \mathbf{v})^{-1} D_{\mathbf{u}}\phi(\mathbf{u}, \mathbf{v})\big|_{\mathbf{v}=f(\mathbf{u})}}$$

for all $\mathbf{u} \in U$.

This is well-known for $m = n = 1$, when ϕ is a function of two real variables (u, v). If $\partial\phi/\partial v \neq 0$ at a certain point (u_0, v_0), then for u close to u_0 and v close to v_0 we can solve the equation $\phi(u, v) = 0$ for v as a function $v = f(u)$ of u, and

$$f' = -\frac{\partial\phi/\partial u}{\partial\phi/\partial v}.$$

Now let us take ϕ to be of the form $\phi(\mathbf{u}, \mathbf{v}) = g(\mathbf{v}) - \mathbf{u}$, where $g\colon W \to \mathbf{R}^n$ is a given function with W open in $\mathbf{R}^n$. Solving $\phi(\mathbf{u}, \mathbf{v}) = \mathbf{0}$ here amounts to *inverting* the function g. Moreover, $D_{\mathbf{v}}\phi = Dg$, so the implicit function theorem yields the following result.

B.5. THEOREM (inverse function theorem). *Let $g\colon W \to \mathbf{R}^n$ be continuously differentiable, where W is open in $\mathbf{R}^n$. Suppose that $\mathbf{v}_0 \in W$ is a point such that $Dg(\mathbf{v}_0)$ is invertible. Then there is an open neighbourhood $U \subseteq \mathbf{R}^n$ of $\mathbf{v}_0$ such that $g(U)$ is an open neighbourhood of $g(\mathbf{v}_0)$ and the function $g\colon U \to g(U)$ is invertible. The inverse $g^{-1}\colon V \to U$ is continuously differentiable with derivative given by*

$$\boxed{Dg^{-1}(\mathbf{u}) = Dg(\mathbf{v})^{-1}\big|_{\mathbf{v}=g^{-1}(\mathbf{u})}}$$

for all $\mathbf{v} \in V$.

Again let us spell out the one-variable case $n = 1$. Invertibility of $Dg(v_0)$ simply means that $g'(v_0) \neq 0$. This implies that near v_0 the function g is strictly monotone increasing (if $g'(v_0) > 0$) or decreasing (if $g'(v_0) < 0$). Therefore if I is a sufficiently small open interval around u_0, then $g(I)$ is an open interval around $g(u_0)$ and the restricted function $g\colon I \to g(I)$ is invertible. The inverse function has derivative

$$(g^{-1})'(u) = \frac{1}{g'(v)},$$

with $v = g^{-1}(u)$.

B.6. EXAMPLE (square roots). Let $g(v) = v^2$. Then $g'(v_0) \neq 0$ whenever $v_0 \neq 0$. For $v_0 > 0$ we can take $I = (0, \infty)$. Then $g(I) = (0, \infty)$, $g^{-1}(u) = \sqrt{u}$, and $(g^{-1})'(u) = 1/(2\sqrt{u})$. For $v_0 < 0$ we can take $I = (-\infty, 0)$. Then $g(I) = (0, \infty)$, $g^{-1}(u) = -\sqrt{u}$, and $(g^{-1})'(u) = -1/(2\sqrt{u})$. In a neighbourhood of 0 it is not possible to invert g.

B.5. The substitution formula for integrals

Let V be an open subset of $\mathbf{R}^n$ and let $f\colon V \to \mathbf{R}$ be a function. Suppose we want to change the variables in the integral $\int_V f(\mathbf{y})\,d\mathbf{y}$. (This is shorthand for an n-fold integral over $y_1, y_2, \ldots, y_n$.) This means we substitute $\mathbf{y} = p(\mathbf{x})$, where $p\colon U \to V$ is a map from an open $U \subseteq \mathbf{R}^n$ to V. Under a suitable hypothesis we can change the integral over $\mathbf{y}$ to an integral over $\mathbf{x}$.

B.7. THEOREM (change of variables formula). *Let U and V be open subsets of $\mathbf{R}^n$ and let $p\colon U \to V$ be a map. Suppose that p is bijective and that p and its inverse are continuously differentiable. Then for any integrable function f we have*

$$\boxed{\int_V f(\mathbf{y})\,d\mathbf{y} = \int_U f(p(\mathbf{x}))|\det Dp(\mathbf{x})|\,d\mathbf{x}.}$$

Again this should look familiar from one-variable calculus: if $p\colon (a,b) \to (c,d)$ is C^1 and has a C^1 inverse, then

$$\int_c^d f(y)\,dy = \begin{cases} \int_a^b f(p(x))p'(x)\,dx & \text{if } p \text{ is increasing,} \\ -\int_a^b f(p(x))p'(x)\,dx & \text{if } p \text{ is decreasing.} \end{cases}$$

This can be written succinctly as $\int_c^d f(y)\,dy = \int_a^b f(p(x))|p'(x)|\,dx$, which looks more similar to the multidimensional case.

Exercises

B.1. Let $g\colon [a,b] \to \mathbf{R}$ be a C^{n+1}-function, where $n \geq 0$. Suppose $a \leq x \leq b$ and put $h = x - a$.

(i) By changing variables in the fundamental theorem of calculus (B.3) show that
$$g(x) = g(a) + h \int_0^1 g'(a+th)\,dt.$$

(ii) Show that
$$\begin{aligned} g(x) &= g(a) + h(t-1)g'(a+th)\big|_0^1 + h^2 \int_0^1 (1-t)g''(a+th)\,dt \\ &= g(a) + hg'(a) + h^2 \int_0^1 (1-t)g''(a+th)\,dt. \end{aligned}$$
(Integrate the formula in part (i) by parts and don't forget to use the chain rule.)

(iii) By induction on n deduce from part (ii) that
$$g(x) = \sum_{k=0}^{n} \frac{g^{(k)}(a)}{k!}h^k + \frac{h^{n+1}}{n!}\int_0^1 (1-t)^n g^{(n+1)}(a+th)\,dt.$$
This is *Taylor's formula with integral remainder term*.

B.2. Let $\mathbf{x}$ and $\mathbf{v}$ be constant vectors in $\mathbf{R}^n$. Define $c(t) = \mathbf{x} + t\mathbf{v}$. Find $c'(t)$.

B.3. Deduce from the chain rule that $D\phi(\mathbf{x})\mathbf{v} = \lim_{t\to 0} \dfrac{\phi(\mathbf{x}+t\mathbf{v}) - \phi(\mathbf{x})}{t}$.

B.4. According to Newton's law of gravitation, a particle of mass m_1 placed at the origin in $\mathbf{R}^3$ exerts a force on a particle of mass m_2 placed at $\mathbf{x} \in \mathbf{R}^3 - \{\mathbf{0}\}$ equal to
$$\mathbf{F} = -\frac{Gm_1m_2}{\|\mathbf{x}\|^3}\mathbf{x},$$
where G is a constant of nature. Show that $\mathbf{F}$ is the gradient of $f(\mathbf{x}) = Gm_1m_2/\|\mathbf{x}\|$.

B.5. A function $f\colon \mathbf{R}^n - \{\mathbf{0}\} \to \mathbf{R}$ is *homogeneous* of degree p if $f(t\mathbf{x}) = t^p f(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{R}^n - \{\mathbf{0}\}$ and $t > 0$. Here p is a real constant.

(i) Show that the functions $f(x,y) = (x^2 - xy)/(x^2+y^2)$, $f(x,y) = \sqrt{x^3+y^3}$, $f(x,y,z) = (x^2z^6 + 3x^4y^2z^2)^{-\sqrt{2}}$ are homogeneous. What are their degrees?

(ii) Assume that f is defined at $\mathbf{0}$ and continuous everywhere. Show that $p \geq 0$. Show that f is constant if $p = 0$.

(iii) Show that if f is homogeneous of degree p and smooth, then
$$\sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i}(x) = pf(\mathbf{x}).$$
(Differentiate the relation $f(t\mathbf{x}) = t^p f(\mathbf{x})$ with respect to t.)

B.6. Define a function $f\colon \mathbf{R} \to \mathbf{R}$ by $f(0) = 0$ and $f(x) = e^{-1/x^2}$ for $x \neq 0$.

(i) Show that f is differentiable at 0 and that $f'(0) = 0$.

(ii) Show that f is smooth and that $f^{(n)}(0) = 0$ for all n.

(iii) Plot the function f over the interval $-5 \leq x \leq 5$. Using software or a graphing calculator is fine, but pay special attention to the behaviour near $x = 0$.

B.7. Define a map ψ from $\mathbf{R}^{n-1}$ to $\mathbf{R}^n$ by
$$\psi(\mathbf{t}) = \frac{1}{\|\mathbf{t}\|^2+1}\big(2\mathbf{t} + (\|\mathbf{t}\|^2 - 1)\mathbf{e}_n\big).$$

(i) Show that $\psi(\mathbf{t})$ lies on the unit sphere S^{n-1} about the origin.

(ii) Show that $\psi(\mathbf{t})$ is the intersection point of the sphere and the line through the points $\mathbf{e}_n$ and $\mathbf{t}$. (Here we regard $\mathbf{t} = (t_1, t_2, \dots, t_{n-1})$ as a point in $\mathbf{R}^n$ by identifying it with $(t_1, t_2, \dots, t_{n-1}, 0)$.)

(iii) Compute $D\psi(\mathbf{t})$.

(iv) Let X be the sphere punctured at the "north pole", $X = S^{n-1} - \{\mathbf{e}_n\}$. *Stereographic projection from the north pole* is the map $\phi\colon X \to \mathbf{R}^{n-1}$ given by $\phi(\mathbf{x}) = (x_n - 1)^{-1}(x_1, x_2, \dots, x_{n-1})$. Show that ϕ is a two-sided inverse of ψ.

(v) Draw diagrams illustrating the maps ϕ and ψ for $n = 2$ and $n = 3$.

(vi) Now let $\mathbf{y}$ be any point on the sphere and let P the hyperplane which passes through the origin and is perpendicular to $\mathbf{y}$. The *stereographic projection from* $\mathbf{y}$ of any point $\mathbf{x}$ in the sphere distinct from $\mathbf{y}$ is defined as the unique intersection point of the line joining $\mathbf{y}$ to $\mathbf{x}$ and the hyperplane P. This defines a map $\phi\colon S^{n-1} - \{\mathbf{y}\} \to P$. The point $\mathbf{y}$ is called the *centre* of the projection. Write a formula for the stereographic projection ϕ from the south pole $-\mathbf{e}_n$ and for its inverse $\psi\colon \mathbf{R}^{n-1} \to S^{n-1}$.

B.8. A map $\phi\colon \mathbf{R}^n \to \mathbf{R}^m$ is called *even* if $\phi(-\mathbf{x}) = \phi(\mathbf{x})$ for all $\mathbf{x}$ in $\mathbf{R}^n$. Find $D\phi(\mathbf{0})$ if ϕ is even and C^1.

B.9. Let $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ be vectors in $\mathbf{R}^n$. A linear combination $\sum_{i=0}^n c_i\mathbf{a}_i$ is *convex* if $\sum_{i=0}^n c_i = 1$. The *simplex* Δ spanned by the $\mathbf{a}_i$'s is the collection of all their convex linear combinations,

$$\Delta = \left\{ \sum_{i=0}^n c_i\mathbf{a}_i \;\middle|\; \sum_{i=0}^n c_i = 1 \right\}.$$

The *standard simplex* in $\mathbf{R}^n$ is the simplex spanned by the vectors $\mathbf{0}, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$.

(i) For $n = 1, 2, 3$ draw pictures of the standard n-simplex as well as a nonstandard n-simplex.

(ii) The *volume* of a region R in $\mathbf{R}^n$ is defined as $\int_R dx_1\,dx_2 \cdots dx_n$. Show that

$$\operatorname{vol}\Delta = \frac{1}{n!}|\det A|,$$

where A is the $n \times n$-matrix with columns $\mathbf{a}_1 - \mathbf{a}_0, \mathbf{a}_2 - \mathbf{a}_0, \dots, \mathbf{a}_n - \mathbf{a}_0$. (First compute the volume of the standard simplex by repeated integration. Then map Δ to the standard simplex by an appropriate substitution and apply the substitution formula for integrals.)

The following two calculus problems are not review problems, but the results are needed in Chapter 9.

B.10. For $x > 0$ define

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt$$

and prove the following assertions.

(i) $\Gamma(x+1) = x\,\Gamma(x)$ for all $x > 0$.

(ii) $\Gamma(n) = (n-1)!$ for positive integers n.

(iii) $\displaystyle\int_0^\infty e^{-u^2}u^a\,du = \frac{1}{2}\Gamma\Big(\frac{a+1}{2}\Big)$.

B.11. Calculate $\Gamma(n + \frac{1}{2})$ (where Γ is the function defined in Exercise B.10) by establishing the following identities. For brevity write $\gamma = \Gamma(\frac{1}{2})$.

(i) $\displaystyle\gamma = \int_{-\infty}^\infty e^{-s^2}\,ds$.

(ii) $\displaystyle\gamma^2 = \int_{-\infty}^\infty\int_{-\infty}^\infty e^{-x^2-y^2}\,dx\,dy$.

(iii) $\gamma^2 = \int_0^{2\pi} \int_0^{\infty} re^{-r^2}\, dr\, d\theta$.

(iv) $\gamma = \sqrt{\pi}$.

(v) $\Gamma\left(n + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi}$ for $n \geq 1$.

Bibliography

[Bac06] D. Bachman, *A geometric approach to differential forms*, Birkhäuser, Boston, MA, 2006.
A recent text, a version of which is available through the author's website `http://pzacad.pitzer.edu/~dbachman/`.

[BS91] P. Bamberg and S. Sternberg, *A course in mathematics for students of physics*, Cambridge University Press, Cambridge, 1991.

[BT82] R. Bott and L. Tu, *Differential forms in algebraic topology*, Springer-Verlag, New York, 1982.
Masterly exposition at beginning graduate level of the uses of differential forms in topology and de Rham cohomology.

[Bre91] D. Bressoud, *Second year calculus*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1991.
Multivariable calculus from the point of view of Newton's law and special relativity with coverage of differential forms.

[Bre05] O. Bretscher, *Linear algebra with applications*, third ed., Pearson Prentice Hall, Upper Saddle River, New Jersey, 2005.
Good reference for the basic linear algebra required in these notes.

[Car94] M. do Carmo, *Differential forms and applications*, Springer-Verlag, Berlin, 1994, translated from the 1971 Portuguese original.
Slightly more advanced than these notes, with some coverage of Riemannian geometry, including the Gauß-Bonnet theorem.

[Dar94] R. Darling, *Differential forms and connections*, Cambridge University Press, Cambridge, 1994.
Advanced undergraduate text on manifolds and differential forms, including expositions of Maxwell theory and its modern generalization, gauge theory.

[Edw94] H. Edwards, *Advanced calculus: A differential forms approach*, Birkhäuser Boston, Boston, Massachusetts, 1994, corrected reprint of the 1969 original.
One of the earliest undergraduate textbooks covering differential forms. Still recommended as an alternative or supplementary source.

[Fla89] H. Flanders, *Differential forms with applications to the physical sciences*, second ed., Dover Publications, New York, 1989.
Written for 1960s engineering graduate students. Concise but lucid (and cheap!).

[Gra98] A. Gray, *Modern differential geometry of curves and surfaces with Mathematica*, second ed., CRC Press, Boca Raton, FL, 1998.
Leisurely and thorough exposition at intermediate undergraduate level with plenty of computer graphics.

[GP74] V. Guillemin and A. Pollack, *Differential topology*, Prentice-Hall, Englewood Cliffs, N.J., 1974.
Written for beginning graduate students, but very intuitive and with lots of interesting applications to topology.

[HH02] J. Hubbard and B. Hubbard, *Vector calculus, linear algebra, and differential forms: A unified approach*, second ed., Prentice Hall, Upper Saddle River, New Jersey, 2002.

[MT03] J. Marsden and A. Tromba, *Vector calculus*, fifth ed., W. H. Freeman, New York, 2003.
Standard multivariable calculus reference.

[Opr03] J. Oprea, *Differential geometry and its applications*, second ed., Prentice Hall, 2003.
Curves, surfaces, geodesics and calculus of variations with plenty of MAPLE programming. Accessible to intermediate undergraduates.

[Sin01] S. Singer, *Symmetry in mechanics. A gentle, modern introduction*, Birkhäuser, Boston, MA, 2001.

[Spi65] M. Spivak, *Calculus on manifolds. A modern approach to classical theorems of advanced calculus*, W. A. Benjamin, New York-Amsterdam, 1965.
Efficient and rigorous treatment of many of the topics in these notes.

[Spi99] ———, *A comprehensive introduction to differential geometry*, third ed., Publish or Perish, Houston, TX, 1999.
Differential geometry textbook at advanced undergraduate level in five massive but fun to read volumes.

[Wei97] S. Weintraub, *Differential forms. A complement to vector calculus*, Academic Press, San Diego, CA, 1997.
Written as a companion to multivariable calculus texts. Contains careful and intuitive explanations of several of the ideas covered in these notes, as well as a number of straightforward exercises.

The Greek alphabet

upper case	lower case	name
A	α	alpha
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ϵ, ε	epsilon
Z	ζ	zeta
H	η	eta
Θ	θ, ϑ	theta
I	ι	iota
K	κ	kappa
Λ	λ	lambda
M	μ	mu
N	ν	nu
Ξ	ξ	xi
O	o	omicron
Π	π, ϖ	pi
P	ρ	rho
Σ	σ	sigma
T	τ	tau
Υ	υ	upsilon
Φ	ϕ, φ	phi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Notation Index

Index

Page numbers in **boldface** refer to definitions or theorems; *italic* page numbers refer to examples or applications. In a few cases ***italic boldface*** is in order.