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

Revised edition, 2017

Contents

Preface

These are the lecture notes for Math 3210 (formerly named 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 [**Spi71**] 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.

These notes, occasionally revised and updated, are available at

`http://www.math.cornell.edu/~sjamaar/manifolds/`.

Corrections, suggestions and comments sent to `sjamaar@math.cornell.edu` will be received gratefully.

Ithaca, New York, December 2017

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 $(x_1 \quad x_2 \quad \cdots \quad x_n)$. Nevertheless, to save space we shall frequently write a column vector $\mathbf{x}$ as an n-tuple

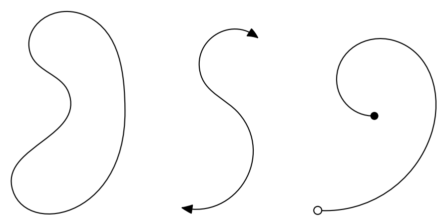
$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

with the entries separated by commas.

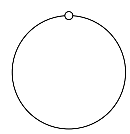
A *manifold* is a certain type of subset of $\mathbf{R}^n$. A precise definition will follow in Chapter 6, but one important consequence of the definition is that at each of its points a manifold has a well-defined tangent space, which is a linear subspace of $\mathbf{R}^n$. This fact enables us to apply the methods of calculus and linear algebra to the study of manifolds. The *dimension* of a manifold is by definition the dimension of any 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*.

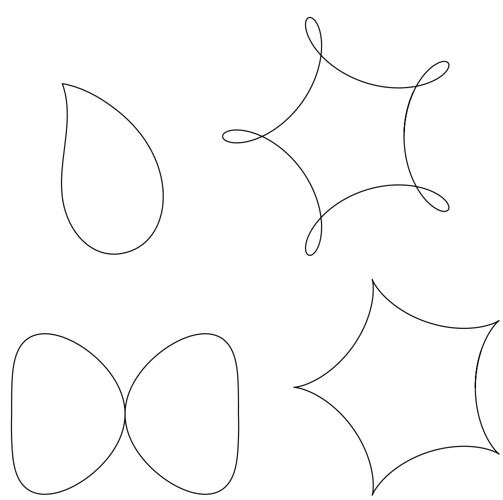


A circle with one point deleted is also an example of a manifold. Think of a torn elastic band.



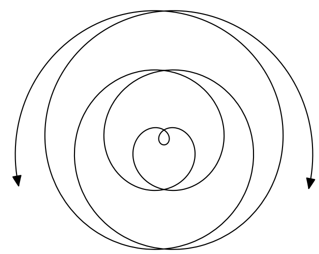
By straightening out the elastic band we see that this manifold is really the same as an open interval.

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*.

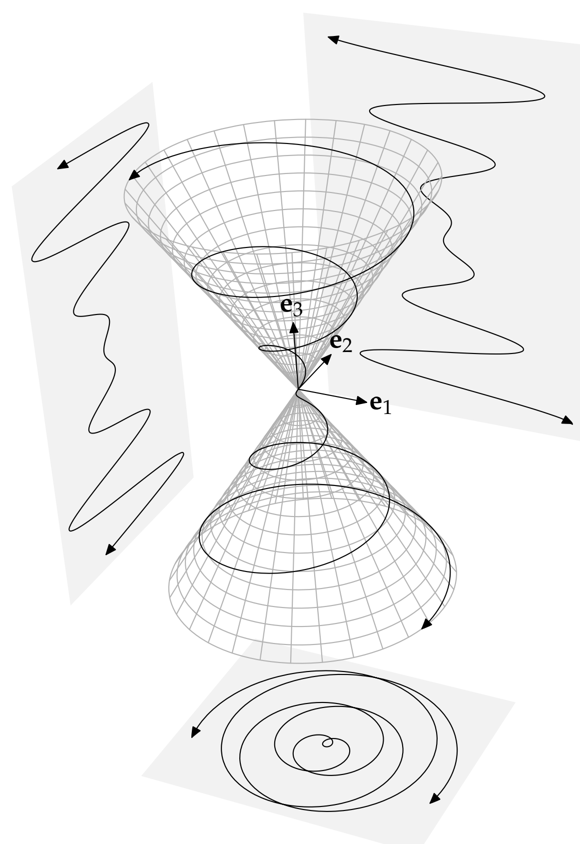


Singularities can sometimes be "resolved". For instance, the self-intersections of the Archimedean spiral, which is given in polar coordinates by r is a constant times

θ, where r is allowed to be negative,

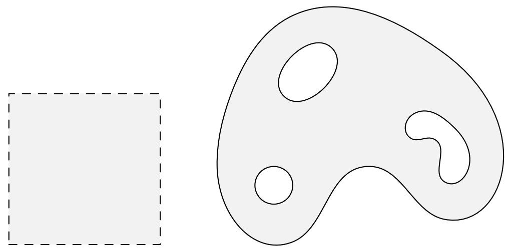


can be removed by uncoiling the spiral and wrapping it around a cone. You can convince yourself that the resulting space curve has no singularities by peeking at it along the direction of the x-axis or the y-axis. What you will see are the smooth curves shown in the (y, z)-plane and the (x, z)-plane.

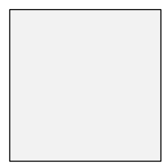


Singularities are very interesting, but in this course we shall focus on gaining a thorough understanding of the smooth points.

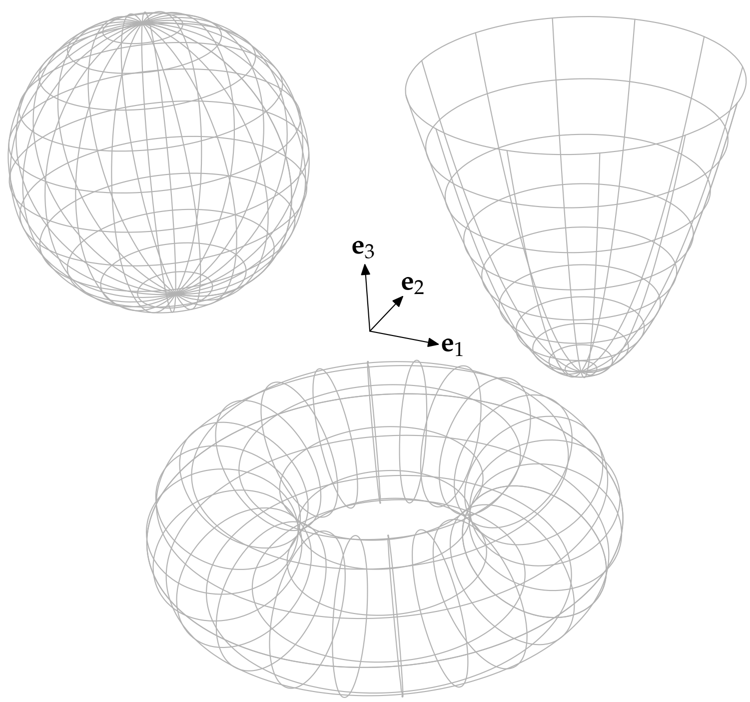
Dimension 2. A two-dimensional manifold is a smooth surface without self-intersections. It may have a boundary, which is always a one-dimensional manifold. You can have two-dimensional manifolds in the plane $\mathbf{R}^2$, but they are relatively boring. Examples are: an arbitrary open subset of $\mathbf{R}^2$,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



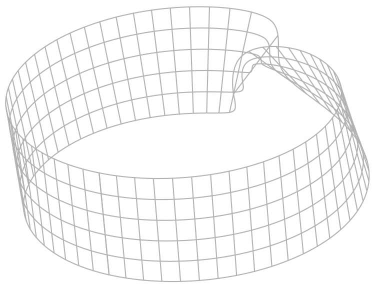
Two-dimensional manifolds in three-dimensional space include a sphere (the surface of a ball), a paraboloid and a torus (the surface of a doughnut).

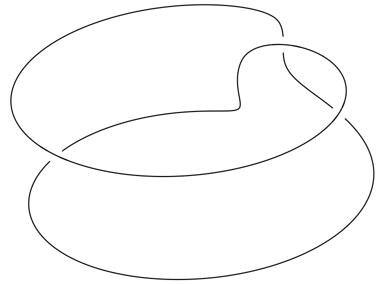


The famous *Möbius band* is made by pasting together the two ends of a rectangular strip of paper giving one end a half twist. The boundary of the band consists of two boundary edges of the rectangle tied together and is therefore a single closed

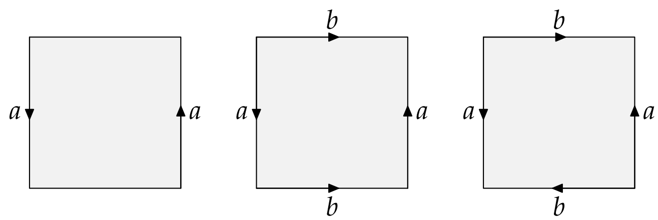
[1]To be strictly accurate, the closed square is a *topological* manifold with boundary, but not a *smooth* manifold with boundary. In these notes we will consider only smooth manifolds.

curve.



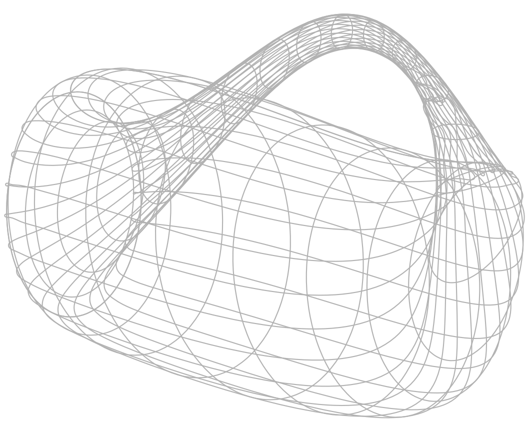


Out of the Möbius band we can create a manifold without boundary by closing it up along the boundary edge. This can be done in two different ways. According to the direction in which we glue the edge to itself, we obtain the *Klein bottle* or the *projective plane*. A simple way to represent these three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

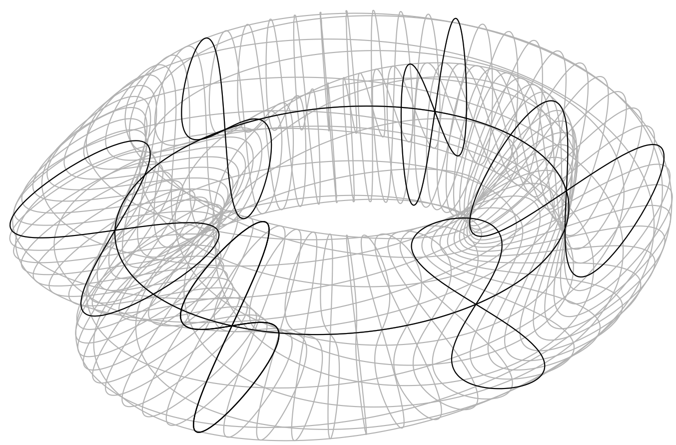


Möbius band Klein bottle projective plane

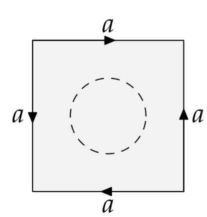
One way to make a model of a Klein bottle is first to paste the top and bottom edges of the square together, which gives a tube, and then to join the resulting boundary circles, making sure the arrows match up. You will notice this cannot be done without passing one end through the wall of the tube.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



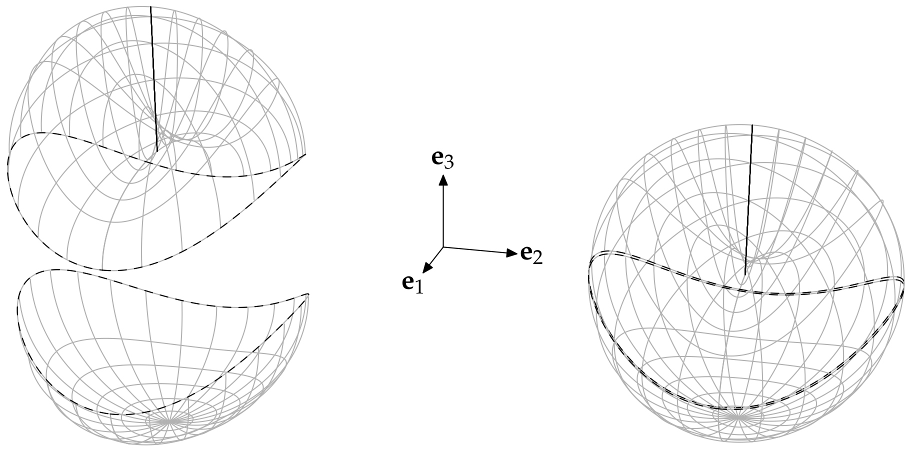
A different model of the Klein bottle can be made by folding over the edge of a Möbius band until it touches the central circle. This creates a Möbius type band with a figure eight cross-section. Equivalently, take a length of tube with a figure eight cross-section and weld the ends together giving one 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 gluing 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.



Cartesian products. The Cartesian product $M \times N$ of two manifolds M and N may fail to be a manifold. (If you don't remember what a Cartesian product is, see Appendix A.1 for a review of set theory.) For instance, if $M = N = [0,1]$, the unit interval, then $M \times N$ is the unit square, which is not a manifold. However, if at least one of the two manifolds M and N has no boundary, then $M \times N$ is a manifold. The dimension of $M \times N$ is the sum of the dimensions of M and N. For instance, if M is an interval and N a circle, then $M \times N$ is a cylinder wall. If both M and N are circles, then $M \times N$ is a torus. We can also form Cartesian products of more than two factors: the product of n copies of a circle with itself is an n-dimensional manifold known as an *n-torus*.

Connected sums. Let M and N be 2-manifolds. The *connected sum* is a 2-manifold $M \# N$ produced by punching a circular hole in each of the manifolds M and N and then gluing the two boundary circles together. For instance, the connected sum of two tori is a pretzel-type surface with two holes.

1.2. Equations

Very commonly a manifold M is 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 $\phi_1, \phi_2, \dots, \phi_m$ 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

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

The solution set M is the set of all vectors $\mathbf{x}$ in $\mathbf{R}^n$ which satisfy $\phi(\mathbf{x}) = \mathbf{c}$ and is denoted by $\phi^{-1}(\mathbf{c})$. (This notation is standard, but a bit unfortunate because it suggests falsely that ϕ is invertible, which it is usually not.) Thus

$$M = \phi^{-1}(\mathbf{c}) = \{\, \mathbf{x} \in \mathbf{R}^n \mid \phi(\mathbf{x}) = \mathbf{c} \,\}.$$

It is in general difficult to find explicit solutions of a system of equations. (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. The simplest case is that of a single equation ($m = 1$), such as

$$x^2 + y^2 - z^2 = 0.$$

Here we have a single scalar-valued function of three variables $\phi(x, y, z) = x^2 + y^2 - z^2$ and $c = 0$. The solution set M of the equation is a cone in $\mathbf{R}^3$, which is not a manifold because it has no well-defined tangent plane at the origin. We can determine the tangent plane at any other point of M by recalling from calculus that the gradient of ϕ is perpendicular to the surface. Hence for any nonzero $\mathbf{x} = (x, y, z) \in M$ the tangent plane to M at $\mathbf{x}$ is the plane perpendicular to $\operatorname{grad}(\phi)(\mathbf{x}) = (2x, 2y, -2z)$.

As we see from this example, the solution set of a system of equations may have singularities and is therefore not necessarily a manifold. In general, if M is given by a single equation $\phi(\mathbf{x}) = c$ and $\mathbf{x}$ is a point of M with the property that $\operatorname{grad}(\phi)(\mathbf{x}) \neq \mathbf{0}$, then $\mathbf{x}$ is a smooth point of M and the tangent space at $\mathbf{x}$ is the orthogonal complement of $\operatorname{grad}(\phi)(\mathbf{x})$. (Conversely, if $\mathbf{x}$ is a singular point of M, we must have $\operatorname{grad}(\phi)(\mathbf{x}) = \mathbf{0}$!) The standard notation for the tangent space to M at $\mathbf{x}$ is $T_{\mathbf{x}}M$. Thus we can write $T_{\mathbf{x}}M = \operatorname{grad}(\phi)(\mathbf{x})^{\perp}$.

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

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

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}. In particular, the one-dimensional unit "sphere" S^1 is the unit circle in the plane, and the zero-dimensional unit "sphere" S^0 is the subset $\{-1, 1\}$ of the real line. To determine the tangent spaces of the unit sphere it is easier to work with the equation $\|\mathbf{x}\|^2 = 1$ instead of $\|\mathbf{x}\| = 1$. In other words, we let $\phi(\mathbf{x}) = \|\mathbf{x}\|^2$. Then $\operatorname{grad}(\phi)(\mathbf{x}) = 2\mathbf{x}$, which is nonzero for all $\mathbf{x}$ in the unit sphere. Therefore S^{n-1} is a manifold and for any $\mathbf{x}$ in S^{n-1} we have

$$T_{\mathbf{x}} S^{n-1} = (2\mathbf{x})^{\perp} = \mathbf{x}^{\perp} = \{\, \mathbf{y} \in \mathbf{R}^n \mid \mathbf{y} \cdot \mathbf{x} = 0 \,\},$$

a linear subspace of $\mathbf{R}^n$. (In Exercise 1.7 you will be asked to find a basis of the tangent space for a particular $\mathbf{x}$ and you will see that $T_{\mathbf{x}}S^{n-1}$ is $n-1$-dimensional.)

1.3. 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. We will discuss in Chapter 6 how to find the tangent spaces to manifolds given by more than one equation.

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 $\|x\| \le r$. 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,

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

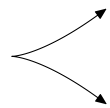
parametrizes the unit circle in $\mathbf{R}^2$ and

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

parametrizes the unit sphere in $\mathbf{R}^3$. (Here ϕ is the angle between a vector and the (x, y)-plane and θ is the polar angle in the (x, y)-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.

1.4. Example. Define $c(t) = (t^2, t^3)$ for $t \in \mathbf{R}$. As t runs through the real line, the point $c(t)$ travels along a curve in the plane, which we call a *path* or *parametrized curve*.



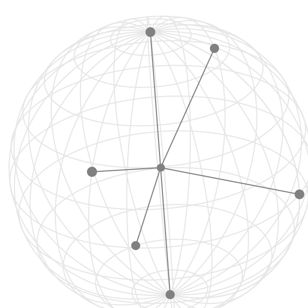
The path c has no self-intersections: if $t_1 \neq t_2$ then $c(t_1) \neq c(t_2)$. The cusp at the origin ($t = 0$) is a singular point, but all other points ($t \neq 0$) are smooth. The tangent line at a smooth point $c(t)$ is the line spanned by the velocity vector $c'(t) = (2t, 3t^2)$. The slope of the tangent line is $3t^2/2t = \frac{3}{2}t$. For $t = 0$ the velocity vector is $c'(0) = \mathbf{0}$, which does not span a line. Nevertheless, the curve has a well-defined (horizontal) tangent line at the origin, which we can think of as the limit of the tangent lines as t tends to 0.

More examples of parametrizations are given in Exercises 1.1–1.3.

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 1.11) and configuration spaces. The *configuration* or *state* of a mechanical system (such as a pendulum, a spinning top, the solar system, a fluid, or a gas etc.) is a complete specification of the position of each of its parts. (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.) 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.

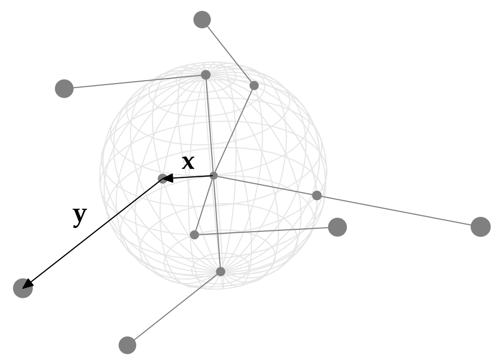
1.5. Example. A *spherical pendulum* is a weight or bob attached to a fixed centre by a rigid rod, free to swing in any direction in three-space.



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, and the spherical pendulum has two degrees of freedom.

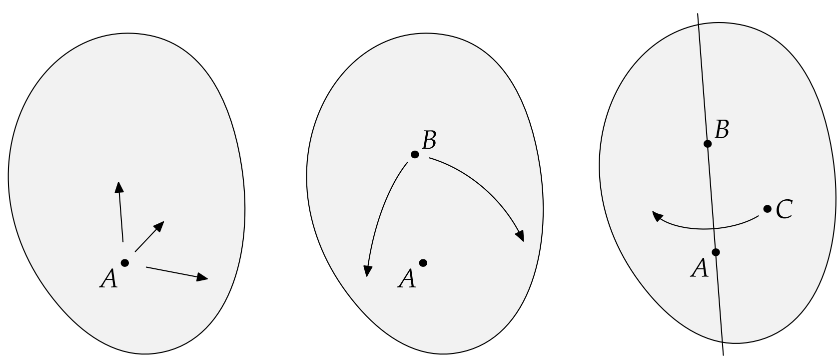
The configuration space of even a fairly small system can be quite complicated. Even if the system is situated in three-dimensional space, it may have many more than three degrees of freedom. This is why higher-dimensional manifolds are common in physics and applied mathematics.

1.6. 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.



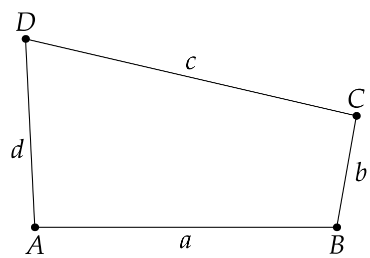
The vector $\mathbf{x}$ is constrained to a sphere of radius r about the centre and $\mathbf{y}$ to a sphere of radius s about the head of $\mathbf{x}$. Aside from this limitation, every pair of vectors can occur (if we suppose the second rod is allowed to swing completely freely and move "through" the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

1.7. 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.

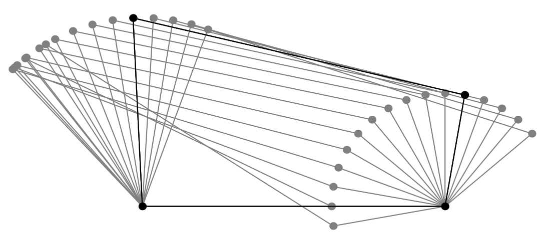


The point A can move about freely and is determined by three coordinates, and so it has three degrees of freedom. But the position of A alone does not determine the position of the whole solid. If A is kept fixed, the point B can perform two independent swivelling motions. In other words, it moves on 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. Let us call this manifold M and try to say something about its shape. Choose an arbitrary reference point in space. Inside the manifold M we have a subset M_0 consisting of all configurations which have the point A placed at the reference point. Configurations in M_0 have three fewer degrees of freedom, because only the points B and C can move, so M_0 is a three-dimensional manifold. Every configuration in M can be moved to a unique configuration in M_0 by a unique parallel translation of the solid which moves A to the reference point. In other words, the points of M can be viewed as pairs consisting of a point in M_0 and a vector in $\mathbf{R}^3$: the manifold M is the Cartesian product $M_0 \times \mathbf{R}^3$. See Exercise 8.6 for more information on M_0.

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

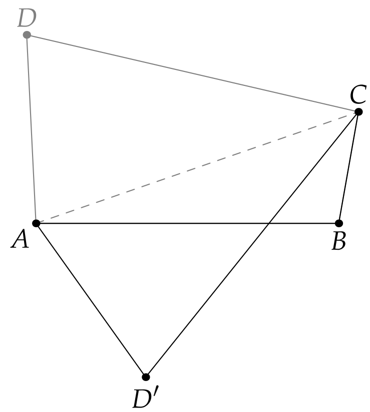


(Think of four rigid rods attached by hinges.) What are all the possibilities? For simplicity let us disregard translations by keeping the first edge AB fixed in one place. Edges are allowed to cross each other, so the short edge BC can spin full circle about the point B. During this motion the point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

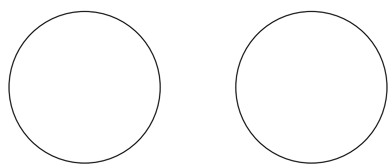


As C moves all the way around, the point D reaches its greatest left- or rightward displacement when the edges BC and CD are collinear. Arrangements such as this are used in engines for converting a circular motion to a pumping motion, or vice versa. The position of the "crank" C wholly determines that of the "rocker" D. 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.

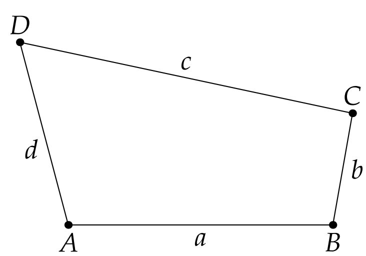


So there is in fact another circle's worth of possible configurations. It is not possible to move continuously from the first set of configurations to the second; in fact they are each other's mirror images. Thus the configuration space is a disjoint union of two circles.

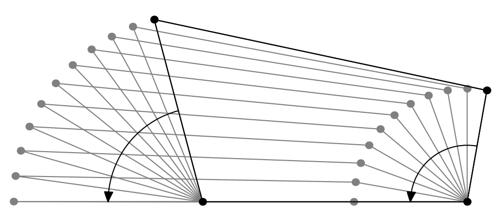


This is an example of a *disconnected* manifold consisting of two *connected components*.

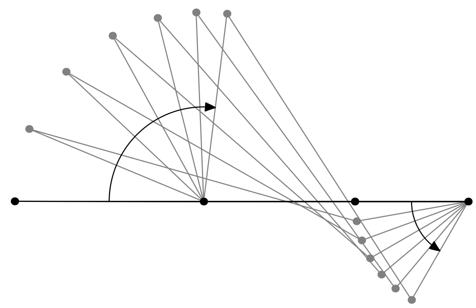
1.9. Example (quadrilaterals, continued). Even this is not the full story: it *is* possible to move from one circle to the other when $b + c = a + d$ (and also when $a + b = c + d$).



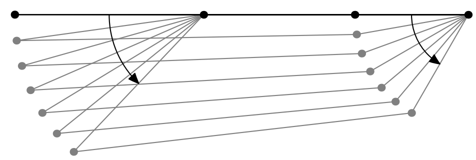
In this case, when BC points straight to the left, the quadrilateral collapses to a line segment:



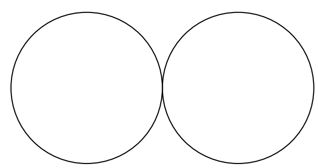
and when C moves further down, there are two possible directions for D to go, back up:



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

Computer software can be helpful with some of the exercises in these notes. Useful free software packages include Microsoft Mathematics for Windows or Grapher for Mac OS. Packages such as Mathematica or MATHLAB are more powerful, but are not free.

1.1. The formulas $x = t - \sin t$, $y = 1 - \cos t$ ($t \in \mathbf{R}$) parametrize a plane curve. Graph this curve. 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.

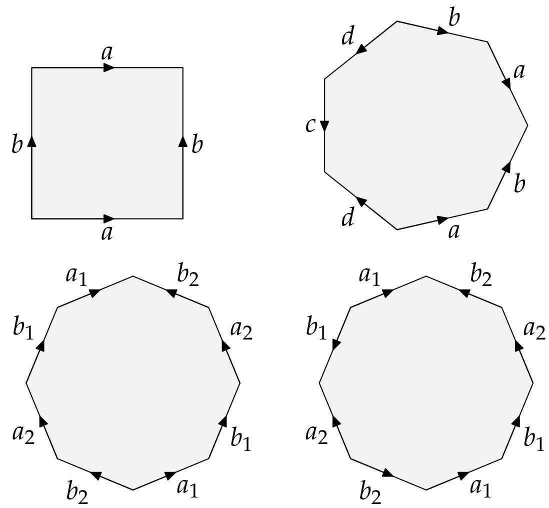
1.2. Same questions as in Exercise 1.1 for the parametrized curve $x = 3t/(1 + t^3)$, $y = 3t^2/(1 + t^3)$, where $0 < t < \infty$.

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

1.4. Draw gluing diagrams of the following surfaces.

(i) A sphere.
(ii) A torus with a hole punched in it.
(iii) The connected sum of two tori.

1.5. Sketch the surfaces defined by the following gluing diagrams. One of these surfaces cannot be embedded in $\mathbf{R}^3$, so use a self-intersection where necessary.



(There are at least two possible strategies. The first is to proceed in stages by gluing the a's, then the b's, etc., and trying to identify what you get at each step. The second is to decompose each diagram into a connected sum of simpler diagrams.)

1.6. Graph the surface in $\mathbf{R}^3$ defined by $x^n = y^2 z$ for the values of n listed below. Determine all the points where the surface does not have a well-defined tangent plane. (You may use computer output. If you want to do it by hand, one useful preliminary step is to determine the intersection of each surface with a general plane parallel to one of the coordinate planes. To investigate the tangent planes, write the equation of the surface as $\phi(x, y, z) = 0$, where $\phi(x, y, z) = x^n - y^2 z$, and then find the gradient of ϕ.)

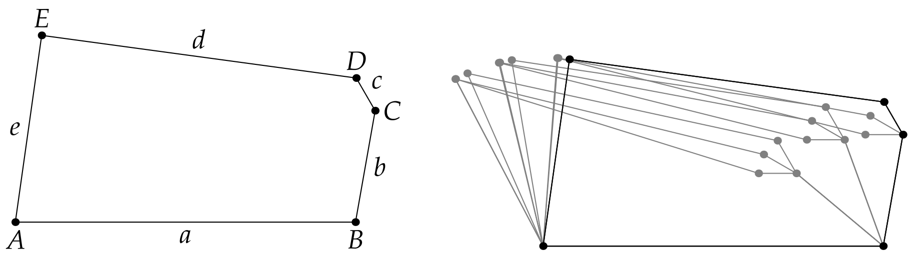
(i) $n = 0$.
(ii) $n = 1$.
(iii) $n = 2$.
(iv) $n = 3$.

1.7. Let M be the sphere of radius $\sqrt{n}$ about the origin in $\mathbf{R}^n$ and let $\mathbf{x}$ be the point $(1, 1, \ldots, 1)$ 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 $y_1, y_2, \ldots, y_n$ of $\mathbf{y}$ and find the general solution by means of linear algebra.)

1.8. 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.)

1.9. Choose two distinct positive real numbers a and b. What is the configuration space of all quadrilaterals $ABCD$ such that AB and CD have length a and BC and AD have length b? (These quadrilaterals include all parallelograms with sides a and b.) What happens if $a = b$? (As in Examples 1.8 and 1.9 assume that the edge AB is kept fixed in place so as to rule out translations.)

1.10. 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.)



1.11. 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 $\frac{1}{g(y)}\,dy = f(x)\,dx$ 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. You may use computer software. An online phase portrait generator can be found at

http://www.onlinesciencetools.com/tools/phaseportrait.

CHAPTER 2

Differential forms on Euclidean space

The notion of a differential form encompasses such ideas as elements of surface area, 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 É. 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, Appendix C.) Here I stands for a *multi-index* $(i_1, i_2, \ldots, i_k)$ 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

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

(Instead of $dx_{i_1} \, dx_{i_2} \cdots dx_{i_k}$ the notation $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$ is used by some authors 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 $U = \mathbf{R}^3 \setminus \{\,(x, y, z) \mid x^2 + y^2 \neq 0\,\}$, 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*, or *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

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

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, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. 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 among the numbers $1, 2, \ldots, n$. 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 $\alpha = \sum_{i=1}^n f_i\, dx_i$ is a 1-form on $\mathbf{R}^n$, then

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

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

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

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 all 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

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

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

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

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

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

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 $g = xy + \sin yz$ 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 2.8.) Hence every closed 1-form on $\mathbf{R}^n$ is exact.

2.10. Example. The 1-form on the punctured plane $\mathbf{R}^2 \setminus \{\mathbf{0}\}$ defined by

$$\alpha_0 = -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy = \frac{-y\,dx + x\,dy}{x^2+y^2}.$$

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 3.8, 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

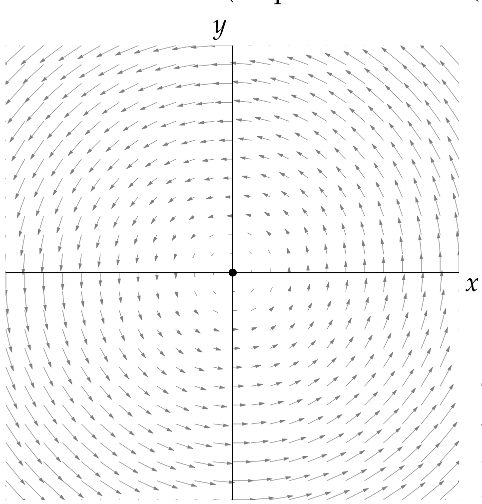
$$\begin{aligned} *dx_i &= (-1)^{i+1} dx_1\,dx_2 \cdots \widehat{dx_i} \cdots dx_n & &\text{for } 1 \le i \le n,\\ *(dx_i\,dx_j) &= (-1)^{i+j+1}\,dx_1\,dx_2 \cdots \widehat{dx_i} \cdots \widehat{dx_j} \cdots dx_n & &\text{for } 1 \le i < j \le n. \end{aligned}$$

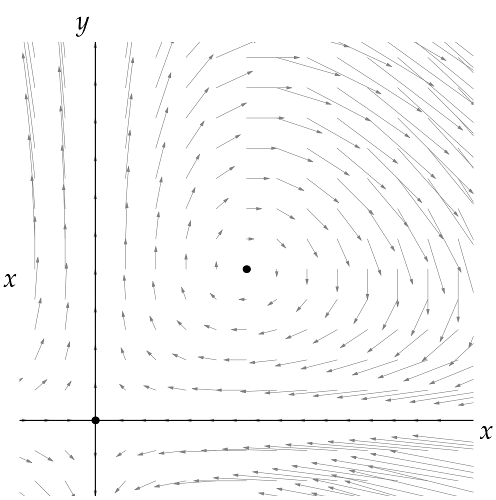
2.5. div, grad and curl

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 1.11). 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 electricity 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 8.16.) 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

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

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

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

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. 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) $d\alpha$, $d\beta$, $d\gamma$.

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

(i) $e^{xy+z^2}\,dx$.
(ii) $\sum_{i=1}^n x_i^2\, dx_1 \cdots \widehat{dx_i} \cdots dx_n$.

2.3. Calculate $d \sin f(\mathbf{x})^2$, where $f \colon \mathbf{R}^n \to \mathbf{R}$ is an arbitrary smooth function.

2.4. Define functions ξ and η by

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

Show that $\alpha_0 = -\eta\, d\xi + \xi\, d\eta$, where α_0 denotes the angle form defined in Example 2.10.

2.5. Write the coordinates on $\mathbf{R}^{2n}$ as $(x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$. 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 $n = 1, 2, 3$.)

2.6. 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 $\alpha(d\alpha)^n = \alpha(d\alpha\,d\alpha\cdots d\alpha)$. (Use the result of Exercise (2.5).)

2.7. 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) $\alpha = 2xy^3z^4\,dx + (3x^2y^2z^4 - ze^y\sin(ze^y))\,dy + (4x^2y^3z^3 - e^y\sin(ze^y) + e^z)\,dz$.

2.8. Let $\alpha = \sum_{i=1}^n f_i\,dx_i$ be a *closed* 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 $d\alpha = 0$. If you get confused, first do the case $n = 2$, where $g(\mathbf{x}) = \int_0^{x_1} f_1(t, x_2)\,dt + \int_0^{x_2} f_2(0, t)\,dt$.)

2.9. Let $\alpha = \sum_{i=1}^n f_i\,dx_i$ be a *closed* 1-form whose coefficients f_i are smooth functions defined on $\mathbf{R}^n \setminus \{\mathbf{0}\}$ 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.6 to each f_i.)

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

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

2.12. Calculate $*\alpha$, $*\beta$, $*\gamma$, $*(\alpha\beta)$, where α, β and γ are as in Exercise 2.1.

2.13. Let $\alpha = x_1\,dx_2 + x_3\,dx_4$, $\beta = x_1x_2\,dx_3\,dx_4 + x_3x_4\,dx_1\,dx_2$ and $\gamma = x_2\,dx_1\,dx_3\,dx_4$ be forms on $\mathbf{R}^4$. Calculate

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

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

(i) Find $*\alpha$ and $*d{*}d\alpha$ (where $*d{*}d\alpha$ is shorthand for $*(d(*(d\alpha)))$.)
(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 $\mathbf{R}^4$.

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

2.16. Recall that for any increasing multi-index $I = (i_1, i_2, \dots, i_k)$ the number $\varepsilon_I = \pm 1$ is determined by the requirement that

$$dx_I\,dx_{I^c} = \varepsilon_I\,dx_1\,dx_2\dots dx_n.$$

Let us define $|I| = i_1 + i_2 + \cdots + i_k$. Show that $\varepsilon_I = (-1)^{|I| + \binom{k+1}{2}}$.

2.17. Let $\alpha = \sum_I a_I\,dx_I$ and $\beta = \sum_J b_J\,dx_J$ be *constant* k-forms on $\mathbf{R}^n$, i.e. forms with constant coefficients a_I and b_J. (We also assume, as usual, that the multi-indices I and J are increasing.) The *inner product* of α and β is the number defined by

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

For instance, if $\alpha = 7\,dx_1\,dx_2 + \sqrt{2}\,dx_1\,dx_3 + 11\,dx_2\,dx_3$ and $\beta = 5\,dx_1\,dx_3 - 3\,dx_2\,dx_3$, then

$$(\alpha, \beta) = \sqrt{2} \cdot 5 + 11 \cdot (-3) = 5\sqrt{2} - 33.$$

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. $(\alpha, \beta) = (*\alpha, *\beta)$.

2.18. The *Laplacian* Δf of a smooth function f 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 2.17(iv).)

2.19. Let $\alpha = \sum_{i=1}^n f_i dx_i$ be a 1-form on $\mathbf{R}^n$.

(i) Find formulas for $*\alpha$, $d*\alpha$, $*d*\alpha$, and $d*d*\alpha$.
(ii) Find formulas for $d\alpha$, $*d\alpha$, $d*d\alpha$, and $*d*d\alpha$.
(iii) Finally compute $d*d*\alpha + (-1)^n *d*d\alpha$. Try to write the answer in terms of the Laplace operator Δ defined in Exercise 2.18.

2.20. Let α be the 1-form $\|\mathbf{x}\|^{2p}\mathbf{x} \cdot *d\mathbf{x}$ on $\mathbf{R}^n$, where p is a real constant. Compute $d\alpha$. Show that α is closed if and only if $p = -\frac{1}{2}n$.

2.21. (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 $\mathbf{R}^n \setminus \{\mathbf{0}\}$, where $\nu = \|\mathbf{x}\|^{-1}\mathbf{x} \cdot *d\mathbf{x}$.

2.22. 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 $1 \leq i < j \leq 4$.

2.23. 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

$$d\alpha = 0,$$
$$d\beta + 4\pi\gamma = 0.$$

(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 2.22.

(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 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 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. The *parallelepiped* spanned by the columns is by definition the set of all linear combinations $\sum_{i=1}^n c_i \mathbf{a}_i$, where the coefficients c_i range over the unit interval $[0, 1]$. A parallelepiped spanned by a single vector is called a *line segment* and a parallelepiped spanned by two vectors is called a *parallelogram*. The determinant of A is variously denoted by

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

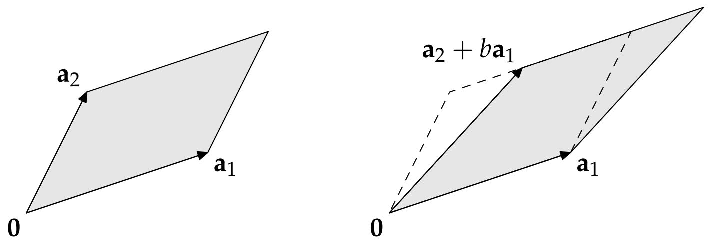
Expansion on the j-th column. You may have seen the following definition of the determinant:

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

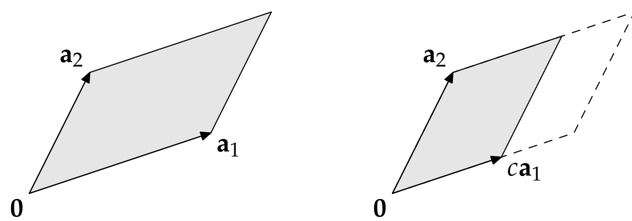
Here $A^{i,j}$ denotes the $(n-1) \times (n-1)$-matrix obtained from A by crossing out the i-th row and the j-th column. This is a recursive definition, which reduces the calculation of any determinant to that of determinants of smaller size. The recursion starts at $n = 1$; the determinant of a 1×1-matrix (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 a few simple axioms, which make good sense in view of its geometrical significance and which comprise an efficient algorithm for calculating any determinant. To motivate these axioms we first consider the case of a 2×2-matrix A ($n = 2$). Then the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ are vectors in the plane, and instead of the oriented "volume" we speak of the oriented area of the parallelogram spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$. (This notion is familiar from calculus: the integral $\int_a^b f(x)\,dx$

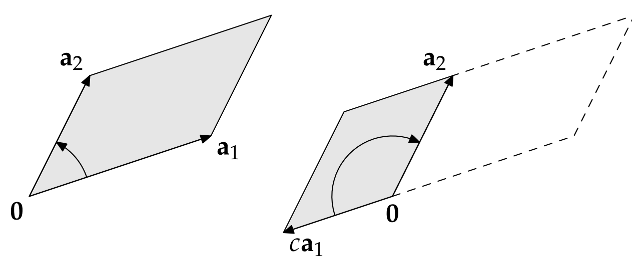
of a function f is the oriented area between its graph and the x-axis.) How is the oriented area affected by various transformations of the vectors? Adding any multiple of $\mathbf{a}_1$ to the second column $\mathbf{a}_2$ has the effect of performing a shear transformation on the parallelogram, which does not change its oriented area:



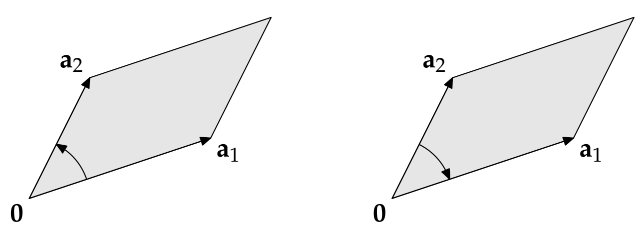
Multiplying the first column by a scalar c has the effect of stretching (if $c > 1$) or compressing (if $0 < c < 1$) the parallelogram and changing its oriented area by a factor of c:



What if c is negative? In the picture below the parallelogram on the left is positively oriented in the sense that the angle from edge $\mathbf{a}_1$ to edge $\mathbf{a}_2$ is counterclockwise (positive). Because c is negative, the parallelogram on the right is negatively oriented in the sense that the angle from edge $c\mathbf{a}_1$ to edge $\mathbf{a}_2$ is clockwise (negative). Therefore multiplying $\mathbf{a}_1$ by a negative c also changes the oriented area by a factor of c:



Similarly, interchanging the columns of A has the effect of reversing the orientation of the parallelogram, which changes the sign of its oriented area:



To generalize this to higher dimensions, recall the *elementary column operations*, which 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). As suggested by the pictures above, type I does not affect the determinant, type II multiplies it by the corresponding constant, and type III causes a sign change. We turn these observations into a definition as follows.

3.1. Definition. A *determinant* is a function det which assigns to every $n \times n$-matrix A a number $\det(A)$ subject to the following axioms:

(i) If E is an elementary column operation, then $\det(E(A)) = k\det(A)$, where
$$k = \begin{cases} 1 & \text{if } E \text{ is of type I,} \\ c & \text{if } E \text{ is of type II (multiplication of a column by } c\text{),} \\ -1 & \text{if } E \text{ is of type III.} \end{cases}$$

(ii) $\det(I) = 1$.

Axiom (ii) is a normalization convention, which is justified by the reasonable requirement that the *unit cube* in $\mathbf{R}^n$ (i.e. the parallelepiped spanned by the columns of the identity matrix I) should have oriented volume 1.

3.2. Example. The following calculation is a sequence of column operations, at the end of which we apply the normalization axiom.

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

As this example suggests, the axioms of Definition 3.1 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.

3.3. Theorem (uniqueness of determinants). *Let* det *and* det′ *be two functions satisfying Axioms* (i)–(ii). *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: $\mathbf{a}_1 = c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$. Repeatedly applying type I column operations gives

$$\begin{aligned} \det(A) &= \det\Bigl(\sum_{i=2}^{n} c_i\mathbf{a}_i, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n\Bigr) \\ &= \det(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n). \end{aligned}$$

Applying a type II operation gives

$$\begin{aligned} \det(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) &= \det(-\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) \\ &= -\det(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n), \end{aligned}$$

and therefore $\det(A) = 0$. 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 $E_m E_{m-1} \cdots E_2 E_1(A) = I$. By Axiom (i) each operation E_i has the effect of multiplying the determinant by a certain factor k_i, so Axiom (ii) 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

3.4. Remark (change of normalization). Suppose that $\det'$ is a function that satisfies Axiom (i) but is normalized differently: $\det'(I) = c$. Then the proof of Theorem 3.3 shows that $\det'(A) = c \det(A)$ for all $n \times n$-matrices A.

The counterpart of this uniqueness theorem is an existence theorem, which states that Axioms 3.1(i)–(ii) are consistent. We will establish consistence 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.

3.5. 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 n-tuples 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 of 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* of 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 of 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 $\{1,2,\dots,n\}$ 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 as $\sigma\tau$ 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 3.5 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 3.5 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 3.5. We need to check that the right-hand side of the determinant formula in Theorem 3.5 obeys Axioms (i)–(ii) of Definition 3.1. Let us for the moment denote the right-hand side by $f(A)$. Axiom (ii) is the easiest to verify: if $A = I$, then

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

and therefore $f(I) = 1$. Next we consider how $f(A)$ behaves when we interchange two columns of A. We assert that 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

$$\begin{aligned}
&f(\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) \\
&\qquad = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\tau\sigma(1)} a_{2,\tau\sigma(2)} \cdots a_{n,\tau\sigma(n)} \\
&\qquad = \sum_{\rho \in S_n} \operatorname{sign}(\tau\rho) a_{1,\rho(1)} a_{2,\rho(2)} \cdots a_{n,\rho(n)} && \text{substitute } \rho = \tau\sigma \\
&\qquad = \sum_{\rho \in S_n} \operatorname{sign}(\tau) \operatorname{sign}(\rho) a_{1,\rho(1)} a_{2,\rho(2)} \cdots a_{n,\rho(n)} && \text{by formula (3.1)} \\
&\qquad = -\sum_{\rho \in S_n} \operatorname{sign}(\rho) a_{1,\rho(1)} a_{2,\rho(2)} \cdots a_{n,\rho(n)} && \text{by Exercise 3.5} \\
&\qquad = -f(\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n).
\end{aligned}$$

To see what happens when we multiply a column of A by c, observe that 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 column of A by c, each term in $f(A)$ is multiplied by c. Therefore

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

By a similar argument we have

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

for any vector $\mathbf{a}_i'$. In particular we can take $\mathbf{a}_i' = \mathbf{a}_j$ for some $j \neq i$, which gives

$$\begin{aligned}
f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i + \mathbf{a}_j, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n) &= f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n) \\
&\qquad + f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_j, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n) \\
&= f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n),
\end{aligned}$$

because $f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n) = 0$. This shows that f satisfies the conditions of Definition 3.1. QED

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. Theorem 3.5 implies that different column reductions lead to the same answer for the determinant.

The following corollary of Theorem 3.5 amounts to a reformulation of Definition 3.1. Recall that $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ denote the standard basis vectors of $\mathbf{R}^n$, i.e. the columns of the identity $n \times n$-matrix.

3.6. Corollary. *The determinant possesses the following properties. These properties characterize the determinant uniquely.*

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

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

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 all vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ *and for all pairs of distinct indices* $i \neq j$*;*

(iii) *normalization:* $\det(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 1$.

Proof. Property (i) was established in the proof of Theorem 3.5, while properties (ii)–(iii) are simply a restatement of part of Definition 3.1. Therefore the determinant has properties (i)–(iii). Conversely, properties (i)–(iii) taken together imply Axioms (i)–(ii) of Definition 3.1. Therefore, by Theorem 3.3, properties (i)–(iii) characterize the determinant uniquely. QED

Here are some further rules obeyed by determinants. Each can be deduced from Definition 3.1 or from Theorem 3.5. (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) $\det(A) = a_{1,1}a_{2,2} \cdots a_{n,n}$ *if* A *is upper triangular (i.e.* $a_{i,j} = 0$ *for* $i > j$*).*

(ii) $\det(AB) = \det(A)\det(B)$.

(iii) $\det(A^T) = \det(A)$.

(iv) *(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 deleting the* i*-th row and the* j*-th column.*

(v) *Let* $\sigma \in S_n$ *be a permutation. Then*

$$\det(\mathbf{a}_{\sigma(1)}, \mathbf{a}_{\sigma(2)}, \dots, \mathbf{a}_{\sigma(n)}) = \operatorname{sign}(\sigma)\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n).$$

When calculating determinants in practice one combines column reductions with these rules. For instance, rule (i) tells us we need not bother reducing A all the way to the identity matrix, like we did in Example 3.2, but that an upper triangular form suffices. Rule (iii) tells us we may use row operations as well as column operations.

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 $\mathbf{R}^n$,

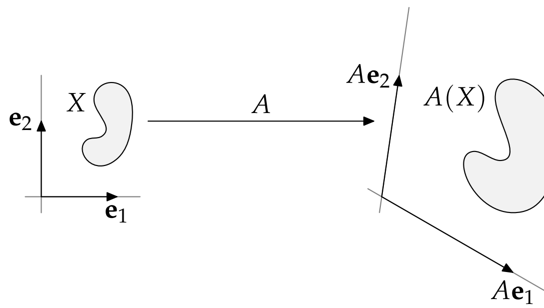
$$[0,1]^n = \{\, \mathbf{x} \in \mathbf{R}^n \mid 0 \leq x_i \leq 1 \quad \text{for } i = 1, 2, \dots, n \,\},$$

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 the parallelepiped spanned by the vectors $A\mathbf{e}_1, A\mathbf{e}_2, \dots, A\mathbf{e}_n$, which are the columns of A. Hence $A([0,1]^n)$ has n-dimensional volume

$$\operatorname{vol}(A([0,1]^n)) = |\det(A| = |\det(A)| \operatorname{vol}([0,1]^n).$$

This rule generalizes as follows: if X is a measurable subset of $\mathbf{R}^n$, then

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



(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.) So $|\det(A)|$ can be interpreted as a *volume change factor*. The sign of the determinant tells you whether A preserves $(+)$ or reverses $(-)$ the orientation of $\mathbf{R}^n$. (See Section 8.2 for more on orientations.)

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 $\mathbf{R}^2 \setminus \{\mathbf{0}\}$ to be

$$\alpha_0 = \frac{-y\,dx + x\,dy}{x^2 + y^2}.$$

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

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

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 $U = \mathbf{R}$, $V = \mathbf{R}^2 \setminus \{\mathbf{0}\}$ 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

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

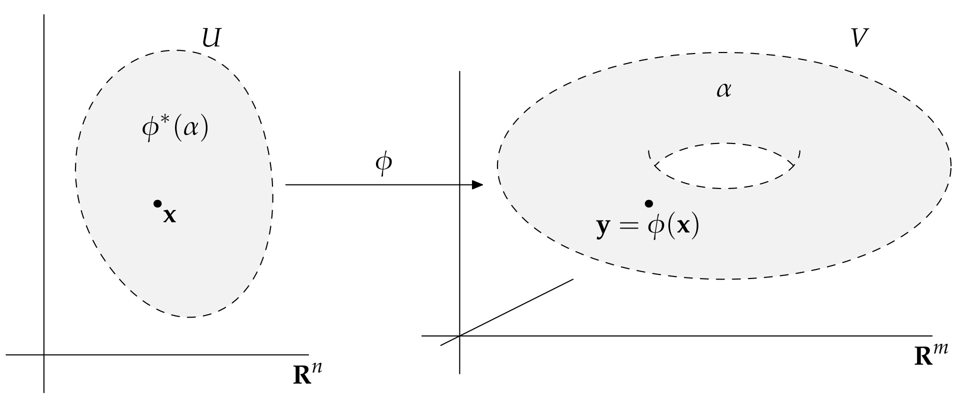
Here $\phi^*(f_I)$ is defined by

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

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 the variables $y_1, y_2, \dots, y_m$; its pullback $\phi^*(\alpha) = \sum_J g_J\, dx_J$ is a k-form in the variables $x_1, x_2, \dots, x_n$. In Theorem 3.13 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: $\phi^*(a\alpha + b\beta) = a\phi^*(\alpha) + b\phi^*(\beta)$ for all scalars a and b and all k-forms α and β on V;*
(ii) *multiplicative: $\phi^*(\alpha\beta) = \phi^*(\alpha)\phi^*(\beta)$ for all k-forms α and l-forms β on V;*
(iii) *natural: $\phi^*(\psi^*(\alpha)) = (\psi \circ \phi)^*(\alpha)$, where $\psi\colon V \to W$ is a second smooth map with W open in $\mathbf{R}^l$, and α is a k-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,

$$\begin{aligned}\phi^*(dy_I\,dy_J) &= \phi^*(dy_{i_1}dy_{i_2}\cdots dy_{i_k}\,dy_{j_1}\,dy_{j_2}\cdots dy_{j_l})\\&= d\phi_{i_1}\,d\phi_{i_2}\cdots d\phi_{i_k}\,d\phi_{j_1}\cdots d\phi_{j_l} = \phi^*(dy_I)\phi^*(dy_J),\end{aligned}$$

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

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

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

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

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

$$\begin{aligned}\phi^*(\psi^*(\alpha)) &= \sum_{k=1}^n \frac{\partial\phi^*(\psi_i)}{\partial x_k}\,dx_k = d\phi^*(\psi_i)\\&= d((\psi\circ\phi)_i) = (\psi\circ\phi)^*(dz_i) = (\psi\circ\phi)^*(\alpha).\end{aligned}$$

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 $\alpha = dz_i$, together with properties (i) and (iii). 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

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

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

$$\phi^*(d\alpha) = \sum_I \phi^*(df_I\,dy_I) = \sum_I \phi^*(df_I)\phi^*(dy_I)$$
$$= \sum_I d\phi^*(f_I)\,d\phi_{i_1}\,d\phi_{i_2}\cdots d\phi_{i_k}, \tag{3.2}$$

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

$$\begin{aligned} d\phi^*(\alpha) &= \sum_I d\bigl(\phi^*(f_I)\phi^*(dy_I)\bigr) = \sum_I d\bigl(\phi^*(f_I)\,d\phi_{i_1}\,d\phi_{i_2}\cdots d\phi_{i_k}\bigr) \\ &= \sum_I d\phi^*(f_I)\,d\phi_{i_1}\,d\phi_{i_2}\cdots d\phi_{i_k} + \sum_I \phi^*(f_I)\,d(d\phi_{i_1}\,d\phi_{i_2}\cdots d\phi_{i_k}) \\ &= \sum_I d\phi^*(f_I)\,d\phi_{i_1}\,d\phi_{i_2}\cdots d\phi_{i_k}. \end{aligned} \tag{3.3}$$

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 2.10.) Comparing (3.2) with (3.3) we see that $\phi^*(d\alpha) = d\phi^*(\alpha)$. QED

Here is an application of Theorem 3.11. An open subset U of $\mathbf{R}^n$ is called *connected* if for every pair of points $\mathbf{x}$ and $\mathbf{y}$ in U there exists a path $c\colon [a,b] \to U$ satisfying $c(a) = \mathbf{x}$ and $c(b) = \mathbf{y}$.

3.12. Lemma. *Let U be a connected open subset of $\mathbf{R}^n$ and let $f\colon U \to \mathbf{R}$ be a smooth function. Then $df = 0$ if and only if f is constant.*

Proof. If f is constant, all its partial derivatives vanish, so $df = 0$. Conversely, suppose $df = 0$. To prove that f is constant it is enough to show that $f(\mathbf{x}) = f(\mathbf{y})$ for any pair of points $\mathbf{x}$, $\mathbf{y}$ in U. Choose a path $c\colon [a,b] \to U$ with the property $c(a) = \mathbf{x}$ and $c(b) = \mathbf{y}$. Let $h\colon [a,b] \to \mathbf{R}$ be the smooth function $h = c^*(f)$. Then $h'(t)\,dt = dh = d(c^*(f)) = c^*(df) = 0$ by Theorem 3.11, so $h'(t) = 0$ for $a \le t \le b$. Therefore h is constant (by one-variable calculus), so $f(\mathbf{x}) = f(c(a)) = h(a) = h(b) = f(c(b)) = f(\mathbf{y})$. 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

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

To express $\phi^*(\alpha)$ in terms of the x-variables we use

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

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

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

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

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

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

$$\begin{aligned} d\phi_{i_1}\, d\phi_{i_2} \cdots d\phi_{i_k} &= \sum_{p_1, p_2, \dots, p_k = 1}^{n} \frac{\partial \phi_{i_1}}{\partial x_{p_1}} \frac{\partial \phi_{i_2}}{\partial x_{p_2}} \cdots \frac{\partial \phi_{i_k}}{\partial x_{p_k}}\, dx_{p_1}\, dx_{p_2} \cdots dx_{p_k} \\ &= \sum_P \frac{\partial \phi_{i_1}}{\partial x_{p_1}} \frac{\partial \phi_{i_2}}{\partial x_{p_2}} \cdots \frac{\partial \phi_{i_k}}{\partial x_{p_k}}\, dx_P, \end{aligned}$$

in which the summation is over *all* n^k multi-indices $P = (p_1, p_2, \dots, p_k)$. If a multi-index P has repeating entries, then $dx_P = 0$. If the entries of P are all distinct, we can rearrange them in increasing order by means of a permutation σ. In other words, we have $P = (p_1, p_2, \dots, p_k) = (j_{\sigma(1)}, j_{\sigma(2)}, \dots, j_{\sigma(k)})$, where $J = (j_1, j_2, \dots, j_k)$ is an *increasing* multi-index and $\sigma \in S_k$ is a permutation. Thus we can rewrite the sum over all multi-indices P as a double sum over all increasing multi-indices J and all permutations σ:

$$\begin{aligned} d\phi_{i_1}\, d\phi_{i_1} \cdots d\phi_{i_k} &= \sum_J \sum_{\sigma \in S_k} \frac{\partial \phi_{i_1}}{\partial x_{j_{\sigma(1)}}} \frac{\partial \phi_{i_2}}{\partial x_{j_{\sigma(2)}}} \cdots \frac{\partial \phi_{i_k}}{\partial x_{j_{\sigma(k)}}}\, dx_{j_{\sigma(1)}}\, dx_{j_{\sigma(2)}} \cdots dx_{j_{\sigma(k)}} \\ &= \sum_J \sum_{\sigma \in S_k} \operatorname{sign}(\sigma) \frac{\partial \phi_{i_1}}{\partial x_{j_{\sigma(1)}}} \frac{\partial \phi_{i_2}}{\partial x_{j_{\sigma(2)}}} \cdots \frac{\partial \phi_{i_k}}{\partial x_{j_{\sigma(k)}}}\, dx_J \qquad (3.4) \\ &= \sum_J \det(D\phi_{I,J})\, dx_J. \qquad (3.5) \end{aligned}$$

In (3.4) we used the result of Exercise 3.8 and in (3.5) we applied Theorem 3.5. 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 $i_1, i_2, \dots, i_k$ and columns $j_1, j_2, \dots, j_k$. 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.

3.13. 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 parallelepipeds 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.

3.14. 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(i) from Theorem 3.5.

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

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

3.3. In this exercise we write C_n for the unit cube $[0,1]^n$.

(i) Draw a picture of (a two-dimensional projection of) C_n for $n = 0, 1, 2, 3, 4, 5$.

(ii) Let $f_k(n)$ be the number of k-dimensional faces of C_n. Show that $f_k(n) = 2^{n-k}\binom{n}{k}$. (Let $p(s) = 2 + s$. Note that the coefficients of p are $f_0(1)$ and $f_1(1)$. Show that $f_k(n)$ is the coefficient of s^k in the polynomial $p(s)^k$ by noting that each face of C_n is a Cartesian product of faces of the unit interval C_1.)

3.4. List all permutations in S_4 with their lengths and signs.

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

3.6. Find all permutations in S_n of length 1.

3.7. 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).

3.8. 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 3.5(i).)

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

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

(ii) Deduce Theorem 3.7(iii) from Theorem 3.5.

3.11. 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 $1 \le i < n$,
(ii) $(\sigma_i\sigma_{i+1})^3 = 1$ for $1 \le i < n-1$,
(iii) $(\sigma_i\sigma_j)^2 = 1$ for $1 \le i, j < n$ and $i+1 < j$.

3.12. 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 $\det(A_\sigma) = \operatorname{sign}(\sigma)$.

3.13. (i) Suppose that A has the shape

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

i.e. all entries below a_{11} are 0. Deduce from Theorem 3.5 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(iv).

3.14. 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.)

3.15. Let $\phi(x_1, x_2, x_3) = (x_1x_2, x_1x_3, x_2x_3)$. Find

(i) $\phi^*(dy_1)$, $\phi^*(dy_2)$, $\phi^*(dy_3)$;
(ii) $\phi^*(y_1y_2y_3)$, $\phi^*(dy_1\,dy_2)$;
(iii) $\phi^*(dy_1\,dy_2\,dy_3)$.

3.16. Let $\phi(x_1, x_2) = (x_1^3, x_1^2x_2, x_1x_2^2, x_2^3)$. 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) $\phi^*(dy_2\,dy_3)$.

3.17. 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.8.

3.18. Let $P_3(r, \theta, \phi) = (r\cos\theta\cos\phi, r\sin\theta\cos\phi, r\sin\phi)$ be spherical coordinates in $\mathbf{R}^3$. Calculate $P_3^*(\alpha)$ for the following forms α:

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

3.19 (SPHERICAL COORDINATES IN n DIMENSIONS). In this problem let us write a point in $\mathbf{R}^n$ as $(r, \theta_1, \dots, \theta_{n-1})$. 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

$$P_{n+1}(r, \theta_1, \dots, \theta_n) = (\cos\theta_n) P_n(r, \theta_1, \dots, \theta_{n-1}), r\sin\theta_n).$$

(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_n.
(iii) Let $\mathbf{p}_n$ be the first column vector of the Jacobi matrix of P_n. Show that $P_n = r\mathbf{p}_n$.
(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

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

(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 $\det(DP_n) \neq 0$ if $r \neq 0$ and $-\frac{1}{2}\pi < \theta_i < \frac{1}{2}\pi$ for $i = 2, 3, \dots, n-1$.

3.20. Let $\phi \colon \mathbf{R}^n \to \mathbf{R}^n$ be an orthogonal linear map. Prove that $\phi^*(*\alpha) = *\phi^*(\alpha)$ for all k-forms α on $\mathbf{R}^n$.

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 *path* or *parametrized curve* in U is a smooth mapping $c\colon I \to U$ from an interval I into U. Our goal is to integrate forms over paths, so to avoid problems with improper integrals we will assume the interval I to be closed and bounded, $I = [a, b]$. Let α be a 1-form on U and let $c\colon [a, b] \to U$ be a path in U. The pullback $c^*(\alpha)$ is a 1-form on $[a, b]$, and can therefore be written as $c^*(\alpha) = h\,dt$, where t is the coordinate on $\mathbf{R}$ and h is a smooth function on $[a, b]$. The *integral* of α over c is now defined by

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

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

$$\int_c \alpha = \sum_{i=1}^n \int_a^b f_i(c(t))c_i'(t)\,dt.$$

4.1. Example. Let U be the punctured plane $\mathbf{R}^2 \setminus \{\mathbf{0}\}$. Let $c\colon [0, 2\pi] \to U$ be the usual parametrization of the circle, $c(t) = (\cos t, \sin t)$, and let α_0 be the angle form,

$$\alpha_0 = \frac{-y\,dx + x\,dy}{x^2 + y^2}.$$

Then $c^*(\alpha_0) = dt$ (see Example 3.8), so $\int_c \alpha_0 = \int_0^{2\pi} dt = 2\pi$.

A path $c\colon [a, b] \to U$ 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 path

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

has the same image as the original path 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 path (or that the paths 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 path in the opposite direction to c.

4.2. Example. The path $c\colon [0, 2\pi] \to \mathbf{R}^2$ 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 path 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 angular velocity.

It turns out that the integral of a form along a path 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 path 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 $c^*(\alpha) = h\,dt$ and $t = p(s)$. Then $p^*(c^*(\alpha)) = p^*(g\,dt) = (p^*(g)dp = p^*(g)(dp/ds)\,ds$, so

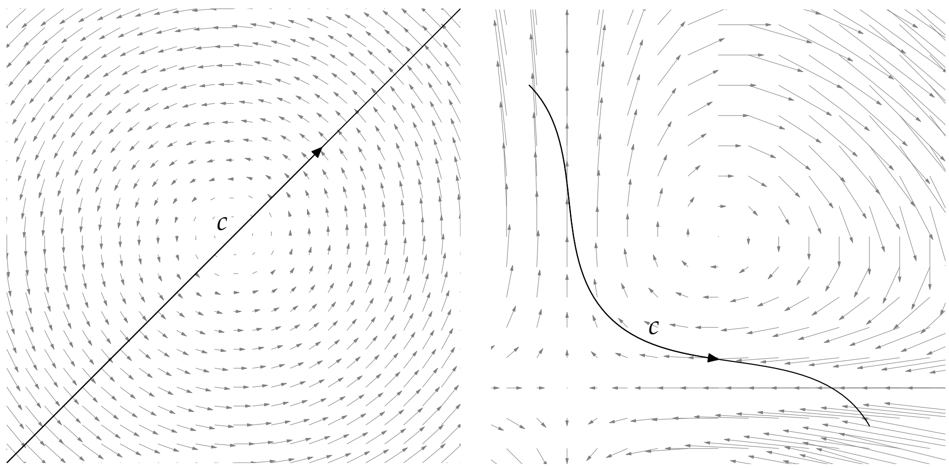
$$\int_{c \circ p} \alpha = \int_{[\bar{a}, \bar{b}]} p^*(g) \frac{dp}{ds}\,ds = \int_{\bar{a}}^{\bar{b}} g(p(s))p'(s)\,ds.$$

On the other hand, $\int_c \alpha = \int_a^b g(t)\,dt$, so by the substitution formula, Theorem B.9, 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 path $c\colon [a, b] \to U$ 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 path. Thus α represents the *work* done by the force field along an infinitesimal vector $d\mathbf{x}$. From (4.1) we see that $c^*(\alpha) = \mathbf{F}(c(t)) \cdot c'(t)\,dt$. We define the *work along the path c* done by the force **F** to be 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 done by the force is zero 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 travels along its path, but only on the path itself and on the direction of travel.

The field **F** is *conservative* if it can be written as the gradient of a function, $\mathbf{F} = \mathrm{grad}(g)$. 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 path. 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.

Theorem 4.4 also gives us a necessary criterion for a 1-form on U to be exact. A path $c\colon [a,b] \to U$ is called *closed* if $c(a) = c(b)$.

4.5. Corollary. *Let α be an exact 1-form defined on an open subset U of $\mathbf{R}^n$. Then $\int_c \alpha = 0$ for every closed path c in U.*

Proof. Let $c\colon [a,b] \to U$ be a closed path and let g be a smooth function on U satisfying $dg = \alpha$. Then $\int_c \alpha = g(c(b)) - g(c(a)) = 0$ by Theorem 4.4. QED

This corollary can be used to detect closed 1-forms that are not exact.

4.6. Example. The angle form $\alpha_0 \in \Omega^1(\mathbf{R}^2 \setminus \{\mathbf{0}\})$ of Example 2.10 is closed, but not exact. Indeed, its integral around the circle is $2\pi \neq 0$, so by Corollary 4.5 α_0 is not exact. Note the contrast with closed 1-forms on $\mathbf{R}^n$, which are always exact! (See Exercise 2.8.)

The main content of the next theorem is that the necessary criterion of Corollary 4.5 is in fact sufficient for a 1-form to be exact.

4.7. Theorem. *Let α be a 1-form on a connected 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 paths c.*
(iii) *$\int_c \alpha$ depends only on the endpoints of c for every path c in U.*

Proof. (i) $\Longrightarrow$ (ii): this is Corollary 4.5.

(ii) $\Longrightarrow$ (iii): assume $\int_c \alpha = 0$ for all closed paths 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 paths 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 path c by

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

(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) $\Longrightarrow$ (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 path $c_{\mathbf{x}}\colon [0,1] \to U$ 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,

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

Now consider a path $\tilde{c}$ composed of two pieces: for $-1 \le t \le 0$ travel from $\mathbf{x}_0$ to $\mathbf{x}$ along the path $c_{\mathbf{x}}$ and then for $0 \le t \le 1$ travel from $\mathbf{x}$ to $\mathbf{x} + h\mathbf{e}_i$ along the straight line given by $l(t) = \mathbf{x} + th\mathbf{e}_i$. 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

$$\frac{\partial g}{\partial x_i}(\mathbf{x}) = \lim_{h\to 0} \frac{1}{h}\Big(\int_{\tilde{c}} \alpha - \int_{c_{\mathbf{x}}} \alpha\Big) = \lim_{h\to 0} \frac{1}{h}\Big(\int_{c_{\mathbf{x}}} \alpha + \int_l \alpha - \int_{c_{\mathbf{x}}} \alpha\Big)$$
$$= \lim_{h\to 0} \frac{1}{h} \int_l \alpha = \lim_{h\to 0} \frac{1}{h} \int_{[0,1]} l^*(\alpha). \quad (4.2)$$

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

$$l^*(\alpha) = \sum_{j=1}^n f_j(\mathbf{x} + th\mathbf{e}_i)\,dl_j = \sum_{j=1}^n f_j(\mathbf{x} + th\mathbf{e}_i) l'_j(t)\,dt$$
$$= \sum_{j=1}^n f_j(\mathbf{x} + th\mathbf{e}_i)\delta_{i,j}h\,dt = hf_i(\mathbf{x} + th\mathbf{e}_i)\,dt. \quad (4.3)$$

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

$$\frac{\partial g}{\partial x_i}(\mathbf{x}) = \lim_{h\to 0} \frac{1}{h} \int_0^1 hf_i(\mathbf{x} + th\mathbf{e}_i)\,dt = \lim_{h\to 0} \int_0^1 f_i(\mathbf{x} + th\mathbf{e}_i)\,dt$$
$$= \int_0^1 \lim_{h\to 0} f_i(\mathbf{x} + th\mathbf{e}_i)\,dt = \int_0^1 f_i(\mathbf{x})\,dt = f_i(\mathbf{x}).$$

This formula shows that g is smooth and that $dg = \alpha$. This proves that (iii) $\Longrightarrow$ (i), and at the same time it proves that the function $g(\mathbf{x}) = \int_{c_{\mathbf{x}}} \alpha$ is an antiderivative of α. QED

Notice that the proof of the theorem tells us how to find an antiderivative of an exact 1-form α, namely by integrating α along an arbitrary path running from a base point $\mathbf{x}_0$ to a variable point $\mathbf{x}$. This useful fact deserves to be recorded separately.

4.8. Theorem. *Let α be an exact 1-form defined on a connected open subset U of $\mathbf{R}^n$. Choose a base point $\mathbf{x}_0$ in U and for each $\mathbf{x}$ in U choose a path $c_{\mathbf{x}}$ joining $\mathbf{x}_0$ to $\mathbf{x}$. Then the function $g(\mathbf{x}) = \int_{c_{\mathbf{x}}} \alpha$ is smooth and satisfies $dg = \alpha$.*

4.9. Example. Let $\alpha = \sum_{i=1}^n f_i\,dx_i$ be a closed 1-form defined on all of $\mathbf{R}^n$. Then we know that α is exact. One method for finding an antiderivative is explained in Exercise 2.8. Theorem 4.8 suggests a quicker way: let us choose the origin $\mathbf{0}$ to be the base point of $\mathbf{R}^n$ and for each $\mathbf{x}$ let $c_{\mathbf{x}} \colon [0,1] \to \mathbf{R}^n$ be the straight path $c_{\mathbf{x}}(t) = t\mathbf{x}$. Then $g(\mathbf{x}) = \int_{c_{\mathbf{x}}} \alpha = \int_0^1 c_{\mathbf{x}}^*(\alpha)$ is an antiderivative of α. Since

$$c_{\mathbf{x}}^*(\alpha) = \sum_{i=1}^n f_i(t\mathbf{x})\,d(tx_i) = \sum_{i=1}^n x_i f_i(t\mathbf{x})\,dt$$

we arrive at

$$g(\mathbf{x}) = \sum_{i=1}^n x_i \int_0^1 f_i(t\mathbf{x})\,dt.$$

(In Exercise 4.3 you will be asked to verify directly that $dg = \alpha$.) For instance, let $\alpha = y\,dx + (z\cos yz + x)\,dy + y\cos yz\,dz$ be the closed 1-form of Example 2.9. Then

$$\begin{aligned} g(\mathbf{x}) &= x\int_0^1 ty\,dt + y\int_0^1 (tz\cos t^2yz + tx)\,dt + z\int_0^1 ty\cos t^2yz\,dt \\ &= 2xy\int_0^1 t\,dt + 2yz\int_0^1 t\cos t^2yz\,dt \\ &= xy + \sin yz. \end{aligned}$$

See Exercises 4.4–4.6 for further applications of this theorem.

4.3. Angle functions 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.

Let $\mathbf{x} = (x, y)$ be a nonzero vector in the plane and let $\theta(\mathbf{x})$ be the angle between the positive x-axis and $\mathbf{x}$. Elementary trigonometry tells us that

$$\cos\theta(\mathbf{x}) = \frac{x}{\sqrt{x^2+y^2}}, \qquad \sin\theta(\mathbf{x}) = \frac{y}{\sqrt{x^2+y^2}}. \tag{4.4}$$

These equations determine $\theta(\mathbf{x})$ up to an integer multiple of 2π, and we will call any particular solution a *choice of angle for* $\mathbf{x}$.

Now let U be an open subset of the punctured plane $\mathbf{R}^2 \setminus \{\mathbf{0}\}$. Is it possible to make a choice of angle $\theta(\mathbf{x})$ for each $\mathbf{x}$ in U which varies smoothly with $\mathbf{x}$? For $U = \mathbf{R}^2 \setminus \{\mathbf{0}\}$ this would appear to be impossible: the usual choice of angle in the punctured plane ($0 \leq \theta(\mathbf{x}) < 2\pi$) has a discontinuity along the positive x-axis, and there seems to be no way to get rid of this discontinuity by making a cleverer choice of angle. But it may be possible if U is a smaller open subset, such as the complement of the positive x-axis in $\mathbf{R}^2 \setminus \{\mathbf{0}\}$. Let us define an *angle function on* U to be a smooth function $\theta\colon U \to \mathbf{R}$ with the property that $\theta(\mathbf{x})$ has property (4.4) for all $\mathbf{x} \in U$. Our next result states that an angle function on U exists if and only if α_0 is exact on U, where α_0 is the angle form

$$\alpha_0 = \frac{-y\,dx + x\,dy}{x^2+y^2}$$

introduced in Example 2.10.

4.10. Theorem. *Let U be a connected open subset of $\mathbf{R}^2 \setminus \{\mathbf{0}\}$.*

(i) *Let $\theta\colon U \to \mathbf{R}$ be an angle function. Then $d\theta = \alpha_0$.*
(ii) *Assume that α_0 is exact on U. Then there exists an angle function θ on U, which can be found as follows: choose a base point $\mathbf{x}_0$ in U, choose an angle θ_0 for $\mathbf{x}_0$, and for each $\mathbf{x} \in U$ choose a path $c_{\mathbf{x}}$ in U from $\mathbf{x}_0$ to $\mathbf{x}$. Then $\theta(\mathbf{x}) = \theta_0 + \int_{c_{\mathbf{x}}} \alpha_0$.*

Proof. (i) Define functions ξ and η on the punctured plane by

$$\xi(x,y) = \frac{x}{\sqrt{x^2+y^2}}, \qquad \eta(x,y) = \frac{y}{\sqrt{x^2+y^2}}. \tag{4.5}$$

Observe that

$$\begin{pmatrix} \xi(\mathbf{x}) \\ \eta(\mathbf{x}) \end{pmatrix} = \frac{\mathbf{x}}{\|\mathbf{x}\|}, \tag{4.6}$$

the unit vector pointing in the direction of $\mathbf{x} = (x, y)$. In particular

$$\xi(\mathbf{x})^2 + \eta(\mathbf{x})^2 = 1. \tag{4.7}$$

We will use the result

$$\alpha_0 = -\eta\, d\xi + \xi\, d\eta \tag{4.8}$$

of Exercise 2.4. If θ is an angle function on U, then $\xi(\mathbf{x}) = \cos\theta(\mathbf{x})$ and $\eta(\mathbf{x}) = \sin\theta(\mathbf{x})$ for all $\mathbf{x}$ in U. Substituting this into (4.8) gives

$$\alpha_0 = -\sin\theta\, d\cos\theta + \cos\theta\, d\sin\theta = \sin^2\theta\, d\theta + \cos^2\theta\, d\theta = d\theta.$$

(ii) Now assume that α_0 is exact on U. It follows from Theorem 4.8 that the function defined by $\theta(\mathbf{x}) = \theta_0 + \int_{c_{\mathbf{x}}} \alpha_0$ is smooth and satisfies

$$d\theta = \alpha_0. \tag{4.9}$$

To prove that θ is an angle function on U it is enough to show that the difference vector

$$\begin{pmatrix} \cos\theta(\mathbf{x}) \\ \sin\theta(\mathbf{x}) \end{pmatrix} - \begin{pmatrix} \xi(\mathbf{x}) \\ \eta(\mathbf{x}) \end{pmatrix}$$

has length 0 for all $\mathbf{x}$. The length squared of the difference vector is equal to

$$\begin{aligned}(\cos\theta - \xi)^2 + (\sin\theta - \eta)^2 = \cos^2\theta + \sin^2\theta - 2(\xi\cos\theta + \eta\sin\theta) + \xi^2 + \eta^2 \\ = 2 - 2(\xi\cos\theta + \eta\sin\theta),\end{aligned}$$

where we used (4.7). Hence we need to show that the function

$$u(\mathbf{x}) = \xi(\mathbf{x})\cos\theta(\mathbf{x}) + \eta(\mathbf{x})\sin\theta(\mathbf{x})$$

is a constant equal to 1. For $\mathbf{x} = \mathbf{x}_0$ we have

$$u(\mathbf{x}_0) = \xi(\mathbf{x}_0)\cos\theta_0 + \eta(\mathbf{x}_0)\sin\theta_0 = \cos^2\theta_0 + \sin^2\theta_0 = 1,$$

because by assumption θ_0 is a choice of angle for $\mathbf{x}_0$. Furthermore, the exterior derivative of u is

$$\begin{aligned} du &= \cos\theta\, d\xi - \xi\sin\theta\, d\theta + \sin\theta\, d\eta + \eta\cos\theta\, d\theta \\ &= \cos\theta\, d\xi + \sin\theta\, d\eta + (-\xi\sin\theta + \eta\cos\theta)\alpha_0 && \text{by (4.9)} \\ &= \cos\theta\, d\xi + \sin\theta\, d\eta + (-\xi\sin\theta + \eta\cos\theta)(-\eta\, d\xi + \xi\, d\eta) && \text{by (4.8)} \\ &= \big((1-\eta^2)\cos\theta + \xi\eta\sin\theta\big)\, d\xi + \big(\xi\eta\cos\theta + (1-\xi^2)\sin\theta\big)\, d\eta) \\ &= (\xi\cos\theta + \eta\sin\theta)\xi\, d\xi + (\xi\cos\theta + \eta\sin\theta))\eta\, d\eta && \text{by (4.7)} \\ &= \frac{1}{2}(\xi\cos\theta + \eta\sin\theta)\, d(\xi^2 + \eta^2) \\ &= 0 && \text{by (4.7).} \end{aligned}$$

Hence, by Lemma 3.12, u is a constant function, so $u(\mathbf{x}) = 1$ for all $\mathbf{x}$ in U. QED

4.11. Example. There exists no angle function on the punctured plane, because α_0 is not exact on $\mathbf{R}^2 \setminus \{\mathbf{0}\}$ (as was shown in Example 4.6).

Angle functions along a path. Now let us modify the angle problem by considering a path $c\colon [a,b] \to \mathbf{R}^2 \setminus \{\mathbf{0}\}$ in the punctured plane and wondering whether we can make a choice of angle for $c(t)$ which depends smoothly on $t \in [a,b]$. We define an *angle function along* c to be a smooth function $\vartheta\colon [a,b] \to \mathbf{R}$ with the property that

$$\cos\vartheta(t) = \xi(c(t)), \qquad \sin\vartheta(t) = \eta(c(t))$$

for all $t \in [a,b]$. The following example shows the difference between this notion and that of an angle function on an open subset.

4.12. Example. Define $c\colon [a,b] \to \mathbf{R}^2 \setminus \{\mathbf{0}\}$ by $c(t) = (\cos t, \sin t)$. This is a path travelling along the unit circle at constant angular velocity 1 from time a to time b. The function $\vartheta(t) = t$ is an angle function along c. The difference $\vartheta(b) - \vartheta(a) = b - a$ is the total angle swept out by the path.

In fact an angle function exists along every path in the punctured plane.

4.13. Theorem. *Let $c\colon [a,b] \to \mathbf{R}^2 \setminus \{\mathbf{0}\}$ be a path. Choose an angle ϑ_a for $c(a)$ and for $a \le t \le b$ define*

$$\vartheta(t) = \vartheta_a + \int_a^t c^*(\alpha_0).$$

The ϑ is an angle function along c.

Proof. This proof is nearly the same as that of Theorem 4.10, so we will skip some details. We will show that the difference vector

$$\begin{pmatrix}\cos\vartheta(t)\\ \sin\vartheta(t)\end{pmatrix} - \begin{pmatrix}\xi(c(t))\\ \eta(c(t))\end{pmatrix}$$

has length 0 for all $t \in [a,b]$. The square of the length is equal to

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

where we introduced the abbreviations

$$f(t) = \xi(c(t)), \qquad g(t) = \eta(c(t)), \qquad u(t) = f(t)\cos\vartheta(t) + g(t)\sin\vartheta(t),$$

and where we used that

$$f^2 + g^2 = 1. \tag{4.10}$$

Therefore it is enough to show that the function u is a constant equal to 1. Now $\vartheta(a) = \vartheta_a$, so $f(a) = \cos\vartheta(a)$ and $g(a) = \sin\vartheta(a)$ (because ϑ_a is a choice of angle for $c(a)$), and hence $u(a) = \cos^2\vartheta_a + \sin^2\vartheta_a = 1$. So we can finish the proof by showing that $u'(t) = 0$ for all t. Using (4.8) we find

$$\begin{aligned}c^*(\alpha_0) = c^*(-\eta\, d\xi + \xi\, d\eta) = -c^*(\eta)\, dc^*(\xi) + c^*(\xi)\, dc^*(\eta) \\ = -g\, df + f\, dg = (-gf' + fg')\, dt,\end{aligned}$$

and therefore, by the fundamental theorem of calculus,

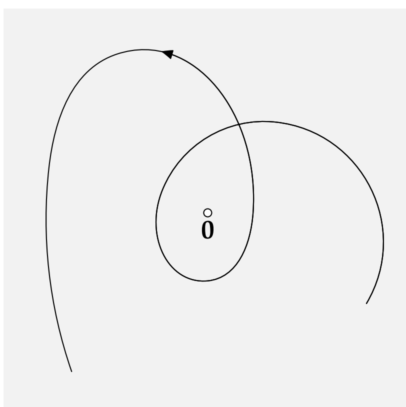
$$\vartheta' = -gf' + fg'. \tag{4.11}$$

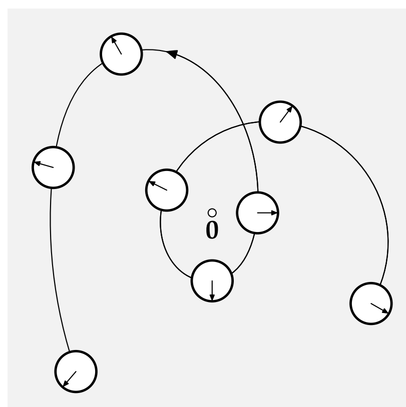
This yields

$$\begin{aligned}
u' &= f'\cos\vartheta + g'\sin\vartheta + (-f\sin\vartheta + g\cos\vartheta)\vartheta' \\
&= f'\cos\vartheta + g'\sin\vartheta + (-f\sin\vartheta + g\cos\vartheta)(-gf' + fg') && \text{by (4.11)} \\
&= ((1-g^2)\cos\vartheta + fg\sin\vartheta)f' + (fg\cos\vartheta + (1-f^2)\sin\vartheta)g' \\
&= (f\cos\vartheta + g\sin\vartheta)ff' + (f\cos\vartheta + g\sin\vartheta))gg' && \text{by (4.10)} \\
&= \frac{1}{2}(f\cos\vartheta + g\sin\vartheta)(f^2+g^2)' \\
&= 0 && \text{by (4.10).}
\end{aligned}$$

QED

It is helpful to think of the vector $(\xi(c(t)), \eta(c(t))) = c(t)/\|c(t)\|$ as a compass held by a traveller wandering through a magnetized punctured plane. The puncture represents the magnetic north pole, so the needle indicates the direction of the traveller's position vector $c(t)$.





As t increases from a to b, the needle starts at the angle $\vartheta(a) = \vartheta_a$, it moves around the compass, and ends up at the final angle $\vartheta(b)$. The difference $\vartheta(b) - \vartheta(a)$ measures the net angle swept out by the needle, where a counterclockwise motion counts as positive and a clockwise motion counts as negative. The formula for ϑ given by Theorem 4.13 shows that this difference can also be expressed as an integral,

$$\vartheta(b) - \vartheta(a) = \int_a^b c^*(\alpha_0) = \int_c \alpha_0. \tag{4.12}$$

4.14. Corollary. *If $c\colon [a,b] \to U$ is a closed path, then $\int_c \alpha_0 = 2\pi k$, where k is an integer.*

Proof. Because of (4.12) it suffices to show that $\vartheta(b) - \vartheta(a) = 2\pi k$. By Theorem 4.13, ϑ is an angle function along c, so the assumption that c is closed ($c(a) = c(b)$) implies

$$\begin{pmatrix}\cos\vartheta(a)\\ \sin\vartheta(a)\end{pmatrix} = \begin{pmatrix}\xi(c(a))\\ \eta(c(a))\end{pmatrix} = \begin{pmatrix}\xi(c(b))\\ \eta(c(b)\end{pmatrix} = \begin{pmatrix}\cos\vartheta(b)\\ \sin\vartheta(b)\end{pmatrix}.$$

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

The integer $k = (2\pi)^{-1} \int_c \alpha_0$ measures how many times the path loops around the origin. It is called the *winding number* of the closed path c about the origin, and we will denote it by $w(c, \mathbf{0})$.

$$\boxed{w(c, \mathbf{0}) = \text{winding number of } c \text{ about origin} = \frac{1}{2\pi} \int_c \alpha_0.} \tag{4.13}$$

4.15. Example. It follows from the calculation in Example 4.1 that the winding number of the circle $c(t) = (\cos t, \sin t)$ $(0 \le t \le 2\pi)$ is equal to 1.

Exercises

4.1. Consider the path $c\colon [0, \frac{1}{2}\pi] \to \mathbf{R}^2$ defined by $c(t) = (a \cos t, b \sin t)$, where a and b are positive constants. Let $\alpha = xy\, dx + x^2 y\, dy$.

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

4.2. Restate Theorem 4.7 in terms of force fields, potentials and energy. Explain why the result is plausible on physical grounds.

4.3. Let $\alpha = \sum_{i=1}^n f_i\, dx_i$ be a *closed* 1-form defined on all of $\mathbf{R}^n$.

(i) Verify that the function $g(\mathbf{x}) = \sum_{i=1}^n x_i \int_0^1 f_i(t\mathbf{x})\, dt$ given in Example 4.9 satisfies $dg = \alpha$.
(ii) In Exercise 2.8 a different formula for the antiderivative g of α was given. Show that that formula can be interpreted as an integral $g(\mathbf{x}) = \int_{c_\mathbf{x}} \alpha$ of α along a suitable path $c_\mathbf{x}$ from the origin to $\mathbf{x}$.

4.4. Consider the 1-form $\alpha = \|\mathbf{x}\|^a \sum_{i=1}^n x_i\, dx_i$ on $\mathbf{R}^n \setminus \{\mathbf{0}\}$, 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 $dg = \alpha$.

4.5. Let $\alpha \in \Omega^1(\mathbf{R}^n \setminus \{\mathbf{0}\})$ be the 1-form of Exercise 4.4. Now let $c_\mathbf{x}$ be the half-line 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.5.)

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

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

4.8. Let $\mathbf{x}_0$ be a point in $\mathbf{R}^2$ and c a closed path which does not pass through $\mathbf{x}_0$. How would you define the winding number $w(c, \mathbf{x}_0)$ of c around $\mathbf{x}_0$? Try to formulate two different definitions: a "geometric" definition and a formula in terms of an integral over c of a certain 1-form analogous to formula (4.13).

4.9. Let $c\colon [0,1] \to \mathbf{R}^2 \setminus \{\mathbf{0}\}$ be a closed path with winding number k. Determine the winding numbers of the following paths $\tilde{c}\colon [0,1] \to \mathbf{R}^2 \setminus \{\mathbf{0}\}$ 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 $\phi(x,y) = (y,x)$;
(v) $\tilde{c}(t) = \phi(c(t))$, where $\phi(x,y) = \dfrac{1}{x^2+y^2}(x,-y)$.

4.10. For each of the following closed paths $c\colon [0,2\pi] \to \mathbf{R}^2 \setminus \{\mathbf{0}\}$ 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 path (the use of software is allowed) and obtain the answer geometrically.

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

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

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

(i) Sketch the path c for $a = \pm b/3$.
(ii) For what values of a and b is the path 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

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

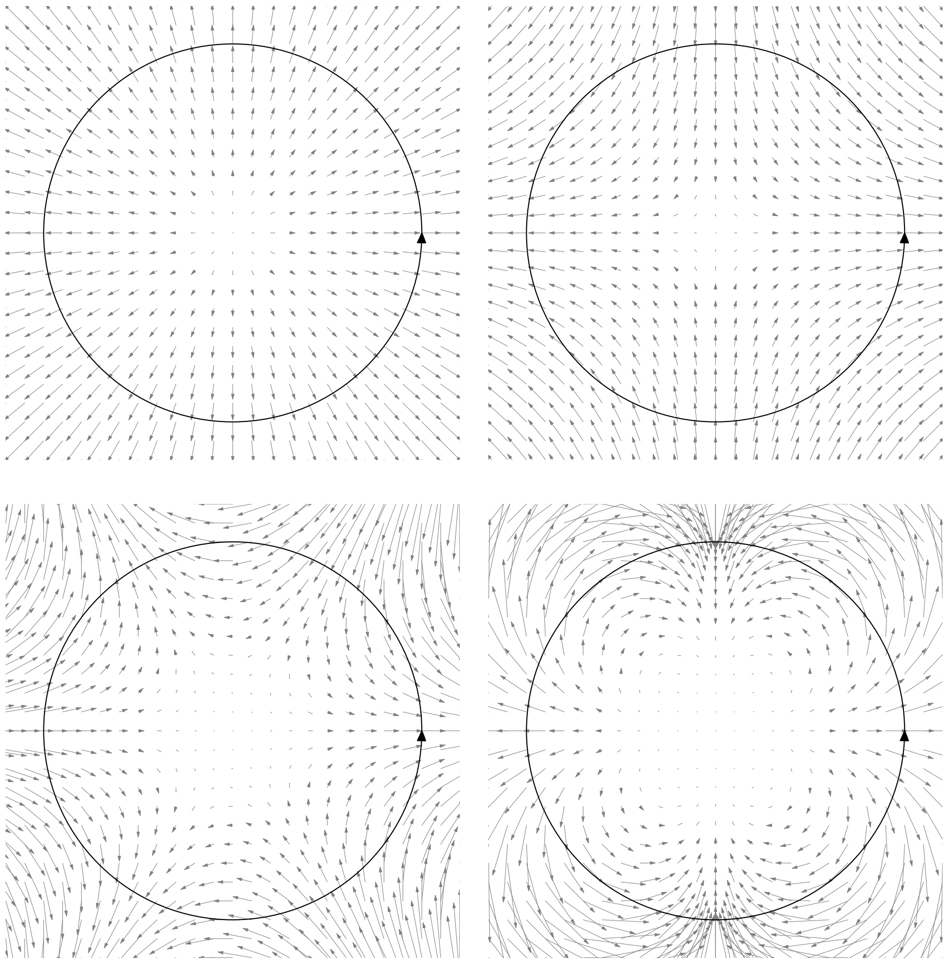
is well-defined at all points $\mathbf{x}$ of U where $\mathbf{F}(\mathbf{x}) \neq 0$. Let c be a parametrized circle contained in U, traversed once in the counterclockwise direction. Assume that $\mathbf{F}(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in c$. The *index* of $\mathbf{F}$ relative to c is

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

Prove the following assertions.

(i) $\beta = \mathbf{F}^*(\alpha_0)$, where α_0 is the angle form $(-y\,dx + x\,dy)/(x^2+y^2)$;
(ii) β is closed;
(iii) $\operatorname{index}(\mathbf{F}, c) = w(\mathbf{F}\circ c, \mathbf{0})$, the winding number of the path $\mathbf{F}\circ c$ 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.

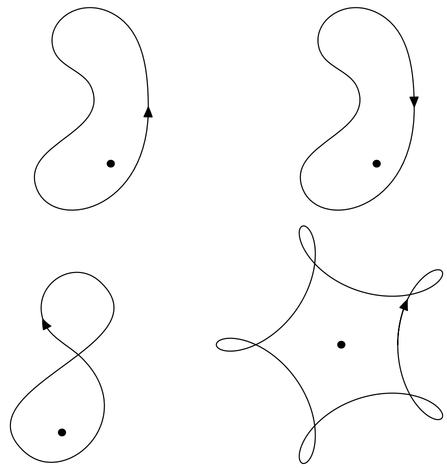
4.14. Let $c\colon [a,b] \to \mathbf{R}^2$ be a path. For each $t \in [a,b]$ the velocity vector $c'(t)$ is tangent to c at the point $c(t)$. The map $c'\colon [a,b] \to \mathbf{R}^2$ is called the *derived path* of c. Let us assume that both c and c' are closed, i.e. $c(a) = c(b)$ and $c'(a) = c'(b)$. Let us also assume that the path c is *regular* in the sense that $c'(t) \neq \mathbf{0}$ for all t. The quantity

$$\tau(c) = \frac{1}{2\pi} \int_{c'} \alpha_0$$

is then well-defined and is called the *turning number* of c. Here α_0 denotes the angle form on $\mathbf{R}^2 \setminus \{\mathbf{0}\}$.

(i) Show that the turning number is an integer. Discuss its geometric meaning and how it differs from the winding number.
(ii) Show that the turning number changes sign under reversal of orientation in the sense that $\tau(\tilde{c}) = -\tau(c)$, where $\tilde{c}(t) = c(a + b - t)$ for $a \leq t \leq b$.

(iii) For each of the following four paths c find the turning number $\tau(c)$, as well as the winding number $w(c, \bullet)$ about the black point $\bullet$.



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

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

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 $h(\mathbf{t})\,dt_1\,dt_2 \cdots dt_k$ for some function $h \colon R \to \mathbf{R}$. The *integral* of α over c is defined by

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

(The definition of the integral 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.)

For $k = 1$ this reproduces the definition given in Chapter 4.

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 completely 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 $c^*(\alpha) = h\, dt_1\, dt_2 \cdots dt_k$ and $\mathbf{t} = p(\mathbf{s})$. Then
$$p^*(c^*(\alpha)) = p^*(h\, dt_1\, dt_2 \cdots dt_k) = p^*(h) \det(Dp)\, ds_1\, ds_2 \cdots ds_k$$
by Theorem 3.14, so
$$\int_{c\circ p} \alpha = \int_{\bar{R}} h(p(\mathbf{s})) \det(Dp(\mathbf{s}))\, ds_1\, ds_2 \cdots ds_k.$$
On the other hand, $\int_c \alpha = \int_R h(t)\, dt_1\, dt_2 \cdots dt_k$, so by the substitution formula, Theorem B.9, 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 path $c\colon [a, b] \to U$ can be reparametrized to a path $c \circ p\colon [0, 1] \to U$ 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
$$A = \begin{pmatrix} b_1 - a_1 & 0 & \dots & 0 \\ 0 & b_2 - a_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_k - a_k \end{pmatrix} \quad \text{and} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}.$$
("Squeeze the unit cube until it has the same edgelengths as R and then translate it to the same position as 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
$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(t_1, t_2)\, dt_1\, dt_2 = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(t_1, t_2)\, dt_2\, dt_1. \tag{5.1}$$
(This follows for instance from the substitution formula, Theorem B.9.) 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

$$\int_c \alpha = \sum_{q=1}^{p} a_q \int_{c_q} \alpha.$$

(In the language of linear algebra, the k-chains form a 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.)

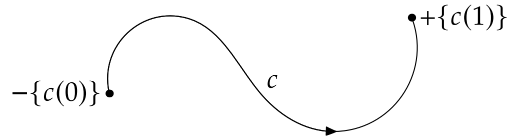
Recall that a 0-cube is nothing but a singleton $\{\mathbf{x}\}$ consisting of a single point $\mathbf{x}$ in U. Thus a 0-chain is a formal linear combination of points, $c = \sum_{q=1}^{p} a_q \{\mathbf{x}_q\}$. A good way to think of a 0-chain c is as a collection of p point charges, with an electric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

$$\int_c f = \sum_{q=1}^{p} a_q \int_{\{\mathbf{x}_q\}} f = \sum_{q=1}^{p} a_q f(\mathbf{x}_q).$$

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

5.2. The boundary of a chain

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

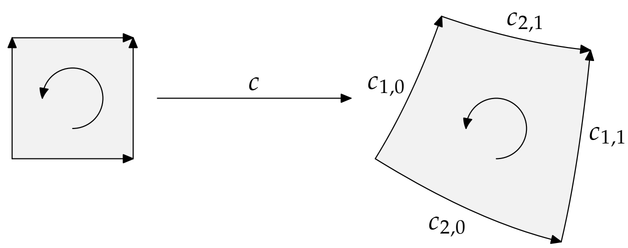


We will define the boundary ∂c of a k-cube $c\colon [0,1]^k \to U$ for $k \geq 1$ similarly, namely as an alternating sum over the $k-1$-dimensional faces of c. There are $2k$ such faces, which can be conveniently labelled as follows. For $\mathbf{t} = (t_1, t_2, \dots, t_{k-1})$ in $[0,1]^{k-1}$ and for $i = 1, 2, \dots, k$ put

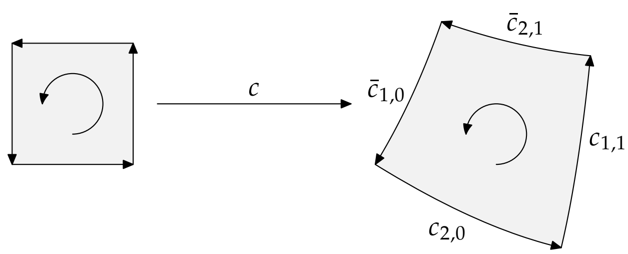
$$\begin{aligned} c_{i,0}(\mathbf{t}) &= c(t_1, t_2, \dots, t_{i-1}, 0, t_i, \dots, t_{k-1}), \\ c_{i,1}(\mathbf{t}) &= c(t_1, t_2, \dots, t_{i-1}, 1, t_i, \dots, t_{k-1}). \end{aligned}$$

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

For instance, a 2-cube $c\colon [0,1]^2 \to U$ has four edges, namely the "left" edge $c_{1,0}$, the "right" edge $c_{1,1}$, the "bottom" edge $c_{2,0}$, and the "top" edge $c_{2,1}$.



The picture suggests that we should define $\partial c = c_{2,0} + c_{1,1} - c_{2,1} - c_{1,0}$, because the orientation of the 2-cube goes "with" the edges $c_{2,0}$ and $c_{1,1}$ and "against" the edges $c_{2,1}$ and $c_{1,0}$. (Alternatively we could reverse the orientations of the top and left edges by defining $\bar{c}_{2,1}(t) = c(1-t, 1)$ and $\bar{c}_{1,0}(t) = c(0, 1-t)$, and then put $\partial c = c_{2,0} + c_{1,1} + \bar{c}_{2,1} + \bar{c}_{1,0}$. This corresponds to the following picture:



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

For any $k \geq 1$ we now define the boundary of a k-cube $c\colon [0,1]^k \to U$ by

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

This definition is consistent with the one- and two-dimensional cases considered above. The boundary of a 0-cube is by convention equal to 0. For an arbitrary

k-chain c we define the boundary ∂c by writing c as a formal linear combination $c = \sum_q a_q c_q$ of k-cubes c_q with real coefficients a_q, and by putting $\partial c = \sum_q a_q \partial c_q$. This definition makes ∂ a linear map from k-chains to $k-1$-chains.

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 ∂ we may assume without loss of generality that c is a single k-cube $c\colon [0,1]^k \to U$. Then 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 as a sum over $i \le j$ and a sum over $i > j$ to give

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

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 $1 \le i \le j \le k-1$. It follows that

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

where in the first line we substituted $(c_{i,\rho})_{j,\sigma} = (c_{j+1,\sigma})_{i,\rho}$ and in the second line we substituted $r = j+1$, $s = i$, $\mu = \sigma$, and $\nu = \rho$. Thus the two terms on the right-hand side of (5.2) cancel out. QED

5.3. Cycles and boundaries

Let $k \geq 1$. A k-cube c is *degenerate* if $c(t_1, \ldots, t_k)$ 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 path. 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

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

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

$$\begin{aligned} f(t_1, \ldots, t_i, \ldots, t_k) &= (t_1, \ldots, \hat{t}_i, \ldots, t_k), \\ g(s_1, \ldots, s_{k-1}) &= c(s_1, \ldots, s_{i-1}, 0, s_{i+1}, \ldots, s_{k-1}). \end{aligned}$$

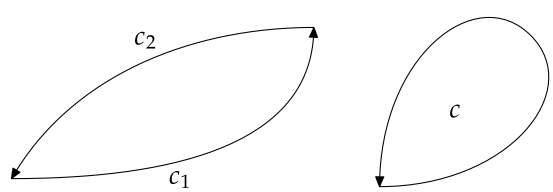
Now $g^*(\alpha)$ is a k-form on $[0,1]^{k-1}$ and hence equal to 0, and so

$$c^*(\alpha) = f^*(g^*(\alpha)) = 0.$$

We conclude that $\int_c \alpha = \int_{[0,1]^k} c^*(\alpha) = 0$. QED

So degenerate chains are irrelevant in so far as 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 $c = \partial b + a$ for some $k+1$-chain b and some degenerate k-chain a.

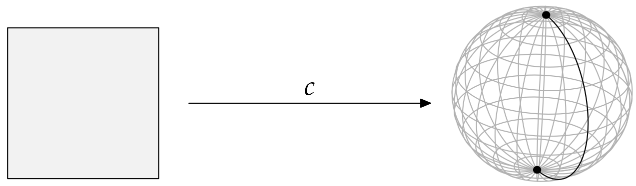
5.6. Example. If c_1 and c_2 are paths arranged head to tail as in the picture below, then $\partial(c_1 + c_2) = 0$, so $c_1 + c_2$ is a 1-cycle. The closed path c satisfies $\partial c = 0$, so it is a 1-cycle as well.



5.7. Example. The 2-cube $c\colon [0,1]^2 \to \mathbf{R}^3$ defined by

$$c(t_1, t_2) = (\cos 2\pi t_1 \sin \pi t_2, \cos 2\pi t_1 \sin \pi t_2, \cos \pi t_2).$$

parametrizes the unit sphere. The left and right edges of the unit square are mapped onto a single meridian of the sphere, the bottom edge is mapped to the north pole, and the top edge to the south pole.



The sphere has no boundary, so one might expect that c is a cycle. Indeed, we have

$$c_{1,0}(t) = c_{1,1}(t) = (\sin \pi t, \sin \pi t, \cos \pi t), \quad c_{2,0}(t) = (0, 0, 1), \quad c_{2,1}(t) = (0, 0, -1),$$

and therefore

$$\partial c = -c_{1,0} + c_{1,1} + c_{2,0} - c_{2,1} = c_{2,0} - c_{2,1}$$

is a degenerate 1-chain and c is a cycle.

5.8. Example. Let $c\colon [0,1] \to \mathbf{R}^2$ be the path $c(t) = (\cos 2\pi t, \sin 2\pi t)$. This is a closed 1-cube, which parametrizes the unit circle. The circle is the boundary of the disc of radius 1 and therefore it is reasonable to expect that the 1-cube c is a boundary. To show that this is the case we will display a 2-cube b and a constant 1-chain a satisfying $c = \partial b + a$. 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 $\partial b = c - a$, where a is the constant path located at the origin. Therefore $c = \partial b + a$, which proves that c is a boundary. You may wonder whether we really need the degenerate path a. Isn't it possible to find a 2-chain b with the property that $c = \partial b$? Exercise 5.2 shows that this is not the case.

5.9. 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 $\mathbf{t} = (t_1, t_2, \ldots, t_{k-1})$. 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

5.10. Corollary. *Every boundary is a cycle.*

Proof. Suppose $c = \partial b + a$ with a degenerate. Then by Lemma 5.5 $\partial c = \partial(\partial b) + \partial a = \partial a$, where we used Proposition 5.4. Lemma 5.9 says that ∂a is degenerate, and therefore so is ∂c. QED

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 5.13.

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".

5.11. 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. Let us assume, as we may, that $c\colon [0,1]^k \to U$ is a single k-cube. 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 (see 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

5.12. Corollary. *Let U be an open subset of $\mathbf{R}^n$, let c be a k-chain in U, and let α be a k-form on U. Then $\int_c \alpha = 0$ if either of the following two conditions holds:*

(i) *c is a cycle and α is exact; or*
(ii) *c is a boundary and α is closed.*

Proof. (i) If c is a cycle, then ∂c is a degenerate $k-1$-chain. If α is exact, then $\alpha = d\beta$ for some $k-1$-form β. Therefore $\int_c \alpha = \int_c d\beta = \int_{\partial c} \beta = 0$ by Stokes' theorem and by Lemma 5.5.

(ii) If c is a boundary, then $c = \partial b + a$ for some $k+1$-chain b and some degenerate k-chain a. If α is closed, then $d\alpha = 0$. Therefore $\int_c \alpha = \int_{\partial b + a} \alpha = \int_b d\alpha + \int_a \alpha = 0$ by Stokes' theorem and by Lemma 5.5. QED

5.13. Example. The unit circle $c(t) = (\cos 2\pi t, \sin 2\pi t)$ $(0 \le t \le 1)$ is a 1-cycle in the punctured plane $U = \mathbf{R}^2 \setminus \{\mathbf{0}\}$. Considered as a chain in $\mathbf{R}^2$ it is also a boundary, as we saw in Example 5.8. However, we claim that it is not a boundary in U: it is impossible to find a 2-chain b and a degenerate 1-chain a *both contained in* U such that $c = \partial b + a$. Indeed, suppose this was possible. Then $\int_c \alpha_0 = 0$ by Corollary 5.12, where α_0 is the angle form, because α_0 is closed. On the other hand, by Example 4.1 we have $\int_c \alpha_0 = 2\pi$. This is a contradiction, so we conclude that c is not a boundary in U. The presence of the puncture in U is responsible both for the existence of the non-exact closed 1-form α_0 (see Example 4.6) and for the non-bounding closed 1-chain c.

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 $\mathbf{R}^n$.

(i) Let b be a $k+1$-chain in U. In Section 5.2 we defined the boundary of b as a certain 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. Prove 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 $c(t_1, t_2) = (t_1^2, t_1 t_2, t_2^2)$, 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 $c(t_1,t_2,t_3) = (t_2t_3, t_1t_3, t_1t_2)$, 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 (see Exercise 3.19) write the $n-1$-dimensional unit sphere S^{n-1} in $\mathbf{R}^n$ as the image of an $n-1$-cube c. (The domain of c will not be the unit cube in $\mathbf{R}^{n-1}$, but a rectangular block R. Choose R in such a way as to cover the sphere as economically as possible.) For $n = 2$, 3, 4, calculate the boundary ∂c and show that c is a cycle. (See also Example 5.7.)

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 path c.
(ii) Green's formula: $\displaystyle\int_c \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx\,dy = \int_{\partial c} (f\,dx + g\,dy)$ 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 8.17.)

5.7. An $\mathbf{R}^r$*-valued* k*-form* on an open subset U of $\mathbf{R}^n$ is a vector

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{pmatrix},$$

whose entries $\alpha_1, \alpha_2, \dots, \alpha_r$ are ordinary k-forms defined on U. The number r can be any nonnegative integer and does not have to be related to n or k. Sometimes we leave the value of r unspecified and speak of $\boldsymbol{\alpha}$ as a *vector-valued k-form*. Examples of vector-valued differential forms are the $\mathbf{R}^n$-valued 1-form $d\mathbf{x}$ and the $\mathbf{R}^n$-valued $n-1$-form $*d\mathbf{x}$ on $\mathbf{R}^n$ introduced in Section 2.5. The *integral* of an $\mathbf{R}^r$-valued k-form $\boldsymbol{\alpha}$ over a k-chain c in U is defined by

$$\int_c \boldsymbol{\alpha} = \begin{pmatrix} \int_c \alpha_1 \\ \int_c \alpha_2 \\ \vdots \\ \int_c \alpha_r \end{pmatrix}.$$

This integral is a *vector* in $\mathbf{R}^r$. Prove the following extensions of Gauss' divergence formula to vector-valued forms.

(i) The gradient version: $\displaystyle\int_c \operatorname{grad}(f)\,dx_1\,dx_2\cdots dx_n = \int_{\partial c} f\,{*d\mathbf{x}}$ for any function f and any n-chain c. (Here $r = n$.)

(ii) The curl version: $\displaystyle\int_c \operatorname{curl}(\mathbf{F})\,dx_1\,dx_2\,dx_3 = -\int_{\partial c} \mathbf{F}\times d\mathbf{x}$ for any vector field $\mathbf{F}$ and any 3-chain c. (Here $r = n = 3$.)

Both formulas can be deduced from Exercise 5.6(iii) by considering each component of the vector-valued integrals separately. Gauss' formulas look more suggestive if we employ the nabla notation of vector calculus. Writing $\nabla f = \operatorname{grad}(f)$, $\nabla\cdot\mathbf{F} = \operatorname{div}(\mathbf{F})$ and $\nabla\times\mathbf{F} = \operatorname{curl}(\mathbf{F})$ we get:

$$\int_c \nabla\cdot\mathbf{F}\,dx_1\,dx_2\cdots dx_n = \int_{\partial c}\mathbf{F}\cdot *d\mathbf{x}$$
$$\int_c \nabla f\,dx_1\,dx_2\cdots dx_n = \int_{\partial c} f *d\mathbf{x}$$
$$\int_c \nabla\times\mathbf{F}\,dx_1\,dx_2\,dx_3 = -\int_{\partial c}\mathbf{F}\times *d\mathbf{x}.$$

5.8. Prove the following extensions of Stokes' formula to $\mathbf{R}^3$-valued forms on $\mathbf{R}^3$. (See Exercise 5.7 for an explanation of vector-valued forms.)

(i) The gradient version: $\displaystyle\int_c \operatorname{grad}(f)\times *d\mathbf{x} = -\int_{\partial c} f\,d\mathbf{x}$ for any function f and any 2-chain c.

(ii) The divergence version: $\displaystyle\int_c (\operatorname{div}(\mathbf{F}) - D\mathbf{F}^T) *d\mathbf{x} = \int_{\partial c}\mathbf{F}\times d\mathbf{x}$ for any vector field $\mathbf{F}$ and any 2-chain c. Here $D\mathbf{F}^T$ denotes the transpose of the Jacobi matrix of $\mathbf{F}$. The expression $(\operatorname{div}(\mathbf{F}) - D\mathbf{F}^T) *d\mathbf{x}$ is shorthand for $\operatorname{div}(\mathbf{F}) *d\mathbf{x} - D\mathbf{F}^T *d\mathbf{x}$. The first term $\operatorname{div}(\mathbf{F}) *d\mathbf{x}$ is the product of the function $\operatorname{div}(\mathbf{F})$ and the $\mathbf{R}^3$-valued form $*d\mathbf{x}$ and the second term $D\mathbf{F}^T *d\mathbf{x}$ is the product of the 3×3-matrix $D\mathbf{F}^T$ and the $\mathbf{R}^3$-valued 2-form $*d\mathbf{x}$.

In nabla notation Stokes' formulas look as follows:

$$\int_c (\nabla\times\mathbf{F})\cdot *d\mathbf{x} = \int_{\partial c}\mathbf{F}\cdot d\mathbf{x}$$
$$\int_c \nabla f\times *d\mathbf{x} = -\int_{\partial c} f\,d\mathbf{x}$$
$$\int_c (\nabla\cdot\mathbf{F} - D\mathbf{F}^T) *d\mathbf{x} = \int_{\partial c}\mathbf{F}\times d\mathbf{x}.$$

CHAPTER 6

Manifolds

6.1. The definition

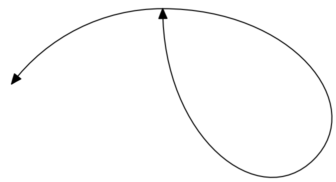
Intuitively, an n-dimensional manifold in the Euclidean space $\mathbf{R}^N$ is a subset that in the neighbourhood of every point "looks like" $\mathbf{R}^n$ up to "smooth distortions". The formal definition is given below and is a bit long. It will help to consider first the basic example of the surface of the earth, which is a two-dimensional sphere placed in three-dimensional space. Geographers describe the earth 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 $\mathbf{t} \in U$. You should think of $\psi(U)$ as an n-dimensional "patch" in $\mathbf{R}^N$ parametrized by the map ψ. The inverse map ψ^{-1} is called a *chart* or *coordinate map*. It maps each point in the patch $\psi(U)$ to an n-tuple of numbers, which we think of as the "coordinates" of the point. 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 condition is imposed to prevent the occurrence of cusps and other singularities in the image $\psi(U)$. Since $D\psi(\mathbf{t})$ has N rows, condition (ii) 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. Condition (iii) can be restated as follows: 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. 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,

6.3. Definition. An *n-dimensional manifold*[1] (or *n-manifold* for short) in $\mathbf{R}^N$ is a subset M of $\mathbf{R}^N$ with the property that for each $\mathbf{x} \in M$ there exist

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

Such an embedding ψ is called a *local parametrization of M at* $\mathbf{x}$. Its inverse $\psi^{-1}\colon \psi(U) \to U$ is a *chart of M at* $\mathbf{x}$. An *atlas* of M is a collection of local parametrizations $\psi_i\colon U_i \to \mathbf{R}^N$ of M with the property that M is the union of all the sets $\psi_i(U_i)$. The *tangent space* to M at a point $\mathbf{x} \in M$ is the column space of $D\psi(\mathbf{t})$,

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

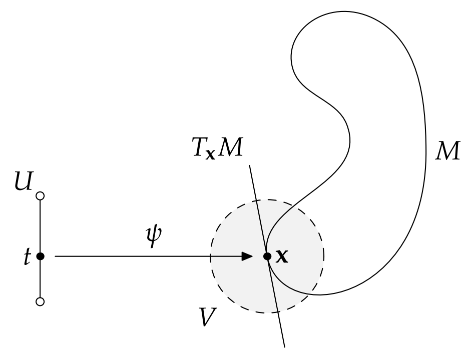
where $\psi\colon U \to \mathbf{R}^N$ is a local parametrization of M at $\mathbf{x}$ and $\mathbf{t}$ is the unique vector in U satisfying $\psi(\mathbf{t}) = \mathbf{x}$. The elements of $T_{\mathbf{x}}M$ are *tangent vectors* to M at $\mathbf{x}$. The *codimension* of M in $\mathbf{R}^N$ is the number $N - n$.

[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.

Note that the tangent space $T_{\mathbf{x}}M$ at each point $\mathbf{x}$ of an n-manifold M is an n-dimensional linear subspace of $\mathbf{R}^N$. The reason is that for every local parametrization ψ of M the Jacobi matrix $D\psi(\mathbf{t})$ has n independent columns.

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.

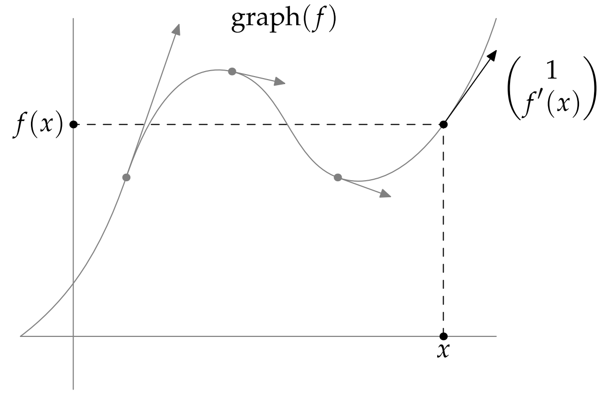
The following picture illustrates the definition. Here M is a curve in the plane, so we have $N = 2$ and $n = 1$. The open set U is an interval in $\mathbf{R}$ and V is an open disc in $\mathbf{R}^2$. The map ψ sends t to $\mathbf{x}$ and parametrizes the portion of the curve inside V. Since $n = 1$, 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.



6.4. 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.

6.5. 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 (x, y)-plane.) 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 the linear subspace $\psi(\mathbf{R}^n)$ of $\mathbf{R}^N$.

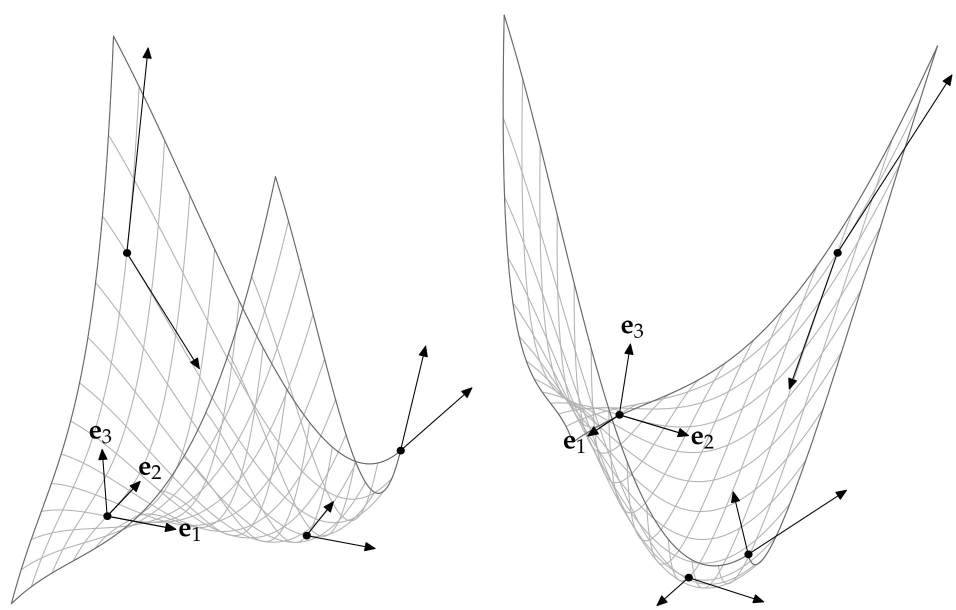
6.6. Example. Let $M = \mathrm{graph}(f)$, where $f\colon U \to \mathbf{R}^m$ is a smooth map. As shown in Example 6.2, M is the image of a single embedding $\psi\colon U \to \mathbf{R}^{n+m}$, so M is an n-dimensional manifold in $\mathbf{R}^{n+m}$, covered by a single chart. At a point $(\mathbf{x}, f(\mathbf{x}))$ in the graph the tangent space is spanned by the columns of $D\psi(\mathbf{x})$. 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 familiar 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

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

The figure below shows the graph of the cubic function $f(x, y) = \frac{1}{2}(x^3 + y^3 - 3xy)$ from two different angles, together with a few points and tangent vectors.

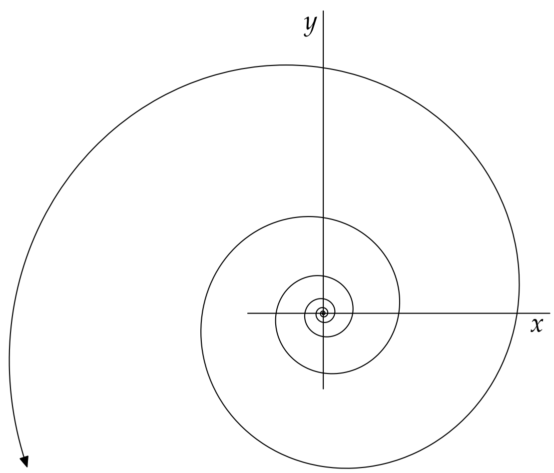


6.7. Example. Consider the path $\psi : \mathbf{R} \to \mathbf{R}^2$ given by $\psi(t) = e^{at}(\cos bt, \sin bt)$, where a and b are nonzero constants. Let us check that ψ is an embedding. Observe first that $\|\psi(t)\| = e^{at}$. Therefore $\psi(t_1) = \psi(t_2)$ implies $e^{at_1} = e^{at_2}$. The exponential function is one-to-one, so $t_1 = t_2$ (since $a \neq 0$). This shows that ψ is one-to-one.

The velocity vector is

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

The 2×2-matrix in this formula is a rotation matrix and hence invertible. The vector $\binom{a}{b}$ is nonzero, and therefore $\psi'(t) \neq \mathbf{0}$ for all t. Moreover we have $t = a^{-1}\ln e^{at} = a^{-1}\ln\|\psi(t)\|$. Hence the inverse of ψ is given by $\psi^{-1}(\mathbf{x}) = a^{-1}\ln\|\mathbf{x}\|$ 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 winds infinitely many times around the origin and which for $t \to -\infty$ converges to the origin.



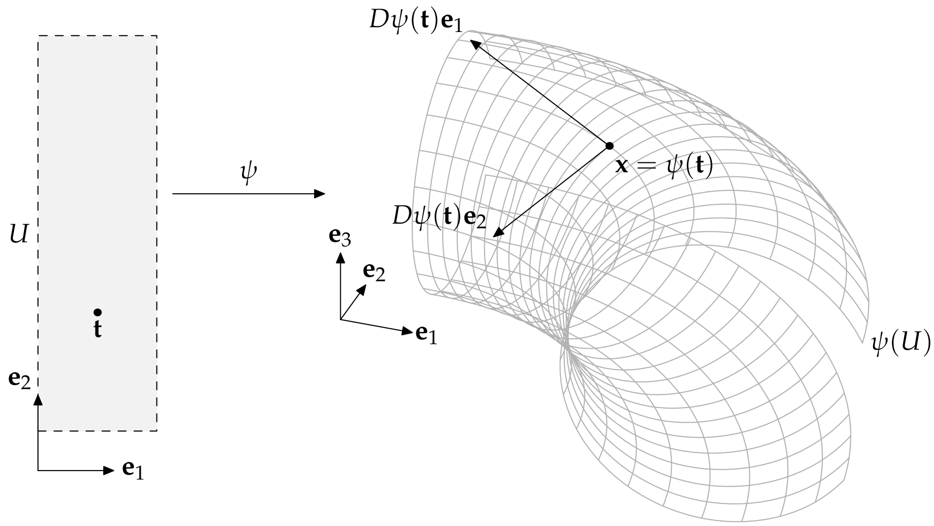
Even though $\psi(\mathbf{R})$ is a manifold, the set $\psi(\mathbf{R}) \cup \{\mathbf{0}\}$ is not: it has a very nasty singularity at the origin!

The manifolds of Examples 6.4–6.7 each have an atlas consisting of one single chart. Here are two examples where one needs more than one chart to cover a manifold.

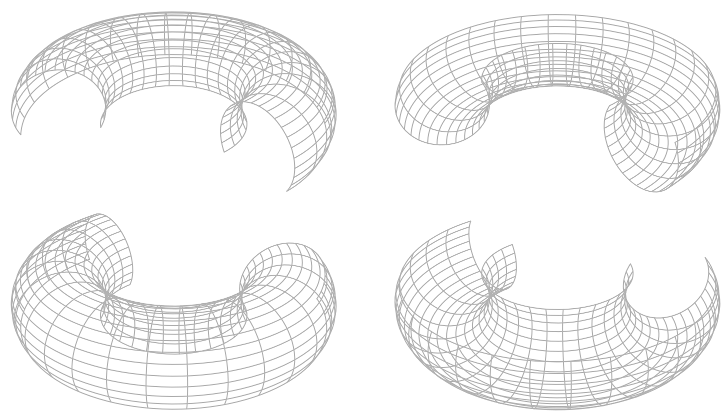
6.8. Example. The picture below shows the map

$$\psi(t_1, t_2) = \big((R + r\cos t_2)\cos t_1, (R + r\cos t_2)\sin t_1, r\sin t_2\big).$$

The domain U is an open rectangle in the plane and the image is a portion of a torus in three-space. One can check that ψ is an embedding, but we will not provide the details here. (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.



In this way we can cover the entire torus with images of rectangles, thus showing that the torus is a 2-dimensional manifold.



6.9. Example. Let M be the unit sphere 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

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

given in Exercise B.8. As we saw in that exercise, the image of ψ is the punctured sphere $M \setminus \{\mathbf{e}_n\}$, so if we let V be the open set $\mathbf{R}^n \setminus \{\mathbf{e}_n\}$, then $\psi(U) = M \cap V$. Also we saw that ψ has an 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 $M \setminus \{-\mathbf{e}_n\} = M \cap V$, where $V = \mathbf{R}^n \setminus \{-\mathbf{e}_n\}$. This finishes the proof that M is an $n-1$-manifold in $\mathbf{R}^n$.

As these examples show, the definition of a manifold can be a little awkward to work with in practice, even for a simple manifold. In practice it can be rather hard to decide whether a given subset is a manifold using the definition alone. In the next section we will give a more manageable criterion for a set to be a manifold.

We conclude this section by taking a second look at tangent vectors. There are many different ways to parametrize a manifold M in the neighbourhood of a point $\mathbf{x}$. (For instance, for a sphere we have a choice among a large number of different cartographic projections.) Let $\psi_1 \colon U_1 \to \mathbf{R}^N$ and $\psi_2 \colon U_2 \to \mathbf{R}^N$ be two local parametrizations of M at $\mathbf{x}$. Then we have $\mathbf{x} = \psi_1(\mathbf{t}_1) = \psi_2(\mathbf{t}_2)$ for some $\mathbf{t}_1 \in U_1$ and $\mathbf{t}_2 \in U_2$. Do $D\psi_1(\mathbf{t}_1)$ and $D\psi_2(\mathbf{t}_2)$ have the same column spaces? In other words, is the tangent space $T_{\mathbf{x}}M$ well-defined? We will answer this question in the affirmative by characterizing tangent vectors to M in language that does not refer to local parametrizations. Namely, we will prove that all tangent vectors to M are velocity vectors of paths in M. By a *path* in a manifold M in $\mathbf{R}^N$ we simply mean a path in $\mathbf{R}^N$ which happens to be contained in M, i.e. a smooth map c from an open interval I to $\mathbf{R}^N$ with the property that $c(t)$ is in M for all $t \in I$.

6.10. Theorem. *Let M be an n-manifold in $\mathbf{R}^N$. Let $\mathbf{x}_0 \in M$ and let $\mathbf{v} \in \mathbf{R}^N$. Then $\mathbf{v}$ is tangent to M at $\mathbf{x}_0$ if and only if there exists a path $c \colon (-\varepsilon, \varepsilon) \to M$ with the properties $c(0) = \mathbf{x}_0$ and $c'(0) = \mathbf{v}$.*

Proof. Let $\psi \colon U \to \mathbf{R}^N$ be a local parametrization of M at $\mathbf{x}_0$, i.e. an embedding with the property that $\psi(U) = M \cap V$ for some open subset V of $\mathbf{R}^N$ containing $\mathbf{x}_0$. Let $\mathbf{t}_0$ be the unique point in U satisfying $\psi(\mathbf{t}_0) = \mathbf{x}_0$. Suppose $\mathbf{v}$ is tangent to M at $\mathbf{x}_0$, i.e. $\mathbf{v} \in T_{\mathbf{x}_0}M$. Then by definition $\mathbf{v} = D\psi(\mathbf{t}_0)(\mathbf{u})$ for some $\mathbf{u} \in \mathbf{R}^n$. Define $c(h) = \psi(\mathbf{t}_0 + h\mathbf{u})$. Then c is a path in M passing through $c(0) = \psi(\mathbf{t}_0) = \mathbf{x}_0$. By the chain rule (see Example B.4) we have $c'(0) = D\psi(\mathbf{t}_0)\mathbf{u} = \mathbf{v}$. Thus $\mathbf{v}$ is the velocity vector at $\mathbf{x}_0$ of some path c in M passing through $\mathbf{x}_0$.

Conversely, assume that $\mathbf{v} = c'(0)$ for some path $c \colon (-\varepsilon, \varepsilon) \to M$ satisfying $c(0) = \mathbf{x}_0$. Can we find a vector $\mathbf{u} \in \mathbf{R}^n$ such that $D\psi(\mathbf{t}_0)\mathbf{u} = \mathbf{v}$? By Lemma 6.11 below, after replacing U and V with smaller open sets if necessary, the map ψ has a smooth left inverse, i.e. a smooth map $\phi \colon V \to U$ satisfying $\phi \circ \psi(\mathbf{t}) = \mathbf{t}$ for all $\mathbf{t} \in U$. If $\mathbf{x} \in M \cap V$, then $\mathbf{x} = \psi(\mathbf{t})$ for some $\mathbf{t} \in U$, so

$$\mathbf{x} = \psi(\mathbf{t}) = \psi(\phi(\psi(\mathbf{t}))) = \psi(\phi(\mathbf{x})).$$

If h is sufficiently small, then $c(h)$ is contained in $M \cap V$, and hence $c(h) = \psi(\phi(c(h)))$. Differentiating this identity with respect to h at $h = 0$ gives

$$c'(0) = D\psi(\phi(c(0)))D(\phi \circ c)(0) = D\psi(\phi(\mathbf{x}_0))\mathbf{u} = D\psi(\mathbf{t}_0)\mathbf{u}.$$

Here we have written $\mathbf{u} = D(\phi \circ c)(0)$ and we have used that $\phi(\mathbf{x}_0) = \mathbf{t}_0$ because $\psi(\mathbf{t}_0) = \mathbf{x}_0$ and ϕ is a left inverse of ψ. This shows that $c'(0)$ is in $D\psi(\mathbf{t}_0)(\mathbf{R}^n) = T_{\mathbf{x}_0}M$. QED

The following technical result, which is a consequence of the implicit function theorem, says that embeddings have smooth left inverses, at least if one suitably restricts the domain and the range.

6.11. Lemma. *Let U be an open subset of $\mathbf{R}^n$, let V be an open subset of $\mathbf{R}^N$, and let $\psi \colon U \to V$ be an embedding. Let $\mathbf{t}_0 \in U$ and $\mathbf{x}_0 = \psi(\mathbf{t}_0)$. Then there exist an open subset*

$\tilde U$ of U containing $\mathbf{t}_0$, an open subset $\tilde V$ of V containing $\mathbf{x}_0$, and a smooth map $\phi\colon \tilde V \to \tilde U$ with the property that $\phi(\psi(\mathbf{t})) = \mathbf{t}$ for all $\mathbf{t} \in \tilde U$.

Proof. Let $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ be the columns of $D\psi(\mathbf{t}_0)$. As ψ is an embedding, these columns are independent, so we can complete them to a basis $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ of $\mathbf{R}^N$, where $k = N - n$. For $\mathbf{t} \in U$ and $\mathbf{s} \in \mathbf{R}^k$ define $\tilde\psi(\mathbf{t}, \mathbf{s}) = \psi(\mathbf{t}) + \sum_{j=1}^k s_j \mathbf{b}_j$. Then $\tilde\psi$ is a map from $U \times \mathbf{R}^k$ to $\mathbf{R}^N$ and its Jacobi matrix at $(\mathbf{t}_0, \mathbf{0})$ is

$$D\tilde\psi(\mathbf{t}_0, \mathbf{0}) = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_k \end{pmatrix}.$$

The columns of this matrix form a basis of $\mathbf{R}^N$, and therefore it is invertible. By the inverse function theorem, Theorem B.7, there exist an open subset $\tilde U$ of U containing $\mathbf{t}_0$ and an open subset W of $\mathbf{R}^k$ containing $\mathbf{0}$ such that $\tilde V = \tilde\psi(\tilde U \times W)$ is an open subset of V and the map $\tilde\psi\colon \tilde U \times W \to \tilde V$ has a smooth inverse $\tilde\phi\colon \tilde V \to \tilde U \times W$. Let $\pi\colon \tilde U \times W \to \tilde V$ be the map defined by $\pi(\mathbf{t}, \mathbf{s}) = \mathbf{t}$, and let $\phi = \pi \circ \tilde\phi$. Then for all $\mathbf{t} \in \tilde U$ we have

$$\phi(\psi(\mathbf{t})) = \pi \circ \tilde\phi \circ \tilde\psi(\mathbf{t}, \mathbf{0}) = \pi(\mathbf{t}, \mathbf{0}) = \mathbf{t},$$

so $\phi\colon \tilde V \to \tilde U$ is a left inverse of $\psi\colon \tilde U \to \tilde V$. QED

6.2. The regular value theorem

Our definition of the notion of a manifold, Definition 6.3, is based on embeddings, which are an "explicit" way of describing manifolds. However, embeddings can be hard to 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 $\mathbf{c}$, or the *preimage* of $\mathbf{c}$ under ϕ. If ϕ is a linear map, the system of equations is inhomogeneous linear and we know from linear algebra that 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 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. (However, it does not apply to every manifold. In other words, it is a sufficient but not a necessary criterion.) The proof rests on the following important fact from linear algebra,

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

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 $\ker(A)$).

6.12. 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.6, 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 6.6 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

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

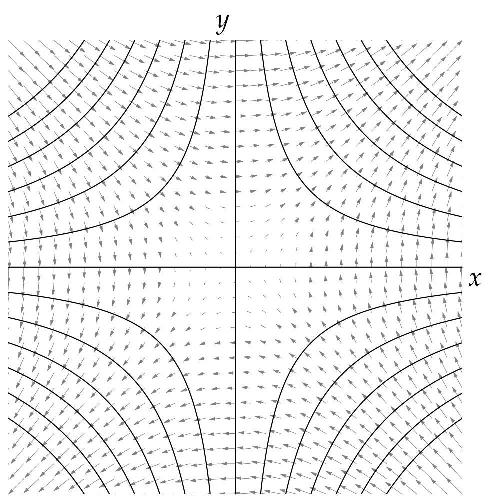
The tangent space $T_{\mathbf{x}}M$ is by definition the column space of $D\psi(\mathbf{u}_0)$, so every tangent vector $\mathbf{v}$ to M at $\mathbf{x}$ is of the form $\mathbf{v} = D\psi(\mathbf{u}_0)\mathbf{a}$ for some $\mathbf{a} \in \mathbf{R}^n$. Therefore $D\phi(\mathbf{x})\mathbf{v} = D\phi(\mathbf{x})D\psi(\mathbf{u}_0)\mathbf{a}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

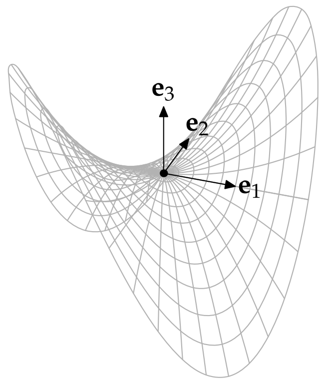
6.13. Corollary (level hypersurfaces). *Let U be open in $\mathbf{R}^N$ and let $\phi\colon U \to \mathbf{R}$ be a smooth function. Suppose that $M = \phi^{-1}(c)$ is nonempty and that $\operatorname{grad}(\phi)(\mathbf{x}) \neq \mathbf{0}$ for all $\mathbf{x}$ in M. Then M is a manifold in $\mathbf{R}^N$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

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

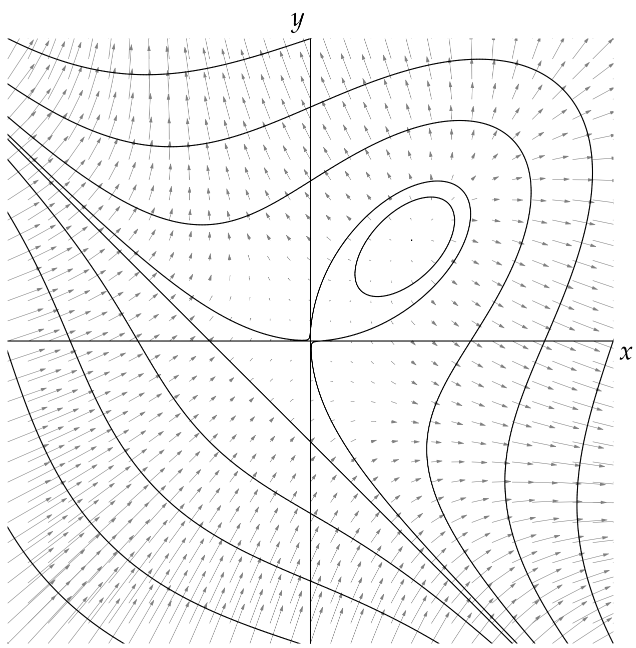
6.14. Example. Let $U = \mathbf{R}^2$ and $\phi(x, y) = xy$. The level curves of ϕ are hyperbolas in the plane and the gradient is $\operatorname{grad}(\phi)(\mathbf{x}) = (y, x)$. The diagram below shows a few level curves as well as the gradient vector field, which as you can see is perpendicular to the level curves.



The gradient vanishes only at the origin, so $\phi(\mathbf{0}) = 0$ is the only singular value of ϕ. By Corollary 6.13 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 $\phi^{-1}(0) \setminus \{\mathbf{0}\}$ 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.



6.15. Example. A more interesting example of an equation in two variables is $\phi(x, y) = x^3 + y^3 - 3xy = c$. Here $\operatorname{grad}(\phi)(\mathbf{x}) = 3(x^2 - y, y^2 - x)$, so $\operatorname{grad}(\phi)$ vanishes at the origin and at $(1, 1)$. 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 $(1, 1)$. 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 6.6.

6.16. Example. Let $U = \mathbf{R}^N$ and $\phi(\mathbf{x}) = \|\mathbf{x}\|^2$. Then $\operatorname{grad}(\phi)(\mathbf{x}) = 2\mathbf{x}$, so as in Example 6.14 $\operatorname{grad}(\phi)$ 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 $\mathrm{grad}(\phi)(\mathbf{x}) = 2\mathbf{x}$. 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 6.15. 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 ($m = 2$).

6.17. 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 $\mathbf{x} = (1, 0, 0, 0)$. 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 $\mathbf{e}_4$.

The orthogonal group. We now come to a more sophisticated example of a manifold determined by a large system of equations. 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. (So they form an ortho*normal* basis of $\mathbf{R}^n$—note the regrettable inconsistency in the terminology.) The orthogonal matrices form a group under matrix multiplication: every orthogonal matrix A is invertible with inverse $A^{-1} = A^T$, the identity matrix I is orthogonal, and the product of orthogonal matrices is orthogonal. This group is called the *orthogonal group* and denoted by $\mathbf{O}(n)$.

6.18. Theorem. *The orthogonal group $\mathbf{O}(n)$ is a manifold of dimension $\frac{1}{2}n(n-1)$. The tangent space to $\mathbf{O}(n)$ at the identity matrix is the space of antisymmetric $n \times n$-matrices.*

Proof. This is an application of the regular value theorem. We start by noting that $\mathbf{O}(n) = \phi^{-1}(I)$, where ϕ is defined by

$$\phi(A) = A^TA.$$

The domain of the map ϕ is $V = \mathbf{R}^{n\times n}$, the vector space of all $n \times n$-matrices. We can regard ϕ as a map from V to itself, but then the identity matrix I is not a regular value! To ensure that I is a regular value we must restrict the range of ϕ. This is done by observing that $(A^TA)^T = A^TA$, so A^TA is a symmetric matrix. In other words, if we let $W = \{ C \in V \mid C = C^T \}$ be the linear subspace of V consisting of all symmetric matrices, then we can regard ϕ as a map

$$\phi \colon V \longrightarrow W.$$

We will show that I is a regular value of this map. To do this we need to compute the total derivative of ϕ. For every matrix $A \in V$ the total derivative at A is a linear map $D\phi(A) \colon V \to W$, which can be computed by using Lemma B.1. This lemma says that for every $B \in V$ the result of applying the linear map $D\phi(A)$ to B is the directional derivative of ϕ at A along B:

$$\begin{aligned} D\phi(A)B &= \lim_{t\to 0} \frac{1}{t}(\phi(A+tB) - \phi(A)) \\ &= \lim_{t\to 0} \frac{1}{t}(A^TA + tA^TB + tB^TA + t^2B^TB - A^TA) \\ &= A^TB + B^TA. \end{aligned}$$

We need to show that for all $A \in \mathbf{O}(n)$ the linear map $D\phi(A) \colon V \to W$ is surjective. This amounts to showing that, given any orthogonal A and any symmetric C, the equation

$$A^TB + B^TA = C \tag{6.1}$$

is solvable for B. Here is a trick for guessing a solution: observe that $C = \frac{1}{2}(C + C^T)$ and first try to solve $A^TB = \frac{1}{2}C$. Left multiplying both sides by A and using $AA^T = I$ gives $B = \frac{1}{2}AC$. We claim that $B = \frac{1}{2}AC$ is a solution of equation (6.1). Indeed,

$$A^TB + B^TA = A^T\frac{1}{2}AC + \frac{1}{2}C^TA^TA = \frac{1}{2}(C + C^T) = C.$$

By the regular value theorem this proves that $\mathbf{O}(n)$ is a manifold. You will be asked to prove the remaining assertions in Exercise 6.12. QED

If A is an orthogonal matrix, then $\det(A^TA) = \det(I) = 1$, so $(\det(A))^2 = 1$, i.e. $\det(A) = \pm 1$. The *special orthogonal group* or *rotation group* is

$$\mathbf{SO}(n) = \{ A \in \mathbf{O}(n) \mid \det(A) = 1 \}.$$

If A and B are elements of $\mathbf{O}(n)$ that are not in $\mathbf{SO}(n)$, then $AB \in \mathbf{O}(n)$ and $\det(AB) = \det(A)\det(B) = 1$, so $AB \in \mathbf{SO}(n)$. An example of an orthogonal matrix which is not a rotation is a reflection matrix, such as

$$A_0 = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Every orthogonal matrix A which is not in $\mathbf{SO}(n)$ can be written as $A = A_0C$ for a unique $C \in \mathbf{SO}(n)$, namely $C = A_0A$. Thus we see that $\mathbf{O}(n)$ is a union of two disjoint pieces,

$$\mathbf{O}(n) = \mathbf{SO}(n) \cup \{ A_0C \mid C \in \mathbf{SO}(n) \},$$

each of which is a manifold of dimension $\frac{1}{2}n(n-1)$.

In two dimenson the rotation group is the group $\mathbf{SO}(2)$ of all matrices of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

As a manifold $\mathbf{SO}(2)$ is a copy of the unit circle S^1. The rotation group in three dimensions $\mathbf{SO}(3)$ is a three-dimensional manifold, which is a little harder to vizualize. Every rotation can be parametrized by a vector $\mathbf{x} \in \mathbf{R}^3$, namely as the rotation about the axis spanned by $\mathbf{x}$ through an angle of $\|\mathbf{x}\|$ radians. But $\mathbf{SO}(3)$ is not the same as $\mathbf{R}^3$, because many different vectors represent the same rotation. Every rotation can be represented by a vector of length $\le \pi$, so let us restrict $\mathbf{x}$ to the (solid) ball B of radius π about the origin. This gets rid of most of the ambiguity, except for the fact that two antipodal points on the boundary sphere represent the same rotation (because rotation through π about a given axis is the same as rotation through $-\pi$). We conclude that $\mathbf{SO}(3)$ is the manifold obtained by identifying opposite points on the boundary of B. This is the *three-dimensional projective space*.

Exercises

6.1. This is a continuation of Exercise 1.1. Define $\psi\colon \mathbf{R} \to \mathbf{R}^2$ by $\psi(t) = (t - \sin t, 1 - \cos t)$. 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 $\psi(t) = (t - a\sin t, 1 - a\cos t)$ 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 $\psi(t) = \frac{1}{2}(e^t + e^{-t}, e^t - e^{-t})$ 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 $\psi(t) = (3at/(1 + t^3), 3at^2/(1+t^3))$, where a is a nonzero constant. Show that ψ is one-to-one and that $\psi'(t) \ne \mathbf{0}$ 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

$$\psi(t) = \begin{cases} (-f(t), f(t)) & \text{if } t \le 0, \\ (f(t), f(t)) & \text{if } t \ge 0, \end{cases}$$

where f is the function given in Exercise B.3. 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 $\psi(t_1, t_2) = (t_1^3, t_1^2 t_2, t_1 t_2^2, t_2^3)$.
- (i) Show that ψ is one-to-one.
- (ii) Show that $D\psi(\mathbf{t})$ is one-to-one for all $\mathbf{t} \ne \mathbf{0}$.
- (iii) Let U be the punctured plane $\mathbf{R}^2 \setminus \{\mathbf{0}\}$. 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.
- (i) Let U be an open subset of $\mathbf{R}^n$ and let $\psi\colon U \to \mathbf{R}^N$ be a smooth map. The *tangent map* of ψ is the map $T\psi\colon U \times \mathbf{R}^n \to \mathbf{R}^{2N}$ defined by $T\psi(\mathbf{t}, \mathbf{u}) = (\psi(\mathbf{t}), D\psi(\mathbf{t})\mathbf{u})$ for $\mathbf{t} \in U$ and $\mathbf{u} \in \mathbf{R}^n$. Prove that the tangent map of an embedding is an embedding.

(ii) Let M be an n-manifold in $\mathbf{R}^N$. The *tangent bundle* of M is the subset TM of $\mathbf{R}^{2N}$ defined by

$$TM = \{\, (\mathbf{x}, \mathbf{v}) \in \mathbf{R}^{2N} \mid \mathbf{v} \in T_{\mathbf{x}}M \,\}.$$

Prove that the tangent bundle of M is a $2n$-manifold.

6.8. Let M be the set of points in $\mathbf{R}^2$ given by the equation $(x^2 + y^2)^2 + y^2 - x^2 = 0$.

(i) Show that $M \setminus \{(0, 0)\}$ is a 1-manifold.
(ii) Determine the points where M has horizontal or vertical tangent lines.
(iii) Sketch M. (Start by finding the intersection points of M with an arbitrary line through the origin, $y = ax$.)
(iv) Is M a manifold at $(0, 0)$? Explain.

6.9. Let $\phi\colon \mathbf{R}^n \setminus \{\mathbf{0}\} \to \mathbf{R}$ be a homogeneous function of degree p as defined in Exercise B.6. 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.6.) Conclude that, if nonempty, $\phi^{-1}(c)$ is an $n-1$-manifold for $c \neq 0$.

6.10. 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.)

6.11. Show that the trajectories of the Lotka-Volterra system of Exercise 1.11 are one-dimensional manifolds.

6.12. Let V be the vector space of $n \times n$-matrices and let W be its linear subspace consisting of all symmetric matrices.

(i) Prove that $\dim(V) = n^2$ and $\dim(W) = \frac{1}{2}n(n + 1)$. (Exhibit explicit bases of V and W, and count the number of elements in each basis.)
(ii) 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. (Use the regular value theorem, which says that the dimension of $\mathbf{O}(n)$ is $\dim(V) - \dim(W)$ and that the tangent space at the identity matrix is the kernel of $D\phi(I)$, where $\phi\colon V \to W$ is defined by $\phi(A) = A^TA$. See the proof of Theorem 6.18.)

6.13. Let V be the vector space of $n \times n$-matrices.

(i) The *general linear group* is the subset of V defined by

$$\mathbf{GL}(n) = \{\, A \in V \mid \det(A) \neq 0 \,\}.$$

Show that $\mathbf{GL}(n)$ is a manifold. What is its dimension?
(ii) Define $\phi\colon V \to \mathbf{R}$ by $\phi(A) = \det(A)$. Show that

$$D\phi(A)B = \sum_{i=1}^{n} \det(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{i-1}, \mathbf{b}_i, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n),$$

where $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$ denote the column vectors of A, resp. B. (Apply Lemma B.1 for the derivative and use the multilinearity of the determinant.)
(iii) 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?
(iv) 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 B satisfying $\operatorname{tr}(B) = 0$.

6.14. (i) Let W be punctured 4-space $\mathbf{R}^4 \setminus \{\mathbf{0}\}$ 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 $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 0 \end{smallmatrix}\right)$.

6.15. Define $\phi\colon \mathbf{R}^4 \to \mathbf{R}^2$ by $\phi(\mathbf{x}) = (x_1 + x_2 + x_3x_4, x_1x_2x_3 + x_4)$.

(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 $t \neq 0$. (Row reduce the matrix $D\phi(\mathbf{x})$ to compute its rank.)

(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 $\mathbf{x}$.)

6.16. 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 λ_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.)

6.17. 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 $f|C$?

6.18 (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 $\mathbf{x} + t\mathbf{e}_j$ may not be contained in M, so we cannot form the limit $\partial f/\partial x_j = \lim_{t\to 0}(f(\mathbf{x}+t\mathbf{e}_j) - f(\mathbf{x}))/t$.

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 $f_j(\mathbf{u}) = f_i(\psi_i^{-1} \circ \psi_j(\mathbf{u}))$. 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* or *transition map*, 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. Appealing to Theorem 3.13 we see that (7.1) is equivalent to the following requirement: 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

$$g_J = \sum_I (\psi_i^{-1} \circ \psi_j)^*\big(f_I \, \det(D(\psi_i^{-1} \circ \psi_j)_{I,J})\big).$$

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):

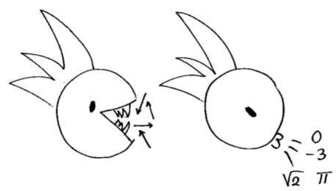
$$(d\alpha)_j = d\alpha_j = d(\psi_i^{-1} \circ \psi_j)^*(\alpha_i) = (\psi_i^{-1} \circ \psi_j)^*(d\alpha_i) = (\psi_i^{-1} \circ \psi_j)^*((d\alpha)_i),$$

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. A covector is a little dinosaur that eats vectors and spits out numbers, in a linear way.



The formal definition goes as follows. Let V be a vector space over the real numbers, for example $\mathbf{R}^n$ or a linear subspace of $\mathbf{R}^n$. A *covector*, or *dual vector*, or *linear functional*, is a linear map from V to $\mathbf{R}$.

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 $\mu(c_1 f_1 + c_2 f_2) = c_1\mu(f_1) + c_2\mu(f_2)$ for all functions f_1, $f_2 \in V$ and all scalars c_1, c_2, so μ is a linear functional on V.

The collection of all covectors on V is denoted by V^* and called the *dual* of V. 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 $(\mu_1+\mu_2)(v) = \mu_1(v)+\mu_2(v)$ and $(c\mu_1)(v) = c\mu_1(v)$ for all $v \in V$.

For the next example, recall that if A is an $m\times n$-matrix and $\mathbf{x}$ an n-vector, then $A\mathbf{x}$ is an m-vector, and the map which sends $\mathbf{x}$ to $A\mathbf{x}$ is linear. Moreover, every linear map from $\mathbf{R}^n$ to $\mathbf{R}^m$ is of this form for a unique matrix A.

7.3. Example. A covector on $\mathbf{R}^n$ is a linear map from $\mathbf{R}^n$ to $\mathbf{R} = \mathbf{R}^1$ and is therefore given by a $1\times n$-matrix, which is nothing but a row vector. Thus $(\mathbf{R}^n)^*$ is the space of row n-vectors. A row vector $\mathbf{y}$ "eating" a column vector $\mathbf{x}$ means multiplying the two, which results in a number:

$$\mathbf{y}\mathbf{x} = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^n y_i x_i.$$

Now suppose that V is a vector space of finite dimension n and choose a basis $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ of V. Then every vector $\mathbf{b} \in V$ can be written in a unique way as a linear combination $\sum_j c_j \mathbf{b}_j$. Define a covector $\beta_i \in V^*$ by $\beta_i(\mathbf{b}) = c_i$. In other words, β_i is determined by the rule $\beta_i(\mathbf{b}_j) = \delta_{i,j}$, where

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}$$

is the *Kronecker delta*. We call β_i the *i-th coordinate function*.

7.4. Lemma. *The coordinate functions $\beta_1, \beta_2, \dots, \beta_n$ form a basis of V^*. It follows that* $\dim(V^*) = n = \dim(V)$.

Proof. Let $\beta \in V^*$. We need to prove that β can be written as a linear combination $\beta = \sum_{i=1}^n c_i\beta_i$ with unique coefficients c_i. First we prove uniqueness. Assuming that β can be expressed as $\beta = \sum_{i=1}^{n} c_i\beta_i$, we can apply both sides to the vector $\mathbf{b}_j$ to obtain

$$\beta(\mathbf{b}_j) = \sum_{i=1}^{n} c_i\beta_i(\mathbf{b}_j) = \sum_{i=1}^{n} c_i\delta_{i,j} = c_j. \tag{7.2}$$

So $c_j = \beta(\mathbf{b}_j)$ is the only possible choice for the coefficient c_j. This argument establishes the uniqueness of the coefficients. Moreover, it tells us what the coefficients should be, which helps us prove that they exist. Namely, let us define $\beta' = \sum_{i=1}^{n} \beta(\mathbf{b}_i)\beta_i$. Then by equation (7.2), $\beta'(\mathbf{b}_j) = \beta(\mathbf{b}_j)$ for all j, so $\beta' = \beta$, and therefore $\beta = \sum_{i=1}^{n} \beta(\mathbf{b}_i)\beta_i$. This proves that $\beta_1, \beta_2, \ldots, \beta_n$ constitute a basis of V^*. The cardinality of the basis is n, so $\dim(V^*) = n$. QED

The basis $\{\beta_1, \beta_2, \ldots, \beta_n\}$ of V^* is said to be *dual* to the basis $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ of V.

7.5. Example. Let $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis of $\mathbf{R}^n$. What is the dual basis $\{\beta_1, \beta_2, \ldots, \beta_n\}$ of $(\mathbf{R}^n)^*$? The β_i's are row vectors determined by the equations $\beta_i\mathbf{b}_j = \delta_{i,j}$. These equations can be written as a single matrix equation: let B be the $n \times n$-matrix with columns $\mathbf{b}_1$, $\mathbf{b}_2, \ldots, \mathbf{b}_n$ and let A be the $n \times n$-matrix with rows β_1, $\beta_2, \ldots, \beta_n$; then $AB = I$. Therefore A is the inverse of B. In other words, β_i is the i-th row of B^{-1}. As a special case consider the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Then $B = I$, so $A = I$, and the dual basis of $(\mathbf{R}^n)^*$ is $\{\mathbf{e}_1^T, \mathbf{e}_2^T, \ldots, \mathbf{e}_n^T\}$.

Dual bases come in handy when writing the matrix of a linear map. Let $L\colon V \to W$ be a linear map between vector spaces V and W. To write the matrix of L we need to start by picking a basis $\mathbf{b}_1$, $\mathbf{b}_2, \ldots, \mathbf{b}_n$ of V and a basis $\mathbf{c}_1$, $\mathbf{c}_2, \ldots, \mathbf{c}_m$ of W. Then for each $j = 1, 2, \ldots, n$ the vector $L\mathbf{b}_j$ can be expanded uniquely in terms of the $\mathbf{c}$'s: $L\mathbf{b}_j = \sum_{i=1}^{m} l_{i,j}\mathbf{c}_i$. 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 γ_1, $\gamma_2, \ldots, \gamma_m \in W^*$ the dual basis of $\mathbf{c}_1$, $\mathbf{c}_2, \ldots, \mathbf{c}_m$. Then the (i, j)-th matrix element of a linear map $L\colon V \to W$ is equal to $l_{i,j} = \gamma_i(L\mathbf{b}_j)$.*

Proof. We have $L\mathbf{b}_j = \sum_{k=1}^{m} l_{k,j}\mathbf{c}_k$, so

$$\gamma_i(L\mathbf{b}_j) = \sum_{k=1}^{m} l_{k,j}\gamma_i(\mathbf{c}_k) = \sum_{k=1}^{m} l_{k,j}\delta_{i,k} = l_{i,j},$$

that is to say $l_{i,j} = \gamma_i(L\mathbf{b}_j)$. QED

1-Forms on $\mathbf{R}^n$ re-examined. Let U be an open subset of $\mathbf{R}^n$. Recall that a vector field on U is a smooth map $\mathbf{F}\colon U \to \mathbf{R}^n$. A 1-form is a type of object "dual" to a vector field. Formally, a 1-*form* or *covector field* on U is defined as a smooth map $\alpha\colon U \to (\mathbf{R}^n)^*$. This means that α is a row vector

$$\alpha = \begin{pmatrix} f_1 & f_2 & \cdots & f_n \end{pmatrix}$$

whose entries are smooth functions on U. The form is called *constant* if the entries $f_1, \ldots, f_n$ are constant. By definition dx_i is the constant 1-form

$$dx_i = \mathbf{e}_i^T = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \tag{7.3}$$

the transpose of $\mathbf{e}_i$, the i-th standard basis vector of $\mathbf{R}^n$. Every 1-form can thus be written as

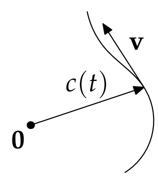
$$\alpha = (f_1 \ \ f_2 \ \cdots \ f_n) = \sum_{i=1}^n f_i \, dx_i.$$

Using this formalism we can write for any smooth function g on U

$$dg = \sum_{i=1}^n \frac{\partial g}{\partial x_i} dx_i = \begin{pmatrix} \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} & \cdots & \frac{\partial g}{\partial x_n} \end{pmatrix},$$

which is simply the Jacobi matrix Dg of g! (This is the reason that many authors use the notation dg for the Jacobi matrix.)

In what sense does the row vector dx_i represent an "infinitesimal increment" along the x_i-axis? Let $\mathbf{v} \in \mathbf{R}^n$ be the velocity vector of a path $c(t)$ at time t.



In an infinitesimal time interval Δt the position changes to $c(t + \Delta t) \approx c(t) + \Delta t \, \mathbf{v}$, so the infinitesimal displacement is $\Delta t \, \mathbf{v}$. The x_i-coordinate changes by an amount $\Delta t \, \mathbf{v}_i = \Delta t \, dx_i(\mathbf{v})$. We conclude that the number $dx_i(\mathbf{v})$ represents the rate of change of the i-th coordinate along the path per unit time.

Multilinear algebra. Multilinear algebra is needed to make sense of differential forms of higher degree.

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 a k-tuple $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ of vectors in V. A *k-multilinear* function on V is a function $\mu \colon V^k \to \mathbf{R}$ which is linear in each vector, i.e.

$$\mu(\mathbf{v}_1, \mathbf{v}_2, \ldots, c\mathbf{v}_i + c'\mathbf{v}_i', \ldots, \mathbf{v}_k) = c\mu(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) + c'\mu(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_k)$$

for all scalars c, c' and all vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i, \mathbf{v}_i', \ldots, \mathbf{v}_k$.

7.7. Example. Let $V = \mathbf{R}^n$ and let $\mu(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$, the inner product of $\mathbf{x}$ and $\mathbf{y}$. Then μ is bilinear (i.e. 2-multilinear).

7.8. Example. Let $V = \mathbf{R}^4$, $k = 2$. The function $\mu(\mathbf{v}, \mathbf{w}) = v_1w_2 - v_2w_1 + v_3w_4 - v_4w_3$ is bilinear on $\mathbf{R}^4$.

7.9. Example. Let $V = \mathbf{R}^n$, $k = n$. It follows from Corollary 3.6 that the determinant $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is an n-multilinear function on $\mathbf{R}^n$.

A k-multilinear function is *alternating* or *antisymmetric* if it has the alternating property,

$$\mu(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_k) = -\mu(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_k)$$

for all $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ in V. More generally, if μ is alternating, then for any permutation $\sigma \in S_k$ we have

$$\mu(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(k)}) = \operatorname{sign}(\sigma)\mu(\mathbf{v}_1, \ldots, \mathbf{v}_k).$$

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 for producing alternating k-multilinear functions starting from k covectors $\mu_1, \mu_2, \dots, \mu_k \in V^*$. The *(wedge) product* is the function

$$\mu_1 \mu_2 \cdots \mu_k \colon V^k \to \mathbf{R}$$

defined by

$$\boxed{\mu_1 \mu_2 \cdots \mu_k(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \det(\mu_i(\mathbf{v}_j))_{1 \le i,j \le k}.}$$

(The determinant on the right is a $k \times k$-determinant.) It follows from the multilinearity and the alternating property of the determinant that $\mu_1 \mu_2 \cdots \mu_k$ is an alternating k-multilinear function. Some authors denote the wedge product by $\mu_1 \wedge \mu_2 \wedge \cdots \wedge \mu_k$ 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 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.

For $k = 1$ the alternating property is vacuous, so an alternating 1-multilinear function is nothing but a linear function. Thus $A^1(V) = V^*$.

A 0-multilinear function is by convention just a number. Thus $A^0(V) = \mathbf{R}$.

There is a nice way to construct a basis of the vector space $A^k(V)$ starting from a basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of V. The idea is to take wedge products of dual basis vectors. Let $\{\beta_1, \dots, \beta_n\}$ be the corresponding dual basis of V^*. Let $I = (i_1, i_2, \dots, i_k)$ be an increasing multi-index, i.e. $1 \le i_1 < i_2 < \cdots < i_k \le n$. Write

$$\beta_I = \beta_{i_1} \beta_{i_2} \cdots \beta_{i_k} \in A^k(V),$$
$$\mathbf{b}_I = (\mathbf{b}_{i_1}, \mathbf{b}_{i_2}, \dots, \mathbf{b}_{i_k}) \in V^k.$$

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. For multi-indices I and J let us define a generalized *Kronecker delta* $\delta_{I,J}$ by

$$\delta_{I,J} = \begin{cases} 1 & \text{if } I = J, \\ 0 & \text{if } I \ne J. \end{cases}$$

7.12. Lemma. *Let I and J be increasing multi-indices of degree k. Then $\beta_I(\mathbf{b}_J) = \delta_{I,J}$.*

Proof. Let $I = (i_1, \dots, i_k)$ and $J = (j_1, \dots, j_k)$. Then

$$\beta_I(\mathbf{b}_J) = \det\big(\beta_{i_r}(\mathbf{b}_{j_s})\big)_{1 \le r,s \le k} = \begin{vmatrix} \delta_{i_1,j_1} & \dots & \delta_{i_1,j_k} \\ \vdots & & \vdots \\ \delta_{i_l,j_1} & \dots & \delta_{i_l,j_k} \\ \vdots & & \vdots \\ \delta_{i_k,j_1} & \dots & \delta_{i_k,j_k} \end{vmatrix}.$$

If $I = J$, then this matrix is the identity $k \times k$-matrix, so $\beta_I(\mathbf{b}_J) = 1$. If $I \neq J$, then there is some $i \in I$ which is not in J, say $i = i_l$, which causes all entries in the l-th row of the matrix to vanish. Hence its determinant is 0, and therefore $\beta_I(\mathbf{b}_J) = 0$. QED

We need one further technical result before showing that the functions β_I are a basis of $A^k(V)$.

7.13. Lemma. *Let $\beta \in A^k(V)$. Suppose $\beta(\mathbf{b}_I) = 0$ for all increasing multi-indices I of degree k. Then $\beta = 0$.*

Proof. The assumption implies

$$\beta(\mathbf{b}_{i_1}, \dots, \mathbf{b}_{i_k}) = 0 \tag{7.4}$$

for *all* multi-indices $(i_1, \dots, i_k)$, because of the alternating property. We need to show that $\beta(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = 0$ for arbitrary vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$. We can expand the $\mathbf{v}_i$ using the basis:

$$\begin{aligned} \mathbf{v}_1 &= a_{1,1}\mathbf{b}_1 + a_{1,2}\mathbf{b}_2 + \dots + a_{1,n}\mathbf{b}_n, \\ \mathbf{v}_2 &= a_{2,1}\mathbf{b}_1 + a_{2,2}\mathbf{b}_2 + \dots + a_{2,n}\mathbf{b}_n, \\ &\vdots \\ \mathbf{v}_k &= a_{k,1}\mathbf{b}_1 + a_{k,2}\mathbf{b}_2 + \dots + a_{k,n}\mathbf{b}_n. \end{aligned}$$

Therefore by multilinearity

$$\beta(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} a_{1,i_1} a_{2,i_2} \cdots a_{k,i_k} \beta(\mathbf{b}_{i_1}, \mathbf{b}_{i_2}, \dots, \mathbf{b}_{i_k}).$$

Each term in the right-hand side is 0 by equation (7.4). QED

7.14. Theorem. *Let V be an n-dimensional vector space with basis $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$. Let $\{\beta_1, \beta_2, \dots, \beta_n\}$ be the corresponding dual basis of V^*. Then the alternating k-multilinear functions $\beta_I = \beta_{i_1}\beta_{i_2}\cdots\beta_{i_k}$, 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 $\beta \in A^k(V)$. We need to write β as a linear combination $\beta = \sum_I c_I \beta_I$. Assuming for the moment that this is possible, we apply both sides to the k-tuple of vectors $\mathbf{b}_J$. Using Lemma 7.12 we obtain

$$\beta(\mathbf{b}_J) = \sum_I c_I \beta_I(\mathbf{b}_J) = \sum_I c_I \delta_{I,J} = c_J.$$

So $c_J = \beta(\mathbf{b}_J)$ is the only possible choice for the coefficient c_J. To show that this choice of coefficients works, let us define $\beta' = \sum_I \beta(\mathbf{b}_I)\beta_I$. Then for all increasing multi-indices I we have $\beta(\mathbf{b}_I) - \beta'(\mathbf{b}_I) = \beta(\mathbf{b}_I) - \beta(\mathbf{b}_I) = 0$. Applying Lemma 7.13

to $\beta - \beta'$ we find $\beta - \beta' = 0$. In other words, $\beta = \sum_I \beta(\mathbf{b}_I)\beta_I$. We have proved that the β_I form a basis. The dimension of $A^k(V)$ is equal to the cardinality of the basis, which is $\binom{n}{k}$. 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 $\{dx_1, \dots, dx_n\}$. (See (7.3) and Example 7.5.) 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 μ on $\mathbf{R}^n$ looks like

$$\mu = \sum_I a_I\,dx_I,$$

with a_I constant. By Lemma 7.12, $\mu(\mathbf{e}_J) = \sum_I a_I\,dx_I(\mathbf{e}_J) = \sum_I a_I \delta_{I,J} = a_J$, so the coefficient a_I is equal to $\mu(\mathbf{e}_I)$.

k-Forms on $\mathbf{R}^n$ re-examined. Let U be an open subset of $\mathbf{R}^n$. We define a k-form α on U to be a smooth map $\alpha\colon U \to A^k(\mathbf{R}^n)$. This means that α can be written as

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

where the coefficients f_I are smooth functions on U. The value of α at $\mathbf{x} \in U$ is denoted by $\alpha_{\mathbf{x}}$, so that we have $\alpha_{\mathbf{x}} = \sum_I f_I(\mathbf{x})\,dx_I$ for all $\mathbf{x} \in U$. For each $\mathbf{x}$ the object $\alpha_{\mathbf{x}}$ is an element of $A^k(\mathbf{R}^n)$, that is to say a k-multilinear function on $\mathbf{R}^n$. So for any k-tuple $\mathbf{v}_I = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ of vectors in $\mathbf{R}^n$ the expression $\alpha_{\mathbf{x}}(v_I) = \alpha_{\mathbf{x}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ is a number. Example 7.15 gives us a useful formula for the coefficients f_I, namely $f_I = \alpha(\mathbf{e}_I)$ (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. We write $\alpha = \sum_I f_I\,dy_I$, where the f_I are smooth functions on V, and $\phi^*(\alpha) = \sum_J g_J\,dx_J$, where the g_J are smooth functions on U. What is the relationship between g_J and f_I? We use the formula $g_J = \phi^*(\alpha)(\mathbf{e}_J)$, our new definition of pullback and the definition of the wedge product to obtain

$$\begin{aligned}
g_J(\mathbf{x}) &= \phi^*(\alpha)_{\mathbf{x}}(\mathbf{e}_J) = \alpha_{\phi(\mathbf{x})}(D\phi(\mathbf{x})\mathbf{e}_{j_1}, D\phi(\mathbf{x})\mathbf{e}_{j_2}, \dots, D\phi(\mathbf{x})\mathbf{e}_{j_k}) \\
&= \sum_I f_I(\phi(\mathbf{x}))\,dy_I(D\phi(\mathbf{x})\mathbf{e}_{j_1}, D\phi(\mathbf{x})\mathbf{e}_{j_2}, \dots, D\phi(\mathbf{x})\mathbf{e}_{j_k}) \\
&= \sum_I \phi^*(f_I)(\mathbf{x})\det\bigl(dy_{i_r}(D\phi(\mathbf{x})\mathbf{e}_{j_s})\bigr)_{1 \le r,s \le k}.
\end{aligned}$$

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 3.13 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 $\mathbf{R}^N$. The book [**BT82**] describes a k-form on M as an animal that inhabits the world M, eats k-tuples of tangent vectors, and spits out numbers. Formally, 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. Let $\psi\colon U \to \mathbf{R}^N$ be a local parametrization of M at $\mathbf{x}$. 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 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 $\mathbf{x}$.

We can calculate the local representative $\psi^*(\alpha)$ of a k-form α for any local parametrization $\psi\colon U \to \mathbf{R}^N$ of M. Suppose we had two different local parametrizations $\psi_i\colon U_i \to \mathbf{R}^N$ and $\psi_j\colon U_j \to \mathbf{R}^N$ of M at $\mathbf{x}$. Then the local expressions $\alpha_i = \psi_i^*(\alpha)$ and $\alpha_j = \psi_j^*(\alpha)$ for α are related by the formula

$$\alpha_j = (\psi_i^{-1} \circ \psi_j)^*(\alpha_i).$$

This is identical to the transformation law (7.1), which shows that Definitions 7.1 and 7.16 of differential forms on a manifold are equivalent.

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

$$\alpha_{\mathbf{x}}(\mathbf{v}) = \begin{cases} \|\mathbf{v}\| & \text{if } \mathbf{v} \text{ is positive,} \\ -\|\mathbf{v}\| & \text{if } \mathbf{v} \text{ is negative.} \end{cases}$$

The form α 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.

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 $\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ be a basis of $\mathbf{R}^n$. Show that $\{\mathbf{b}_1^T, \mathbf{b}_2^T, \dots, \mathbf{b}_n^T\}$ is the corresponding dual basis of $(\mathbf{R}^n)^*$ if and only if the basis is orthonormal.

7.3. Let μ be a k-multilinear function on a vector space V. Suppose that μ satisfies $\mu(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = 0$ whenever two of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are equal, i.e. $\mathbf{v}_i = \mathbf{v}_j$ for some pair of distinct indices $i \neq j$. Prove that μ is alternating.

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

$$\mathbf{x} \times \mathbf{y} = \left(*(\mathbf{x}^T \wedge \mathbf{y}^T)\right)^T$$

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 $\mu_1, \mu_2, \ldots, \mu_k \in V^*$ 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

$$\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_k(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \mu_1(\mathbf{v}_1)\mu_2(\mathbf{v}_2)\cdots\mu_k(\mathbf{v}_k).$$

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 $\mathrm{Alt}(\mu) \colon V^k \to \mathbf{R}$ by

$$\mathrm{Alt}(\mu)(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k) = \frac{1}{k!}\sum_{\sigma \in S_k} \mathrm{sign}(\sigma)\mu(\mathbf{v}_{\sigma(1)}, \mathbf{v}_{\sigma(2)}, \ldots, \mathbf{v}_{\sigma(k)}).$$

Prove the following.

(i) $\mathrm{Alt}(\mu)$ is an *alternating* k-multilinear function.
(ii) $\mathrm{Alt}(\mu) = \mu$ if μ is alternating.
(iii) $\mathrm{Alt}(\mathrm{Alt}(\mu)) = \mathrm{Alt}(\mu)$ for all k-multilinear μ.
(iv) Let $\mu_1, \mu_2, \ldots, \mu_k \in V^*$. Then

$$\mu_1\mu_2\cdots\mu_k = k!\,\mathrm{Alt}(\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_k).$$

7.8. Show that $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = dx_1\,dx_2\cdots dx_n(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ for all vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbf{R}^n$. 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 $\lambda, \mu \in W^*$.

CHAPTER 8

Volume forms

8.1. n-Dimensional volume in $\mathbf{R}^N$

The *parallelepiped* spanned by n vectors $\mathbf{a}_1$, $\mathbf{a}_2, \ldots, \mathbf{a}_n$ in $\mathbf{R}^N$ is the set of all linear combinations $\sum_{i=1}^n c_i \mathbf{a}_i$, where the coefficients c_i range over the unit interval $[0, 1]$. This is the same as the definition given in Section 3.1, except that here we allow the number of vectors n to be different from the dimension N. (Think of a parallelogram in three-space.) We will need a formula for the volume of a parallelepiped. If $n < N$ there is no coherent way of defining an orientation on all n-parallelepipeds in $\mathbf{R}^N$, so this volume will be not an oriented but an absolute volume. (The reason is that for $n < N$ an n-dimensional parallelepiped in $\mathbf{R}^N$ can be rotated onto its mirror image through the extra dimensions. This is impossible for $n = N$.) It turns out that n-dimensional volume in $\mathbf{R}^N$, like the determinant, can be characterized by a few reasonable axioms.

8.1. Definition. An *(absolute) n-dimensional Euclidean volume function on* $\mathbf{R}^N$ 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:

$$\mathrm{vol}_n(\mathbf{a}_1, \ldots, \mathbf{a}_i + c\mathbf{a}_j, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) = \mathrm{vol}_n(\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n)$$

for all scalars c and all pairs of indices $i \neq j$;

(iii) invariance under Euclidean motions:

$$\mathrm{vol}_n(Q\mathbf{a}_1, Q\mathbf{a}_2, \ldots, Q\mathbf{a}_n) = \mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n)$$

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. *Let $\mathbf{a}_1$, $\mathbf{a}_2, \ldots, \mathbf{a}_n$ be vectors in $\mathbf{R}^N$.*

(i) $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = 0$ *if the vectors $\mathbf{a}_1$, $\mathbf{a}_2, \ldots, \mathbf{a}_n$ are linearly dependent.*

(ii) $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n) = \|\mathbf{a}_1\| \, \|\mathbf{a}_2\| \cdots \|\mathbf{a}_n\|$ *if $\mathbf{a}_1$, $\mathbf{a}_2, \ldots, \mathbf{a}_n$ are orthogonal vectors.*

Proof. (i) Assume $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are linearly 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

$$\begin{aligned}\operatorname{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) &= \operatorname{vol}_n\left(\sum_{i=2}^n c_i \mathbf{a}_i, \mathbf{a}_2, \dots, \mathbf{a}_n\right) \\ &= \operatorname{vol}_n\left(\sum_{i=3}^n c_i \mathbf{a}_i, \mathbf{a}_2, \dots, \mathbf{a}_n\right) = \dots = \operatorname{vol}_n(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n).\end{aligned}$$

Now by Axiom (i),

$$\operatorname{vol}_n(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n) = \operatorname{vol}_n(0\,\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n) = 0\operatorname{vol}_n(\mathbf{0}, \mathbf{a}_2, \dots, \mathbf{a}_n) = 0,$$

which proves property (i).

(ii) Suppose $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are orthogonal. First assume they are nonzero. Then we can define $\mathbf{q}_i = \|\mathbf{a}_i\|^{-1}\mathbf{a}_i$. The vectors $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$ are ortho*normal*. Complete them to an orthonormal basis $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n, \mathbf{q}_{n+1}, \dots, \mathbf{q}_N$ 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

$$\begin{aligned}\operatorname{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) &= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\|\operatorname{vol}_n(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n) && \text{by Axiom (i)} \\ &= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\|\operatorname{vol}_n(Q\mathbf{e}_1, Q\mathbf{e}_2, \dots, Q\mathbf{e}_n) \\ &= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\|\operatorname{vol}_n(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) && \text{by Axiom (iii)} \\ &= \|\mathbf{a}_1\|\,\|\mathbf{a}_2\|\cdots\|\mathbf{a}_n\| && \text{by Axiom (iv)},\end{aligned}$$

which proves part (ii) 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 then the statement follows from part (i). QED

A special case of Lemma 8.2(i) is the following observation: if $n > N$ then every set $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ of n vectors in $\mathbf{R}^N$ is dependent, so $\operatorname{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = 0$. This makes sense: a degenerate parallelepiped spanned by three vectors in the plane has three-dimensional volume equal to 0.

This brings us to the volume formula. We can form an $N \times n$-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 *Gram matrix* A^TA of A is square and we can take *its* determinant.

8.3. Lemma. *We have* $\det(A^TA) \geq 0$ *for all* $N \times n$*-matrices* A*, and* $\det(A^TA) = 0$ *if and only if the columns of* A *are dependent.*

Proof. The inner product of two vectors $\mathbf{u}$ and $\mathbf{v}$ can be written as $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v}$. In particular,

$$(A^TA\mathbf{v}) \cdot \mathbf{v} = (A^TA\mathbf{v})^T\mathbf{v} = \mathbf{v}^TA^TA\mathbf{v} = (A\mathbf{v}) \cdot (A\mathbf{v}) \geq 0 \tag{8.1}$$

for all vectors $\mathbf{v} \in \mathbf{R}^n$. The Gram matrix is symmetric and therefore, by the spectral theorem, has an eigenbasis consisting of real eigenvectors. If $\mathbf{v}$ is an eigenvector of A^TA with eigenvalue λ, then $\lambda\mathbf{v} \cdot \mathbf{v} = (A^TA\mathbf{v}) \cdot \mathbf{v}$, which is nonnegative by (8.1). Hence $\lambda \geq 0$: all eigenvalues of the Gram matrix are nonnegative, and therefore its determinant is non-negative. If the columns of A are dependent, then A has a nontrivial nullspace, so $A\mathbf{v} = \mathbf{0}$ for some nonzero $\mathbf{v}$. Hence $A^TA\mathbf{v} = \mathbf{0}$, so the columns of A^TA are dependent as well, so $\det(A^TA) = 0$. Conversely, if $\det(A^TA) = 0$ then A^TA has a nontrivial nullspace, so $A^TA\mathbf{w} = \mathbf{0}$ for some nonzero

$\mathbf{w}$. Therefore $(A\mathbf{w}) \cdot (A\mathbf{w}) = 0$ by (8.1), i.e. $A\mathbf{w} = \mathbf{0}$, so the columns of A are dependent. QED

It follows from the lemma that the formula in the next theorem makes sense.

8.4. Theorem. *There exists a unique n-dimensional volume function on $\mathbf{R}^N$. It is given by the following formula:*

$$\boxed{\operatorname{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \sqrt{\det(A^T A)}}$$

for all vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ in $\mathbf{R}^N$, where A is the $N \times n$-matrix whose i-th column is $\mathbf{a}_i$.

Proof. The existence is proved by checking that the function $\sqrt{\det(A^T A)}$ satisfies the axioms of Definition 8.1. You will be asked to do this in Exercise 8.2. The uniqueness is proved by verifying that the formula holds, which we proceed to do now.

First assume that $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are dependent. Then $\operatorname{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = 0$ by Lemma 8.2(i) and $\det(A^T A) = 0$ by Lemma 8.3, so the formula holds in this case.

Next consider a sequence of independent vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. Recall that such a sequence can be transformed into an orthogonal sequence $\mathbf{a}_1^\perp, \mathbf{a}_2^\perp, \dots, \mathbf{a}_n^\perp$ 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 $\mathbf{a}_i^\perp = \mathbf{a}_i - \mathbf{b}_i$. (See illustration below.) Each $\mathbf{b}_i$ is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{i-1}$, so by repeated applications of Axiom (ii) we get

$$\begin{aligned}
\operatorname{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) &= \operatorname{vol}_n(\mathbf{a}_1^\perp + \mathbf{b}_1, \mathbf{a}_2^\perp + \mathbf{b}_2, \dots, \mathbf{a}_n^\perp + \mathbf{b}_n) \\
&= \operatorname{vol}_n(\mathbf{a}_1^\perp, \mathbf{a}_2^\perp + \mathbf{b}_2, \dots, \mathbf{a}_n^\perp + \mathbf{b}_n) \\
&= \operatorname{vol}_n(\mathbf{a}_1^\perp, \mathbf{a}_2^\perp, \dots, \mathbf{a}_n^\perp + \mathbf{b}_n) \\
&\;\;\vdots \\
&= \operatorname{vol}_n(\mathbf{a}_1^\perp, \mathbf{a}_2^\perp, \dots, \mathbf{a}_n^\perp) \\
&= \|\mathbf{a}_1^\perp\| \|\mathbf{a}_2^\perp\| \cdots \|\mathbf{a}_n^\perp\|,
\end{aligned} \tag{8.2}$$

where the last equality follows from Lemma 8.2(ii). The Gram-Schmidt process can be expressed in matrix form by letting $\mathbf{q}_i$ be the normalized vector $\mathbf{q}_i = \mathbf{a}_i^\perp / \|\mathbf{a}_i^\perp\|$ and Q the $N \times n$-matrix with columns $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$. Then we have the QR-decomposition $A = QR$, where R is an $n \times n$-matrix of the form

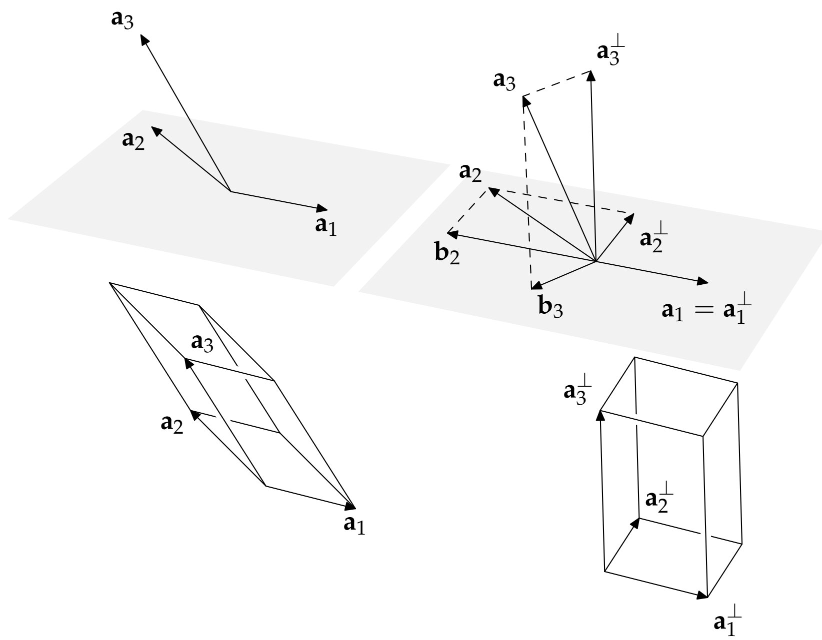
$$R = \begin{pmatrix} \|\mathbf{a}_1^\perp\| & * & * & \cdots & * \\ 0 & \|\mathbf{a}_2^\perp\| & * & \cdots & * \\ 0 & 0 & \|\mathbf{a}_3^\perp\| & \cdots & * \\ \vdots & \vdots & & \ddots & * \\ 0 & 0 & \cdots & 0 & \|\mathbf{a}_n^\perp\| \end{pmatrix}.$$

Since Q is orthogonal, $A^T A = R^T Q^T Q R = R^T R$, and therefore

$$\det(A^T A) = \det(R^T R) = \|\mathbf{a}_1^\perp\|^2 \|\mathbf{a}_2^\perp\|^2 \cdots \|\mathbf{a}_n^\perp\|^2.$$

Comparing this with (8.2) gives the desired conclusion. 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 $\mathbf{a}_1^\perp, \mathbf{a}_2^\perp, \dots, \mathbf{a}_n^\perp$. (The horizontal "floor" represents the plane spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$.) The parallelepiped spanned by the $\mathbf{a}$'s has the same volume as the rectangular block spanned by the $\mathbf{a}^\perp$'s.



8.5. Corollary. $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \geq 0$ *for all vectors* $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ *in* $\mathbf{R}^N$.

For $n = N$ Theorem 8.4 gives the following result.

8.6. 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. Since A is square, we have $\det(A^TA) = \det(A^T)\det(A) = (\det(A))^2$ by Theorem 3.7(iii) and therefore $\mathrm{vol}_n(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \sqrt{(\det(A))^2} = |\det(A)|$ by Theorem 8.4. QED

8.2. Orientations

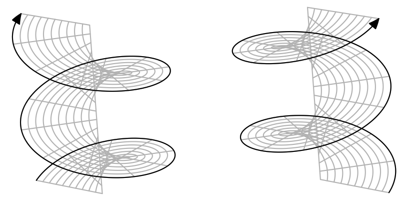
Oriented vector spaces. You are probably familiar with orientations of vector spaces of dimension ≤ 3. An orientation of a point is a sign, positive or negative.

$$\oplus \qquad \ominus$$

An orientation of a line is a direction, an arrow pointing either way.

An orientation of a plane is a direction of rotation, clockwise versus counterclockwise.

An orientation of a three-dimensional space is a "handedness" convention, left hand versus right hand.



These notions can be generalized as follows. Let V be an n-dimensional vector space over the real numbers. A *frame* or *ordered basis* of V is an n-tuple $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ consisting of vectors $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ which form a basis of V. In other words, a frame is a basis together with a specified ordering among the basis vectors. An *oriented frame* of V is an ordered $n+1$-tuple $\mathscr{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ consisting of a frame $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ together with a sign $\varepsilon = \pm 1$. Suppose that $\mathscr{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ and $\mathscr{B}' = (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n; \varepsilon')$ are two oriented frames of V. Then we can write $\mathbf{b}'_j = \sum_{i=1}^n a_{i,j}\mathbf{b}_i$ and $\mathbf{b}_j = \sum_{i=1}^n a'_{i,j}\mathbf{b}'_i$ with unique coefficients $a_{i,j}$ and $a'_{i,j}$. The $n \times n$-matrices $A = (a_{i,j})$ and $A' = (a'_{i,j})$ satisfy $AA' = A'A = I$ and are therefore invertible. In particular, the determinant of A is nonzero. If $\varepsilon' = \operatorname{sign}(\det(A))\varepsilon$ we say that the oriented frames $\mathscr{B}$ and $\mathscr{B}'$ define the *same orientation* of V. If $\varepsilon' = -\operatorname{sign}(\det(A))\varepsilon$ we say that the oriented frames $\mathscr{B}$ and $\mathscr{B}'$ define *opposite orientations*.

For instance, if $(\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n) = (\mathbf{b}_2, \mathbf{b}_1, \dots, \mathbf{b}_n)$, 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 oriented frames

$$(\mathbf{b}_2, \mathbf{b}_1, \dots, \mathbf{b}_n; 1) \quad \text{and} \quad (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; 1)$$

define opposite orientations, while $(\mathbf{b}_2, \mathbf{b}_1, \dots, \mathbf{b}_n; 1)$ and $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; -1)$ define the same orientation.

We know now what it means for two bases to have "the same orientation", but what about the concept of an orientation itself? We define the orientation of V determined by the oriented frame $\mathscr{B}$ to be the collection of all oriented frames that have the same orientation as $\mathscr{B}$. (This is analogous to the definition of the number 29 as being the collection of all sets that contain twenty-nine elements.) The orientation determined by $\mathscr{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ is denoted by $[\mathscr{B}]$ or $[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon]$. So if $\mathscr{B}$ and $\mathscr{B}'$ define the same orientation then $[\mathscr{B}] = [\mathscr{B}']$. If they define opposite orientations we write $[\mathscr{B}] = -[\mathscr{B}']$. There are two possible orientations of V. An *oriented vector space* is a vector space together with a specified orientation. This preferred orientation is then called *positive*.

8.7. Example. The *standard orientation* on $\mathbf{R}^n$ is the orientation $[\mathbf{e}_1, \dots, \mathbf{e}_n; 1]$, where $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ is the standard ordered basis. 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 frame $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ of V. Because L is invertible, the $n+1$-tuple $(L\mathbf{b}_1, L\mathbf{b}_2, \dots, L\mathbf{b}_n; \varepsilon)$ is an oriented frame of W. If this frame 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 $(\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_n; \varepsilon')$ is another positively oriented frame of V, then $\mathbf{b}'_i = \sum_j a_{i,j}\mathbf{b}_j$ with $\varepsilon' = \operatorname{sign}(\det(a_{i,j}))\varepsilon$. Therefore $L\mathbf{b}'_i = L\bigl(\sum_j a_{i,j}\mathbf{b}_j\bigr) = \sum_j a_{i,j}L\mathbf{b}_j$, and hence the two oriented frames $(L\mathbf{b}_1, L\mathbf{b}_2, \dots, L\mathbf{b}_n; \varepsilon)$ and $(L\mathbf{b}'_1, L\mathbf{b}'_2, \dots, L\mathbf{b}'_n; \varepsilon')$ of W determine the same orientation of W.

Oriented manifolds. Now let M be a manifold in $\mathbf{R}^N$. 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 $\psi\colon U \to \mathbf{R}^N$ of M at $\mathbf{x}$ with the property that $D\psi(\mathbf{t})\colon \mathbf{R}^n \to T_{\psi(\mathbf{t})}M$ preserves the orientation for all $\mathbf{t} \in U$. (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 smooth 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 $\mathbf{x} \in M$.

8.8. Proposition. *A smooth hypersurface in $\mathbf{R}^N$ is orientable if and only if it possesses a unit normal vector field.*

Proof. Let M be a smooth hypersurface in $\mathbf{R}^N$ and put $n = \dim(M) = N-1$. Suppose M possesses a unit normal vector field. Let $\mathbf{b}_1$, $\mathbf{b}_2, \dots, \mathbf{b}_n$ be a basis of $T_\mathbf{x}M$ for some $\mathbf{x} \in M$. Then $\mathbf{n}(\mathbf{x}), \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$ is a basis of $\mathbf{R}^N$, because $\mathbf{n}(\mathbf{x}) \perp \mathbf{b}_i$ for all i. Choose $\varepsilon = \pm 1$ such that $(\mathbf{n}(\mathbf{x}), \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ is a positively oriented frame of $\mathbf{R}^N$. Then we call $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ a positively oriented frame of $T_\mathbf{x}M$. 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 smooth hypersurface in $\mathbf{R}^N$. For each $\mathbf{x} \in M$ the tangent space $T_\mathbf{x}M$ is n-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 $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ be a positively oriented frame of $T_\mathbf{x}M$. The *positive* unit normal vector is that unit normal vector $\mathbf{n}(\mathbf{x})$ that makes $(\mathbf{n}(\mathbf{x}), \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon)$ a positively oriented frame of $\mathbf{R}^n$. In Exercise 8.7 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

8.9. Example. Let us regard $\mathbf{R}^{N-1}$ as the linear 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 orientations on $\mathbf{R}^N$ and on $\mathbf{R}^{N-1}$ are $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N; 1]$, resp. $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N-1}; 1]$. What is the positive unit normal $\mathbf{n}$ to $\mathbf{R}^{N-1}$? According to the proof of Proposition 8.8 we must choose $\mathbf{n}$ in such a way that

$$[\mathbf{n}, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N-1}; 1) = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N; 1].$$

The only two possibilities are $\mathbf{n} = \rho\mathbf{e}_N$ with $\rho = \pm 1$. By parts (i) and (iv) of Exercise 8.5 we have

$$\begin{aligned}[\mathbf{n}, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N-1}; 1] &= (-1)^{N+1}[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N-1}, \mathbf{n}; 1] \\ &= (-1)^{N+1}[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N-1}, \rho\mathbf{e}_N; 1] = (-1)^{N+1}\rho[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N-1}, \mathbf{e}_N; 1],\end{aligned}$$

so we want $(-1)^{N+1}\rho = 1$. We conclude that 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 smooth 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.

8.10. 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 smooth hypersurface $\phi^{-1}(c)$ 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 smooth hypersurface in $\mathbf{R}^N$ (if nonempty), and also that $T_\mathbf{x}M = \ker(D\phi(\mathbf{x})) = \operatorname{grad}(\phi)(\mathbf{x})^\perp$. 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 8.8 we conclude that M is orientable. QED

8.11. Example. Taking $\phi(\mathbf{x}) = \|\mathbf{x}\|^2$ and $c = r^2$ we obtain that the 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 an atlas of M consisting of local parametrizations $\psi_i\colon U_i \to \mathbf{R}^N$ with the property 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

$$\mu_i = \psi_i^*(\mu) = \sqrt{\det(D\psi_i(\mathbf{t})^T D\psi_i(\mathbf{t}))}\, dt_1\, dt_2 \cdots dt_n.$$

Theorem 8.4 tells us that the square-root factor measures the volume of the n-dimensional parallelepiped 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 parallelepipeds inside M.

8.12. 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 3.14 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

$$\begin{aligned}
\phi^*(\mu_i) &= \sqrt{\det\bigl(D\psi_i(\phi(\mathbf{u}))^T D\psi_i(\phi(\mathbf{u}))\bigr)}\,\lvert\det(D\phi(\mathbf{u}))\rvert\,du_1\,du_2\cdots du_n \\
&= \sqrt{\det(D\phi(\mathbf{u}))^T \det\bigl(D\psi_i(\phi(\mathbf{u}))^T D\psi_i(\phi(\mathbf{u}))\bigr)\det(D\phi(\mathbf{u}))}\,du_1\,du_2\cdots du_n \\
&= \sqrt{\det\bigl((D\psi_i(\phi(\mathbf{u}))D\phi(\mathbf{u}))^T D\psi_i(\phi(\mathbf{u}))D\phi(\mathbf{u})\bigr)}\,du_1\,du_2\cdots du_n \\
&= \sqrt{\det\bigl((D\psi_j(\mathbf{u}))^T D\psi_j(\mathbf{u})\bigr)}\,du_1\,du_2\cdots du_n = \mu_j,
\end{aligned}$$

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 these old-fashioned notations: volume forms are seldom exact! Another thing to remember is that the volume form μ_M depends on the embedding of M into $\mathbf{R}^N$. It changes if we dilate or shrink or otherwise deform M.

8.13. Example. Let U be an open subset of $\mathbf{R}^n$. Recall from Example 6.4 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 $\mathbf{R}^n$.

8.14. 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 6.6 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.

8.15. 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*

$$\mu_{M,\mathbf{x}}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = \begin{cases} \operatorname{vol}_n(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) & \text{if } (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n; 1) \text{ is a positively oriented frame}, \\ -\operatorname{vol}_n(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) & \text{if } (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n; 1) \text{ is a negatively oriented frame}, \\ 0 & \text{if } \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \text{ are linearly dependent}, \end{cases}$$

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 local parametrization of

M. Choose $\mathbf{t} \in U$ satisfying $\psi(\mathbf{t}) = \mathbf{x}$. 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

$$g(\mathbf{t}) = \psi^*(\omega_{\mathbf{x}})(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = \omega_{\mathbf{x}}(D\psi(\mathbf{t})\mathbf{e}_1, D\psi(\mathbf{t})\mathbf{e}_2, \ldots, D\psi(\mathbf{t})\mathbf{e}_n),$$

where in the second equality we used the definition of pullback. Then the tuple $(D\psi(\mathbf{t})\mathbf{e}_1, D\psi(\mathbf{t})\mathbf{e}_2, \ldots, D\psi(\mathbf{t})\mathbf{e}_n; 1)$ is a positively oriented frame of $T_{\mathbf{x}}M$ and, moreover, are the columns of the matrix $D\psi(\mathbf{t})$, so by Theorem 8.4 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. Recall the vector-valued forms

$$d\mathbf{x} = \begin{pmatrix} dx_1 \\ \vdots \\ dx_N \end{pmatrix} \qquad \text{and} \qquad *d\mathbf{x} = \begin{pmatrix} *dx_1 \\ \vdots \\ *dx_N \end{pmatrix}$$

on $\mathbf{R}^N$, which were introduced in Section 2.5. We will use these forms to give a convenient expression for the volume form on a hypersurface. Let M be an oriented hypersurface in $\mathbf{R}^N$. 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-form on M, where $n = \dim(M) = N - 1$. On the other hand we have the n-form $*(\mathbf{F} \cdot d\mathbf{x}) = \mathbf{F} \cdot *d\mathbf{x}$.

8.16. 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 make an orthogonal change of coordinates in $\mathbf{R}^N$ in such a way that $(\mathbf{e}_1, \mathbf{e}_2 \ldots, \mathbf{e}_{N-1}; 1)$ is a positively oriented frame of $T_{\mathbf{x}}M$. Then, according to Example 8.9, 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 $\mu_{M,\mathbf{x}}(\mathbf{e}_1, \ldots, \mathbf{e}_{N-1}) = 1$. 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 $(\mathbf{F} \cdot *d\mathbf{x})(\mathbf{e}_1, \ldots, \mathbf{e}_{N-1}) = (-1)^{N+1}F_N$. This proves that

$$(\mathbf{F} \cdot *d\mathbf{x})_{\mathbf{x}}(\mathbf{e}_1, \ldots, \mathbf{e}_{N-1}) = (\mathbf{F}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}))\mu_M(\mathbf{e}_1, \ldots, \mathbf{e}_{N-1}),$$

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 a local parametrization $\psi\colon U \to \mathbf{R}^N$ of M at $\mathbf{x}$. 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 $(\mathbf{e}_1, \mathbf{e}_2 \dots, \mathbf{e}_{N-1}; 1)$ is a positively oriented frame 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

$$D\tilde{\psi}(\tilde{\mathbf{t}})\mathbf{e}_i = D\psi(\mathbf{t})A\mathbf{e}_i = D\psi(\mathbf{t})\mathbf{a}_i = \mathbf{e}_i \tag{8.3}$$

for $i = 1, 2, \dots, N-1$. (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 $\tilde{\psi}^*((\mathbf{F}\cdot\mathbf{n})\mu_M)$ 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

$$\tilde{\psi}^*((\mathbf{F}\cdot\mathbf{n})\mu_M) = \Bigl(\sum_{i=1}^N \tilde{\psi}^*(F_i n_i)\Bigr)\sqrt{\det(D\tilde{\psi}^T D\tilde{\psi})}\, d\tilde{t}_1\, d\tilde{t}_2 \cdots d\tilde{t}_{N-1}.$$

From formula (8.3) we have $\det(D\tilde{\psi}(\tilde{\mathbf{t}})^T D\tilde{\psi}(\tilde{\mathbf{t}})) = 1$. So evaluating this expression at the point $\tilde{\mathbf{t}}$ and using $\mathbf{n}(\mathbf{x}) = (-1)^{N+1}\mathbf{e}_N$ we get

$$(\tilde{\psi}^*(\mathbf{F}\cdot\mathbf{n})\mu_M)_{\tilde{\mathbf{t}}} = (-1)^{N+1}F_N(\mathbf{x})\, d\tilde{t}_1\, d\tilde{t}_2 \cdots d\tilde{t}_{N-1}.$$

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 (8.3) 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

$$(\tilde{\psi}^*(\mathbf{F}\cdot *d\mathbf{x}))_{\tilde{\mathbf{t}}} = (-1)^{N+1}F_N(\mathbf{x})\, d\tilde{t}_1\, d\tilde{t}_2 \cdots d\tilde{t}_{N-1}.$$

We conclude that $(\tilde{\psi}^*(\mathbf{F}\cdot\mathbf{n})\mu_M)_{\tilde{\mathbf{t}}} = (\tilde{\psi}^*(\mathbf{F}\cdot *d\mathbf{x}))_{\tilde{\mathbf{t}}}$, in other words $((\mathbf{F}\cdot\mathbf{n})\mu_M)_{\mathbf{x}} = (\mathbf{F}\cdot *d\mathbf{x})_{\mathbf{x}}$. 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 on $\mathbf{R}^N$. 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 8.16 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 are the following formulas involving the volume form on a hypersurface. The formulas provide 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 $\mathbf{n}$.

8.17. Corollary. *Let* $\mathbf{n}$ *be the unit normal vector field and* μ_M *the volume form of the oriented hypersurface* M*. Then*

$$\boxed{\mu_M = \mathbf{n}\cdot *d\mathbf{x} \quad \text{and} \quad *d\mathbf{x} = \mathbf{n}\mu_M.}$$

Proof. Set $\mathbf{F}=\mathbf{n}$ in Theorem 8.16. Then $\mathbf{F}\cdot\mathbf{n}=1$, because $\|\mathbf{n}\|=1$, and hence $\mu_M=\mathbf{n}\cdot *d\mathbf{x}$. Now take $\mathbf{F}$ to be any vector field that is tangent to M. Then $\mathbf{F}\cdot\mathbf{n}=0$, because $\mathbf{n}$ is normal to M, so $\mathbf{F}\cdot *d\mathbf{x}=0$. It follows that on the hypersurface M the $\mathbf{R}^n$-valued $n-1$-form $*d\mathbf{x}$ is equal to the product of $\mathbf{n}$ and a scalar $n-1$-form ν: $*d\mathbf{x}=\mathbf{n}\nu$. Taking the dot product of both sides with $\mathbf{n}$ we obtain $\nu=\mathbf{n}\cdot\mathbf{n}\nu=\mathbf{n}\cdot *d\mathbf{x}=\mu_M$, i.e. $*d\mathbf{x}=\mathbf{n}\mu_M$. QED

8.18. 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 8.10 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. Let $\mathbf{a}$ and $\mathbf{b}$ be vectors in $\mathbf{R}^N$.

(i) Deduce from Theorem 8.4 that the area of the parallelogram spanned by $\mathbf{a}$ and $\mathbf{b}$ 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).

(ii) Show that for $N=3$ we have $\|\mathbf{a}\|\,\|\mathbf{b}\|\,\sin\phi=\|\mathbf{a}\times\mathbf{b}\|$. (Consider $\det(\mathbf{a},\mathbf{b},\mathbf{a}\times\mathbf{b})$.)

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. Let V be an n-dimensional vector space with basis $\mathbf{b}_1$, $\mathbf{b}_2, \dots, \mathbf{b}_n$. Let $\varepsilon = \pm 1$. Prove the following identities concerning orientations of V.

(i) $[\mathbf{b}_1, \dots, c\mathbf{b}_i, \dots, \mathbf{b}_n; \varepsilon] = \operatorname{sign}(c)[\mathbf{b}_1, \dots, \mathbf{b}_i, \dots, \mathbf{b}_n; \varepsilon]$ for all nonzero scalars c.
(ii) $[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon] = \varepsilon[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; 1]$.
(iii) $[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; -\varepsilon] = -[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon]$.
(iv) $[\mathbf{b}_{\sigma(1)}, \mathbf{b}_{\sigma(2)}, \dots, \mathbf{b}_{\sigma(n)}; \varepsilon] = \operatorname{sign}(\sigma)[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n; \varepsilon]$ for all permutations σ in S_n.

8.6. A frame $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ of $\mathbf{R}^n$ is *orthonormal* if the vectors $\mathbf{b}_1$, $\mathbf{b}_2, \dots, \mathbf{b}_n$ form an orthonormal basis of $\mathbf{R}^n$. Denote the set of orthonormal frames by F_n and the set of positively oriented orthonormal frames by F_n^+.

(i) Explain how to identify F_n with the orthogonal group $\mathbf{O}(n)$ and F_n^+ with the special orthogonal group $\mathbf{SO}(n)$. Conclude that F_n and F_n^+ are manifolds of dimension $\frac{1}{2}n(n-1)$. (Use Theorem 6.18).
(ii) Explain how to identify F_3^+ with the configuration space M_0 discussed in Example 1.7 and conclude that M_0 is a three-dimensional projective space. (See the discussion following Theorem 6.18).

8.7. Show that the unit normal vector field $\mathbf{n}\colon M \to \mathbf{R}^N$ defined in the proof of Proposition 8.8 is smooth. (Compute $\mathbf{n}$ in terms of an orientation-preserving local parametrization ψ of M.)

8.8. 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. We 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 $\psi^*(\mu_M) = \sqrt{1 + \|\operatorname{grad} f(\mathbf{x})\|^2}\, dx_1\, dx_2 \cdots dx_n$.

8.9. Let $M = \operatorname{graph}(f)$ be the oriented hypersurface of Exercise 8.8.

(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 $\psi^*(\mu_M) = \sqrt{1 + \|\operatorname{grad} f(\mathbf{x})\|^2}\, dx_1\, dx_2 \cdots dx_n$ of Exercise 8.8 from Corollary 8.17 by substituting $x_{n+1} = f(x_1, x_2, \dots, x_n)$. (You must replace N with $n+1$ in Corollary 8.17.)

8.10. Let $\psi\colon U \to \mathbf{R}^N$ be an embedding of an open subset U of $\mathbf{R}^n$ into $\mathbf{R}^N$. Let M be the image of ψ and let μ be the volume form of M. Now let R be a nonzero number, let ψ_R be the embedding $\psi_R(\mathbf{t}) = R\psi(\mathbf{t})$, let M_R be the image of ψ_R, and let μ_R be the volume form of M_R. Show that $\psi_R^*(\mu_R) = R^n\psi^*(\mu)$. (Use $\psi^*(\mu) = \sqrt{\det(D\psi_i(\mathbf{t})^T D\psi_i(\mathbf{t}))}\, dt_1\, dt_2 \cdots dt_n$.)

CHAPTER 9

Integration and Stokes' theorem for 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 is defined as follows. The *n-dimensional half-space* is

$$\mathbf{H}^n = \{\, x \in \mathbf{R}^n \mid x_n \geq 0 \,\}.$$

The *boundary* of $\mathbf{H}^n$ is $\partial\mathbf{H}^n = \{\, x \in \mathbf{R}^n \mid x_n = 0 \,\}$ 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 each $\mathbf{x} \in M$ 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 $\psi(U \cap \mathbf{H}^n) = M \cap V$.

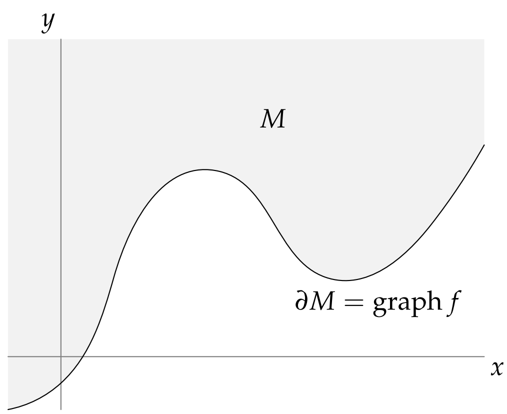
You should compare this definition carefully with Definition 6.3 of a manifold. If $\mathbf{x} = \psi(\mathbf{t})$ with $\mathbf{t} \in \partial\mathbf{H}^n$, then $\mathbf{x}$ is a *boundary point* of M. The *boundary* of M is the set of all boundary points and is denoted by ∂M. Its complement $M \setminus \partial M$ is the *interior* of M and is denoted by $\operatorname{int}(M)$.

Definition 9.1 does not rule out the possibility that the boundary of an n-manifold with boundary might be empty! If the boundary of M is empty, then M is a manifold in the sense of Definition 6.3. If the boundary ∂M is nonempty, then ∂M is an $n-1$-dimensional manifold. The interior $\operatorname{int}(M)$ is an n-manifold.

The most obvious example of an n-manifold with boundary is the half-space $\mathbf{H}^n$ itself. Its boundary is the hyperplane $\partial\mathbf{H}^n$, which is a copy of $\mathbf{R}^{n-1}$, and its interior is the open half-space $\{\, \mathbf{x} \in \mathbf{R}^n \mid x_n > 0 \,\}$. 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 above the graph* or the *supergraph of f* is the set consisting of all $\binom{\mathbf{x}}{y}$ in U such that $y \geq f(\mathbf{x})$.



We assert that the supergraph is an n-manifold with boundary, 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 6.2 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 the image $M = \psi(U \cap \mathbf{H}^n)$ is therefore 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 $u \geq 0$, this is equivalent to $\mathbf{x} \in U'$ and $y \geq f(\mathbf{x})$. Thus M is exactly the supergraph. 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 "above" 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:

$$y_1 = f_1(\mathbf{x}), \quad y_2 = f_2(\mathbf{x}), \quad \dots, \quad y_{m-1} = f_{m-1}(\mathbf{x}), \quad y_m \geq f_m(\mathbf{x}).$$

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 6.12, to manifolds with boundary.

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*

$$\phi_1(\mathbf{x}) = c_1, \quad \phi_2(\mathbf{x}) = c_2, \quad \dots, \quad \phi_{m-1}(\mathbf{x}) = c_{m-1}, \quad \phi_m(\mathbf{x}) \geq c_m.$$

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 $\partial M = \phi^{-1}(\mathbf{c})$.

We will not spell out the proof, which is similar to that of Theorem 6.12. The statement remains true if we replace "$\geq$" with "$\leq$", as one sees by replacing ϕ with $-\phi$.

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}.

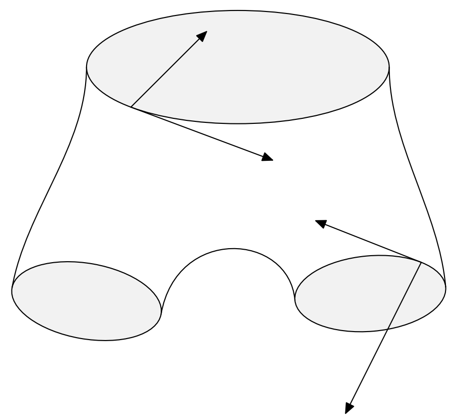
If more than one inequality is involved, singularities often arise. A simple example is the closed quadrant in $\mathbf{R}^2$ given by the pair of inequalities $x \geq 0$ and $y \geq 0$. This is not a manifold with boundary because its edge has a sharp angle at the origin. Similarly, a closed square 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 $\mathbf{R}^n$.

The *tangent space* to a manifold with boundary M at a point $\mathbf{x}$ is defined in the usual way: choose a local parametrization ψ of M at $\mathbf{x}$ and put

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

As in the case of a manifold, the tangent space does not depend on the choice of the embedding ψ. At boundary points we can distinguish between three different types of tangent vectors. 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 a unique vector $\mathbf{u} \in \mathbf{R}^n$. We say that the tangent vector $\mathbf{v}$

$$\begin{cases} \textit{points inward} & \text{if } u_n > 0, \\ \textit{is tangent to } \partial M & \text{if } u_n = 0, \\ \textit{points outward} & \text{if } u_n < 0. \end{cases}$$

The tangent space to the boundary at $\mathbf{x}$ is

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

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 of M gives rise to an orientation of the boundary ∂M by a method very similar to the one which produces an orientation of a hypersurface. (See Proposition 8.8.) Namely, for $\mathbf{x} \in \partial M$ we let $\mathbf{n}(\mathbf{x}) \in T_{\mathbf{x}}M$ 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* $\mathbf{n}$ on ∂M. Let $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{n-1}$ be a basis of $T_{\mathbf{x}}\partial M$. Then $\mathbf{n}(\mathbf{x}), \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{n-1}$ is a basis of $T_{\mathbf{x}}M$. Choose $\varepsilon = \pm 1$ such that $(\mathbf{n}(\mathbf{x}), \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{n-1}; \varepsilon)$ is a positively oriented frame of $T_{\mathbf{x}}M$. Then we define $(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{n-1}; \varepsilon)$ to be a positively oriented frame of $T_{\mathbf{x}}\partial M$. The resulting orientation of ∂M is called the *induced orientation*.

9.6. Example. Consider the upper half-space $\mathbf{H}^n$ with its standard orientation $[\mathbf{e}_1, \dots, \mathbf{e}_n; 1]$. At each point of $\partial\mathbf{H}^n$ the outward pointing normal is $-\mathbf{e}_n$. Since

$$\begin{aligned}[-\mathbf{e}_n, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}; 1] &= -[\mathbf{e}_n, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}; 1] \\ &= -(-1)^{n-1}[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}, \mathbf{e}_n; 1] = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}, \mathbf{e}_n; (-1)^n],\end{aligned}$$

the induced orientation on $\partial\mathbf{H}^n$ is $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}; (-1)^n]$.

You may wonder why we didn't get rid of the $(-1)^n$ by adopting a different convention for the orientation of the boundary. The justification for our convention is that it avoids the need for sign corrections in the statement of Stokes' theorem, Theorem 9.9.

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 suggests the possibility of integrating an n-form over an oriented n-manifold. One might try to do this by breaking up the manifold into n-chains and then integrating over each of the chains, but that turns out to be not so easy. Instead we shall employ the simpler method of breaking the differential form into small pieces and then integrating each of the pieces.

In the remainder of this section $M \subseteq \mathbf{R}^N$ denotes an n-manifold with boundary and α denotes a differential form on M.

Support. The *support* of α is defined as the set of all points $\mathbf{x}$ in M with the property that for every open ball B around $\mathbf{x}$ there is a $\mathbf{y} \in B \cap M$ such that $\alpha_{\mathbf{y}} \neq 0$. The support of α is denoted by $\operatorname{supp}(\alpha)$. If $\alpha_{\mathbf{x}}$ is nonzero, then $\mathbf{x}$ is in the support of α (because we can take $\mathbf{y} = \mathbf{x}$ for all B). But for $\mathbf{x}$ to be in the support it is not necessary for $\alpha_{\mathbf{x}}$ to be nonzero; we only need to be able to find points arbitrarily close to $\mathbf{x}$ where α is nonzero. In other words, $\mathbf{x}$ is *not* in the support if and only if there exists a ball B around $\mathbf{x}$ such that $\alpha_{\mathbf{y}} = 0$ for all $\mathbf{y} \in B \cap M$.

9.7. Example. Let $M = \mathbf{R}$ and $\alpha_i = f_i\,dx$, where f_i is one of the following smooth functions.

(i) $f_1(x) = \sin x$. This function has infinitely many zeroes, but they are all isolated: $\sin \pi k = 0$, but $\sin y \neq 0$ for y close to but distinct from πk. Thus $\operatorname{supp}(\alpha_1) = \mathbf{R}$.

(ii) f_2 is a nonzero polynomial function. Again f_2 has isolated zeroes, so $\operatorname{supp}(\alpha_2) = \mathbf{R}$.

(iii) $f_3(x) = \exp(-1/x)$ for $x > 0$ and $f_3(x) = 0$ for $x \leq 0$. (This function is similar to the function of Exercise B.3.) We have $f_3(x) > 0$ for all $x > 0$. It follows that $x \in \operatorname{supp}(\alpha_3)$ for all $x \geq 0$. On the other hand, negative x are not in the support, so $\operatorname{supp}(\alpha_3) = [0, \infty)$.

(iv) $f_4(x) = f_3(x-a)f_3(b-x)$, where $a < b$ are constants. We have $f_4(x) > 0$ for $a < x < b$ and $f_4(x) = 0$ for $x < a$ and $x > b$. Hence $\operatorname{supp}(\alpha_4) = [a, b]$.

9.8. Example. The volume form μ of M is nowhere 0, so $\operatorname{supp}(\mu) = M$.

Partitions of unity. Chopping differential forms into "little pieces" requires a device known as a partition of unity. Let $\psi_i\colon U_i \to \mathbf{R}^N$ be an atlas of M, where i runs over an indexing set $\mathscr{I}$. A *partition of unity subordinate to the atlas* is a collection of smooth functions $\lambda_i\colon M \to \mathbf{R}$ with the following properties:

(i) $\lambda_i \geq 0$ for all i;
(ii) $\operatorname{supp}(\lambda_i)$ is contained in $\psi_i(U_i)$ for all i;
(iii) for every $\mathbf{x} \in M$ there exists a ball B around $\mathbf{x}$ with the property that $\operatorname{supp}(\lambda_i) \cap B$ is empty for all but finitely many $i \in \mathscr{I}$;
(iv) $\sum_{i\in\mathscr{I}} \lambda_i = 1$.

Condition (iv) says that the functions λ_i add up to the constant function 1; it is in this sense that they "partition" the "unit" function. Together with the positivity condition (i) this implies that every λ_i takes values between 0 an 1. Condition (ii) expresses that λ_i is "small" in another sense as well: for every point $\mathbf{x}$ of M which is not contained in the coordinate patch $\psi_i(U_i)$ the function λ_i vanishes identically in a neighbourhood of $\mathbf{x}$. Condition (iii) is imposed to ensure that even if the indexing set $\mathscr{I}$ is infinite the sum in condition (iv) is finite at every point and is a well-defined smooth function.

It is a very useful technical fact that partitions of unity exist subordinate to any atlas of M. See Chapter 3 of the book [**Spi71**] for a proof, or see Exercise 9.3 for a special case.

Defining the integral. From now on we assume that M is oriented and that α is of degree $n = \dim(M)$. Moreover, we assume that the support of α is a compact set. (A subset of $\mathbf{R}^N$ is called *compact* if it is closed and bounded; see Appendix A.2. The compactness assumption is made to ensure that the integral of α is a proper integral and therefore converges. For instance, if we let $M = \mathbf{R}$ and α_i one of the 1-forms of Example 9.7, then only α_4 has a well-defined integral over M. Also note that the support of α is certainly compact if the manifold M itself is compact.)

Step 1. Assume there exists an orientation-preserving local parametrization $\psi\colon U \to \mathbf{R}^N$ of M with the property that the support of α is contained in $\psi(U\cap\mathbf{H}^n)$. Then we define the integral of α over M by

$$\int_M \alpha = \int_{U\cap\mathbf{H}^n} \psi^*(\alpha).$$

The right-hand side is well-defined because the integrand is of the form $\psi^*(\alpha) = g\, dt_1\, dt_2 \cdots dt_n$, where g is a smooth function on $U \cap \mathbf{H}^n$ which vanishes outside a compact subset. Moreover, the integral does not depend on the choice of ψ: if $\psi'\colon U' \to \mathbf{R}^N$ is another orientation-preserving local parametrization of M such that $\operatorname{supp}(\alpha)$ is contained in $\psi'(U' \cap \mathbf{H}^n)$, then, letting $\zeta = \psi^{-1} \circ \psi'$, we have $\psi \circ \zeta = \psi'$, so

$$\int_{U'\cap\mathbf{H}^n} (\psi')^*\alpha = \int_{U'\cap\mathbf{H}^n} (\psi\circ\zeta)^*(\alpha) = \int_{U'\cap\mathbf{H}^n} \zeta^*(\psi^*(\alpha)) = \int_{U\cap\mathbf{H}^n} \psi^*(\alpha),$$

where the last step uses Theorem 5.1 and the fact that ζ preserves the orientation.

Step 2. In the general case we choose an atlas of M consisting of orientation-preserving local parametrizations $\psi_i\colon U_i \to \mathbf{R}^N$, and we choose a partition of unity subordinate to this atlas consisting of functions $\lambda_i\colon M \to \mathbf{R}$. Let $\alpha_i = \lambda_i\alpha$. Then α_i is an n-form with support contained in $\psi_i(U_i \cap \mathbf{H}^n)$, so its integral is well-defined by step 1. Moreover, $\sum_{i\in\mathscr{I}} \alpha_i = \sum_{i\in\mathscr{I}} \lambda_i\alpha = (\sum_{i\in\mathscr{I}} \lambda_i)\alpha = \alpha$. We now define the integral of α by

$$\int_M \alpha = \sum_{i\in\mathscr{I}} \int_M \alpha_i = \sum_{i\in\mathscr{I}} \int_{U_i\cap\mathbf{H}^n} \psi_i^*(\alpha_i). \tag{9.1}$$

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 5.11.

9.9. Theorem (Stokes' theorem for manifolds). *Let α be an $n-1$-form with compact support on an oriented n-manifold with boundary M. Give the boundary ∂M the induced orientation. Then*

$$\boxed{\int_M d\alpha = \int_{\partial M} \alpha.}$$

Proof. *Step* 1. Suppose $M = \mathbf{H}^n$. Then we can write

$$\alpha = \sum_{i=1}^n g_i\, dt_1\, dt_2 \cdots \widehat{dt}_i \cdots dt_n$$

for certain smooth functions g_i defined on $\mathbf{H}^n$. We have

$$d\alpha = \sum_{i=1}^n (-1)^{i+1} \frac{\partial g_i}{\partial t_i}\, dt_1\, dt_2 \cdots dt_n.$$

The support of α is a compact subset of $\mathbf{H}^n$ and so is enclosed in a box of the shape

$$[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_{n-1}, b_{n-1}] \times [0, c]. \tag{9.2}$$

Therefore

$$\int_{\mathbf{H}^n} d\alpha = \sum_{i=1}^n (-1)^{i+1} \int_0^c \int_{a_{n-1}}^{b_{n-1}} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} \frac{\partial g_i}{\partial t_i}\, dt_1\, dt_2 \cdots dt_n.$$

The coefficients g_i of α are smooth functions on $\mathbf{H}^n$ which vanish outside the box (9.2). In particular the g_i and their partial derivatives vanish along all the walls of the box except possibly the "floor" $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_{n-1}, b_{n-1}] \times \{0\}$. Hence, by the fundamental theorem of calculus,

$$\int_{a_i}^{b_i} \frac{\partial g_i}{\partial t_i}\, dt_i = g_i(t_1, \ldots, b_i, \ldots, t_n) - g_i(t_1, \ldots, a_i, \ldots, t_n) = 0$$

for $i \le n-1$, while

$$\int_0^c \frac{\partial g_n}{\partial t_n}\, dt_n = g_n(t_1, \ldots, t_{n-1}, c) - g_n(t_1, \ldots, t_{n-1}, 0) = -g_n(t_1, \ldots, t_{n-1}, 0).$$

Hence

$$\int_{\mathbf{H}^n} d\alpha = (-1)^n \int_{a_{n-1}}^{b_{n-1}} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} g_n(t_1, \ldots, t_{n-1}, 0)\, dt_1\, dt_2 \cdots dt_{n-1} = \int_{\partial\mathbf{H}^n} \alpha,$$

where the sign $(-1)^n$ is accounted for by Example 9.6, which says that the orientation of $\partial\mathbf{H}^n = \mathbf{R}^{n-1}$ is $(-1)^n$ times the standard orientation of $\mathbf{R}^{n-1}$.

Step 2. In the general case we choose an atlas of M consisting of orientation-preserving local parametrizations $\psi_i \colon U_i \to \mathbf{R}^N$, and a subordinate partition of unity consisting of functions $\lambda_i \colon M \to \mathbf{R}$. Let $\alpha_i = \lambda_i \alpha$. Then

$$\int_M d\alpha = \sum_{i \in \mathscr{I}} \int_{U_i \cap \mathbf{H}^n} \psi_i^*(d\alpha) = \sum_{i \in \mathscr{I}} \int_{U_i \cap \mathbf{H}^n} d\psi_i^*(\alpha) = \sum_{i \in \mathscr{I}} \int_{U_i \cap \partial \mathbf{H}^n} \psi_i^*(\alpha) = \int_{\partial M} \alpha,$$

where the first and last equalities follow from the definition (9.1) of the integral, and the third equality uses step 1. QED

We conclude this section by considering a few special cases of the integral. 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 of 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 $\bar{f} = (\operatorname{vol}(M))^{-1} \int_M f\mu$. We can think of f as representing an electric charge density distributed over the manifold. Then the integral of f is the total charge of M and the mean of f is the average charge per unit volume. If f is nonnegative, we can think of f as a mass density, the integral as the total mass, and the mean as the average mass per unit volume. The *barycentre* or *centre of mass* of M with respect to a mass density f is the point $\bar{\mathbf{x}}$ in R^n defined as

$$\bar{\mathbf{x}} = \frac{\int_M f\mathbf{x}\mu}{\int_M f\mu}.$$

This is a vector-valued integral as discussed in Exercise 5.7. The i-th coordinate of $\mathbf{x}$ is i.e.

$$\bar{x}_i = \frac{\int_M f x_i \mu}{\int_M f\mu}.$$

The *centroid* of M is its barycentre with respect to a constant mass density f.

The volume form depends on the embedding of M into $\mathbf{R}^N$. If we change the embedding (i.e. distort the shape of M), then the volume form, the volume and the centroid of M will usually change as well.

9.3. Gauss and Stokes

Stokes' theorem, Theorem 9.9, 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 9.9 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 9.9. 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 Z is a oriented hypersurface in $\mathbf{R}^n$ with positive unit normal $\mathbf{n}$, then $*\alpha = (\mathbf{F}\cdot\mathbf{n})\mu_Z$ on Z by Theorem 8.16. 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 Z. For this reason the $n-1$-form $*\alpha$ is also called the *flux* of $\mathbf{F}$, and its integral over Z the *total flux* through Z.

Now let $Z = \partial M$, the boundary of a compact domain M in $\mathbf{R}^n$. Applying Stokes' theorem to M and $d{*\alpha}$ 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. Next 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 $f(\mathbf{x}) \leq y \leq g(\mathbf{x})$.

(i) 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?

(ii) 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 $g(x, y) = 2 - x^2 - y^2$.

(iii) Give an example showing that M is not necessarily a manifold with boundary if the condition $f(\mathbf{x}) < g(\mathbf{x})$ fails.

9.2. Let α and β be differential forms on a manifold M. Show that $\operatorname{supp}(\alpha\beta)$ is contained in $\operatorname{supp}(\alpha) \cap \operatorname{supp}(\beta)$. Give an example to show that we can have $\operatorname{supp}(\alpha\beta) \neq \operatorname{supp}(\alpha) \cap \operatorname{supp}(\beta)$.

9.3. Let $I_1 = [a_1, b_1)$ and $I_2 = (a_2, b_2]$ be two half-open intervals, where $a_1 < a_2 < b_1 < b_2$, and let $M = I_1 \cup I_2 = [a_1, b_2]$. Show that there exist two smooth functions $\lambda_1, \lambda_2\colon M \to \mathbf{R}$ with the following properties: (i) $\lambda_1(x) \geq 0$ and $\lambda_2(x) \geq 0$ for all $x \in M$; (ii) $\lambda_1 + \lambda_2 = 1$; (iii) $\operatorname{supp}(\lambda_1) \subseteq I_1$ and $\operatorname{supp}(\lambda_2) \subseteq I_2$. (Use the function of Example 9.7(iii) as a building block. First show there exist smooth functions $\chi_1, \chi_2\colon M \to \mathbf{R}$ with properties (i), (iii), and property (ii)′ $\chi_1(x) + \chi_2(x) > 0$ for all $x \in M$.)

9.4. Let $\psi\colon (a, b) \to \mathbf{R}^n$ be an embedding. Then $M = \psi((a, b))$ is a smooth 1-manifold. Let us call the direction of the tangent vector $\psi'(t)$ positive; this defines an orientation of M. Let μ be the element of arc length of M.

(i) Show that $\psi^*(\mu) = \|\psi'(t)\|\,dt = \sqrt{\psi_1'(t)^2 + \psi_2'(t)^2 + \cdots + \psi_n'(t)^2}\,dt$, where t denotes the coordinate on $\mathbf{R}$. Conclude that the arc length ("volume") of M is $\int_a^b \|\psi'(t)\|\,dt$.

(ii) Compute the arc length of the astroid $x = \cos^3 t$, $y = \sin^3 t$, where $t \in [0, \pi/2]$.

(iii) Consider a plane curve given in polar coordinates by an equation $r = f(\theta)$. Show that its element of arc length is $\sqrt{f'(\theta)^2 + f(\theta)^2}\,d\theta$. (Apply the result of part (i) to $\psi(\theta) = (f(\theta)\cos\theta, f(\theta)\sin\theta)$.)

(iv) Compute the arc length of the cardioid given by $r = 1 + \cos\theta$.

9.5. (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?

9.6. Deduce the following gradient and curl theorems from Gauss' divergence theorem:

$$\int_M \operatorname{grad}(f)\,dx_1\,dx_2 \cdots dx_n = \int_{\partial M} f\mathbf{n}\mu_{\partial M},$$

$$\int_M \operatorname{curl}(\mathbf{F})\,dx_1\,dx_2\,dx_3 = -\int_{\partial M} \mathbf{F} \times {*d\mathbf{x}}.$$

In the first formula M denotes a compact domain in $\mathbf{R}^n$, f a smooth function defined on M, and $\mathbf{n}$ the outward pointing unit normal vector field on the boundary ∂M of M. In the second formula M denotes a compact domain in $\mathbf{R}^3$ and $\mathbf{F}$ a smooth vector field defined on M. (Compare with Exercise 5.7. Use the second formula in Corollary 8.17.)

9.7. Suppose the region $x_3 \leq 0$ in $\mathbf{R}^3$ is filled with a stationary fluid of (not necessarily constant) density ρ. The gravitational force causes a pressure p inside the fluid. The gradient of p is equal to $\operatorname{grad}(p) = -g\rho\mathbf{e}_3$, where g is the gravitational acceleration (assumed to be constant). A 3-dimensional solid M is submerged in the fluid and kept stationary. At every point of the boundary ∂M the pressure imposes a force on the solid which is orthogonal to ∂M and points into M. The magnitude of this force per unit surface area is equal to the pressure p.

(i) Write the gradient version of Gauss' theorem (see Exercise 9.6) for the function $-p$.
(ii) Deduce Archimedes' Law: the buoyant force exerted on the submerged body is equal to the weight of the displaced fluid. $E\acute{\upsilon}\rho\eta\kappa\alpha$!
(iii) Let $\mathbf{F}_B$ denote the buoyant force and

$$\mathbf{x}_B = \frac{\int_M \rho\mathbf{x}\,dx_1\,dx_2\,dx_3}{\int_M \rho\,dx_1\,dx_2\,dx_3}$$

the *centre of buoyancy*, i.e. the barycentre of the fluid displaced by M. Fix any point $\mathbf{x}_0$ in $\mathbf{R}^3$. Recall that the *torque* about $\mathbf{x}_0$ of a force $\mathbf{F}$ acting at a point $\mathbf{x}$ is $(\mathbf{x} - \mathbf{x}_0) \times \mathbf{F}$. Show that the total torque about $\mathbf{x}_0$ produced by the fluid pressure on the surface of the solid is equal to $(\mathbf{x}_B - \mathbf{x}_0) \times \mathbf{F}_B$. (Apply the curl version of Gauss' theorem, Exercise 9.6, to the vector field $p(\mathbf{x} - \mathbf{x}_0)$.)

Part (iii) explains the principle of a self-righting boat: a boat with a heavy keel has its centre of mass below its centre of buoyancy, so that the force of gravity (which acts at the centre of mass) and the buoyant force (which acts at the centre of buoyancy) create a net torque that keeps the boat upright.

9.8. Let $R_1 \geq R_2 \geq 0$ 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.

9.9. 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,

$$\int_{\partial M}\Bigl(f\frac{\partial g}{\partial \mathbf{n}} - g\frac{\partial f}{\partial \mathbf{n}}\Bigr)\mu_{\partial M} = \int_M (f\Delta g - g\Delta f)\mu.$$

9.10. 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 $V_n(R) = \mathrm{vol}_n(B(R))$ and $A_n(R) = \mathrm{vol}_{n-1}(S(R))$. Also put $V_n = V_n(1)$ and $A_n = A_n(1)$.

(i) Deduce from Corollary 8.17 that the volume form on $S(R)$ is the restriction of ν to $S(R)$, where ν is as in Exercise 2.21. 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 2.21(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

$$\Bigl(\int_{-\infty}^{\infty} e^{-r^2}\,dr\Bigr)^n = A_n\int_0^{\infty} e^{-r^2}r^{n-1}\,dr.$$

(Take $f(r) = e^{-r^2}$ in part (iii) and let $R \to \infty$.)

(v) Using Exercises B.15 and B.16 conclude that

$$A_n = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}, \quad\text{whence}\quad A_{2m} = \frac{2\pi^m}{(m-1)!} \quad\text{and}\quad A_{2m+1} = \frac{2^{m+1}\pi^m}{1\cdot 3\cdot 5\cdots(2m-1)}.$$

(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

$$V_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}, \quad\text{whence}\quad V_{2m} = \frac{\pi^m}{m!} \quad\text{and}\quad V_{2m+1} = \frac{2^{m+1}\pi^m}{1\cdot 3\cdot 5\cdots(2m+1)}.$$

(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

In this chapter, to avoid endless repetitions it will be convenient to make a slight change in terminology. By a "manifold" we will now mean a "manifold with boundary". It is understood that the boundary of a "manifold with boundary" may be empty. If we specifically require the boundary to be empty, we will speak of a "manifold without boundary".

10.1. Brouwer's fixed point theorem

Let M be a manifold (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,

$$M = \{\, \mathbf{x} \in \mathbf{R}^n \mid 0 < \|\mathbf{x}\| \le 1 \,\}.$$

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

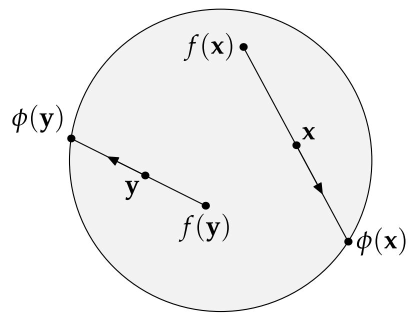
$$0 = \int_{\partial M} \alpha = \int_{\partial M} \beta = \operatorname{vol}(\partial M) \neq 0,$$

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 $M = \{\, \mathbf{x} \in \mathbf{R}^n \mid \|\mathbf{x}\| \le 1 \,\}$ 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 half-line starting at $f(\mathbf{x})$ and pointing in the direction of $\mathbf{x}$. This half-line 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 stirring 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. (See Theorems 10.8 and 10.10 below.) The notion of a smooth deformation can be defined formally as follows. The maps 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. 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.

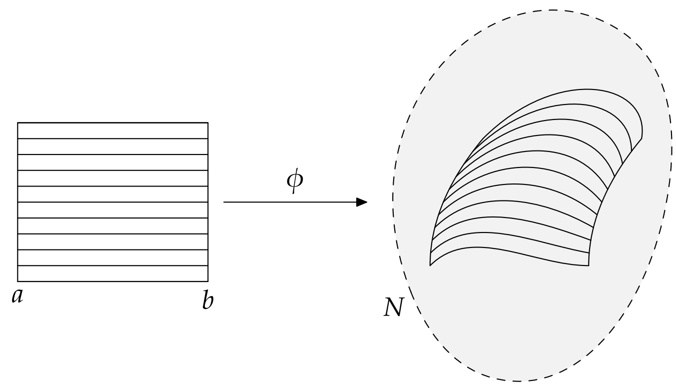
10.3. Example. Let $M = N = \mathbf{R}^n$ and $\phi_0(\mathbf{x}) = \mathbf{x}$ (identity map) and $\phi_1(\mathbf{x}) = \mathbf{0}$ (constant map). Then ϕ_0 and ϕ_1 are homotopic. A homotopy is given by $\phi(\mathbf{x}, t) = (1-t)\mathbf{x}$. This homotopy collapses Euclidean space onto the origin by moving each point radially inward at a speed equal to its distance to the origin. There are many other homotopies from ϕ_0 to ϕ_1, such as $(1-t)^2\mathbf{x}$ and $(1-t^2)\mathbf{x}$. We can also interchange the roles of 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 $\mathbf{R}^n \setminus \{\mathbf{0}\}$ 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 the punctured

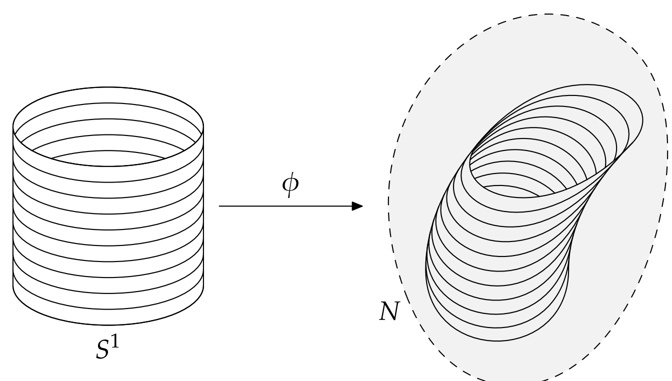
Euclidean space onto the unit sphere by smoothly stretching or shrinking each vector until it has length 1.

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 $\phi_1(\mathbf{x}) = \mathbf{x}$. 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 $\mathbf{x}_0$ by means of a contraction given by a very similar formula.) 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 $\mathbf{R}^n \setminus \{\mathbf{0}\}$ is *not* contractible.

Homotopy of paths. If M is an interval $[a,b]$ and N any manifold, then maps from M to N are nothing but paths (parametrized curves) in N. A homotopy of paths can be visualized as a piece of string moving through the manifold N.

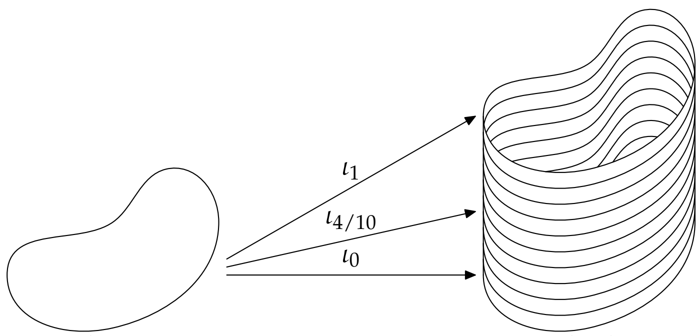


Homotopy of loops. A *loop* in a manifold N is a smooth map from the unit circle S^1 into N. This can be visualized as a thin rubber band sitting in N. A homotopy of loops $\phi\colon S^1 \times [0,1] \to N$ can be pictured as a rubber band floating through N from time 0 until time 1.



10.6. Example. Consider the two loops ϕ_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 $\mathbf{R}^2 \setminus \{\mathbf{0}\}$? 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-form α on the cylinder can be written as

$$\alpha = \sum_I f_I(x,t)\,dx_I + \sum_J g_J(x,t)\,dt\,dx_J,$$

with I running over increasing multi-indices of degree k and J over increasing multi-indices of degree $k-1$. (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,

$$\kappa\colon \Omega^k(M \times [0,1]) \to \Omega^{k-1}(M),$$

by taking the piece of α involving dt and integrating it over the unit interval,

$$\kappa(\alpha) = \sum_J \left(\int_0^1 g_J(x,t)\,dt \right) dx_J.$$

(In particular $\kappa(\alpha) = 0$ for any α that does not involve dt.) For a general manifold M we can write a k-form on the cylinder as $\alpha = \beta + dt\,\gamma$, where β and γ are forms on $M \times [0,1]$ (of degree k and $k-1$ respectively) that do not involve dt. We then define $\kappa(\alpha) = \int_0^1 dt\,\gamma$.

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 $\iota_1^*(\alpha) - \iota_0^*(\alpha) = \kappa(d\alpha) + d\kappa(\alpha)$ for all k-forms α 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 M of $\mathbf{R}^n$. The proof for arbitrary manifolds is similar. It suffices to consider two cases: $\alpha = f\,dx_I$ and $\alpha = g\,dt\,dx_J$.

Case 1. If $\alpha = f\,dx_I$, then $\kappa(\alpha) = 0$ and $d\kappa(\alpha) = 0$. Also

$$d\alpha = \frac{\partial f}{\partial t}\,dt\,dx_I + \sum_i \frac{\partial f}{\partial x_i}\,dx_i\,dx_I = \frac{\partial f}{\partial t}\,dt\,dx_I + \text{terms not involving } dt,$$

so

$$\begin{aligned} d\kappa(\alpha) + \kappa(\,d\alpha) = \kappa(\,d\alpha) &= \left(\int_0^1 \frac{\partial f}{\partial t}(\mathbf{x}, t)\,dt\right) dx_I \\ &= \big(f(\mathbf{x}, 1) - f(\mathbf{x}, 0)\big)\,dx_I = \iota_1^*(\alpha) - \iota_0^*(\alpha). \end{aligned}$$

Case 2. If $\alpha = g\,dt\,dx_J$, then $\iota_0^*(\alpha) = \iota_1^*(\alpha) = 0$ and

$$d\alpha = \sum_i \frac{\partial g}{\partial x_i}\,dx_i\,dt\,dx_J = -\sum_i \frac{\partial g}{\partial x_i}\,dt\,dx_i\,dx_J,$$

so

$$\kappa(d\alpha) = -\sum_{i=1}^n \left(\int_0^1 \frac{\partial g}{\partial x_i}(\mathbf{x}, t)\,dt\right) dx_i\,dx_J.$$

Also $\kappa(\alpha) = \left(\int_0^1 g(\mathbf{x}, t)\,dt\right) dx_J$, so

$$d\kappa(\alpha) = \sum_{i=1}^n \frac{\partial}{\partial x_i}\left(\int_0^1 g(\mathbf{x}, t)\,dt\right) dx_i\,dx_J = \sum_{i=1}^n \left(\int_0^1 \frac{\partial g}{\partial x_i}(\mathbf{x}, t)\,dt\right) dx_i\,dx_J.$$

Hence $d\kappa(\alpha) + \kappa(d\alpha) = 0 = \iota_1^*(\alpha) - \iota_0^*(\alpha)$. 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-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 the form $\phi^*(\alpha)$ on $M \times [0, 1]$ 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-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 without 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 $M \times [0,1]$. 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,

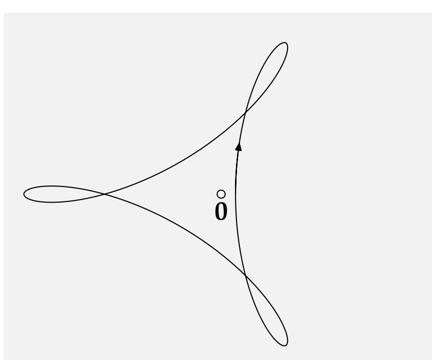
$$\begin{aligned} 0 = \int_{M\times[0,1]} \phi^*(d\alpha) = \int_{M\times[0,1]} d\phi^*(\alpha) = \int_{\partial(M\times[0,1])} \phi^*(\alpha) \\ = \int_M \phi_1^*(\alpha) - \int_M \phi_0^*(\alpha), \end{aligned}$$

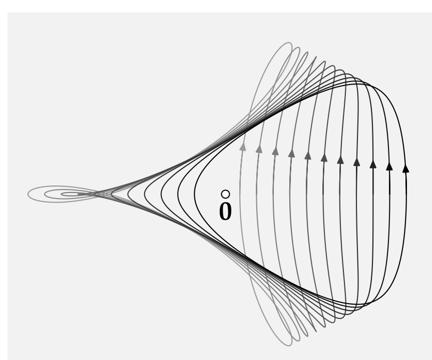
so $\int_M \phi_1^*(\alpha) = \int_M \phi_0^*(\alpha)$. QED

10.11. Corollary. *Homotopic loops in $\mathbf{R}^2 \setminus \{\mathbf{0}\}$ have the same winding number about the origin.*

Proof. Let M be the circle S^1, N the punctured plane $\mathbf{R}^2 \setminus \{\mathbf{0}\}$, and α_0 the angle form $(-y\,dx + x\,dy)/(x^2+y^2)$. Then a map ϕ from M to N is nothing but a loop in the punctured plane, and the integral $\int_M \phi^*(\alpha_0)$ is 2π times the winding number $w(\phi, \mathbf{0})$. (See Section 4.3.) Thus, if ϕ_0 and ϕ_1 are homotopic loops in N, Theorem 10.10 tells us that $w(\phi_0, \mathbf{0}) = w(\phi_1, \mathbf{0})$. QED

10.12. Example. Unfolding the three self-intersections in the path 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 about the origin 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 question 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. It is to make these statements precise that de Rham developed his cohomology theory. We shall not develop this theory in detail, 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 4.7 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 4.7. QED

10.16. Example. The punctured plane $\mathbf{R}^2 \setminus \{\mathbf{0}\}$ 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 $\mathbf{R}^n \setminus \{\mathbf{0}\}$ 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 forms of degree $k \geq 1$ on a contractible manifold are exact.*

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 $\alpha = \sum_i f_i \, dx_i$ be a 1-form and let g be the function $\kappa\phi^*(\alpha)$. Then

$$\phi^*(\alpha) = \sum_i f_i(t\mathbf{x}) d(tx_i) = \sum_i f_i(t\mathbf{x})(x_i \, dt + t \, dx_i),$$

so

$$g = \kappa\phi^*(\alpha) = \sum_i x_i \int_0^1 f_i(t\mathbf{x}) \, dt.$$

According to the proof of the Poincaré lemma the function g satisfies $dg = \alpha$, provided that $d\alpha = 0$. We checked this directly in Exercise 4.3.

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 $\mathbf{R}^2 \setminus \{\mathbf{0}\}$ is not contractible because it possesses a nonexact closed 1-form, namely the angle form. (See Example 4.6.) The *generalized angle form* is the $n-1$-form α_0 on punctured n-space $\mathbf{R}^n \setminus \{\mathbf{0}\}$ defined by

$$\alpha_0 = \frac{\mathbf{x} \cdot *d\mathbf{x}}{\|\mathbf{x}\|^n}.$$

10.19. Theorem. *The generalized angle form α_0 is a closed but non-exact $n-1$-form on punctured n-space. Hence punctured n-space is not contractible.*

Proof. That $d\alpha_0 = 0$ follows from Exercise 2.20. The $n-1$-sphere $M = S^{n-1}$ has unit normal vector field $\mathbf{x}$, so by Corollary 8.17 on M we have $\alpha_0 = \mu$, the volume form. Hence $\int_M \alpha_0 = \mathrm{vol}(M) \neq 0$. On the other hand, suppose α_0 was exact, $\alpha_0 = d\beta$ for an $n-1$-form β. Then

$$\int_M \alpha_0 = \int_M d\beta = \int_{\partial M} \beta = 0$$

by Stokes' theorem, Theorem 9.9. This is a contradiction, so α_0 is not exact. It now follows from the Poincaré lemma, Theorem 10.17, that $\mathbf{R}^n \setminus \{\mathbf{0}\}$ is not contractible. QED

A similar argument gives the next result.

10.20. Theorem. *Compact orientable manifolds without boundary of dimension ≥ 1 are not contractible.*

Proof. Let M be a compact orientable manifold of dimension $n \geq 1$. Let $\mu = \mu_M$ be the volume form of M. Then $\int_M \mu = \mathrm{vol}(M) \neq 0$. On the other hand, if M was contractible, then μ would be exact by the Poicaré lemma, so $\mu = d\nu$ and $\int_M \mu = \int_M d\nu = \int_{\partial M} \nu = 0$ because M has no boundary. This is a contradiction, so M is not contractible. QED

In particular the unit sphere S^n is not contractible for $n \geq 1$: it has a closed nonexact n-form. But how about forms of degree not equal to $n-1$? Without proof we state the following fact.

10.21. Theorem. *On $\mathbf{R}^n \setminus \{\mathbf{0}\}$ 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 $\mathbf{R}^n \setminus \{\mathbf{0}\}$ the integral

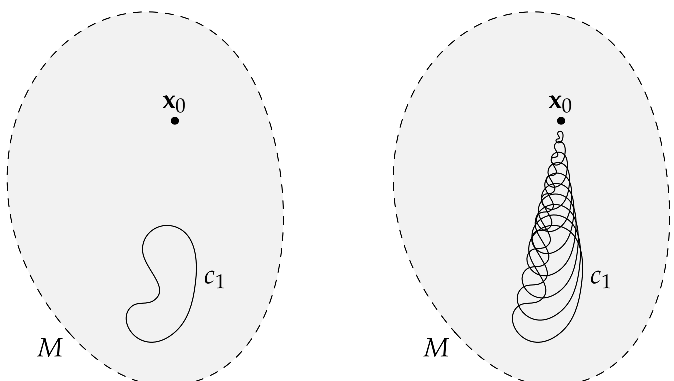
$$\boxed{w(M, \mathbf{0}) = \frac{1}{\operatorname{vol}_{n-1}(S^{n-1})} \int_M \frac{\mathbf{x} \cdot {*d\mathbf{x}}}{\|\mathbf{x}\|^n}}$$

is called the *winding number* of M about the origin. It generalizes the winding number of a closed path in $\mathbf{R}^2 \setminus \{\mathbf{0}\}$ 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.

10.22. 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 $\mathbf{R}^n \setminus \{\mathbf{0}\}$ 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 notion of homotopy 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 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 [**Mil03**] listed in the bibliography.

De Rham cohomology. The distinction between closed and exact differential forms on a manifold can be encoded in an invariant called de Rham cohomology. To explain this we need to review a little set theory and linear algebra.

Let X be a set. A binary relation $\sim$ (i.e. a relation among pairs of elements) on X is an *equivalence relation* if it is

(i) reflexive: $x \sim x$;
(ii) symmetric: if $x \sim y$ then $y \sim x$;
(iii) transitive: if $x \sim y$ and $y \sim z$ then $x \sim z$

for all $x, y, z \in X$.

10.23. Example. Consider the following binary relations.

(i) Let $X = \mathbf{R}$ with the order relation $\leq$ ("less than or equal to").
(ii) Let $X = \mathbf{Z}$. Fix $n \in \mathbf{Z}$. Define $x \equiv y$ if $x - y$ is divisible by n ("congruence modulo n").
(iii) Let X be the set of all people.
  (a) Define $x \bowtie y$ if x is a blood relative of y.
  (b) Define $x \smile y$ if x is a friend of y.
(iv) Let X be the set of straight lines in the plane. Define $x \parallel y$ if x is parallel to y.

Then we have the following table.

	reflexive	symmetric	transitive
(i)	Y	N	Y
(ii)	Y	Y	Y
(iiia)	Y	Y	?
(iiib)	N	N	N
(iv)	Y	Y	Y

Whether relation (iiia) is transitive is a matter of taste. In a broad sense we are all blood relatives of mitochondrial Eve, but perhaps this stretches the definition too far. Only relations (ii) and (iv) are true equivalence relations.

Let $\sim$ be an equivalence relation on a set X. The *equivalence class* of $x \in X$ is the set of all $y \in X$ that are equivalent to x. Notation:

$$[x] = \{\, y \in X \mid y \sim x \,\}.$$

Note that we have $[x] = [y]$ if $x \sim y$. Any element of an equivalence class is called a *representative* of that class. An equivalence class usually has many different representatives. The set of all equivalence classes is called the *quotient* of X by the equivalence relation and denoted by $X/\sim$.

In Example 10.23(ii) we write $\mathbf{Z}/n$ for the quotient. An element of $\mathbf{Z}/n$ is a *remainder class modulo n*,

$$[i] = \{i, i+n, i-n, i+n, i+2n, i-2n, \dots\} = \{\, i+kn \mid k \in \mathbf{Z}\,\}.$$

Example 10.23(iv) comes up in projective geometry. The equivalence class of a line in the plane is a *point at infinity* in the plane. If one adds the points at infinity to the plane one arrives at the *projective plane*.

10.24. Example. Let E be a real vector space. Fix a linear subspace F. Vectors $\mathbf{x}$, $\mathbf{y} \in E$ are *congruent modulo F* if $\mathbf{x} - \mathbf{y} \in F$. Congruence modulo F is an equivalence relation on E. The equivalence class $[\mathbf{x}]$ of $\mathbf{x}$ is the *affine subspace through* $\mathbf{x}$ *parallel to* F. The equivalence class $[\mathbf{0}]$ is equal to the linear subspace F itself. We denote the quotient by E/F and call it the *quotient of E by F*. A basic fact is that E/F is a vector space in its own right. The vector space operations are defined by $[\mathbf{x}]+[\mathbf{y}] = [\mathbf{x}+\mathbf{y}]$ and $c[\mathbf{x}] = [c\mathbf{x}]$ for all $\mathbf{x}$, $\mathbf{y} \in E$ and $c \in \mathbf{R}$. One can check that these operations are well-defined and obey the axioms of a real vector space. The origin of E/F is the equivalence class $[\mathbf{0}]$, and the opposite of a class $[\mathbf{x}]$ is the class $[-\mathbf{x}]$.

We have the following special case of Example 10.24 in mind. Let M be a manifold and $\Omega^k(M)$ the vector space of all k-forms on M. Let E be the linear subspace of $\Omega^k(M)$ consisting of all *closed* k-forms,

$$E = \{\, \alpha \in \Omega^k(M) \mid d\alpha = 0\,\},$$

and let F be the subspace of E consisting of all *exact* k-forms,

$$F = \{\, \alpha \in \Omega^k(M) \mid \alpha = d\beta \text{ for some } \beta \in \Omega^{k-1}(M)\,\}.$$

The quotient

$$H^k_{\mathrm{DR}}(M) = E/F = \frac{\text{closed } k\text{-forms on } M}{\text{exact } k\text{-forms on } M}$$

is the *de Rham cohomology* of M in degree k. Elements of $H^k_{\mathrm{DR}}(M)$ are equivalence classes $[\alpha]$ where α is a closed k-form and $\alpha \sim \beta$ if $\alpha - \beta$ is exact. Both vector spaces E and F are usually infinite-dimensional, but it may very well happen that the quotient $H^k_{\mathrm{DR}}(M)$ is finite-dimensional. For any n-manifold M we have $H^k(M) = 0$ if $k < 0$ or $k > n$. The reason is simply that we don't have any nonzero forms on M in degrees below 0 or above n. More interesting are the following assertions.

10.25. Theorem. *If M is contractible then*

$$H^k_{\mathrm{DR}}(M) = \begin{cases} \mathbf{R} & \text{if } k = 0, \\ 0 & \text{if } k \geq 1. \end{cases}$$

Proof. This is a restatement of the Poincaré lemma, Theorem 10.17. QED

10.26. Theorem. *Let M be a compact connected oriented n-manifold without boundary. Then $H^n_{\mathrm{DR}}(M)$ is 1-dimensional. A basis of $H^n_{\mathrm{DR}}(M)$ is $[\mu]$, the class of the volume form of M.*

We shall not prove this theorem in general (except to note that the class $[\mu]$ is nonzero by the proof of Theorem 10.20), but only for the unit circle.

10.27. THEOREM. *$H^1_{\mathrm{DR}}(S^1)$ is 1-dimensional. A basis of $H^1_{\mathrm{DR}}(S^1)$ is $[\mu]$, the class of the element of arc length of S^1.*

PROOF. The 1-form μ is closed but not exact and therefore defines a nonzero class in $H^1_{\mathrm{DR}}(S^1)$. We need to prove that for every 1-form α on S^1 (which is necessarily closed because $\dim(S^1) = 1$) there exists a constant k such that $[\alpha] = k[\mu]$. Assuming we could do this, let us first guess what k should be. The equality $[\alpha] = k[\mu]$ means $[\alpha - k\mu] = [0]$, i.e. $\alpha - k\mu$ is exact, i.e. $\alpha = k\mu + dg$ for some smooth function g on S^1. Integrating over the circle and using Stokes gives

$$\int_{S^1} \alpha = k \int_{S^1} \mu + \int_{S^1} dg = 2\pi k.$$

Let us therefore define $k = (2\pi)^{-1} \int_{S^1} \alpha$, as we must. Next we determine what g should be by solving $dg = \alpha - k\mu$ for g. We do this indirectly by first solving the equation $dh = c^*(\alpha - k\mu)$, where $c(t) = (\cos t, \sin t)$ is the usual parametrization of the circle. We have $c^*(\alpha) = f\,dt$, where $f : \mathbf{R} \to \mathbf{R}$ is a 2π-periodic function. The constant k is given in terms of f by

$$k = \frac{1}{2\pi} \int_{S^1} \alpha = \frac{1}{2\pi} \int_0^{2\pi} c^*(\alpha) = \frac{1}{2\pi} \int_0^{2\pi} f(t)\,dt. \tag{10.1}$$

A solution of the equation $dh = c^*(\alpha - k\mu) = (f - k)\,dt$ is the function $h(t) = \int_0^t (f(s) - k)\,ds = \int_0^t f(s)\,ds - kt$. This function has the property

$$\begin{aligned} h(t + 2\pi) = \int_0^{t+2\pi} f(s)\,ds - k(t + 2\pi) &= \int_0^t f(s)\,ds - kt + \int_t^{t+2\pi} f(s)\,ds - 2\pi k \\ &= h(t) + \int_0^{2\pi} f(s)\,ds - 2\pi k = h(t), \end{aligned}$$

where the last step follows from the periodicity of f and from (10.1). This shows that h is 2π-periodic, and therefore of the form $h(t) = g(c(t))$, i.e. $h = c^*(g)$, for some smooth function g on S^1. The equation $dh = c^*(\alpha - k\mu)$ then becomes $c^*(dg) = c^*(\alpha - k\mu)$, which implies $dg = \alpha - k\mu$ because c is a parametrization of the circle. QED

Exercises

10.1. Write a formula for the map ϕ occurring 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) The open ball $B^\circ(\varepsilon, \mathbf{x})$ of radius ε centred at $\mathbf{x}$ is convex.

10.5. Recall that $\mathbf{GL}(n, \mathbf{R})$ denotes the general linear group and $\mathbf{O}(n)$ the orthogonal group. (See Theorem 6.18 and Exercise 6.13.) Define a map $\phi\colon \mathbf{GL}(n, \mathbf{R}) \to \mathbf{O}(n)$ by $\phi(A) = Q$, where Q is the first factor in the QR-decomposition of A. (See the proof of Theorem 8.4.) Prove the following assertions.

(i) ϕ is a retraction.
(ii) ϕ is homotopic to the identity mapping of M.

10.6. Compute the turning number (see Exercise 4.14) of the loops ϕ_0 and ϕ_1 of Example 10.12. Despite the fact that the loops are homotopic they do not have the same turning number. Why does this not contradict Theorem 10.10?

10.7. Let $k \geq 1$ and let α be the k-form $f\,dx_I = f\,dx_{i_1}\,dx_{i_2}\cdots dx_{i_k}$ on $\mathbf{R}^n$. 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

$$\kappa\phi^*(\alpha) = \sum_{m=1}^{k} (-1)^{m+1} x_{i_m} \Big(\int_0^1 t^{k-1} f(t\mathbf{x})\,dt\Big)\,dx_{I\setminus i_m},$$

where we have written $dx_{I\setminus i_m}$ as an abbreviation for $dx_{i_1}\,dx_{i_2}\cdots \widehat{dx}_{i_m}\cdots dx_{i_k}$. Check directly that $d\kappa\phi^*(\alpha) + \kappa d\phi^*(\alpha) = \alpha$.

10.8. 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

$$\begin{aligned}\kappa\phi^*(\alpha) = \Big(\int_0^1 f(tx,ty,tz)t\,dt\Big)(x\,dy - y\,dx) + \Big(\int_0^1 g(tx,ty,tz)t\,dt\Big)(z\,dx - x\,dz)\\ + \Big(\int_0^1 h(tx,ty,tz)t\,dt\Big)(y\,dz - z\,dy).\end{aligned}$$

10.9. Let $\alpha = \sum_I f_I\,dx_I$ be a *closed* k-form whose coefficients f_I are smooth functions defined on $\mathbf{R}^n \setminus \{\mathbf{0}\}$ 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.6 to each f_I; see also Exercise 2.9.)

10.10. Let $\alpha = (2xyz - x^2y)\,dy\,dz + (xz^2 - y^2z)\,dz\,dx + (2xyz - xy^2)\,dx\,dy$.

(i) Check that α is closed.
(ii) Find a 1-form β such that $d\beta = \alpha$.

10.11. 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.

10.12. 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 $\phi(\mathbf{x}) = \mathbf{y}_0$.)

10.13. Let $\mathbf{x}_0 = (2,0)$ and let M be the twice-punctured plane $\mathbf{R}^2 \setminus \{\mathbf{0}, \mathbf{x}_0\}$. 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.)

10.14. Let M be a manifold. Let $[\alpha]$ be a class in $H^k_{\mathrm{DR}}(M)$ and let $[\beta]$ be a class in $H^l_{\mathrm{DR}}(M)$. Define the *product* of $[\alpha]$ and $[\beta]$ to be the class $[\alpha]\cdot[\beta] = [\alpha\beta] \in H^{k+l}_{\mathrm{DR}}(M)$. Show that $[\alpha]\cdot[\beta]$ is well-defined, i.e. independent of the choice of the representatives of the classes $[\alpha]$ and $[\beta]$.

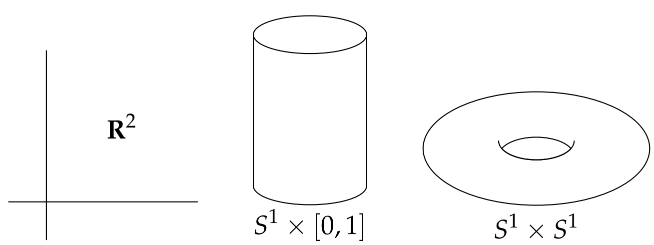
APPENDIX A

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* $x \in Y$.
$X \setminus Y$: the complement of Y in X. This is defined as the set of x in X such that x is not in Y.
(x, y): the ordered pair consisting of two elements x and 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; and $S^1 \times S^1$ is a torus.



$(x_1, x_2, \ldots, x_k)$: the k-tuple, i.e. ordered list, consisting of the k elements x_1, $x_2, \ldots, x_k$.
$X_1 \times X_2 \times \cdots \times X_k$: the k-fold Cartesian product of sets $X_1, X_2, \ldots, X_k$. This is by definition the set of all k-tuples $(x_1, x_2, \ldots, x_k)$ with $x_i \in X_i$. Examples: $\mathbf{R}^{n_1} \times \mathbf{R}^{n_2} \times \cdots \times \mathbf{R}^{n_k} = \mathbf{R}^{n_1+n_2+\cdots+n_k}$; and $S^1 \times S^1 \times \cdots \times S^1$ (k times) is a k-torus.

$\{ x \in X \mid P(x) \}$: the set of all $x \in X$ which have the property $P(x)$. Examples:

$$\{ x \in \mathbf{R} \mid 1 \le x < 3 \} \text{ is the interval } [1, 3),$$
$$\{ x \mid x \in X \text{ and } x \in Y \} \text{ is the intersection } X \cap Y,$$
$$\{ x \mid x \in X \text{ or } x \in Y \} \text{ is the union } X \cup Y,$$
$$\{ x \in X \mid x \notin Y \} \text{ is the complement } X \setminus Y.$$

$f : X \to Y$: 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

$$(f|A)(x) = \begin{cases} f(x) & \text{if } x \in A, \\ \text{not defined} & \text{if } x \notin A. \end{cases}$$

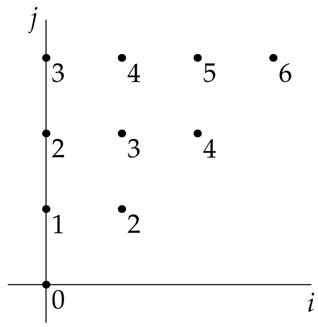
In other words, $f|A$ is equal to f on A, but "forgets" the values of f at points outside A.

$g \circ f$: the *composition* of f and g. If $f : X \to Y$ and $g : Y \to Z$ are functions, then $g \circ f : X \to Z$ 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 $f : X \to Y$ 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* or *invertible* if it is both injective and surjective. The function f is bijective if and only if it has an *inverse*, that is a map $g : Y \to X$ satisfying both $g(f(x)) = x$ for all $x \in X$ and $f(g(y)) = y$ for all $y \in Y$. Both conditions must be satisfied; that's why an inverse is sometimes called a *two-sided inverse* for emphasis. The inverse of a bijective map f is unique and is denoted by f^{-1}.

Let $f : X \to Y$ be an injective map. We can consider f as a bijection from X onto the image $f(X)$ and then form the inverse $f^{-1} : f(X) \to X$. We can extend f^{-1} to a map $g : Y \to X$ as follows: for $y \in f(X)$ put $g(y) = f^{-1}(y)$. For each $y \in Y$ which is not in $f(X)$ we choose at random an element $x \in X$ and define $g(y) = x$. (For instance, we could choose an arbitrary $x_0 \in X$ and send all y's not in $f(X)$ to the same x_0.) The map g satisfies $g(f(x)) = x$ for all $x \in X$, but not $f(g(y)) = y$ for all $y \in Y$ (unless f is bijective). We call g a *left inverse* of f.

If X is a finite set and $f : X \to \mathbf{R}$ 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 $\{1, 2, \dots, n\}$, 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 $0 \le i \le j \le n$, 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

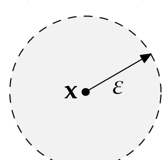
$$\sum_{0 \le i \le j \le n} (i + j).$$

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,

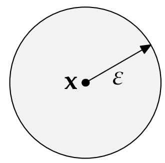
$$B^\circ(\varepsilon, \mathbf{x}) = \{\, \mathbf{y} \in \mathbf{R}^n \mid \|\mathbf{y} - \mathbf{x}\| < \varepsilon \,\}.$$



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 $\mathbf{R}^n \setminus C$ is open. This definition is equivalent to the following: C is closed if and only if for every sequence of points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$ 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 (ii) and (iii) of this problem require the use of an atlas or the Web. 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) Determine the sets $f(X)$, $f^{-1}(0)$, $f^{-1}(90)$, and $f^{-1}(-90)$.
(ii) Let A be the contiguous United States. Find $f(A)$. Round the numbers to whole degrees.
(iii) Let $B = f(A)$, where A is as in part (ii). 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 $S(n) = \sum_{0 \leq i \leq j \leq n} (i + j)$. 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 $\|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{x}\|$.)

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**], [**HH09**] and [**MT11**].

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

$$\boxed{g(x) = g(a) + \int_a^x g'(t)\,dt.} \tag{B.3}$$

Formulæ (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 (oriented) 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.) The most basic way to differentiate a map ϕ from $\mathbf{R}^n$ to $\mathbf{R}^m$ is to form a limit of the kind

$$\left.\frac{d\phi(\mathbf{x}+t\mathbf{v})}{dt}\right|_{t=0} = \lim_{t\to 0}\frac{\phi(\mathbf{x}+t\mathbf{v})-\phi(\mathbf{x})}{t}. \tag{B.4}$$

If this limit exists, it is called the *directional derivative of ϕ at $\mathbf{x}$ along $\mathbf{v}$*. The expression $\mathbf{x}+t\mathbf{v}$ as a function of t parametrizes a straight line in $\mathbf{R}^n$ passing through the point $\mathbf{x}$ in the direction of the vector $\mathbf{v}$. The directional derivative (B.4) measures the rate of change of ϕ along this straight line. Since $\phi(\mathbf{x}+t\mathbf{v})-\phi(\mathbf{x})$ is a vector in $\mathbf{R}^m$ for all t, the directional derivative is likewise a vector in $\mathbf{R}^m$.

If the map ϕ is not defined everywhere but only on a subset U of $\mathbf{R}^n$, it may be difficult to make sense of the limit (B.4). Let $\mathbf{x}$ be in the domain U of ϕ. The problem is that $\mathbf{x}+t\mathbf{v}$ may not be in U for all $t \neq 0$, so that we cannot evaluate $\phi(\mathbf{x}+t\mathbf{v})$. The solution is to assume that U is an *open* subset of $\mathbf{R}^n$. Given any direction vector $\mathbf{v} \in \mathbf{R}^n$, we can then find a number $\varepsilon > 0$ such that the points $\mathbf{x}+t\mathbf{v}$ are contained in U for $-\varepsilon < t < \varepsilon$. Therefore $\phi(\mathbf{x}+t\mathbf{v})$ is well-defined for $-\varepsilon < t < \varepsilon$ and we can legitimately ask whether the limit (B.4) exists.

The *partial derivatives* of ϕ at $\mathbf{x}$ are by definition the directional derivatives

$$\frac{\partial\phi}{\partial x_j}(\mathbf{x}) = \left.\frac{d\phi(\mathbf{x}+t\mathbf{e}_j)}{dt}\right|_{t=0}, \tag{B.5}$$

where

$$\mathbf{e}_1 = \begin{pmatrix}1\\0\\0\\\vdots\\0\\0\end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix}0\\1\\0\\\vdots\\0\\0\end{pmatrix}, \qquad \dots, \qquad \mathbf{e}_n = \begin{pmatrix}0\\0\\0\\\vdots\\0\\1\end{pmatrix}$$

are the standard basis vectors of $\mathbf{R}^n$, i.e. the columns of the identity $n \times n$-matrix. We can write the partial derivative in components as follows:

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

The *total derivative* or *Jacobi matrix* of ϕ at $\mathbf{x}$ is obtained by lining these columns up in an $m \times n$-matrix

$$D\phi(\mathbf{x}) = \begin{pmatrix}\frac{\partial\phi_1}{\partial x_1}(\mathbf{x}) & \frac{\partial\phi_1}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial\phi_1}{\partial x_n}(\mathbf{x})\\ \frac{\partial\phi_2}{\partial x_1}(\mathbf{x}) & \frac{\partial\phi_2}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial\phi_2}{\partial x_n}(\mathbf{x})\\ \vdots & \vdots & & \vdots\\ \frac{\partial\phi_m}{\partial x_1}(\mathbf{x}) & \frac{\partial\phi_m}{\partial x_2}(\mathbf{x}) & \dots & \frac{\partial\phi_m}{\partial x_n}(\mathbf{x})\end{pmatrix}.$$

We say that the map ϕ is *continuously differentiable* or C^1 if the partial derivatives

$$\frac{\partial\phi_i}{\partial x_j}(\mathbf{x})$$

are well-defined for all $\mathbf{x}$ in the domain U of ϕ and depend continuously on $\mathbf{x}$ for all $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$. If the second partial derivatives

$$\frac{\partial^2 \phi_i}{\partial x_j \partial x_k}(\mathbf{x})$$

exist for all $\mathbf{x} \in U$ and are continuous 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*, or C^∞, or *smooth*. This means that ϕ can be differentiated arbitrarily many times with respect to any of the variables. Smooth functions include such familiar one-variable functions as polynomials, exponentials, logarithms and trig functions. (Of course for log and tan one must make the proviso that they are smooth only on their domain of definition.)

The following useful fact says that any directional derivative can be expressed as a linear combination of partial derivatives. The proof is Exercise B.4.

B.1. Lemma. *Let U be an open subset of $\mathbf{R}^n$ and let $\phi \colon U \to \mathbf{R}^m$ be a C^1 map. Let $\mathbf{x} \in U$ and $\mathbf{v} \in \mathbf{R}^n$. Then the directional derivative of ϕ at $\mathbf{x}$ along $\mathbf{v}$ exists and is equal to*

$$\left.\frac{d\phi(\mathbf{x} + t\mathbf{v})}{dt}\right|_{t=0} = D\phi(\mathbf{x})\mathbf{v},$$

the vector in $\mathbf{R}^m$ obtained by multiplying the matrix $D\phi(\mathbf{x})$ by the vector $\mathbf{v}$.

Velocity vectors. Suppose $n = 1$. Then ϕ is a vector-valued function of one variable x, called a *path* or *parametrized curve* in $\mathbf{R}^m$. The matrix $D\phi(x)$ consists of a single column vector, called the *velocity vector*, and is usually denoted simply by $\phi'(x)$.

Gradients. Suppose $m = 1$. Then 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, 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,

$$D\phi(\mathbf{x})\mathbf{v} = \operatorname{grad}(\phi)(\mathbf{x})^T \mathbf{v} = \operatorname{grad}(\phi)(\mathbf{x}) \cdot \mathbf{v}.$$

There is an important characterization of the gradient, which is based on the familiar 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 $\mathbf{b}$. Let us fix a point $\mathbf{x}$ in the domain of ϕ and let us consider all possible directional derivatives of ϕ at $\mathbf{x}$ along *unit* vectors $\mathbf{v}$ (i.e. vectors of length 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 $\operatorname{grad}(\phi)(\mathbf{x})$.

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 maps, we can apply ψ after ϕ to obtain the composite map $(\psi \circ \phi)(x) = \psi(\phi(x))$.

B.2. 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 $\psi\colon V \to \mathbf{R}^l$ 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 $l \times m$-matrix $D\psi(\phi(\mathbf{x}))$ and the $m \times n$-matrix $D\phi(\mathbf{x})$.

B.3. Example. In the one-variable case $n = m = l = 1$ 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.4. Example. Let U be an open subset of $\mathbf{R}^n$ and let $\phi\colon U \to \mathbf{R}^m$ be C^1. Let I be an open interval and let $c\colon I \to U$ be a path in U. Then $\phi \circ c$ is a path in $\mathbf{R}^m$. Suppose that at time $t_0 \in I$ the path c passes through the point $c(t_0) = \mathbf{x}$ at velocity $c'(t_0) = \mathbf{v}$. How to compute the velocity vector of the composite path $\phi \circ c$ at time t_0? The chain rule gives

$$(\phi \circ c)'(t) = D(\phi \circ c)(t) = D\phi(c(t))Dc(t) = D\phi(c(t))c'(t)$$

for all $t \in I$. Setting $t = t_0$ gives $(\phi \circ c)'(t_0) = D\phi(\mathbf{x})\mathbf{v}$.

Let us write out a formula for the (i, j)-th entry of the Jacobi matrix of $\psi \circ \phi$. The i-th component of the map $\psi \circ \phi$ is $\psi_i \circ \phi$, so the (i, j)-th entry of $D(\psi \circ \phi)(\mathbf{x})$ is the partial derivative

$$\frac{\partial(\psi_i \circ \phi)}{\partial x_j}(\mathbf{x}).$$

According to Theorem B.2 this entry can be computed by multiplying the i-th row of $D\psi(\phi(\mathbf{x}))$, which is

$$\begin{pmatrix} \frac{\partial \psi_i}{\partial y_1}(\phi(\mathbf{x})) & \frac{\partial \psi_i}{\partial y_2}(\phi(\mathbf{x})) & \cdots & \frac{\partial \psi}{\partial y_m}(\phi(\mathbf{x})) \end{pmatrix},$$

by the j-th column of $D\phi(\mathbf{x})$, which is

$$\begin{pmatrix} \frac{\partial \phi_1}{\partial x_j}(\mathbf{x}) \\ \frac{\partial \phi_2}{\partial x_j}(\mathbf{x}) \\ \vdots \\ \frac{\partial \phi_m}{\partial x_j}(\mathbf{x}) \end{pmatrix}.$$

This gives the formula

$$\frac{\partial(\psi_i \circ \phi)}{\partial x_j}(\mathbf{x}) = \sum_{k=1}^{m} \frac{\partial \psi_i}{\partial y_k}(\phi(\mathbf{x}))\frac{\partial \phi_k}{\partial x_j}(\mathbf{x}).$$

This is perhaps the form in which the chain rule is most often used. Sometimes we are sloppy and abbreviate this identity to

$$\frac{\partial(\psi_i \circ \phi)}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial \psi_i}{\partial y_k} \frac{\partial \phi_k}{\partial x_j}.$$

Even sloppier, but nevertheless quite common, notations are

$$\frac{\partial \psi_i}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial \psi_i}{\partial y_k} \frac{\partial \phi_k}{\partial x_j},$$

or even

$$\frac{\partial \psi_i}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial \psi_i}{\partial y_k} \frac{\partial y_k}{\partial x_j}.$$

In these notes we frequently prefer the so-called "pullback" notation. Instead of $\psi \circ \phi$ we often write $\phi^*(\psi)$, so that $\phi^*(\psi)(\mathbf{x})$ stands for $\psi(\phi(\mathbf{x}))$. Similarly, $\phi^*(\partial\psi_i/\partial y_k)(\mathbf{x})$ stands for $\partial\psi_i/\partial y_k(\phi(\mathbf{x}))$. In this notation we have

$$\boxed{\frac{\partial \phi^*(\psi_i)}{\partial x_j} = \sum_{k=1}^{m} \phi^*\Big(\frac{\partial \psi_i}{\partial y_k}\Big) \frac{\partial \phi_k}{\partial x_j}.} \tag{B.6}$$

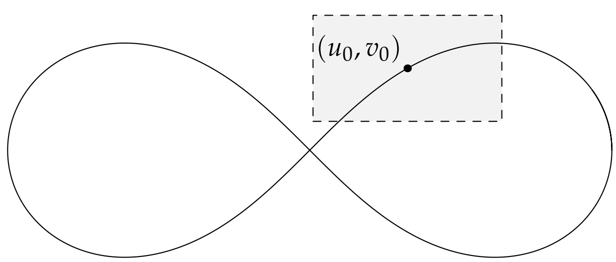
B.4. The implicit function theorem

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.

B.5. Example. To motivate the general result let us consider the case $m = n = 1$. Then ϕ is a function of two real variables (u, v), and the equation $\phi(u, v) = 0$ represents a curve in the plane, such as the lemniscate $(u^2 + v^2)^2 - 2(u^2 - v^2) = 0$.



Suppose we manage to find a special solution (u_0, v_0) of the equation. The gradient $\operatorname{grad}(\phi) = (\partial\phi/\partial u, \partial\phi/\partial v)$ is perpendicular to the curve at every point, so if $\partial\phi/\partial v \neq 0$ at (u_0, v_0), then the curve has a nonvertical tangent line at (u_0, v_0). Then for u close to u_0 and v close to v_0 (i.e. for (u, v) in a box centred at (u_0, v_0), such as the little grey box in the picture above) the curve looks like the graph of a function, so we can solve the equation $\phi(u, v) = 0$ for v as a function $v = f(u)$ of u. How to find the derivative of f? By differentiating the relation $\phi(u, f(u)) = 0$. We rewrite this as $\phi(\psi(u)) = 0$, where ψ is defined by $\psi(u) = (u, f(u))$ for u

in an open interval I containing u_0. According to the chain rule, Theorem B.2, $D(\phi \circ \psi)(u) = D\phi(\psi(u)D\psi(u)$. We have

$$D\phi\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{\partial\phi}{\partial u}(u,v) & \frac{\partial\phi}{\partial v}(u,v) \end{pmatrix}, \qquad D\psi(u) = \begin{pmatrix} 1 \\ f'(u) \end{pmatrix},$$

and therefore

$$D(\phi \circ \psi)(u) = \frac{\partial\phi}{\partial u}(u, f(u)) + \frac{\partial\phi}{\partial v}(u, f(u))f'(u).$$

The relation $\phi(u, f(u)) = 0$ (which holds for all $u \in I$) tells us that $D(\phi \circ \psi)(u) = 0$ for $u \in I$. Solving for f' gives the implicit differentiation formula

$$f'(u) = -\frac{\partial\phi}{\partial u}(u, f(u)) \Big/ \frac{\partial\phi}{\partial v}(u, f(u)).$$

For general m and n we form the Jacobi matrices of ϕ with respect to the **u**- and **v**-variables separately,

$$D_{\mathbf{u}}\phi = \begin{pmatrix} \frac{\partial\phi_1}{\partial u_1} & \dots & \frac{\partial\phi_1}{\partial u_n} \\ \vdots & & \vdots \\ \frac{\partial\phi_m}{\partial u_1} & \dots & \frac{\partial\phi_m}{\partial u_n} \end{pmatrix}, \qquad D_{\mathbf{v}}\phi = \begin{pmatrix} \frac{\partial\phi_1}{\partial v_1} & \dots & \frac{\partial\phi_1}{\partial v_m} \\ \vdots & & \vdots \\ \frac{\partial\phi_m}{\partial v_1} & \dots & \frac{\partial\phi_m}{\partial v_m} \end{pmatrix}.$$

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.6. 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 $\mathbf{u} \in U$.

As a special case we 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.7. 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 map $g \colon U \to g(U)$ 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.8. 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 $y_1, y_2, \dots, y_n$.) 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.9. 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

$$\int_c^d f(y)\,dy = \begin{cases} \int_a^b f(p(x))p'(x)\,dx & \text{if } p \text{ is increasing,} \\ \int_b^a f(p(x))p'(x)\,dx & \text{if } p \text{ is decreasing.} \end{cases}$$

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 $\phi\colon \mathbf{R} \to \mathbf{R}$ be the function defined by $\phi(x) = x|x|$. Show that ϕ is C^1 but not C^2. For each $r \geq 1$ give an example of a function which is C^r but not C^{r+1}.

B.3. 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 $f'(0) = 0$. (Use the definition of the derivative,
$$f'(0) = \lim_{x \to 0} \frac{f(x) - f(0)}{x}.)$$
(ii) Show that f is smooth and that $f^{(n)}(0) = 0$ for all n. (By induction on n, prove the following assertion:
$$f^{(n)}(x) = \begin{cases} 0 & \text{if } x = 0, \\ g_n(1/x)e^{-1/x^2} & \text{if } x \neq 0, \end{cases}$$
where, for each $n \geq 0$, $g_n(1/x)$ is a certain polynomial in $1/x$, which you need not determine explicitly.)

(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.4. (i) Let $\mathbf{x}$ and $\mathbf{v}$ be vectors in $\mathbf{R}^n$. Define a path $c\colon \mathbf{R} \to \mathbf{R}^n$ by $c(t) = \mathbf{x} + t\mathbf{v}$. Find $c'(t)$.

(ii) Prove Lemma B.1. (Let $\phi\colon U \to \mathbf{R}^m$ be a C^1 map, where U is open in $\mathbf{R}^n$. Combine the result of part (i) with the formula of Example B.4.)

B.5. 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 $\mathbf{x} \in \mathbf{R}^3 \setminus \{\mathbf{0}\}$ 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.6. A function $f\colon \mathbf{R}^n \setminus \{\mathbf{0}\} \to \mathbf{R}$ is *homogeneous* of degree p if $f(t\mathbf{x}) = t^p f(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{R}^n \setminus \{\mathbf{0}\}$ and $t > 0$. Here p is a real constant.

(i) Show that the functions $f_1(x,y) = (x^2 - xy)/(x^2 + y^2)$, $f_2(x,y) = \sqrt{x^3 + y^3}$, $f_3(x,y,z) = (x^2z^6 + 3x^4y^2z^2)^{-\sqrt{2}}$ are homogeneous. What are their degrees?

(ii) Let f be a homogeneous function of degree p. Assume that f is defined at $\mathbf{0}$ and continuous everywhere. Also assume $f(\mathbf{x}) \neq 0$ for at least one $\mathbf{x} \in \mathbf{R}^n$. 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
$$\sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i}(\mathbf{x}) = pf(\mathbf{x}).$$
(Differentiate the relation $f(t\mathbf{x}) = t^p f(\mathbf{x})$ with respect to t.)

B.7. A function $f\colon \mathbf{R}^n \setminus \{\mathbf{0}\} \to \mathbf{R}$ is *quasihomogeneous* of degree p if there exist real constants $a_1, a_2, \ldots, a_n$ with the property that
$$f(t^{a_1}x_1, t^{a_2}x_2, \ldots, t^{a_n}x_n) = t^p f(\mathbf{x})$$

for all $\mathbf{x} \in \mathbf{R}^n \setminus \{\mathbf{0}\}$ and $t > 0$. Suppose that f is quasihomogeneous of degree p and smooth. Show that $\sum_{i=1}^n a_i x_i \frac{\partial f}{\partial x_i}(\mathbf{x}) = pf(\mathbf{x})$.

B.8. Define a map ψ from $\mathbf{R}^{n-1}$ to $\mathbf{R}^n$ by

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

(Here we regard a point $\mathbf{t} = (t_1, t_2, \ldots, t_{n-1})$ in $\mathbf{R}^{n-1}$ as a point in $\mathbf{R}^n$ by identifying it with $(t_1, t_2, \ldots, t_{n-1}, 0)$.)

(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 $\mathbf{t}$.
(iii) Compute $D\psi(\mathbf{t})$.
(iv) Let X be the sphere punctured at the "north pole", $X = S^{n-1} \setminus \{\mathbf{e}_n\}$. *Stereographic projection from the north pole* is the map $\phi\colon X \to \mathbf{R}^{n-1}$ given by $\phi(\mathbf{x}) = (1 - x_n)^{-1}(x_1, x_2, \ldots, x_{n-1})^T$. 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 perpendicular to $\mathbf{y}$. (A *hyperplane* in $\mathbf{R}^n$ is a linear subspace of dimension $n - 1$.) 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 $\phi\colon S^{n-1} \setminus \{\mathbf{y}\} \to P$. 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.9. Let $\phi\colon \mathbf{R}^n \to \mathbf{R}^m$ be a C^1 map. Prove the following assertions.

(i) ϕ is constant if and only if $D\phi(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbf{R}^n$.
(ii) ϕ is linear if and only if $D\phi(\mathbf{x})\mathbf{v} = \phi(\mathbf{v})$ for all $\mathbf{x}$ and $\mathbf{v}$ in $\mathbf{R}^n$.

B.10. 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.11. Let $f\colon \mathbf{R}^2 \to \mathbf{R}$ be a smooth function which satisfies $f(x, y) = -f(y, x)$ for all x and y. Show that

$$\frac{\partial f}{\partial x}(a, b) = -\frac{\partial f}{\partial y}(b, a)$$

for all a and b.

B.12. Let $f\colon \mathbf{R}^2 \to \mathbf{R}$ and $g\colon \mathbf{R} \to \mathbf{R}$ be smooth functions. Show that

$$\frac{d}{dx}\int_0^{g(x)} f(x, y)\,dy = f(x, g(x))g'(x) + \int_0^{g(x)} \frac{\partial f(x, y)}{\partial x}\,dy.$$

B.13. Thermodynamicists like to use rules such as

$$\frac{\partial y}{\partial x}\frac{\partial x}{\partial y} = 1.$$

Explain the rule and show that it is correct. (Assume that the variables are subject to a relation $F(x, y) = 0$ defining functions $x = f(y)$, $y = g(x)$, and apply the multivariable chain rule. See also Example B.5.) Similarly, explain why

$$\frac{\partial y}{\partial x}\frac{\partial z}{\partial y}\frac{\partial x}{\partial z} = -1.$$

Naively cancelling numerators against denominators gives the wrong answer!

B.14. 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 all coefficients are nonnegative and their sum is 1: $c_i \geq 0$ and $\sum_{i=0}^n c_i = 1$. The *simplex* Δ spanned by the $\mathbf{a}_i$'s is the collection of all their convex linear combinations,

$$\Delta = \left\{ \sum_{i=0}^n c_i\mathbf{a}_i \,\middle|\, c_1 \geq 0, \dots, c_n \geq 0, \quad \sum_{i=0}^n c_i = 1 \right\}.$$

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 $\operatorname{vol}(R) = \displaystyle\int_R dx_1\,dx_2\cdots dx_n$. Show that the volume of the standard n-simplex is $1/n!$.

(iii) Show that

$$\operatorname{vol}(\Delta) = \frac{1}{n!}|\det(A)|,$$

where A is the $n \times n$-matrix with columns $\mathbf{a}_1 - \mathbf{a}_0, \mathbf{a}_2 - \mathbf{a}_0, \dots, \mathbf{a}_n - \mathbf{a}_0$. (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.15. 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) $\displaystyle\int_0^\infty e^{-u^2}u^a\,du = \frac{1}{2}\Gamma\left(\frac{a+1}{2}\right)$.

B.16. Calculate $\Gamma(n+\frac{1}{2})$ (where Γ is the function defined in Exercise B.15) 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) $\displaystyle\gamma^2 = \int_0^{2\pi}\int_0^\infty re^{-r^2}\,dr\,d\theta$.

(iv) $\gamma = \sqrt{\pi}$.

(v) $\displaystyle\Gamma\left(n+\frac{1}{2}\right) = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^n}\sqrt{\pi}$ for $n \geq 1$.

APPENDIX C

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, ς	sigma
T	τ	tau
Υ	υ	upsilon
Φ	ϕ, φ	phi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Bibliography

[Bac06] D. Bachman, *A geometric approach to differential forms*, second ed., Birkhäuser, Boston, MA, 2012.

[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.
Perhaps the best exposition at the graduate level of the applications of differential forms to topology.

[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.

[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, NJ, 1974.
Written for beginning graduate students, but very intuitive and with lots of interesting applications to topology.

[HH09] J. Hubbard and B. Hubbard, *Vector calculus, linear algebra, and differential forms: A unified approach*, fourth ed., Matrix Editions, Ithaca, NY, 2009.

[Lay12] D. Lay, *Linear algebra and its applications*, fourth ed., Pearson, 2012.
Good reference for the basic linear algebra required in these notes.

[MT11] J. Marsden and A. Tromba, *Vector calculus*, sixth ed., W. H. Freeman, New York, 2011.
Standard multivariable calculus reference.

[Mil03] J. Milnor, *Towards the Poincaré conjecture and the classification of 3-manifolds*, Notices of the American Mathematical Society **50** (2003), no. 10, 1226–1233.
A discussion of one of the great problems of differential topology. Available online at http://www.ams.org/notices/200310/fea-milnor.pdf.

[ONe06] B. O'Neill, *Elementary differential geometry*, revised second ed., Elsevier/Academic Press, Amsterdam, 2006.

[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.

[Sam01] H. Samelson, *Differential forms, the early days; or the stories of Deahna's theorem and of Volterra's theorem*, American Mathematical Monthly **108** (2001), no. 6, 522–530.
A brief history of the subject, available online at http://www.jstor.org/stable/2695706.

[Sin01] S. Singer, *Symmetry in mechanics. A gentle, modern introduction*, Birkhäuser, Boston, MA, 2001.

[Spi71] M. Spivak, *Calculus on manifolds. A modern approach to classical theorems of advanced calculus*, Westview Press, 1971.
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.

[Tu11] L. Tu, *An introduction to manifolds*, second ed., Springer-Verlag, New York, 2011.

Notation Index

Index

Page numbers in **boldface** refer to definitions; page numbers in *italic* refer to theorems, examples, or applications. ***Italic boldface*** is used when both types of items occur on the same page.