Differential Forms Lecture Notes

Liam Mazurowski

Contents

1 Introduction

Differential forms are a certain class of objects that can be integrated. Hence to understand differential forms it's helpful to start with the simplest possible notion of integration: the single variable Riemann integral. Fundamentally, the single variable integral can solve two different kinds of problems.

Example 1. Let I be an interval of the real line and let $f\colon I \to \mathbb{R}$ be a continuous function. One can then try to determine the area under the graph of f over the interval I. Notice that this problem depends only on the interval I as a set of points. Of course, the solution is just the integral of the function f over the interval I. We can write this as

$$\int_I f(x)\,dx$$

to emphasize that this problem depends only on which points are in I.

Example 2. Consider a particle that travels along the real line with velocity $v(t)$. Suppose the particle is at the origin at time $t = 0$. One can then try to determine the position of the particle at all other times t. Again the solution is given by an integral: the position of the particle at time t is

$$p(t) = \int_0^t v(s)\,ds.$$

Note however that this integral depends both on the interval of points between 0 and t as a set as well as an orientation of this interval.

To compute the value of this integral, we need to know whether t is the left endpoint of the interval or the right endpoint of the interval!

Of course this is a pretty pedantic distinction in one variable. If J denotes the interval between 0 and t then

$$\int_0^t f(s)\,ds = \begin{cases} \int_J f(s)\,ds, & \text{if } t \geq 0 \\ -\int_J f(s)\,ds, & \text{if } t \leq 0. \end{cases}$$

However, in higher dimensions these concepts start to diverge rather dramatically.

Example 3. Consider a continuous function $f\colon \mathbb{R}^3 \to \mathbb{R}$. What is the average value of f on the unit sphere S^2? Note that this problem depends on S^2 only as a set of points. The branch of math called measure theory was developed to solve problems like this.

Example 4. Suppose gas particles are circulating in $\mathbb{R}^3$. In a given time interval, how much gas flows across the unit sphere S^2? Notice that the problem is not well-posed as stated. Do we want to measure the amount of gas flowing into the sphere or the amount of gas flowing out of the sphere? Hence the problem depends on an orientation. Developing a theory of oriented integration leads to differential forms.

2 Differential 1-forms

2.1 Intuition

We begin with an intuitive discussion.

Definition 5. (Informal) A differential 1-form on $\mathbb{R}^n$ is a rule that assigns a number to each oriented line segment in $\mathbb{R}^n$ in a suitable way.

Example 6. Every oriented line segment in $\mathbb{R}$ can be represented by a pair of numbers p and q. The notation $\vec{pq}$ represents the oriented line segment that originates at p and ends at q. Given a function $v\colon \mathbb{R} \to \mathbb{R}$ we can then define a 1-form $v(x)\,dx$ by the following formula:

$$\Big[v(x)\,dx\Big](\vec{pq}) = v(p)(q-p).$$

If v represents the velocity of a particle moving in $\mathbb{R}$ and $\vec{pq}$ is short, then the value of $v(x)\,dx$ on $\vec{pq}$ is very nearly the position of the particle at time q minus the position of the particle at time p.

Example 7. Consider a vector field $F\colon \mathbb{R}^n \to \mathbb{R}^n$. Think of F as a force field: a particle at position p experiences a force of magnitude and direction $F(p)$. When a particle moves in the presence of such a force field, the force performs a certain amount of work on the particle. The amount of work done by F is positive if the force F aids the motion, and the amount of work done by F is negative if the force F resists the motion.

Again an oriented line segment in $\mathbb{R}^n$ can be represented by a pair of points p and q. The notation $\vec{pq}$ again stands for the line segment beginning at p and terminating at q. We can define a 1-form ω_F by

$$\omega_F(\vec{pq}) = F(p) \cdot (\vec{pq}),$$

where on the right hand side we are thinking of $\vec{pq}$ as a vector pointing from p to q. When $\vec{pq}$ is short, the value of ω_F on $\vec{pq}$ is very nearly the work done by F on a particle that moves in a straight line from p to q

Example 8. Consider a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$. We can define a 1-form df on $\mathbb{R}^n$ by

$$df(\vec{pq}) = \nabla f(p) \cdot (\vec{pq}).$$

Notice that the right hand side is exactly the directional derivative of f at the point p in the direction $\vec{pq}$, i.e.,

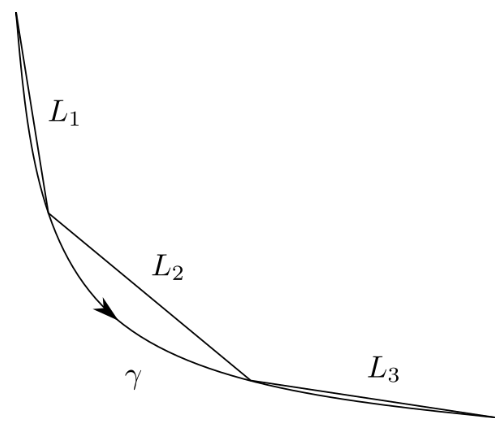
$$df(\vec{pq}) = \left.\frac{d}{dt}\right|_{t=0} f(p + t(q-p)).$$

Hence when $\vec{pq}$ is short, the value of df on $\vec{pq}$ is very nearly the difference $f(q) - f(p)$.

Note that by definition 1-forms only assign values to line segments. However, in principle, they should also be able to assign values to smooth curves. For example, given any curve γ connecting p to q the 1-form ω_F evaluated on γ should tell us the work done by F on a particle that moves from p to q along the path γ. Likewise, the 1-form df evaluated on γ should tell us the change in the value of a function f as we go from p to q along γ. Integration is the process by which 1-forms can assign values to curves.

Intuitively, integration of 1-forms works as follows. Consider a 1-form α on $\mathbb{R}^n$ and let γ be a smooth, oriented curve in $\mathbb{R}^n$. We begin by approximating

γ by a sequence of oriented line segments $L_1, \ldots, L_k$.



Then we evaluate α on each of these line segments and add up the resulting numbers. This gives a quantity

$$\sum_{i=1}^{k} \alpha(L_i)$$

associated to the approximation. If this quantity converges as the approximation gets better and better, the resulting limit should be the integral of α over γ. Let's see how this works in each of the previous examples.

Example 9. Consider the 1-form $v(x)\,dx$ from Example 6. Take two numbers $a < b$ and let γ be the line segment from a to b. Pick a partition

$$a = x_0 < x_1 < \ldots < x_k = b.$$

Then the line segments $L_i = [x_{i-1}, x_i]$ for $i = 1, \ldots, k$ form an approximation to $[a, b]$. If we evaluate $v(x)\,dx$ on each of these segments and add everything up we get the quantity

$$\sum_{i=1}^{k} \Big[v(x)\,dx \Big](L_i) = \sum_{i=1}^{k} v(x_{i-1})(x_i - x_{i-1}).$$

But this is exactly a Riemann sum for the usual integral $\int_a^b v(x)\,dx$! Hence taking the limit of the above quantity as the partition gets finer and finer yields the usual integral of the function v over the interval $[a, b]$.

Example 10. Consider the work 1-form ω_F associated to a vector field F as in Example 7. Consider a smooth curve γ in $\mathbb{R}^n$ starting at p and ending at

q. Approximate γ by oriented line segments $L_1, \ldots, L_k$. Assuming these line segments are sufficiently short, each number $\omega_F(L_i)$ is very nearly the amount of work F does on the particle as it moves along the line segment L_i. Thus

$$\sum_{i=1}^{k} \omega_F(L_i)$$

is very nearly the amount of work F does on the particle as it moves from p to q along the piecewise linear path obtained by concatenating the L_i. As we approximate γ by more and more line segments, this quantity should converge to the amount of work F does on the particle as it moves from p to q along γ. Hence the integral of ω_F over γ should represent the amount of work F does on a particle moving from p to q along γ.

Example 11. Consider the 1-form df associated to a function f as in Example 8. Consider a smooth curve γ in $\mathbb{R}^n$ starting at p and ending at q. Choose points $p = p_0, p_1, \ldots, p_k = q$ along γ and let L_i be the line segment from p_{i-1} to p_i. Assuming these line segments L_i are sufficiently short, each number $df(L_i)$ is very nearly equal to $f(p_i) - f(p_{i-1})$. Hence we have

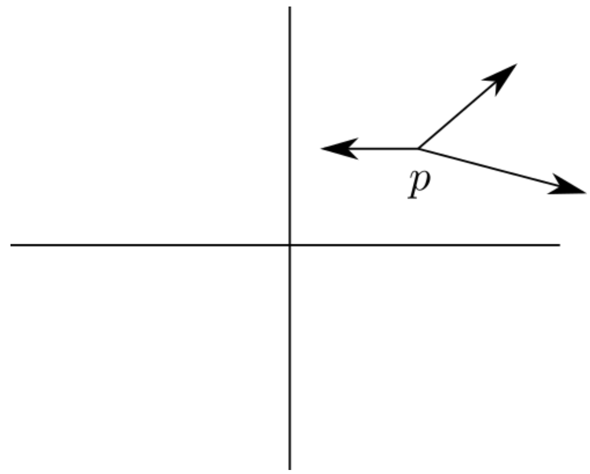
$$\sum_{i=1}^{k} df(L_i) \approx \sum_{i=1}^{k} f(p_i) - f(p_{i-1}) = f(q) - f(p).$$

Taking a limit as the approximation gets better and better, we see that the integral of df over γ should be equal to $f(q) - f(p)$.

2.2 Formal Definitions

Let's now try to formalize the preceding constructions. First we give another description of the space of oriented line segments on $\mathbb{R}^n$.

Definition 12. Fix a point $p \in \mathbb{R}^n$. The tangent space to $\mathbb{R}^n$ at p is the set $T_p\mathbb{R}^n = \{(p, \vec{v}) : \vec{v} \in \mathbb{R}^n\}$.



We can think of $T_p\mathbb{R}^n$ as a set of vectors based at p. It represents all the different speeds and directions we can travel starting from p. This tangent space $T_p\mathbb{R}^n$ is a vector space: we can add and scalar multiply vectors based at p to get new vectors based at p. Any oriented line segment $\vec{pq}$ can be equivalently thought of as the vector $q - p$ based at p, i.e., the element $(p, q-p) \in T_p\mathbb{R}^n$. When the context is clear, we will often abuse notation and simply write $v \in T_p\mathbb{R}^n$ for the vector v thought of as being based at p.

Notation. Often we write α_p to indicate the restriction of α to $T_p\mathbb{R}^n$. We write $\alpha_p(v)$ to mean the value of α on the tangent vector $(p, v) \in T_p\mathbb{R}^n$, i.e., the value of α on the oriented line segment from p to $p+v$.

Example 13. Define a 1-form α on $\mathbb{R}^2$ by

$$\alpha(\vec{pq}) = x_1(y_2 - y_1), \quad \text{for } p = (x_1, y_1),\ q = (x_2, y_2).$$

In the tangent vector notation, $\vec{pq}$ is represented by $(p, v) \in T_p\mathbb{R}^n$ where $v = (x_2 - x_1, y_2 - y_1)$. Hence we could equivalently specify α by the formula $\alpha_{(x,y)}(v) = xv_2$.

To get a formal definition for differential 1-forms, we need to explain what is meant by "suitable" in the informal Definition 5. To see that some condition is required on the way 1-forms assign numbers to line segments, recall that we would like to be able to integrate 1-forms over curves. The following example shows that this may not be possible for very silly rules assigning numbers to line segments.

Example 14. Consider the rule $\alpha_p(v) = 1$ for all p and v in $\mathbb{R}^n$, i.e., α assigns the number 1 to every single oriented line segment in $\mathbb{R}^n$. Now suppose we try to integrate α over a curve γ by the procedure described above. We approximate γ by line segments $L_1, \ldots, L_k$. Then we add up the value of α on each of these segments to get

$$\sum_{i=1}^{k} \alpha(L_i) = k.$$

As our approximation to γ gets better and better k goes to infinity, and hence the above quantity diverges to infinity. Thus it doesn't make sense to integrate this rule α.

The problem with the preceding example is that α assigns large numbers to very short line segments. In order for a 1-form to be integrable, we need to impose some requirement that forces α to be small on small line segments. We can do this with the following scaling condition:

$$\alpha_p(t\vec{v}) = t\alpha_p(\vec{v}), \quad \text{for all } t \in \mathbb{R} \text{ and } p, \vec{v} \in \mathbb{R}^n.$$

Any α satisfying the above condition will be small on short vectors. There is also a second requirement we are going to impose on 1-forms:

$$\alpha_p(v + w) = \alpha_p(v) + \alpha_p(w), \qquad \text{for all } v, w \in \mathbb{R}^n.$$

This additivity condition is somewhat harder to motivate, and we will not give an extended discussion of it here. (The interested reader can consult Whitney's book Geometric Integration Theory for a heroic attempt to explain why we impose this condition. Essentially if we assume that α undergoes some amount of cancellation when we integrate it over small loops then this forces additivity.)

The two conditions above say that α_p is a linear functional $T_p\mathbb{R}^n \to \mathbb{R}$. Next we briefly recall some facts about linear functionals.

Definition 15. A linear functional on $\mathbb{R}^n$ is a map $A\colon \mathbb{R}^n \to \mathbb{R}$ such that $A(tv) = tA(v)$ and $A(v + w) = Av + Aw$ for all $v, w \in \mathbb{R}^n$ and all $t \in \mathbb{R}$.

Any such A can be represented by a row matrix $A = \begin{pmatrix} A_1 & A_2 & \cdots & A_n \end{pmatrix}$. The functional A with the above matrix acts on vectors v by the formula

$$Av = \begin{pmatrix} A_1 & A_2 & \cdots & A_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = A_1v_1 + A_2v_2 + \ldots + A_nv_n.$$

Note that if A and B are both linear functionals then so are the functionals tA and $A+B$ defined by

$$(tA)(v) = t(Av), \quad \text{and} \quad (A+B)(v) = (Av) + (Bv).$$

In this way, the space of linear functionals itself forms a vector space.

There is a special way of writing linear functionals that is typically used in differential forms. Define functionals dx_1, dx_2, ..., dx_n on $\mathbb{R}^n$ by

$$dx_i(v) = v_i.$$

That is, dx_i takes a vector v and outputs its ith component. Every linear functional can be written as a linear combination of the dx_i. Indeed if $A = \begin{pmatrix} A_1 & A_2 & \cdots & A_n \end{pmatrix}$ then we can equivalently write $A = A_1\,dx_1 + \ldots + A_n\,dx_n$ because

$$\begin{aligned} (A_1\,dx_1 + \ldots + A_n\,dx_n)(v) &= A_1\,dx_1(v) + \ldots + A_n\,dx_n(v) \\ &= A_1v_1 + \ldots + A_nv_n = Av. \end{aligned}$$

Another way to say this is that the functionals $dx_1, \ldots, dx_n$ form a basis for the space of all linear functionals on $\mathbb{R}^n$.

Example 16. Consider $\mathbb{R}^3$ with coordinates x, y, z. Then

$$dx\begin{pmatrix}1\\2\\3\end{pmatrix} = 1, \quad dy\begin{pmatrix}1\\2\\3\end{pmatrix} = 2, \quad dz\begin{pmatrix}1\\2\\3\end{pmatrix} = 3.$$

If we think of v as a vector in $T_p\mathbb{R}^3$ telling us how to get from p to q then $dx(v)$ tells us the change in the x coordinate moving from p to q. In a similar fashion, $dy(v)$ tells us the change in the y coordinate when we move from p to q, and $dz(v)$ tells us the change in the z coordinate.

In light of the above discussion, given a 1-form α, its restriction $\alpha_p : T_p\mathbb{R}^n \to \mathbb{R}$ is a linear functional. Any linear functional can be written in terms of the dx_i and therefore we know

$$\alpha_p = A_1dx_1 + \ldots + A_ndx_n$$

for some constants A_i. Actually, it's better to write

$$\alpha_p = A_1(p)dx_1 + \ldots + A_n(p)dx_n$$

to emphasize that these constants depend on the point p. The final requirement we will impose on our 1-forms is that these coefficients $A_i(p)$ depend smoothly on p.

We now summarize the above discussion with the formal definition of a differential 1-form.

Definition 17. A differential 1-form α on $\mathbb{R}^n$ is a rule assigning a number to each tangent vector in $\mathbb{R}^n$ which can be expressed in the form

$$\alpha = f_1\, dx_1 + \ldots + f_n\, dx_n$$

for some smooth functions $f_i \colon \mathbb{R}^n \to \mathbb{R}$.

Example 18. Consider the 1-form $\alpha = y\, dx + dz$ on $\mathbb{R}^3$. We have

$$\alpha_{(1,2,3)}((4,5,6)^T) = 2dx((4,5,6)^T) + dz((4,5,6)^T) = (2)(4) + 6 = 14.$$

Example 19. Consider a force field

$$F(x,y,z) = (F_1(x,y,z), F_2(x,y,z), F_3(x,y,z))$$

on $\mathbb{R}^3$. Recall that the work form associated to F is the 1-form ω_F given by

$$(\omega_F)_p(v) = F(p) \cdot v = F_1(p)v_1 + F_2(p)v_2 + F_3(p)v_3.$$

Hence we can write $\omega_F = F_1\, dx + F_2\, dy + F_3\, dz$.

Definition 20. Consider a smooth function $f \colon \mathbb{R}^n \to \mathbb{R}$. The differential of f is the 1-form

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

The value of this 1-form on $v \in T_p\mathbb{R}^n$ is $(df)_p(v) = \nabla f(p) \cdot v$.

Example 21. Consider the function $f(x,y,z) = xyz$. The gradient of f is $\nabla f = (yz, xz, xy)$ and

$$df = yz\, dx + xz\, dy + xy\, dz.$$

Hence, for example, $(df)_{(4,5,6)}((1,2,3)^T) = (5)(6)(1) + (4)(6)(2) + (4)(5)(3) = 138$.

Example 22. Define a 2-form α_{rot} by

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

Notice that this is not defined on all of $\mathbb{R}^2$ since the coefficients blow up at 0. Nevertheless, α_{rot} is still a 1-form on $\mathbb{R}^2 \setminus \{0\}$. To understand what α_{rot} computes, notice that α_{rot} is the work form associated to the vector field

$$F(x,y) = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2}\right).$$

What does this vector field look like? First observe that

$$\left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$$

is a vector that points in the (x,y) direction with length $(x^2+y^2)^{-1/2}$. Now

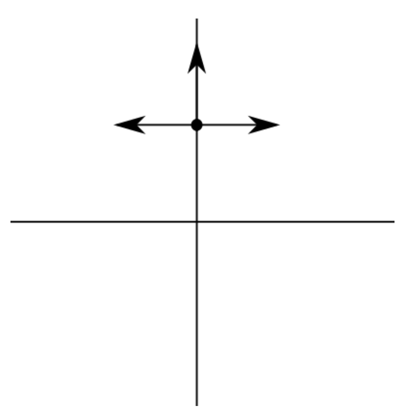
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{x}{x^2+y^2} \\ \frac{y}{x^2+y^2} \end{pmatrix} = \begin{pmatrix} \frac{-y}{x^2+y^2} \\ \frac{x}{x^2+y^2} \end{pmatrix}$$

and this matrix represents a 90° counterclockwise rotation. Thus $F(x,y)$ is a vector tangent to the circle of radius $(x^2+y^2)^{1/2}$ with length equal to $(x^2+y^2)^{-1/2}$. Thus

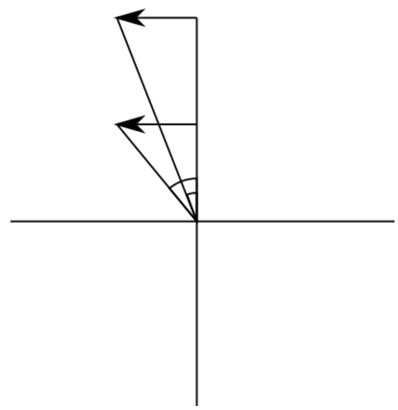
$$(\alpha_{\text{rot}})_{(x,y)}(v) = F(x,y)\cdot v$$

measures the extent to which a displacement v induces counterclockwise rotation about the origin.

For instance, we have $(\alpha_{\text{rot}})_{(0,1)}(0,1) = 0$, and $(\alpha_{\text{rot}})_{(0,1)}(1,0) = -1$, and $(\alpha_{\text{rot}})_{(0,1)}(-1,0) = 1$.



The first formula reflects the fact that starting at $(0,1)$ and moving up does not induce rotation about the origin. The second formula reflects the fact that starting at $(0,1)$ and moving right induces clockwise rotation, and the third reflects the fact that starting at $(0,1)$ and moving left induces counterclockwise rotation. Notice also that the length of F increases as (x,y) gets closer to the origin.



This reflects the fact that undergoing a displacement v from (x,y) will result in a greater change in angle when (x,y) is close to the origin.

3 Curves

3.1 Definition of a Curve

Differential 1-forms can be integrated over curves. In order to make this precise, we first need to decide what we mean by curves.

Definition 23. (Informal) A 1-manifold is a subset of $\mathbb{R}^n$ which looks locally like a line near each of its points. This line is called a tangent line.

Remark 24. Manifolds can be specified in many different ways. For example, we can describe the unit circle as the set of points (x,y) satisfying the equation $x^2+y^2=1$. However, we can also describe the unit circle as the image of the parameterization $(\cos t, \sin t)$, $t \in [0,2\pi]$.

For the purpose of working with differential forms, its most convenient to describe curves via parameterizations.

Definition 25. A parameterized curve is a smooth map $\gamma\colon [a,b] \to \mathbb{R}^n$.

Definition 26. We say that γ is regular provided $\gamma'(t) = (\gamma_1'(t), \ldots, \gamma_n'(t)) \neq 0$ for all t.

Example 27. The map $\gamma\colon [0,1] \to \mathbb{R}^2$ given by $\gamma(t) = (t, 2t)$ parameterizes the line segment from $(0,0)$ to $(1,2)$. It has $\gamma'(t) = (1,2)$ for all t and so γ is regular.

Example 28. The map $\gamma\colon [0,2\pi] \to \mathbb{R}^2$ given by $\gamma(t) = (\cos t, \sin t)$ parameterizes the unit circle. Note that $\gamma'(t) = (-\sin t, \cos t)$ is always non-zero. In fact $\gamma'(t)$, thought of as an element of $T_{\gamma(t)}\mathbb{R}^2$, is tangent to the curve γ for every t.

Example 29. Consider $\gamma\colon [-2,2] \to \mathbb{R}^2$ given by $\gamma(t) = (t^3, t^2)$. Note that γ is a parameterized curve, but γ is not regular since $\gamma'(0) = (0,0)$. In fact, every point of γ lies on the graph of $y = x^{2/3}$ and so γ has a cusp at the origin.

3.2 Integrating 1-Forms over Curves

We can now formalize the procedure for integrating a 1-form over a curve. Recall that intuitively, to integrate a 1-form α over a curve γ, we approximate γ by line segments $L_1, \ldots, L_k$, compute the quantity

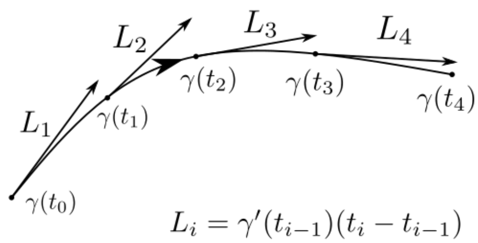
$$\sum_{i=1}^{k} \alpha(L_i),$$

and then try to take a limit of this quantity as the approximation gets better and better. Unfortunately, it would be very cumbersome if we had to construct an ad hoc scheme for approximating a curve every time we wanted to compute an integral.

Fortunately, there is a nice way to use a parameterization to build an approximation to a given curve. Suppose we're given a curve $\gamma\colon [a,b] \to \mathbb{R}^n$ and a 1-form α on $\mathbb{R}^n$. Take a partition $a = t_0 < t_1 < \ldots < t_k = b$ of the interval $[a,b]$. Such a partition naturally gives rise to a partition of γ. Now Taylor's theorem says that

$$\gamma(t_i) - \gamma(t_{i-1}) = \gamma'(t_{i-1})(t_i - t_{i-1}) + o(|t_i - t_{i-1}|).$$

Therefore the line segment L_i from $\gamma(t_{i-1})$ to $\gamma(t_{i-1}) + \gamma'(t_{i-1})(t_i - t_{i-1})$ is a good approximation to the portion of γ between $\gamma(t_{i-1})$ and $\gamma(t_i)$ when $t_i - t_{i-1}$ is small.



We now use these line segments L_i as our approximation to γ. Assume that $\alpha = f_1\,dx_1 + \ldots + f_n\,dx_n$. Then evaluating α on each of these line segments and adding the results gives

$$\begin{aligned}\sum_{i=1}^{k} \alpha(L_i) &= \sum_{i=1}^{k} \alpha_{\gamma(t_{i-1})}\Big(\gamma'(t_{i-1})(t_i - t_{i-1})\Big)\\ &= \sum_{i=1}^{k}\sum_{j=1}^{n} f_j(\gamma(t_{i-1}))\gamma_j'(t_{i-1})(t_i - t_{i-1}).\end{aligned}$$

But this is exactly a Riemann sum for integral

$$\int_a^b \sum_{j=1}^{n} f_j(\gamma(t))\gamma_j'(t)\,dt!$$

So the quantity $\sum_{i=1}^{k} \alpha(L_i)$ converges to the above integral as our approximation gets better and better. Based on this we make the following definition.

Definition 30. The integral of a 1-form $\alpha = f_1\,dx_1 + \ldots + f_n\,dx_n$ over a parameterized curve $\gamma\colon [a,b] \to \mathbb{R}^n$ is

$$\int_\gamma \alpha = \int_a^b f_1(\gamma(t))\gamma_1'(t) + \ldots + f_n(\gamma(t))\gamma_n'(t)\,dt.$$

Remark 31. Notice that $f_1(\gamma(t))\gamma_1'(t) + \ldots + f_n(\gamma(t))\gamma_n'(t) = \alpha_{\gamma(t)}(\gamma'(t))$. Hence we could equivalently write

$$\int_\gamma \alpha = \int_a^b \alpha_{\gamma(t)}(\gamma'(t))\,dt.$$

Example 32. Let $\alpha = x^2\,dx + dy$ and let $\gamma(t) = (t, t^2)$ for $t \in [0,1]$. Then $\gamma_1'(t) = 1$ and $\gamma_2'(t) = 2t$ and so
$$\int_\gamma \alpha = \int_0^2 t^2 \cdot 1 + 2t\,dt = \int_0^1 t^2 + 2t\,dt = \frac{4}{3}.$$
Remark 33. Really we would like the integral $\int_\gamma \alpha$ to depend only on the image of γ and its orientation and not the particular parameterization we picked. In other words, we'd like the integral to depend on the curve as an oriented geometric object, and not the particular way it's traversed.

Of course there may be multiple different parameterizations of the same curve.

Example 34. Both $\gamma(t) = (\cos t, \sin t)$, $t \in [0,\pi]$ and $\phi(t) = (-t, \sqrt{1-t^2})$, $t \in [-1,1]$ parameterize the upper half of the unit circle.

The following proposition shows that the integral of a 1-form does not change if we reparameterize a given curve.

Proposition 35. *Assume we have parameterized curves $\gamma\colon [a,b] \to \mathbb{R}^n$ and $\phi\colon [c,d] \to \mathbb{R}^n$. Moreover, assume that γ and ϕ are reparameterizations of each other in the sense that there exists a map $\psi\colon [a,b] \to [c,d]$ such that*

(i) ψ is invertible and both ψ and ψ^{-1} are continuously differentiable,

(ii) $\gamma(t) = \phi(\psi(t))$ for all $t \in [a,b]$,

(iii) $\psi'(t) > 0$ for all $t \in [a,b]$.

Then $\int_\gamma \alpha = \int_\phi \alpha$ for every 1-form α.

Proof. Fix a 1-form $\alpha = \sum_{j=1}^n f_j\,dx_j$. Then
$$\begin{aligned}
\int_\gamma \alpha &= \int_a^b \sum_{j=1}^n f_j(\gamma(t))\gamma_j'(t)\,dt \\
&= \int_a^b \sum_{j=1}^n f_j(\phi(\psi(t)))\phi_j'(\psi(t))\psi'(t)\,dt \\
&= \int_c^d \sum_{j=1}^n f_j(\phi(u))\phi_j'(u)\,du \quad \Big(u = \psi(t),\ du = \psi'(t)\,dt\Big) \\
&= \int_\phi \alpha.
\end{aligned}$$

Here we've used assumption (iii) to guarantee that $\psi(a) = c$ and $\psi(b) = d$. $\square$

Remark 36. Assumption (iii) says that ψ is orientation preserving. If we instead assume that ψ is orientation reversing in the sense that $\psi'(t) < 0$ for $t \in [a, b]$ then the same argument shows that $\int_\gamma \alpha = -\int_\phi \alpha$.

Example 37. Returning to Example 34 we have $\gamma(t) = \phi(\psi(t))$ where $\psi(t) = -\cos t$. So if $\alpha = x\,dx + y^2\,dy$ we get

$$\begin{aligned}\int_\gamma \alpha &= \int_0^\pi (\cos t)(-\sin t) + (\sin^2 t)(\cos t)\,dt \\ &= \int_{-1}^1 u - u\sqrt{1-u^2}\,du \quad \Big(u = -\cos t,\ du = \sin t\,dt\Big) \\ &= \int_\phi \alpha,\end{aligned}$$

as promised by Proposition 35.

We can now prove a simple generalization of the fundamental theorem of calculus.

Proposition 38. *Consider a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$. For any curve $\gamma\colon [a, b] \to \mathbb{R}^n$ we have*

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

Proof. Recall that the chain rule says

$$\frac{d}{dt}[f(\gamma(t))] = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\gamma(t))\gamma_j'(t).$$

Therefore we have

$$\begin{aligned}\int_\gamma df &= \int_\gamma \sum_{j=1}^n \frac{\partial f}{\partial x_j}\,dx_j \\ &= \int_a^b \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\gamma(t))\gamma_j'(t)\,dt \\ &= \int_a^b \frac{d}{dt}[f(\gamma(t))]\,dt = f(\gamma(b)) - f(\gamma(a)),\end{aligned}$$

where we've used the ordinary fundamental theorem of calculus in the last line. $\square$

Remark 39. Of course this makes sense intuitively: df measures how f changes under small displacements and so integrating df along a curve should tell us the net change in f induced by traveling along the curve.

4 The Winding Number

4.1 Definition and Basic Properties

Recall that the rotation form α_{rot} on $\mathbb{R}^2 \setminus \{0\}$ is given by

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

Definition 40. Let $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ be a closed curve, i.e., $\gamma(a) = \gamma(b)$. Then the winding number of γ about 0 is

$$w(\gamma, 0) = \frac{1}{2\pi}\int_\gamma \alpha_{\mathrm{rot}}.$$

Intuitively, α_{rot} tells us how much rotation is induced by a small displacement. So integrating α_{rot} over a closed curve γ should tell us the total amount of rotation induced by following γ. Hence the winding number should compute the total number of times γ winds around the origin in a coutnerclockwise sense.

Example 41. Let $\gamma_k(t) = (\cos(kt), \sin(kt))$, $t \in [0, 2\pi]$. Then γ_k winds around the origin k times in a counterclockwise sense. We have

$$\begin{aligned}
\int_{\gamma_k} \alpha_{\mathrm{rot}} &= \int_{\gamma_k} \frac{-y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy \\
&= \int_0^{2\pi} [-\sin(kt)][-k\sin(kt)] + [\cos(kt)][k\cos(kt)]\,dt \\
&= \int_0^{2\pi} k\sin^2(kt) + k\cos^2(kt)\,dt = \int_0^{2\pi} k\,dt = 2\pi k,
\end{aligned}$$

and therefore $w(\gamma_k, 0) = k$, as expected.

Next we will try to rigorously justify this intuition for the winding number. This will require a fair bit of effort. We begin with a definition.

Definition 42. Let $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ be a curve. An angle function for γ is a continuous function $\theta\colon [a,b] \to \mathbb{R}$ such that $\gamma(t) = \|\gamma(t)\|(\cos\theta(t), \sin\theta(t))$ for all $t \in [a,b]$.

Example 43. Let $\gamma(t) = (\cos t, \sin t)$, $t \in [0, 4\pi]$. Then $\theta(t) = t$ is an angle function for γ. Actually $\theta(t) = t + 2\pi k$ is also an angle function for any fixed integer k. Note that the function

$$\tilde{\theta}(t) = \begin{cases} t, & \text{if } 0 \leq t < 2\pi \\ t - 2\pi, & \text{if } 2\pi \leq t \leq 4\pi \end{cases}$$

also has the property that $\gamma(t) = \|\gamma(t)\|(\cos\tilde{\theta}(t), \sin\tilde{\theta}(t))$ for all $t \in [0, 4\pi]$. However, $\tilde{\theta}$ is *not* an angle function for γ since $\tilde{\theta}$ is not continuous. Hence it is important that an angle function is allowed to take values in $\mathbb{R}$ and not just $[0, 2\pi]$.

Lemma 44. *Let $\gamma\colon [a,b] \to \mathbb{R}^2$ be a closed curve. Assume θ and $\tilde{\theta}$ are both angle functions for γ. Then there is an integer k such that $\theta(t) = \tilde{\theta}(t) + 2\pi k$ for all $t \in [a,b]$.*

Proof. Consider the function

$$f(t) = \frac{\theta(t) - \tilde{\theta}(t)}{2\pi}.$$

Note that f is continuous and that $f(t)$ is an integer for all $t \in [a,b]$. Therefore f must be constant. Thus there is an integer k such that $f(t) = k$ for all $t \in [a,b]$, i.e., such that $\theta(t) = \tilde{\theta}(t) + 2\pi k$ for all $t \in [a,b]$. □

Notice that if $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ is a closed curve, and θ is an angle function for γ then the number of times γ winds about the origin is

$$\frac{\theta(b) - \theta(a)}{2\pi}.$$

By Lemma 44, this quantity does not depend on the choice of angle function. It remains to show that every curve actually has an angle function. We can prove this by integrating the rotation form.

Lemma 45. *Let $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ be any curve. Choose θ_0 so that $\gamma(a) = \|\gamma(a)\|(\cos\theta_0, \sin\theta_0)$. Then*

$$\theta(t) = \theta_0 + \int_{\gamma|_{[a,t]}} \alpha_{rot}$$

is an angle function for γ. Here $\gamma|_{[a,t]}$ is the restriction of γ to the interval $[a,t]$.

Proof. We need to check that

$$\frac{\gamma(t)}{\|\gamma(t)\|} = (\cos\theta(t), \sin\theta(t))$$

for all t. It's equivalent to check that

$$\left(\frac{\gamma_1}{\|\gamma\|} - \cos\theta\right)^2 + \left(\frac{\gamma_2}{\|\gamma\|} - \sin\theta\right)^2 = 0.$$

Expanding the left hand side we find

$$\begin{aligned}
&\left(\frac{\gamma_1}{\|\gamma\|} - \cos\theta\right)^2 + \left(\frac{\gamma_2}{\|\gamma\|} - \sin\theta\right)^2 \\
&= \frac{\gamma_1^2}{\|\gamma\|^2} - 2\frac{\gamma_1}{\|\gamma\|}\cos\theta + \cos^2\theta + \frac{\gamma_2^2}{\|\gamma\|^2} - 2\frac{\gamma_2}{\|\gamma\|}\sin\theta + \sin^2\theta \\
&= 2 - 2\left(\frac{\gamma_1}{\|\gamma\|}\cos\theta + \frac{\gamma_2}{\|\gamma\|}\sin\theta\right).
\end{aligned}$$

Hence it's enough to check that

$$f(t) = \left(\frac{\gamma_1}{\|\gamma\|}\cos\theta + \frac{\gamma_2}{\|\gamma\|}\sin\theta\right)(t) = 1$$

for all $t \in [a,b]$. By choice of θ_0, we have $f(a) = 1$. Therefore, we're reduced to showing that $f'(t) = 0$ for all $t \in [a,b]$.

Now let's compute some derivatives. We have

$$\theta'(t) = \frac{d}{dt}\int_a^t \left(\frac{-\gamma_2\gamma_1'}{\|\gamma\|^2} + \frac{\gamma_1\gamma_2'}{\|\gamma\|^2}\right)(s)\,ds = \left(\frac{-\gamma_2\gamma_1'}{\|\gamma\|^2} + \frac{\gamma_1\gamma_2'}{\|\gamma\|^2}\right)(t).$$

Also by the quotient rule we have

$$\left(\frac{\gamma_1}{\|\gamma\|}\right)' = \frac{d}{dt}\left(\frac{\gamma_1}{(\gamma_1^2+\gamma_2^2)^{1/2}}\right) = \frac{\|\gamma\|\gamma_1' - \gamma_1\left(\frac{2\gamma_1\gamma_1'+2\gamma_2\gamma_2'}{2\|\gamma\|}\right)}{\|\gamma\|^2}$$
$$= \frac{\gamma_1'(\gamma_1^2+\gamma_2^2) - \gamma_1^2\gamma_1' - \gamma_1\gamma_2\gamma_2'}{\|\gamma\|^3} = \frac{\gamma_1'\gamma_2^2 - \gamma_1\gamma_2\gamma_2'}{\|\gamma\|^3}.$$

By a similar computation,

$$\left(\frac{\gamma_2}{\|\gamma\|}\right)' = \frac{\gamma_2'\gamma_1^2 - \gamma_1\gamma_2\gamma_1'}{\|\gamma\|^3}.$$

Hence we get

$$f'(t) = \left[\frac{\gamma_1'\gamma_2^2 - \gamma_1\gamma_2\gamma_2'}{\|\gamma\|^3}\right]\cos\theta - \frac{\gamma_1}{\|\gamma\|}\left[\frac{-\gamma_2\gamma_1'}{\|\gamma\|^2} + \frac{\gamma_1\gamma_2'}{\|\gamma\|^2}\right]\sin\theta$$
$$+ \frac{\gamma_2}{\|\gamma\|}\left[\frac{-\gamma_2\gamma_1'}{\|\gamma\|^2} + \frac{\gamma_1\gamma_2'}{\|\gamma\|^2}\right]\cos\theta + \left[\frac{\gamma_2'\gamma_1^2 - \gamma_1\gamma_2\gamma_1'}{\|\gamma\|^3}\right]\sin\theta.$$

But this is zero, as needed, since the coefficients of $\cos\theta$ cancel to 0 and likewise the coefficients of $\sin\theta$ cancel to 0. □

We can now prove the main theorem about the winding number.

Theorem 46. *Let $\gamma\colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ be a closed curve. Then the winding number $w(\gamma, 0)$ is an integer.*

Proof. From the lemma we know that

$$\theta(t) = \theta_0 + \int_{\gamma|_{[a,t]}} \alpha_{\mathrm{rot}}$$

is an angle function for γ. Now $\theta(b) = \theta(a) + 2\pi k$ for some integer k since γ is closed. Hence

$$\int_\gamma \alpha_{\mathrm{rot}} = \int_{\gamma|_{[a,b]}} \alpha_{\mathrm{rot}} - \int_{\gamma|_{[a,a]}} \alpha_{\mathrm{rot}}$$
$$= (\theta(b) - \theta_0) - (\theta(a) - \theta_0) = \theta(b) - \theta(a) = 2\pi k,$$

and it follows that $w(\gamma, 0) = k$ is an integer. □

4.2 Homotopy Invariance

Next we'd like to discuss how winding number changes when we deform curves.

Definition 47. Fix a subset $U \subset \mathbb{R}^2$. Assume $\gamma, \eta \colon [a,b] \to U$ are closed curves. A homotopy in U between γ and η is a smooth map $h \colon [0,1]\times[a,b] \to U$ such that

(i) $h(0,t) = \gamma(t)$ for all $t \in [a,b]$,

(ii) $h(1,t) = \eta(t)$ for all $t \in [a,b]$,

(iii) $h(s,a) = h(s,b)$ for all $s \in [0,1]$.

If such a homotopy h exists, we say that γ and η are homotopic in U.

Example 48. Let $\gamma(t) = (\cos t, \sin t)$ and let $\eta(t) = (2\cos t, 2\sin t)$ for $t \in [0, 2\pi]$. Then $h(s,t) = ((1+s)\cos t, (1+s)\sin t)$ is a homotopy in $\mathbb{R}^2$ between γ and η.

One can think of a homotopy as a smooth deformation interpolating between γ and η. When considering questions about homotopy, it is essential to keep the set U in mind. For example, the notion of homotopy is not very interesting when $U = \mathbb{R}^2$, as the following theorem shows.

Theorem 49. *Every pair of closed curves $\gamma, \eta \colon [a,b] \to \mathbb{R}^2$ are homotopic in $\mathbb{R}^2$.*

Proof. The idea of the proof is that we can deform γ to η along straight lines connecting $\gamma(t)$ to $\eta(t)$. More formally, define $h \colon [0,1] \times [a,b] \to \mathbb{R}^2$ by

$$h(s,t) = (1-s)\gamma(t) + s\eta(t).$$

It is easy to see that h satisfies all the requirements for a homotopy. $\square$

Homotopy is more interesting when we take $U = \mathbb{R}^2 \setminus \{0\}$. Indeed, in $\mathbb{R}^2 \setminus \{0\}$ the previous straight line homotopy may no longer work if one of the lines passes through the origin. In fact, there are closed curves in $\mathbb{R}^2 \setminus \{0\}$ that are not homotopic.

Theorem 50. *Assume the closed curves* $\gamma, \eta \colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ *are homotopic in* $\mathbb{R}^2 \setminus \{0\}$. *Then* $w(\gamma, 0) = w(\eta, 0)$.

Proof. By assumption there exists a homotopy $h \colon [0,1] \times [a,b] \to \mathbb{R}^2 \setminus \{0\}$ between γ and η. Define a function $f \colon [0,1] \to \mathbb{R}$ by

$$f(s) = \frac{1}{2\pi} \int_a^b \frac{-\partial_t h_2(s,t) h_1(s,t)}{\|h(s,t)\|^2} + \frac{h_2(s,t)\partial_t h_1(s,t)}{\|h(s,t)\|^2}\, dt. \tag{1}$$

Then $f(s) = w(h_s, 0)$ where $h_s \colon [a,b] \to \mathbb{R}^2 \setminus \{0\}$ is the closed curve $h_s(t) = h(s,t)$. In particular, $f(s)$ is an integer for all $s \in [0,1]$. On the other hand, formula (1) implies that f is a continuous function of s. Therefore f must be constant, and so $w(\gamma, 0) = f(0) = f(1) = w(\eta, 0)$. □

5 Determinants

Next we're going to move on to discussing n-forms on $\mathbb{R}^n$. These are the easiest forms to understand after 1-forms. Before we can define n-forms, however, we need to discuss determinants.

5.1 Definition and Basic Properties

The determinant is a function from $n \times n$ matrices to real numbers. Equivalently, it takes in n vectors in $\mathbb{R}^n$ and spits out a number.

$$\det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \leftrightarrow \det(v_1, \ldots, v_n), \quad \text{where } v_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}.$$

You may have seen various definitions of the determinant in the past. For instance,

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \left[\det \begin{pmatrix} e & f \\ h & i \end{pmatrix}\right] - b \left[\det \begin{pmatrix} d & f \\ g & i \end{pmatrix}\right] + c \left[\det \begin{pmatrix} d & e \\ g & h \end{pmatrix}\right].$$

These computational rules certainly have their place, but for us the following axiomatic definition is also useful.

Definition 51. An n-dimensional determinant is a function

$$D\colon \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}} \to \mathbb{R}$$

which satisfies the following properties.

(i) $D(v_1, \ldots, tv_i, \ldots, v_n) = tD(v_1, \ldots, v_i, \ldots, v_n)$

(ii) $D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) = -D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n)$

(iii) $D(v_1, \ldots, v_i + tv_j, \ldots, v_n) = D(v_1, \ldots, v_i, \ldots, v_n)$

(iv) $D(e_1, \ldots, e_n) = 1$.

When faced with such an axiomatic definition it is natural to ask whether a function satisfying the axioms actually exists. Moreover, if such a function does exist, must it be unique? We will now address these questions. We begin with the issue of uniqueness.

Proposition 52. *There is at most one n-dimensional determinant function.*

Proof. Suppose D_1 and D_2 both satisfy all the properties of an n-dimensional determinant function. Consider n vectors $v_1, \ldots, v_n$ in $\mathbb{R}^n$. There are two cases to consider.

First suppose that $v_1, \ldots, v_n$ are linearly dependent. For simplicity, we can assume that

$$v_1 = a_2 v_2 + \ldots + a_n v_n$$

for some constants $a_2, \ldots, a_n \in \mathbb{R}$. Then repeatedly applying property (iii) shows that

$$\begin{aligned} D_1(v_1, \ldots, v_n) &= D_1(\sum_{j=2}^n a_j v_j, v_2, \ldots, v_n) = D_1(\sum_{j=3}^n a_j v_j, v_2, \ldots, v_n) \\ &= D_1(\sum_{j=4}^n a_j v_j, v_2, \ldots, v_n) = \cdots = D_1(0, v_2, \ldots, v_n). \end{aligned}$$

Now by properties (i) and (ii),

$$D_1(0, v_2, \ldots, v_n) = D_1(-0, v_2, \ldots, v_n) = -D_1(0, v_2, \ldots, v_n)$$

and therefore $D_1(0, v_2, \ldots, v_n) = 0$. Thus $D_1(v_1, \ldots, v_n) = 0$. By the exact same argument we must have $D_2(v_1, \ldots, v_n) = 0$ and hence $D_1(v_1, \ldots, v_n) = D_2(v_1, \ldots, v_n)$.

Now suppose instead that $v_1, \ldots, v_n$ are linearly independent. Consider the $n \times n$ matrix A whose ith column is v_i. Then A is invertible. Recall that an elementary matrix E is an $n \times n$ matrix such that right multiplying another matrix $n \times n$ matrix B by E has one of the following effects:

(i) it scales a column of B by t, or

(ii) it interchanges two columns of B, or

(iii) it adds a multiple of one column of B to another column of B.

Since A is invertible, there exists a sequence of elementary matrices $E_1, \ldots, E_m$ such that $AE_1E_2 \cdots E_m = I$.

Now observe that properties (i)-(iii) of a determinant function exactly tell us how the value of D_1 changes when we right multiply by an elementary matrix. Therefore, there are non-zero numbers a_i, uniquely determined by E_i, such that

$$\begin{aligned} 1 = D_1(I) &= D_1(AE_1 \cdots E_{m-1}E_m) \\ &= D_1(AE_1 \cdots E_{m-1})a_m = \cdots = D_1(A)a_1 \cdots a_m. \end{aligned}$$

Since right multiplying by elementary matrices has exactly the same effect on D_2, we likewise get

$$\begin{aligned} 1 = D_2(I) &= D_2(AE_1 \cdots E_{m-1}E_m) \\ &= D_2(AE_1 \cdots E_{m-1})a_m = \cdots = D_2(A)a_1 \cdots a_m, \end{aligned}$$

and therefore $D_1(A) = D_2(A)$, as needed. □

Example 53. As a concrete example of the ideas in the previous proof, we show how to compute

$$\det \begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}$$

by using column operations. First, using the fact that determinant is unchanged by adding a multiple of one column to another we get

$$\det\begin{pmatrix}1&2&2\\2&3&4\\3&4&5\end{pmatrix}=\det\begin{pmatrix}1&1&2\\2&1&4\\3&1&5\end{pmatrix}=\det\begin{pmatrix}1&1&0\\2&1&0\\3&1&-1\end{pmatrix}=\det\begin{pmatrix}0&1&0\\1&1&0\\2&1&-1\end{pmatrix}$$
$$=\det\begin{pmatrix}0&1&0\\1&1&0\\0&1&-1\end{pmatrix}=\det\begin{pmatrix}0&1&0\\1&0&0\\0&1&-1\end{pmatrix}=\det\begin{pmatrix}0&1&0\\1&0&0\\0&0&-1\end{pmatrix}.$$

Next using the fact that determinant changes sign when we interchange columns we get

$$\det\begin{pmatrix}0&1&0\\1&0&0\\0&0&-1\end{pmatrix}=-\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&-1\end{pmatrix}.$$

Then using the scaling property we get

$$-\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&-1\end{pmatrix}=\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}.$$

But the determinant of the identity matrix is 1, and therefore

$$\det\begin{pmatrix}1&2&2\\2&3&4\\3&4&5\end{pmatrix}=1.$$

In the same way, we could compute the determinant of any other matrix by column reducing it to the identity.

Next we need to show that a determinant function actually exists. We can do this by writing down an explicit formula. In fact,

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

To make sense of this, we need to explain what all the symbols in the formula mean. Here S_n stands for all permutations of the numbers $1,\ldots,n$. In other words, S_n consists of all ordered lists where each number $1,\ldots,n$ appears exactly once.

Example 54. The two elements of S_2 are $(1,2)$ and $(2,1)$. The six elements of S_3 are $(1,2,3)$, $(1,3,2)$, $(2,1,3)$, $(2,3,1)$, $(3,1,2)$, and $(3,2,1)$. In general, S_n has $n!$ elements.

Notation. We write $\sigma = (\sigma(1), \ldots, \sigma(n))$ for a permutation in S_n. For example, if $\sigma = (1,3,2)$ then $\sigma(1) = 1$, $\sigma(2) = 3$, and $\sigma(3) = 2$.

Definition 55. It is possible to obtain any permutation $\sigma \in S_n$ by starting with $(1,2,\ldots,n)$ and then repeatedly swapping pairs of numbers. We say σ is even if it takes an even number of swaps to get from $(1,2,\ldots,n)$ to σ, and σ is odd if it takes an odd number of swaps to get from $(1,2,\ldots,n)$ to σ.

Example 56. The permutation $(1,2,4,3)$ is odd since we can obtain it from the identity via one swap:

$$(1,2,3,4) \to (1,2,4,3).$$

The permutation $(3,2,4,1)$ is even since we can obtain it from the identity via two swaps:

$$(1,2,3,4) \to (1,2,4,3) \to (3,2,4,1).$$

Of course, there are many different ways to get from the identity to σ by performing swaps. For example, we can also get to $(1,2,4,3)$ by the following sequence of five swaps:

$$(1,2,3,4) \to (2,1,3,4) \to (4,1,3,2) \to (4,1,2,3) \to (4,2,1,3) \to (1,2,4,3).$$

It's a theorem (that we won't prove) that for every $\sigma \in S_n$, it's impossible to get from the identity to σ by an even number of swaps and to also get from the identity to σ by an odd number of swaps. Thus the notion of σ being even or odd is well-defined.

Definition 57. For $\sigma \in S_n$ the sign of σ is

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

We've now defined all the symbols appearing in the formula

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

For concreteness, we now expand the formula for $n = 2$ and $n = 3$.

Example 58. Assume that $n = 2$ and let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

There are exactly two permutations $\sigma = (1, 2)$ and $\tau = (2, 1)$ in S_2. Hence the formula for $\det(A)$ becomes

$$\begin{aligned} \det(A) &= \operatorname{sign}(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} + \operatorname{sign}(\tau) a_{\tau(1),1} a_{\tau(2),2} \\ &= a_{11}a_{22} - a_{21}a_{12}, \end{aligned}$$

which is the familiar rule for the determinant of a 2×2 matrix.

Example 59. Assume that $n = 3$ and let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

There are six permutations $\sigma_1 = (1, 2, 3)$, $\sigma_2 = (1, 3, 2)$, $\sigma_3 = (2, 1, 3)$, $\sigma_4 = (2, 3, 1)$, $\sigma_5 = (3, 1, 2)$, and $\sigma_6 = (3, 2, 1)$ in S_3. Writing out the formula for $\det(A)$ yields

$$\begin{aligned} \det(A) &= \operatorname{sign}(\sigma_1) a_{11}a_{22}a_{33} + \operatorname{sign}(\sigma_2) a_{11}a_{32}a_{23} + \operatorname{sign}(\sigma_3) a_{21}a_{12}a_{33} \\ &\qquad + \operatorname{sign}(\sigma_4) a_{21}a_{32}a_{13} + \operatorname{sign}(\sigma_5) a_{31}a_{12}a_{23} + \operatorname{sign}(\sigma_6) a_{31}a_{22}a_{13} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} \\ &\qquad + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}. \end{aligned}$$

A convenient way to remember this is to form an augmented matrix by copying the first two rows of A.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

The three terms appearing with a plus sign are the products of the diagonals running down and to the right, and the three terms appearing with a minus sign are the products of the diagonals running up and to the right.

Theorem 60. *The function*

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

satisfies all the properties of a determinant function.

Proof. We will check each of the four properties.

(i) Let B be the matrix obtained from A by scaling the ith column by t. Then for each $\sigma \in S_n$ we have

$$b_{\sigma(j),j} = \begin{cases} a_{\sigma(j),j} & \text{if } j \neq i, \\ t a_{\sigma(j),j} & \text{if } j = i. \end{cases}$$

Therefore

$$\begin{aligned} \det(B) &= \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) b_{\sigma(1),1} \cdots b_{\sigma(n),n} \\ &= \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) t a_{\sigma(1),1} \cdots a_{\sigma(n),n} \\ &= t \det(A), \end{aligned}$$

as needed.

(ii) Let B be the matrix obtained from A by swapping the ith and jth columns. Also for each $\sigma \in S_n$, let $\tilde{\sigma}$ be the permutation obtained by swapping the ith and jth entries in σ. Notice that $\tilde{\sigma}$ and σ have opposite signs and that $b_{\sigma(i),i} = a_{\tilde{\sigma}(i),i}$. Therefore

$$\begin{aligned} \det(B) &= \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) b_{\sigma(1),1} \cdots b_{\sigma(n),n} \\ &= \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) a_{\tilde{\sigma}(1),1} \cdots a_{\tilde{\sigma}(n),n} \\ &= \sum_{\tilde{\sigma} \in S_n} -\mathrm{sign}(\tilde{\sigma}) a_{\tilde{\sigma}(1),1} \cdots a_{\tilde{\sigma}(n),n} \\ &= -\det(A), \end{aligned}$$

as needed.

(iii) Let B be the matrix obtained from A by adding t times the jth column to the ith column. Let C be the matrix obtained from A by replacing the ith column with a copy of the jth column. Then we have

$$\begin{aligned}
\det(B) &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) b_{\sigma(1),1} \cdots b_{\sigma(n),n} \\
&= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} + t \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) c_{\sigma(1),1} \cdots c_{\sigma(n),n} \\
&= \det(A) + t\det(C).
\end{aligned}$$

In this calculation, we've used the fact that $b_{\sigma(i),i} = a_{\sigma(i),i} + c_{\sigma(i),i}$ and that $b_{\sigma(k),k} = a_{\sigma(k),k} = c_{\sigma(k),k}$ for $k \neq i$. Finally note that C has two identical columns and so $\det(C) = 0$ by property (ii). Thus $\det(B) = \det(A)$, as needed.

(iv) Let A be the $n \times n$ identity matrix. Notice that the product $a_{\sigma(1),1} \cdots a_{\sigma(n),n}$ will be zero unless $\sigma(i) = i$ for every i. Thus

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} \\
&= \operatorname{sign}((1, 2, \ldots, n)) a_{11} \cdots a_{nn} = 1,
\end{aligned}$$

as needed. □

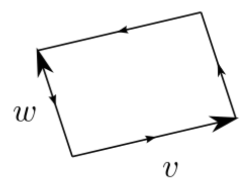
5.2 Relationship with Volume

We have now shown that there exists a unique determinant function satisfying properties (i)-(iv). Of course the question still remains as to why a function satisfying these properties is interesting. To answer this, let's look at another function that also satisfies these properties. For simplicity, we'll just describe the situation in $\mathbb{R}^2$.

Definition 61. Given $v, w \in \mathbb{R}^2$, the parallelogram $P(v, w)$ spanned by v and w is the set of all points of the form $sv + tw$ with $0 \leq s \leq 1$ and $0 \leq t \leq 1$.

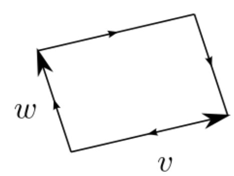
Note that we can give an orientation to $P(v, w)$ in the following way. Travel along the boundary of $P(v, w)$ starting in the v direction. Then we say $P(v, w)$ is positively oriented if we traverse the boundary counterclockwise and we say

$P(v, w)$ is negatively oriented if we traverse the boundary clockwise. For example, the following parallelogram $P(v, w)$ is positively oriented.



$P(v, w)$ positively oriented

On the other hand, interchanging the order of v and w yields a negatively oriented parallelogram $P(w, v)$.



$P(w, v)$ negatively oriented

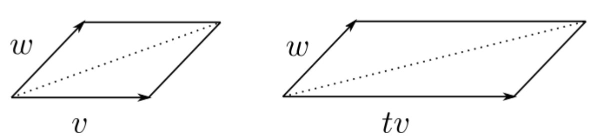
Definition 62. The oriented area function $A\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is

$$A(v, w) = \begin{cases} \operatorname{Area}(P(v, w)) & \text{if } P(v, w) \text{ is positively oriented} \\ -\operatorname{Area}(P(v, w)) & \text{if } P(v, w) \text{ is negatively oriented.} \end{cases}$$

Proposition 63. *The oriented area function A satisfies all the axioms of a determinant function. Therefore $A(v, w) = \det(v, w)$.*

Proof. We need to check that A satisfies properties (i)-(iv).

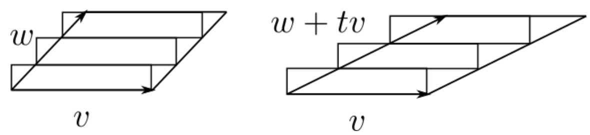
(i) Note that $P(tv, w)$ is obtained from $P(v, w)$ by scaling one side by a factor of t.



If we think of $P(v, w)$ as composed of two triangles, this scales the base of the triangles by a factor of t and leaves the height unchanged. Consequently it scales the area by a factor of t as well: $A(P(tv, w)) = tA(P(v, w))$.

(ii) Note that $P(v, w)$ and $P(w, v)$ have opposite orientations, and hence $A(v, w) = -A(w, v)$.

(iii) Note that $P(v, w + tv)$ is obtained from $P(v, w)$ by shearing.



Since shearing doesn't change area, we get that $A(v, w + tv) = A(v, w)$.

(iv) Note that $P(e_1, e_2)$ is a positively oriented unit square. Hence it has oriented area 1, and so $A(e_1, e_2) = 1$. $\square$

Remark 64. The same thing is true in higher dimensions for the same reasons. The determinant $\det(v_1, \ldots, v_n)$ computes the oriented volume of the n-dimensional parallelepiped spanned by $v_1, \ldots, v_n$.

We close this section by recording one more useful fact about determinants. The determinant is actually additive in each of its arguments.

Proposition 65. *The determinant satisfies*

$$\det(v_1, \ldots, v_i + \tilde{v}_i, \ldots, v_n) = \det(v_1, \ldots, v_i, \ldots, v_n) + \det(v_1, \ldots, \tilde{v}_i, \ldots, v_n)$$

Proof. Let A be the matrix with columns $v_1, \ldots, v_i, \ldots, v_n$. Let B be the matrix with columns $v_1, \ldots, \tilde{v}_i, \ldots, v_n$. Let C be the matrix with columns $v_1, \ldots, v_i + \tilde{v}_i, \ldots, v_n$. Then

$$\begin{aligned}
\det(C) &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) c_{\sigma(1),1} \cdots c_{\sigma(n),n} \\
&= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} + \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) b_{\sigma(1),1} \cdots b_{\sigma(n),n} \\
&= \det(A) + \det(B),
\end{aligned}$$

as needed. In the above calculation, we've used the fact that $c_{\sigma(i),i} = a_{\sigma(i),i} + b_{\sigma(i),i}$ and that $c_{\sigma(j),j} = a_{\sigma(j),j} = b_{\sigma(j),j}$ for $j \neq i$. $\square$

6 Top Dimensional Forms

6.1 Definition of Top Dimensional Forms

We are now ready to define n-forms on $\mathbb{R}^n$.

Definition 66. (Informal) An n-form ω on $\mathbb{R}^n$ is a rule that assigns a number to each oriented n-dimensional parallelepiped in $\mathbb{R}^n$ in a suitable fashion.

As before, we will place some requirements on ω. Namely, on each individual tangent space $T_p\mathbb{R}^n$ we require

(i) $\omega_p(v_1, \ldots, tv_i, \ldots, v_n) = t\omega_p(v_1, \ldots, v_i, \ldots, v_n)$

(ii) $\omega_p(v_1, \ldots, v_i + \tilde{v}_i, \ldots, v_n) = \omega_p(v_1, \ldots, v_i, \ldots, v_n) + \omega_p(v_1, \ldots, \tilde{v}_i, \ldots, v_n)$

(iii) $\omega_p(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -\omega_p(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n)$.

The scaling property (i) and additivity property (ii) are similar to the requirements we placed on 1-forms. The new alternating property (iii) essentially says that ω respects orientation.

But now notice that

$$\begin{aligned}
&\omega_p(v_1, \ldots, v_i + tv_j, \ldots, v_j, \ldots, v_n) \\
&\quad = \omega_p(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) + t\omega_p(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n) \\
&\quad = \omega_p(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n),
\end{aligned}$$

where the second term on the second line is zero by the alternation property. Therefore ω_p satisfies all the properties of the determinant except the normalization condition: $\omega_p(e_1, \ldots, e_n)$ may not equal 1. However,

$$\frac{\omega_p}{\omega_p(e_1, \ldots, e_n)}$$

does satisfy all properties of the determinant. It follows that there is some constant a such that $\omega_p(v_1, \ldots, v_n) = a \det(v_1, \ldots, v_n)$.

We've now shown that on each tangent space $T_p\mathbb{R}^n$, the form ω acts by some multiple $a(p)$ of determinant. As in the case of 1-forms, we will require this function $a(p)$ to depend smoothly on p. This leads to the following formal definition of an n-form.

Definition 67. A differential n-form on $\mathbb{R}^n$ is a rule assigning a number to each oriented n-dimensional parallelepiped in $\mathbb{R}^n$ which can be expressed in the form

$$\omega_p(v_1, \ldots, v_n) = f(p) \det(v_1, \ldots, v_n)$$

for some smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$.

Notation. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function. We write $\omega = f\, dx_1 \ldots dx_n$ for the n-form ω that acts by $\omega_p(v_1, \ldots, v_n) = f(p) \det(v_1, \ldots, v_n)$. In other words, we write $dx_1 \ldots dx_n$ instead of det.

Example 68. Consider the 3-form $\omega = xyz^2\, dx\, dy\, dz$ on $\mathbb{R}^3$. Then

$$\omega_{(3,1,2)}(e_1, e_2, e_3) = (3)(1)(2)^2 \det(e_1, e_2, e_3) = 12.$$

6.2 Integrating Top Dimensional Forms

Now let's talk about integration. We integrate n-forms over n-dimensional subsets of $\mathbb{R}^n$. As in the 1-dimensional case, we will work with parameterized n-dimensional subsets of $\mathbb{R}^n$.

Definition 69. An n-cell in $\mathbb{R}^n$ is a smooth map $c\colon [a_1, b_1] \times \ldots \times [a_n, b_n] \to \mathbb{R}^n$.

Remark 70. An n-cell can be singular. For example, it could just collapse everything to a point. Like with curves, if we want to rule out singularities we can impose some extra condition on the derivative. An n-cell c is called regular if the Jacobian matrix $(Dc)(p)$ is invertible at all points p in the domain of c.

Example 71. The map $c\colon [0, 1]^2 \to \mathbb{R}^2$ given by $c(x, y) = (x, y)$ parameterizes the unit square in the obvious way.

Example 72. The map $c\colon [0, 1] \times [0, 2\pi] \to \mathbb{R}^2$ given by $c(r, \theta) = (r \cos\theta, r \sin\theta)$ parameterizes the unit disk in polar coordinates. Note that c is not regular.

Example 73. The map $c\colon [0, 1] \times [0, 2\pi] \times [-1, 1] \to \mathbb{R}^3$ given by

$$c(r, \theta, h) = (r \cos\theta, r \sin\theta, h)$$

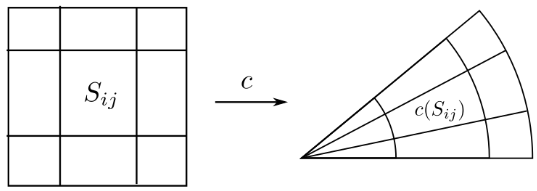
parameterizes a solid cylinder using cylindrical coordinates.

Example 74. The map $c\colon [0, 1] \times [0, 2\pi] \times [0, \pi] \to \mathbb{R}^3$ given by

$$c(r, \theta, \phi) = (r \cos\theta \sin\phi, r \sin\theta \sin\phi, \cos\phi)$$

parameterizes the unit ball in $\mathbb{R}^3$ using spherical coordinates.

The intuition for integrating n-forms is very similar to the intuition for integrating 1-forms. For simplicity, we describe the picture in the case $n = 2$. Consider a map $c\colon [0,1]^2 \to \mathbb{R}^2$. Choose partitions $0 = s_0 < s_1 < \ldots < s_\ell = 1$ and $0 = t_0 < t_1 < \ldots < t_m = 1$. These give a decomposition of $[0,1]^2$ into small squares $S_{ij} = [s_{i-1}, s_i] \times [t_{j-1}, t_j]$. The images $c(S_{ij})$ cut up the image of c into small pieces.



Notice that by Taylor's theorem

$$\begin{aligned} c(s_i, t_{j-1}) - c(s_{i-1}, t_{j-1}) &= \left[\frac{\partial c}{\partial s}(s_{i-1}, t_{j-1})\right](s_i - s_{i-1}) + o(|s_i - s_{i-1}|), \\ c(s_{i-1}, t_j) - c(s_{i-1}, t_{j-1}) &= \left[\frac{\partial c}{\partial t}(s_{i-1}, t_{j-1})\right](t_j - t_{j-1}) + o(|t_j - t_{j-1}|). \end{aligned}$$

Therefore, when $s_i - s_{i-1}$ and $t_j - t_{j-1}$ are small, the set $c(S_{ij})$ is very nearly a parallelogram P_{ij} based at $c(s_{i-1}, t_{j-1})$ with sides $\frac{\partial c}{\partial s}(s_{i-1}, t_{j-1})$ and $\frac{\partial c}{\partial t}(s_{i-1}, t_{j-1})$.

The collection of parallelograms P_{ij}, $1 \le i \le \ell$, $1 \le j \le m$ forms an approximation to the image of c. Now consider an n-form $\omega = f\, dx\, dy$. To integrate ω over c, we add up the value of ω over all the parallelograms P_{ij} to get a quantity $\sum_{i=1}^{\ell} \sum_{j=1}^{m} \omega(P_{ij})$. Then we try to take a limit of this quantity as the approximation gets better and better. Now we have

$$\begin{aligned} &\sum_{i=1}^{\ell} \sum_{j=1}^{m} \omega(P_{ij}) \\ &\quad = \sum_{i=1}^{\ell} \sum_{j=1}^{m} (s_i - s_{i-1})(t_j - t_{j-1}) \omega_{c(s_{i-1}, t_{j-1})} \left(\frac{\partial c}{\partial s}(s_{i-1}, t_{j-1}), \frac{\partial c}{\partial t}(s_{i-1}, t_{j-1})\right) \\ &\quad = \sum_{i=1}^{\ell} \sum_{j=1}^{m} (s_i - s_{i-1})(t_j - t_{j-1}) f(c(s_{i-1}, t_{j-1})) \begin{vmatrix} \frac{\partial c_1}{\partial s}(s_{i-1}, t_{j-1}) & \frac{\partial c_1}{\partial t}(s_{i-1}, t_{j-1}) \\ \frac{\partial c_2}{\partial s}(s_{i-1}, t_{j-1}) & \frac{\partial c_2}{\partial t}(s_{i-1}, t_{j-1}) \end{vmatrix}. \end{aligned}$$

But this is exactly a Riemann sum for the integral

$$\int_0^1 \int_0^1 f(c(s,t)) \det(Dc(s,t))\, ds\, dt.$$

Hence the quantity $\sum_{i=1}^{\ell} \sum_{j=1}^{m} \omega(P_{ij})$ converges to this integral as the approximation gets better and better. Based on this, we make the following definition.

Definition 75. The integral of an n-form $\omega = f\, dx_1 \dots dx_n$ over an n-cell $c\colon [a_1, b_1] \times \dots \times [a_n, b_n] \to \mathbb{R}^n$ is

$$\int_c \omega = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(c(t_1, \dots, t_n)) \det(Dc(t_1, \dots, t_n))\, dt_1 \dots dt_n.$$

Remark 76. Notice that

$$f(c(t_1, \dots, t_n)) \det(Dc(t_1, \dots, t_n)) = \omega_{c(t_1, \dots, t_n)} \left(\frac{\partial c}{\partial t_1}, \dots, \frac{\partial c}{\partial t_n} \right).$$

Hence we could equivalently write

$$\int_c \omega = \int_{a}^{b_1} \cdots \int_{a_n}^{b_n} \omega_{c(t_1, \dots, t_n)} \left(\frac{\partial c}{\partial t_1}, \dots, \frac{\partial c}{\partial t_n} \right) dt_1 \dots dt_n.$$

Example 77. Let $\omega = x^2 y\, dx\, dy$ and let $c\colon [0,1]^2 \to \mathbb{R}^2$ be given by $c(s,t) = (s,t)$. Then

$$Dc = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and so $\det(Dc) = 1$. Thus

$$\int_c \omega = \int_0^1 \int_0^1 s^2 t\, ds\, dt = \frac{1}{3} \int_0^1 t\, dt = \frac{1}{6}.$$

Note that this is just the usual integral of the function $f(x,y) = x^2 y$ over the unit square.

Example 78. Let $\omega = (x^2 + y^2)\, dx\, dy$ and let $c\colon [0,1] \times [0, 2\pi] \to \mathbb{R}^2$ be given by $c(r, \theta) = (r \cos\theta, r \sin\theta)$. Then

$$Dc = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and so $\det(Dc) = r$. It follows that

$$\int_c \omega = \int_0^1 \int_0^{2\pi} r^2 \cdot r \, dr \, d\theta = \frac{\pi}{2}.$$

Again, this is just the usual integral of the function $f(x, y) = x^2 + y^2$ over the unit circle, computed in polar coordinates.

As with 1-forms, we can show that the integral of an n-form does not change if we reparameterize an n-cell. To give the proof, we need to recall the chain rule from multivariable calculus, as well as the change of variables formula for multiple integrals.

Theorem 79 (Chain Rule)**.** *Assume that $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^m \to \mathbb{R}^\ell$ are smooth maps. Then*

$$D(g \circ f)(p) = \Big[(Dg)(f(p))\Big]\Big[Df(p)\Big].$$

Here the right hand side is a product of two matrices.

Theorem 80 (Change of Variables)**.** *Assume U and V are subsets of $\mathbb{R}^n$ and $\psi\colon U \to V$ is invertible and both ψ and ψ^{-1} are continuously differentiable. Then*

$$\int_V f(y)\, dy = \int_U f(\psi(x)) |\det(D\psi(x))| \, dx.$$

We also need the following lemma about determinants.

Lemma 81. *For any matrices A and B we have $\det(AB) = \det(A)\det(B)$.*

Proof. We can assume A and B are both invertible because otherwise both $\det(AB)$ and $\det(A)\det(B)$ are 0. Choose elementary matrices $E_1, \ldots, E_m$ such that $AE_1 \cdots E_m = I$. Likewise choose elementary matrices $F_1, \ldots, F_\ell$ such that $BF_1 \cdots F_\ell = I$. Then

$$1 = \det(I) = \det(AE_1 \cdots E_m) = \det(A) a_1 \cdots a_m$$

where a_i is uniquely determined by E_i. Likewise,

$$1 = \det(I) = \det(BF_1 \cdots F_\ell) = \det(B) b_1 \cdots b_\ell$$

where b_i is uniquely determined by F_i. But $ABF_1 \cdots F_\ell E_1 \cdots E_m = I$ and so

$$1 = \det(I) = \det(ABF_1 \cdots F_\ell E_1 \cdots E_m) = \det(AB) b_1 \cdots b_\ell a_1 \cdots a_m.$$

Therefore

$$\det(AB) = \frac{1}{a_1 \cdots a_m b_1 \cdots b_\ell} = \det(A)\det(B),$$

as needed. □

We can now prove invariance under reparameterizations.

Proposition 82. *Suppose* $c\colon [a_1, b_1] \times \ldots \times [a_n, b_n] \to \mathbb{R}^n$ *and* $\tilde{c}\colon [\tilde{a}_1, \tilde{b}_1] \times \ldots \times [\tilde{a}_n, \tilde{b}_n] \to \mathbb{R}^n$ *are reparameterizations of each other in the following sense: there exists a map*

$$\psi\colon [a_1, b_1] \times \ldots \times [a_n, b_n] \to [\tilde{a}_1, \tilde{b}_1] \times \ldots \times [\tilde{a}_n, \tilde{b}_n]$$

such that

(i) ψ *is invertible and both* ψ *and* ψ^{-1} *are continuously differentiable,*

(ii) $c(p) = \tilde{c}(\psi(p))$ *for all* p,

(iii) $\det(D\psi) > 0$ *at all points* p.

Then for any n*-form* ω *we have* $\int_c \omega = \int_{\tilde{c}} \omega$.

Proof. We have

$$\begin{aligned}
\int_c \omega &= \int_{[a_1,b_1]\times\ldots\times[a_n,b_n]} f(c(p)) \det(Dc(p))\, dp \\
&= \int_{[a_1,b_1]\times\ldots\times[a_n,b_n]} f(\tilde{c}(\psi(p))) \det(D(\tilde{c}\circ\psi)(p))\, dp \\
&= \int_{[a_1,b_1]\times\ldots\times[a_n,b_n]} f(\tilde{c}(\psi(p))) \det(D\tilde{c}(\psi(p))) \det(D\psi(p))\, dp \\
&= \int_{[\tilde{a}_1,\tilde{b}_1]\times\ldots\times[\tilde{a}_n,\tilde{b}_n]} f(\tilde{c}(q)) \det(D\tilde{c}(q))\, dq \\
&= \int_{\tilde{c}} \omega.
\end{aligned}$$

Here we used the chain rule and Lemma 81 to get from the second to the third line, and we made the change of variables $q = \psi(p)$ to get from the third line to the fourth line. □

7 Differential k-forms

7.1 Definition of k-forms

We are now ready to handle the general case of differential k-forms on $\mathbb{R}^n$.

Definition 83. (Informal) A k-form α on $\mathbb{R}^n$ is a rule that assigns a number to each oriented k-dimensional parallelepiped in $\mathbb{R}^n$ in a suitable way.

As before, we can specify an oriented k-dimensional parallelepiped based at p by giving a list of vectors $v_1, \ldots, v_k$ in $T_p\mathbb{R}^n$. We write $\alpha_p(v_1, \ldots, v_k)$ for the value of α on this parallelepiped. Again we shall require α_p to satisfy the scaling, additivity, and alternation requirements:

(i) $\alpha_p(v_1, \ldots, tv_i, \ldots, v_k) = t\alpha_p(v_1, \ldots, v_i, \ldots, v_k)$

(ii) $\alpha_p(v_1, \ldots, v_i + \tilde{v}_i, \ldots, v_k) = \alpha_p(v_1, \ldots, v_i, \ldots, v_k) + \alpha_p(v_1, \ldots, \tilde{v}_i, \ldots, v_k)$

(iii) $\alpha_p(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_k) = -\alpha_p(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)$.

Also we require α_p to depend smoothly on p.

Example 84. Define a 2-form $dx\,dy$ on $\mathbb{R}^3$ by setting

$$(dx\,dy)_p(v, w) = \det \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}.$$

Thus $(dx\,dy)_p(v, w)$ tells us the oriented area of the parallelogram obtained by projecting v and w to the xy-plane. This inherits properties (i)-(iii) from the fact that determinant satisfies these properties.

Likewise we can define 2-forms $dx\,dz$ and $dy\,dz$ by

$$(dx\,dz)_p(v, w) = \det \begin{pmatrix} v_1 & w_1 \\ v_3 & w_3 \end{pmatrix},$$
$$(dy\,dz)_p(v, w) = \det \begin{pmatrix} v_2 & w_2 \\ v_3 & w_3 \end{pmatrix}.$$

These compute the oriented area of the parallelogram obtained by projecting v and w to the xz-plane and the yz-plane respectively.

Example 85. Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. Think of F as the describing the velocity of a fluid at each point in space. For example, the fluid could be air and then F would say which way the wind is blowing at each point of space.

Now fix some parallelogram. In a small instant of time, some amount of fluid flows through this parallelogram. The amount of fluid flowing through is proportional to $(F \cdot \vec{n})\,\mathrm{Area}(P)$ where $\vec{n}$ is the normal vector to the parallelogram (oriented according to the right hand rule). Notice that $(F(p)\cdot\vec{n})\,\mathrm{Area}(P)$ is also equal to the oriented volume of the parallelepiped spanned by v, w, and $F(p)$. We call this quantity the flux of F through P.

The flux form ω_F associated to F is the 2-form

$$(\omega_F)_p(v, w) = \det(v, w, F(p)) = \det \begin{pmatrix} v_1 & w_1 & F_1(p) \\ v_2 & w_2 & F_2(p) \\ v_3 & w_3 & F_3(p) \end{pmatrix}$$
$$= F_1(p) \det \begin{pmatrix} v_2 & w_2 \\ v_3 & w_3 \end{pmatrix} - F_2(p) \det \begin{pmatrix} v_1 & w_1 \\ v_3 & w_3 \end{pmatrix} + F_3(p) \det \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}.$$

Again this satisfies the scaling, additivity, and alternation properties because determinant satisfies these properties. Also, notice that we can write $\omega_F = F_1\, dy\, dz - F_2\, dxdz + F_3\, dx\, dy$. Later we'll see that every 2-form on $\mathbb{R}^3$ can be expressed in the form

$$f_1\, dx\, dy + f_2\, dx\, dz + f_3\, dy\, dz$$

for some functions $f_1, f_2, f_3\colon \mathbb{R}^3 \to \mathbb{R}$.

Example 86. Consider the vector field F given by

$$F(x, y, z) = \frac{(x, y, z)}{\|(x, y, z)\|^3}.$$

This is defined on $\mathbb{R}^3 \setminus \{0\}$ and represents the electric field generated by a point charge placed at the origin.

Now suppose Σ is a closed surface in $\mathbb{R}^3 \setminus \{0\}$. Then Gauss's law from physics says that the flux of the electric field over Σ depends only on the net charge enclosed by Σ. We'll prove this later. But, taking it as a given for now, note that this says that the flux of F over Σ is 0 is Σ does not enclose the origin, and the flux of F over Σ is 4π if Σ does enclose the origin. Hence if we let α_F be the flux form for F, we see a similar sort of behavior to the rotation form in $\mathbb{R}^2 \setminus \{0\}$. Namely, the integral of α_F over a closed surface Σ can only take a discrete set of values. Moreover, this value depends only on whether or not Σ encloses the origin.

Example 87. Think of $\mathbb{R}^4$ as $\mathbb{R}^2 \times \mathbb{R}^2$ with coordinates x_1, x_2, y_1, y_2. We can think of a point in $\mathbb{R}^4$ as recording the position and momentum of a particle moving in $\mathbb{R}^2$. The x coordinates tell us the position, and the y coordinates tell us the momentum.

On this $\mathbb{R}^4$, define a 2-form

$$\omega = dx_1\, dy_1 + dx_2\, dy_2.$$

So given a parallelogram in $\mathbb{R}^4$, ω tells us the oriented area of the projection to the x_1y_1-plane plus the oriented area of the projection to the x_2y_2-plane. This form ω is called the symplectic form. It plays a very important role in Hamiltonian mechanics.

7.2 Writing k-forms in coordinates

Next we'll derive a general formula for k-forms in $\mathbb{R}^n$.

Definition 88. A multi-index of length k in $\mathbb{R}^n$ is a list $(i_1, \ldots, i_k)$ consisting of k integer entries each between 1 and n. Often we write $I = (i_1, \ldots, i_k)$ for a multi-index.

Notation. Given a multi-index $I = (i_1, \ldots, i_k)$ we write dx_I as an abbreviation for $dx_{i_1} \ldots dx_{i_k}$.

Example 89. If we're working in $\mathbb{R}^5$ and $I = (1, 5, 2)$ then dx_I means $dx_1\, dx_5\, dx_2$.

Definition 90. Let $I = (i_1, \ldots, i_k)$ be a multi-index. Then dx_I is the k-form on $\mathbb{R}^n$ defined by

$$(dx_I)_p(v^1, \ldots, v^k) = \det \begin{pmatrix} v^1_{i_1} & v^2_{i_1} & \cdots & v^k_{i_k} \\ v^1_{i_2} & v^2_{i_2} & & \vdots \\ \vdots & & \ddots & \\ v^1_{i_k} & \cdots & & v^k_{i_k} \end{pmatrix}.$$

In other words, dx_I projects $v^1, \ldots, v^k$ to the $x_{i_1} \ldots x_{i_k}$-plane and then computes the oriented volume of the projection.

Notice that if I contains a repeated index, then $dx_I = 0$. Also if I is a multi-index and J is obtained from I by swapping a single pair of indices then $dx_I = -dx_J$.

Example 91. For each multi-index I of length k, let $f_I \colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Then

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

is a k-form on $\mathbb{R}^n$. By the previous observation, we can always rewrite the previous sum to be taken only over increasing multi-indices: those I such that $i_1 < i_2 < \ldots < i_k$.

Definition 92. A multi-index $I = (i_1, \ldots, i_k)$ is called increasing provided $i_1 < i_2 < \ldots < i_k$.

Example 93. The multi-index $(2, 4, 8)$ is increasing, but $(2, 1, 3)$ and $(2, 2, 2)$ are not increasing.

Proposition 94. *On $\mathbb{R}^n$ there are $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ increasing multi-indices of length k.*

Proof. There are $n(n-1)(n-2)\cdots(n-k+1)$ ways to select an ordered list of k numbers from $\{1, \ldots, n\}$ without replacements. Of these lists, only one out of every $k!$ is in increasing order. Therefore, the total number of increasing lists is

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

as claimed. $\square$

Example 95. When $n = 3$ and $k = 2$ we have $\binom{3}{2} = 3$. The three increasing multi-indices of length 2 are $(1, 2)$, $(1, 3)$, and $(2, 3)$.

Example 96. When $n = 4$ and $k = 2$ we have $\binom{4}{2} = 6$. The six increasing multi-indices of length 2 are $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 3)$, $(2, 4)$, and $(3, 4)$.

Lemma 97. *Let $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$ be increasing multi-indices. Let $e^1, \ldots, e^n$ be the standard basis vectors in $\mathbb{R}^n$. Then*

$$dx_I(e^{j_1}, \ldots, e^{j_k}) = \begin{cases} 1, & \text{if } I = J \\ 0, & \text{if } I \neq J. \end{cases}$$

Proof. Note that

$$e^{\ell}_m = \begin{cases} 1, & \text{if } \ell = m \\ 0, & \text{if } \ell \neq m. \end{cases}$$

First suppose that $I = J$. Then $i_\ell = j_\ell$ for all ℓ and so

$$\begin{aligned} dx_I(e^{j_1}, \ldots, e^{j_k}) = dx_I(e^{i_1}, \ldots, e^{i_k}) &= \det \begin{pmatrix} e^{i_1}_{i_1} & e^{i_2}_{i_1} & \cdots & e^{i_k}_{i_1} \\ e^{i_1}_{i_2} & e^{i_2}_{i_2} & & \\ \vdots & & \ddots & \\ e^{i_1}_{i_k} & \cdots & & e^{i_k}_{i_k} \end{pmatrix} \\ &= \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & \cdots & & 1 \end{pmatrix} = 1. \end{aligned}$$

On the other hand, suppose that $I \neq J$. Then since I and J are increasing, there is some ℓ such that i_ℓ does not appear in J. But then the ℓth row of the matrix in the formula for $dx_I(e^{j_1}, \ldots, e^{j_k})$ is

$$\begin{pmatrix} e^{j_1}_{i_\ell} & e^{j_2}_{i_\ell} & \cdots & e^{j_k}_{i_\ell} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}$$

and hence the determinant of this matrix is 0. □

Remark 98. We can also understand this lemma geometrically. Note that $e^{j_1}, \ldots, e^{j_k}$ spans a unit cube in the $x_{j_1} \cdots x_{j_k}$-plane. If $I = J$ then we project this cube to the $x_{j_1} \cdots x_{j_k}$-plane (which changes nothing) and then we take the oriented k-dimensional volume which is 1. On the other hand, if $I \neq J$ then we project to some other coordinate plane and this collapses the cube to a lower dimensional object which has k-dimensional volume 0.

Proposition 99. *For each increasing multi-index I of length k, let a_I be a constant. If $\sum_I a_I \, dx_I = 0$, then $a_I = 0$ for all I. In other words, the forms dx_I are linearly independent.*

Proof. Fix some arbitrary increasing multi-index J. Then

$$0 = \left(\sum_I a_I \, dx_I \right) (e^{j_1}, \ldots, e^{j_k}) = \sum_I a_I \, dx_I(e^{j_1}, \ldots, e^{j_k}) = a_J.$$

Thus $a_J = 0$. But J was arbitrary, and so all the coefficients are zero. □

Proposition 100. *Assume*

$$\omega\colon \underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}$$

satisfies the scaling, additivity, and alternation properties. Then there exist constants a_I *such that* $\omega = \sum_I a_I \, dx_I$*.*

Proof. First we need to make a guess for the constants a_I. Note that if the formula $\omega = \sum_I a_I \, dx_I$ were actually valid, then a_I would be the value of ω on $(e^{i_1}, \ldots, e^{i_k})$. Based on this, we define

$$a_I = \omega(e^{i_1}, \ldots, e^{i_k}).$$

It remains to show that this works.

Define $\alpha = \sum_I a_I \, dx_I$. We want to verify that $\omega = \alpha$. Note that both ω and α satisfy the scaling, additivity, and alternation properties and hence so does their difference $\omega - \alpha$. Moreover, by the definition of the constants a_I, we see that

$$(\omega - \alpha)(e^{i_1}, \ldots, e^{i_k}) = 0$$

for all increasing multi-indices I. Actually, by the alternation property, this implies that

$$(\omega - \alpha)(e^{i_1}, \ldots, e^{i_k}) = 0$$

for all multi-indices I (not just the increasing ones).

Now consider k vectors $v_1, \ldots, v_k$. We can write

$$\begin{aligned} v_1 &= b_{11}e^1 + b_{12}e^2 + \ldots + b_{1n}e^n \\ v_2 &= b_{21}e^1 + b_{22}e^2 + \ldots + b_{2n}e^n \\ &\vdots \\ v_k &= b_{k1}e^1 + b_{k2}e^2 + \ldots + b_{kn}e^n. \end{aligned}$$

Then by scaling and additivity

$$\begin{aligned}
(\omega-\alpha)(v_1,\dots,v_k) &= (\omega-\alpha)(\sum_{i_1=1}^{n} b_{1i_1}e^{i_1}, v_2,\dots,v_k) \\
&= \sum_{i_1=1}^{n} b_{1i_1}(\omega-\alpha)(e^{i_1}, v_2,\dots,v_k) \\
&= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} b_{1i_1}b_{2i_2}(\omega-\alpha)(e^{i_1}, e^{i_2}, v_3,\dots,v_k) \\
&= \cdots \\
&= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} b_{1i_1}b_{2i_2}\cdots b_{ki_k}(\omega-\alpha)(e^{i_1}, e^{i_2},\dots,e^{i_k}).
\end{aligned}$$

But every term in this last sum vanishes, and hence ω agrees with α, as needed. $\square$

Combing the previous two propositions yields the following corollary.

Corollary 101. *Every*

$$\omega\colon \underbrace{\mathbb{R}^n\times\ldots\times\mathbb{R}^n}_{k\ times}\to\mathbb{R}$$

that satisfies the scaling, additivity, and alternation properties can be written uniquely in the form $\omega=\sum_I a_I\,dx_I$ *where the* a_I *are constants and the sum is taken over all increasing multi-indices of length* k*.*

Definition 102. A k-form ω on $\mathbb{R}^n$ is a rule assigning a number to each oriented k-dimensional parallelepiped in $\mathbb{R}^n$ which can be expressed in the form

$$\omega=\sum_I f_I\,dx_I$$

for some smooth functions $f_I\colon\mathbb{R}^n\to\mathbb{R}$.

Example 103. Every 2-form ω on $\mathbb{R}^3$ can be written in the form

$$\omega=f_1\,dx\,dy+f_2\,dx\,dz+f_3\,dy\,dz$$

for some smooth functions $f_1,f_2,f_3\colon\mathbb{R}^3\to\mathbb{R}$.

7.3 Integrating k-forms

It remains to discuss the integration of k-forms. As before, k-forms will be integrated over parameterized k-dimensional subsets of $\mathbb{R}^n$.

Definition 104. A k-cell in $\mathbb{R}^n$ is a smooth map $c\colon [a_1, b_1] \times \ldots \times [a_k, b_k] \to \mathbb{R}^n$.

Example 105. The map $c\colon [0, 2\pi] \times [0, 1] \to \mathbb{R}^3$ given by $c(\theta, h) = (\cos\theta, \sin\theta, h)$ parameterizes the side of a cylinder.

Example 106. The map $c\colon [0, 2\pi] \times [0, \pi/2] \to \mathbb{R}^3$ given by

$$c(\theta, \phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi)$$

parameterizes a hemisphere.

Definition 107. The integral of a k-form $\omega = \sum_I f_I \, dx_I$ over a k-cell $c\colon [a_1, b_1] \times \ldots \times [a_k, b_k] \to \mathbb{R}^n$ is

$$\int_c \omega = \sum_I \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} f_I(c(t_1, \ldots, t_k)) \det \begin{pmatrix} \frac{\partial c_{i_1}}{\partial t_1} & \cdots & \frac{\partial c_{i_1}}{\partial t_k} \\ \vdots & & \vdots \\ \frac{\partial c_{i_k}}{\partial t_1} & \cdots & \frac{\partial c_{i_k}}{\partial t_k} \end{pmatrix} dt_1 \ldots dt_k.$$

Remark 108. Notice that

$$\sum_I f_I(c(t_1, \ldots, t_k)) \det \begin{pmatrix} \frac{\partial c_{i_1}}{\partial t_1} & \cdots & \frac{\partial c_{i_1}}{\partial t_k} \\ \vdots & & \vdots \\ \frac{\partial c_{i_k}}{\partial t_1} & \cdots & \frac{\partial c_{i_k}}{\partial t_k} \end{pmatrix} = \omega_{c(t_1, \ldots, t_k)} \left(\frac{\partial c}{\partial t_1}, \ldots, \frac{\partial c}{\partial t_k} \right).$$

Hence we could equivalently write

$$\int_c \omega = \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} \omega_{c(t_1, \ldots, t_k)} \left(\frac{\partial c}{\partial t_1}, \ldots, \frac{\partial c}{\partial t_k} \right) dt_1 \ldots dt_k.$$

Again the motivation for this definition is very similar to what we have already seen in the case of 1-forms and n-forms. We will describe it only in the case $k = 2$ and $n = 3$ for simplicity. Consider a map $c\colon [0, 1]^2 \to \mathbb{R}^3$. Pick partitions $0 = s_0 < s_1 < \ldots, s_\ell = 1$ and $0 = t_0 < t_1 < \ldots < t_m = 1$. These induce a partition of $[0, 1]^2$ into small squares S_{ij}. The sets $c(S_{ij})$ cut the image of c into small pieces. As before, each set $c(S_{ij})$ is very nearly a parallelogram P_{ij} based at $c(s_{i-1}, t_{j-1})$ with sides $\frac{\partial c}{\partial s}(s_{i-1}, t_{j-1})$ and $\frac{\partial c}{\partial t}(s_{i-1}, t_{j-1})$.

The collection of parallelograms P_{ij}, $1 \le i \le \ell$, $1 \le j \le m$ therefore forms a good approximation to the image of c. Let $\omega = f\,dx\,dy + g\,dx\,dz + h\,dy\,dz$ be a 2-form on $\mathbb{R}^3$. To integrate ω over c, we evaluate ω on each of the parallelograms P_{ij} and add up the results to form a quantity $\sum_{i=1}^{\ell}\sum_{j=1}^{m}\omega(P_{ij})$. We then try to take a limit of this quantity as our approximation gets better and better. Now we have

$$\begin{aligned}
&\sum_{i=1}^{\ell}\sum_{j=1}^{m}\omega(P_{ij})\\
&\quad=\sum_{i=1}^{\ell}\sum_{j=1}^{m}(s_i-s_{i-1})(t_j-t_{j-1})\omega_{c(s_{i-1},t_{j-1})}\left(\frac{\partial c}{\partial s}(s_{i-1},t_{j-1}),\frac{\partial c}{\partial t}(s_{i-1},t_{j-1})\right)\\
&\quad=\sum_{i=1}^{\ell}\sum_{j=1}^{m}(s_i-s_{i-1})(t_j-t_{j-1})f(c(s_{i-1},t_{j-1}))\begin{vmatrix}\frac{\partial c_1}{\partial s} & \frac{\partial c_1}{\partial t}\\ \frac{\partial c_2}{\partial s} & \frac{\partial c_2}{\partial t}\end{vmatrix}(s_{i-1},t_{j-1})\\
&\qquad+\sum_{i=1}^{\ell}\sum_{j=1}^{m}(s_i-s_{i-1})(t_j-t_{j-1})g(c(s_{i-1},t_{j-1}))\begin{vmatrix}\frac{\partial c_1}{\partial s} & \frac{\partial c_1}{\partial t}\\ \frac{\partial c_3}{\partial s} & \frac{\partial c_3}{\partial t}\end{vmatrix}(s_{i-1},t_{j-1})\\
&\qquad+\sum_{i=1}^{\ell}\sum_{j=1}^{m}(s_i-s_{i-1})(t_j-t_{j-1})h(c(s_{i-1},t_{j-1}))\begin{vmatrix}\frac{\partial c_2}{\partial s} & \frac{\partial c_2}{\partial t}\\ \frac{\partial c_3}{\partial s} & \frac{\partial c_3}{\partial t}\end{vmatrix}(s_{i-1},t_{j-1}).
\end{aligned}$$

But this is exactly a Riemann sum for the integral

$$\begin{aligned}
&\int_0^1\int_0^1 f(c(s,t))\det\begin{pmatrix}\frac{\partial c_1}{\partial s} & \frac{\partial c_1}{\partial t}\\ \frac{\partial c_2}{\partial s} & \frac{\partial c_2}{\partial t}\end{pmatrix}ds\,dt+\int_0^1\int_0^1 g(c(s,t))\det\begin{pmatrix}\frac{\partial c_1}{\partial s} & \frac{\partial c_1}{\partial t}\\ \frac{\partial c_3}{\partial s} & \frac{\partial c_3}{\partial t}\end{pmatrix}ds\,dt\\
&\qquad+\int_0^1\int_0^1 h(c(s,t))\det\begin{pmatrix}\frac{\partial c_2}{\partial s} & \frac{\partial c_2}{\partial t}\\ \frac{\partial c_3}{\partial s} & \frac{\partial c_3}{\partial t}\end{pmatrix}ds\,dt.
\end{aligned}$$

Hence the quantity $\sum_{i=1}^{\ell}\sum_{j=1}^{m}\omega(P_{ij})$ converges to this integral as the approximation gets better and better.

Example 109. Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field and let Σ be a surface in $\mathbb{R}^3$. Choose a parameterization c of Σ and let α_F be the flux form for F. Then c gives rise to an approximation of Σ by parallelograms P_{ij} as described above. For each i and j, we know that $\alpha_F(P_{ij})$ is the flux of F over P_{ij} and hence

$$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\alpha_F(P_{ij})$$

is the net flux of F over this entire collection of parallelograms. In the limit, as this collection of parallelograms becomes a better and better approximation to Σ, this should converge to the flux of F over Σ. Thus the integral $\int_c \alpha_F$ computes the flux of F over Σ.

Example 110. Let Σ be the boundary of the solid cylinder

$$\{(x, y, z) : x^2 + y^2 \leq 9,\ -2 \leq z \leq 5\}.$$

Let $F(x, y, z) = (x, y, z)$. Let's use differential forms to find the flux of F through Σ with respect to the outward pointing normal vector.

First, recall that the flux form associated to F is the 2-form

$$\alpha_F = x\,dy\,dz - y\,dx\,dz + z\,dx\,dy.$$

In order to integrate this over Σ, we need to parameterize Σ. We'll parameterize the top, bottom, and sides separately. Here we need to take some care to make sure we get the orientations correct.

We start with the top of the cylinder. This can be parameterized by $c\colon [0, 3] \times [0, 2\pi] \to \mathbb{R}^3$ where

$$c(r, \theta) = (r\cos\theta, r\sin\theta, 5).$$

We have

$$\begin{pmatrix} \frac{\partial c}{\partial r} & \frac{\partial c}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \\ 0 & 0 \end{pmatrix}.$$

Hence we get

$$\begin{aligned} \int_c \alpha_F = \int_0^3 \int_0^{2\pi} & r\cos\theta \det\begin{pmatrix} \sin\theta & r\cos\theta \\ 0 & 0 \end{pmatrix} - r\sin\theta \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ 0 & 0 \end{pmatrix} \\ & + 5\det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} dr\,d\theta \\ = \int_0^3 \int_0^{2\pi} & 5r\,dr\,d\theta = 45\pi. \end{aligned}$$

As a sanity check, note that this part of the cylinder is horizontal, so it makes sense that we only see the z component of F in the calculation. Also, the flux of F through the top of the cylinder is definitely positive, so we got the orientation correct.

Next we deal with the bottom of the cylinder. This can be parameterized by $c\colon [0,3]\times[0,2\pi]\to\mathbb{R}^3$ where

$$c(\theta,r) = (r\cos(-\theta), r\sin(-\theta), -2).$$

We have

$$\begin{pmatrix}\frac{\partial c}{\partial r} & \frac{\partial c}{\partial \theta}\end{pmatrix} = \begin{pmatrix}\cos(-\theta) & r\sin(-\theta)\\ \sin(-\theta) & -r\cos(-\theta)\\ 0 & 0\end{pmatrix}.$$

Hence we get

$$\begin{aligned}\int_c \alpha_F = \int_0^3\int_0^{2\pi} & r\cos(-\theta)\det\begin{pmatrix}\sin(-\theta) & -r\cos(-\theta)\\ 0 & 0\end{pmatrix}\\ &- r\sin(-\theta)\det\begin{pmatrix}\cos(-\theta) & r\sin(-\theta)\\ 0 & 0\end{pmatrix}\\ &- 2\det\begin{pmatrix}\cos(-\theta) & r\sin(-\theta)\\ \sin(-\theta) & -r\cos(-\theta)\end{pmatrix}\,dr\,d\theta\\ = \int_0^3\int_0^{2\pi} & (-2)(-r)\,dr\,d\theta = 18\pi.\end{aligned}$$

Again the flux through the bottom of the cylinder is definitely positive, so we got the orientation correct.

Finally we handle the sides. These can be parameterized by $c\colon [0,2\pi]\times[-2,5]\to\mathbb{R}^3$ where

$$c(\theta,h) = (3\cos\theta, 3\sin\theta, h).$$

We have

$$\begin{pmatrix}\frac{\partial c}{\partial \theta} & \frac{\partial c}{\partial h}\end{pmatrix} = \begin{pmatrix}-3\sin\theta & 0\\ 3\cos\theta & 0\\ 0 & 1\end{pmatrix}.$$

It follows that

$$\begin{aligned}\int_c \alpha_F = \int_0^{2\pi}\int_{-2}^5 & 3\cos\theta\det\begin{pmatrix}3\cos\theta & 0\\ 0 & 1\end{pmatrix} - 3\sin\theta\det\begin{pmatrix}-3\sin\theta & 0\\ 0 & 1\end{pmatrix}\\ &+ h\det\begin{pmatrix}-3\sin\theta & 0\\ 3\cos\theta & 0\end{pmatrix}\,d\theta\,dh\\ = \int_0^{2\pi}\int_{-2}^5 & 9\cos^2\theta + 9\sin^2\theta\,d\theta\,dh = 126\pi.\end{aligned}$$

Finally we get that the total flux of F over Σ is $45\pi + 18\pi + 126\pi = 189\pi$.

8 Stokes' Theorem

Stokes' Theorem is a vast generalization of the fundamental theorem of calculus. It includes all of the classical theorems of vector calculus as special cases. Consider a k-form ω on $\mathbb{R}^n$ and a $(k+1)$-cell c in $\mathbb{R}^n$. Stokes' Theorem says that

$$\int_{\partial c} \omega = \int_c d\omega.$$

Here ∂c is the boundary of c and $d\omega$ is the exterior derivative of ω. In order for this formula to make sense, we need to define both of these objects.

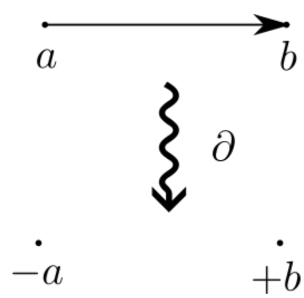
8.1 Classical Theorems of Vector Calculus

Before defining these objects in general, let's look at some familiar examples.

Example 111. The usual fundamental theorem of calculus says that

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

To interpret this in the above language, think of the function f as a zero form. It assigns values to 0-dimensional parallelograms (i.e. points) in $\mathbb{R}^n$. The differential of f is the 1-form $df = f'(x)\, dx$. Let $c\colon [a,b] \to [a,b]$ be given by $c(t) = t$. Then the boundary of c consists of the two points a and b. To remember the orientation of c, we assign the point b a $+$ sign and the point a a $-$ sign.



In this set up, the above formula becomes

$$\int_c df = \int_{\partial c} f,$$

where to integrate a 0-form over a collection of points with signs, we just add up plus or minus the values of the function at the points.

Example 112. We've already seen a slight generalization of this: if $f\colon \mathbb{R}^n \to \mathbb{R}$ and γ is a curve in $\mathbb{R}^n$ then

$$\int_\gamma df = f(\gamma(b)) - f(\gamma(a)) = \int_{\partial\gamma} f,$$

where $\partial\gamma$ consists of the point $\gamma(b)$ with a plus sign and the point $\gamma(a)$ with a minus sign.

Example 113 (Green's Theorem)**.** Let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field. Let Ω be a nice region in the plane. Then

$$\int_{\partial\Omega} F \cdot d\vec{\ell} = \int_\Omega \operatorname{circ}(F)\, dA.$$

Here the integral on the left is the line integral of F over $\partial\Omega$, and

$$\operatorname{circ}(F) = \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}$$

denotes the circulation of F. This formula fits into the general paradigm of Stokes' theorem: it relates the integral of F over the boundary of a region, with the integral of some derivatives of F over the region itself.

Example 114 (The Divergence Theorem)**.** Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. Let Ω be a nice region in 3-space. Then

$$\int_{\partial\Omega} F \cdot \nu\, dA = \int_\Omega \operatorname{div}(F)\, dV.$$

Here ν denotes the outward pointing unit normal to $\partial\Omega$, and

$$\operatorname{div}(F) = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}$$

is the divergence of F. Again this formula relates an integral of F over the boundary of a region with an integral of some derivatives of F over the region itself.

Example 115 (Classical Stokes' Theorem)**.** Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field and let Σ be an oriented surface in $\mathbb{R}^3$ with boundary $\partial\Sigma$. Then

$$\int_{\partial\Sigma} F \cdot d\vec{\ell} = \int_\Sigma \operatorname{curl}(F) \cdot \nu\, dA.$$

Here ν is a unit normal vector to Σ and

$$\operatorname{curl}(F) = \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right)$$

is the curl of F. This formula relates an integral of F over the boundary of a surface with the integral of some derivatives of F over the surface itself.

8.2 Motivating the Exterior Derivative

To motivate the definition of the exterior derivative, let's investigate the relationship between the "ω" and the "$d\omega$" in each of the above examples.

Example 116. Let $f\colon \mathbb{R}^n \to \mathbb{R}$. Then

$$df = \frac{\partial f}{\partial x_1}\,dx_1 + \ldots + \frac{\partial f}{\partial x_n}\,dx_n$$

and $df_p(v) = \nabla f(p) \cdot v = \frac{d}{dt}\big|_{t=0} f(p+tv)$. Let γ_ε be the line segment from p to $p + \varepsilon v$ so that $\partial\gamma_\varepsilon$ consists of the point $p + \varepsilon v$ with a $+$ sign and the point p with a minus sign. Then

$$\int_{\partial\gamma_\varepsilon} f = f(p+\varepsilon v) - f(p)$$

and so

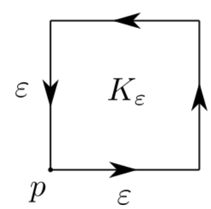
$$df_p(v) = \frac{d}{dt}\bigg|_{t=0} f(p+tv) = \lim_{\varepsilon\to 0} \frac{f(p+\varepsilon v) - f(p)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\partial\gamma_\varepsilon} f.$$

Thus df_p tells us about the integral of f over the boundary of very small line segments based at p.

Example 117. Let $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ and let Ω be a nice region of $\mathbb{R}^2$. We can use Green's theorem

$$\int_{\partial\Omega} F \cdot d\vec{\ell} = \int_\Omega \mathrm{circ}(F)\,dA$$

to derive a formula for the circulation of F as follows. Fix a point $p \in \mathbb{R}^2$ and let K_ε be a square based at p with side length ε.



Note that since F is smooth, the circulation of F will be approximately equal to $\mathrm{circ}(F)(p)$ over all of K_ε. Thus we get an approximation

$$\int_{\partial K_\varepsilon} F \cdot d\vec{\ell} = \int_{K_\varepsilon} \mathrm{circ}(F)\,dA \approx \varepsilon^2 \mathrm{circ}(F)(p).$$

This suggests the following formula for circulation:

$$\operatorname{circ}(F)(p) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \int_{\partial K_\varepsilon} F \cdot d\vec{\ell}.$$

Thus geometrically, $\operatorname{circ}(F)$ tells us about the line integral of F around very small loops based at p. In other words, it tells us about the infinitesimal circulation of F at p. Note also the formal similarity to the previous formula for df.

Example 118. Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ and let Ω be a nice region in $\mathbb{R}^3$. Then the divergence theorem says

$$\int_{\partial\Omega} F \cdot \nu \, dA = \int_{\Omega} \operatorname{div}(F) \, dV.$$

As above, we can use this to derive a formula for divergence. Indeed, fix a point $p \in \mathbb{R}^3$ and let K_ε be a cube based at p with side length ε. Then since F is smooth, $\operatorname{div}(F)$ is approximately equal to $\operatorname{div}(F)(p)$ on all of K_ε. Therefore, we get an approximation

$$\int_{\partial K_\varepsilon} F \cdot \nu \, dA = \int_{K_\varepsilon} \operatorname{div}(F) \, dV \approx \varepsilon^3 \operatorname{div}(F)(p).$$

This suggests that

$$\operatorname{div}(F)(p) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^3} \int_{\partial K_\varepsilon} F \cdot \nu \, dA.$$

Thus $\operatorname{div}(F)$ tells us about the flux of F through very tiny cubes based at p. Again, note the similarity of this formula to the previous ones.

Example 119. Finally consider the case of the classical Stokes' theorem. Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ and let Σ be a surface in $\mathbb{R}^3$ with boundary $\partial\Sigma$. Then Stokes' theorem says

$$\int_{\partial\Sigma} F \cdot d\vec{\ell} = \int_{\Sigma} \text{curl(F)} \cdot \nu \, dA.$$

Now fix a point $p \in \mathbb{R}^3$, and let K_ε^{xy} be a small square based at p, parallel to the xy-plane, with side length ε. Again $\operatorname{curl}(F)$ is approximately equal to $\operatorname{curl}(F)(p)$ on all of K_ε^{xy} and hence we get an approximation

$$\int_{\partial K_\varepsilon^{xy}} F \cdot d\vec{\ell} = \int_{K_\varepsilon^{xy}} \operatorname{curl}(F) \cdot \vec{k} \, dA \approx \varepsilon^2 \left[\operatorname{curl}(F)(p) \cdot \vec{k} \right].$$

This suggests the formula

$$\operatorname{curl}(F)(p) \cdot \vec{k} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \int_{\partial K_\varepsilon^{xy}} F \cdot d\vec{\ell}.$$

Hence the z-component of curl tells us about the line integral of F over very small squares perpendicular to the z-axis. We could repeat the argument with small squares perpendicular to the x-axis and y-axis to get the x and y components of curl, respectively.

8.3 The Boundary Operator

The point of all these examples is that in each case the "$d\omega$" tells us about the integral of the "ω" over the boundary of very small line segments, or squares, or cubes, etc. This motivates the definition of $d\omega$. Given a k-form ω, the exterior derivative $d\omega$ will be a $(k+1)$-form $d\omega$ such that $d\omega(v_1, \ldots, v_{k+1})$ tells us about the integral of ω over the boundary of the parallelepiped spanned by $\varepsilon v_1, \ldots, \varepsilon v_{k+1}$ when ε is small. In order to formalize this, we first need to discuss boundaries.

Definition 120. A k-chain in $\mathbb{R}^n$ is a formal sum $a_1 c_1 + \ldots + a_m c_m$ where each c_i is a k-cell and each a_i is an integer.

Definition 121. Let ω be a k-form on $\mathbb{R}^n$ and let $a_1 c_1 + \ldots + a_m c_m$ be a k-chain. Then we define

$$\int_{a_1 c_1 + \ldots + a_m c_m} \omega = a_1 \int_{c_1} \omega + \ldots + a_m \int_{c_m} \omega.$$

Remark 122. One can think of the k-chain $2c$ as two copies of the cell c, and the k-chain $-c$ as the cell c but with the opposite orientation.

Definition 123. Let $c \colon [0,1]^k \to \mathbb{R}^n$ be a k-cell. The faces of c are the following collection of maps:

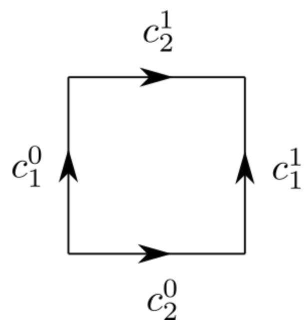
$$\begin{aligned} c_i^1(t_1, \ldots, t_{k-1}) &= c(t_1, \ldots, t_{i-1}, 1, t_i, \ldots, t_{k-1}), \quad i = 1, \ldots, k \\ c_i^0(t_1, \ldots, t_{k-1}) &= c(t_1, \ldots, t_{i-1}, 0, t_i, \ldots, t_{k-1}), \quad i = 1, \ldots, k. \end{aligned}$$

Thus each face is a $(k-1)$-cell in $\mathbb{R}^n$, and there are $2k$ faces in total.

Example 124. Let $c\colon [0,1]^2 \to \mathbb{R}^2$ be the standard cube $c(x,y) = (x,y)$. Then the four edges of c are

$$\begin{aligned} c_1^1(t) &= c(1,t) = (1,t), \quad & c_1^0(t) &= c(0,t) = (0,t), \\ c_2^1(t) &= c(t,1) = (t,1), \quad & c_2^0(t) &= c(t,0) = (t,0). \end{aligned}$$

Note that the orientation of these edges does not agree with the standard counterclockwise orientation.



Hence, in order to get the orientation correct, the boundary of c should be the 1-chain $c_2^0 + c_1^1 - c_2^1 - c_1^0$.

Definition 125. Let $c\colon [0,1]^k \to \mathbb{R}^n$ be a k-cell. The boundary of c is the $(k-1)$-chain

$$\partial c = \sum_{i=1}^{k} \sum_{j=0}^{1} (-1)^{i+j} c_i^j.$$

Example 126. Let $c\colon [0,1]^3 \to \mathbb{R}^3$ be the standard cube $c(x,y,z) = (x,y,z)$. The six faces of c are

$$\begin{aligned} c_1^1(s,t) &= c(1,s,t) = (1,s,t), \quad & c_1^0(s,t) &= c(0,s,t) = (0,s,t), \\ c_2^1(s,t) &= c(s,1,t) = (s,1,t), \quad & c_2^0(s,t) &= c(s,0,t) = (s,0,t), \\ c_3^1(s,t) &= c(s,t,1) = (s,t,1), \quad & c_3^0(s,t) &= c(s,t,0) = (s,t,0). \end{aligned}$$

The boundary of c is the 2-chain $c_1^1 - c_1^0 - c_2^1 + c_2^0 + c_3^1 - c_3^0$. One can check that this choice of signs makes the unit normal associated to each face by the right hand rule point outward.

8.4 Definition of the Exterior Derivative

We can now define the exterior derivative.

Definition 127. Let ω be a k-form on $\mathbb{R}^n$. Fix a point $p \in \mathbb{R}^n$ and let $v_1, \ldots, v_{k+1}$ be vectors in $\mathbb{R}^n$. Let $P_\varepsilon \colon [0,1]^{k+1} \to \mathbb{R}^n$ be given by

$$P_\varepsilon(t_1, \ldots, t_{k+1}) = p + \varepsilon v_1 + \ldots + \varepsilon v_{k+1}.$$

Then, by definition,

$$(d\omega)_p(v_1, \ldots, v_{k+1}) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^{k+1}} \int_{\partial P_\varepsilon} \omega,$$

provided that this limit exists. In other words, "$d\omega$ is the thing that makes Stokes' theorem true infinitesimally."

Remark 128. We will refer to this as the geometric formula for $d\omega$. It makes it clear what $d\omega$ is actually computing. On the other hand, this formula is very cumbersome computationally. In each of the model examples we've seen, there are nicer computational formulas for $d\omega$ involving some partial derivatives:

$$\operatorname{circ}(F) = \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y},$$

$$\operatorname{div}(F) = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z},$$

$$\operatorname{curl}(F) = \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right).$$

Fortunately, it turns out that this is always the case. The above limit can always be rewritten in terms of a certain combination of partial derivatives.

Theorem 129. *Let $\omega = \sum_I f_I \, dx_I$ be a k-form. Then the exterior derivative of ω is the $(k+1)$-form $d\omega = \sum_I (df_I) \, dx_I$.*

Before proving the theorem, we note that this formula is true in each of our model examples.

Example 130. Let $F \colon \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field and consider the work form $\omega_F = F_1 \, dx + F_2 \, dy$. Then the above formula says

$$\begin{aligned} d\omega_F &= (dF_1) \, dx + (dF_2) \, dy \\ &= \left(\frac{\partial F_1}{\partial x} \, dx + \frac{\partial F_1}{\partial y} \, dy \right) dx + \left(\frac{\partial F_2}{\partial x} \, dx + \frac{\partial F_2}{\partial y} \, dy \right) dy \\ &= \frac{\partial F_1}{\partial x} \, dx \, dx + \frac{\partial F_1}{\partial y} \, dy \, dx + \frac{\partial F_2}{\partial x} \, dx \, dy + \frac{\partial F_2}{\partial y} \, dy \, dy \\ &= \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx \, dy = \operatorname{circ}(F) \, dx \, dy. \end{aligned}$$

Hence we can write Green's theorem as $\int_{\partial\Omega} \omega_F = \int_\Omega d\omega_F$.

Example 131. Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field and consider the flux form $\alpha_F = F_1\,dy\,dz - F_2\,dx\,dz + F_3\,dx\,dy$. The above formula says

$$\begin{aligned}
d\alpha_F &= (dF_1)\,dy\,dz - (dF_2)\,dx\,dz + (dF_3)\,dx\,dy \\
&= \left(\frac{\partial F_1}{\partial x}\,dx + \frac{\partial F_1}{\partial y}\,dy + \frac{\partial F_1}{\partial z}\,dz\right)\,dy\,dz \\
&\quad - \left(\frac{\partial F_2}{\partial x}\,dx + \frac{\partial F_2}{\partial y}\,dy + \frac{\partial F_2}{\partial z}\,dz\right)\,dx\,dz \\
&\quad + \left(\frac{\partial F_3}{\partial x}\,dx + \frac{\partial F_3}{\partial y}\,dy + \frac{\partial F_3}{\partial z}\,dz\right)\,dx\,dy \\
&= \frac{\partial F_1}{\partial x}\,dx\,dy\,dz - \frac{\partial F_2}{\partial y}\,dy\,dx\,dz + \frac{\partial F_3}{\partial z}\,dz\,dx\,dy \\
&= \operatorname{div}(F)\,dx\,dy\,dz.
\end{aligned}$$

Hence we can write the divergence theorem as $\int_{\partial\Omega} \alpha_F = \int_{\Omega} d\alpha_F$.

Example 132. Let $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field and consider the work form $\omega_F = F_1\,dx + F_2\,dy + F_3\,dz$. The above formula says

$$\begin{aligned}
d\omega_F &= (dF_1)\,dx - (dF_2)\,dy + (dF_3)\,dz \\
&= \left(\frac{\partial F_1}{\partial x}\,dx + \frac{\partial F_1}{\partial y}\,dy + \frac{\partial F_1}{\partial z}\,dz\right)\,dx \\
&\quad + \left(\frac{\partial F_2}{\partial x}\,dx + \frac{\partial F_2}{\partial y}\,dy + \frac{\partial F_2}{\partial z}\,dz\right)\,dy \\
&\quad + \left(\frac{\partial F_3}{\partial x}\,dx + \frac{\partial F_3}{\partial y}\,dy + \frac{\partial F_3}{\partial z}\,dz\right)\,dz \\
&= \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\,dx\,dy + \left(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}\right)\,dx\,dz + \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\,dy\,dz.
\end{aligned}$$

But this is exactly the flux form for $\operatorname{curl}(F)$. Hence we can write the classical Stokes' theorem as $\int_{\partial\Sigma} \omega_F = \int_{\Sigma} d\omega_F$.

We now return to the proof of Theorem 129. For the proof, we need to use the multivariable Taylor theorem.

Theorem 133 (Taylor)**.** *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Then for any p and v in $\mathbb{R}^n$ we have*

$$f(p+v) = f(p) + \nabla f(p)\cdot v + \mathit{error}$$

where $|\mathit{error}| \leq C\|v\|^2$. Here C is a positive constant.

Proof. (Theorem 129) For simplicity, we give the proof only in the case $k = 1$ and $n = 2$. Assume that $\omega = f\,dx + g\,dy$ is a 1-form on $\mathbb{R}^2$ and fix vectors $v, w \in \mathbb{R}^2$. Let P_ε be a parallelogram based at p with sides εv and εw. We need to evaluate the limit

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \int_{\partial P_\varepsilon} \omega.$$

First observe that $\partial P_\varepsilon = \gamma_1 + \gamma_2 - \gamma_3 - \gamma_4$ where $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \colon [0,1] \to \mathbb{R}^2$ are given by

$$\begin{aligned} \gamma_1(t) &= p + \varepsilon t v, & \gamma_3(t) &= p + \varepsilon w + \varepsilon t v \\ \gamma_2(t) &= p + \varepsilon v + \varepsilon t w, & \gamma_4(t) &= p + \varepsilon t w. \end{aligned}$$

We will use Taylor's theorem to estimate the integral of ω over each of these four segments.

Observe that

$$\int_{\gamma_1} \omega = \int_0^1 f(p + \varepsilon t v)\varepsilon v_1 + g(p + \varepsilon t v)\varepsilon v_2 \, dt.$$

Now Taylor's theorem says

$$\begin{aligned} f(p + \varepsilon t v) &= f(p) + \varepsilon t \nabla f(p) \cdot v + \text{error} \\ g(p + \varepsilon t v) &= g(p) + \varepsilon t \nabla g(p) \cdot v + \text{error} \end{aligned}$$

where $|\text{error}| \leq C t^2 \varepsilon^2 \|v\|^2 \leq C\varepsilon^2$. Using these formulas to estimate the integral, we get

$$\int_{\gamma_1} \omega = \varepsilon v_1 f(p) + \frac{\varepsilon^2 v_1}{2} \nabla f(p) \cdot v + \varepsilon v_2 g(p) + \frac{\varepsilon^2 v_2}{2} \nabla g(p) \cdot v + \text{error}$$

where $|\text{error}| \leq C\varepsilon^3$.

Next consider γ_3. We have

$$\int_{\gamma_3} \omega = \int_0^1 f(p + \varepsilon w + \varepsilon t v)\varepsilon v_1 + g(p + \varepsilon w + \varepsilon t v)\varepsilon v_2 \, dt.$$

Taylor expanding gives

$$\begin{aligned} f(p + \varepsilon w + \varepsilon t v) &= f(p) + \varepsilon \nabla f(p) \cdot w + \varepsilon t \nabla f(p) \cdot v + \text{error} \\ g(p + \varepsilon w + \varepsilon t v) &= g(p) + \varepsilon \nabla f(p) \cdot w + \varepsilon t \nabla g(p) \cdot v + \text{error} \end{aligned}$$

where $|\text{error}| \leq C\varepsilon^2$. Therefore

$$\int_{\gamma_3} \omega = \varepsilon v_1 f(p) + \varepsilon^2 v_1 \nabla f(p) \cdot w + \frac{\varepsilon^2 v_1}{2} \nabla f(p) \cdot v$$
$$+ \varepsilon v_2 g(p) + \varepsilon^2 v_2 \nabla g(p) \cdot w + \frac{\varepsilon^2 v_2}{2} \nabla g(p) \cdot v + \text{error}$$

where $|\text{error}| \leq C\varepsilon^3$.

Hence we get that

$$\begin{aligned}\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \left[\int_{\gamma_1} \omega - \int_{\gamma_3} \omega \right] &= \lim_{\varepsilon \to 0} \left[-v_1 \nabla f(p) \cdot w - v_2 \nabla g(p) \cdot w + \frac{\text{error}}{\varepsilon^2} \right] \\ &= -v_1 \nabla f(p) \cdot w - v_2 \nabla g(p) \cdot w\end{aligned}$$

since $|\text{error}| \leq C\varepsilon^3$. Repeating the same arguments with γ_2 and γ_4 gives that

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \left[\int_{\gamma_2} \omega - \int_{\gamma_4} \omega \right] = w_1 \nabla f(p) \cdot v + w_2 \nabla g(p) \cdot v.$$

It follows that

$$\begin{aligned}\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \int_{\partial P_\varepsilon} \omega &= w_1 \nabla f(p) \cdot v - v_1 \nabla f(p) \cdot w + w_2 \nabla g(p) \cdot v - v_2 \nabla g(p) \cdot w \\ &= w_1 \left(\frac{\partial f}{\partial x} v_1 + \frac{\partial f}{\partial y} v_2 \right) - v_1 \left(\frac{\partial f}{\partial x} w_1 + \frac{\partial f}{\partial y} w_2 \right) \\ &\quad + w_2 \left(\frac{\partial g}{\partial x} v_1 + \frac{\partial g}{\partial y} v_2 \right) - v_2 \left(\frac{\partial g}{\partial x} w_1 + \frac{\partial g}{\partial y} w_2 \right) \\ &= \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) (v_1 w_2 - v_2 w_1).\end{aligned}$$

Thus

$$d\omega = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx\, dy = (df)\, dx + (dg)\, dy,$$

as needed. $\square$

Remark 134. The proof of the theorem for arbitrary k and n is exactly the same conceptually. We Taylor expand the function f and use this to estimate the integral of ω over each face of ∂c. We then add up these estimates and observe that almost everything cancels, leaving the desired formula. The only difficulty is that the notation is much more involved.

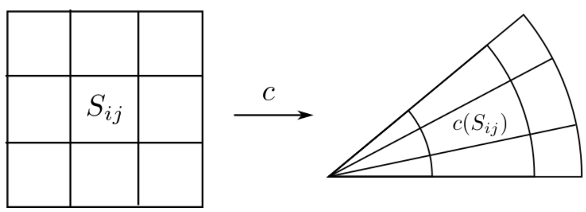
8.5 Stokes' Theorem

We close this section by giving the proof of Stokes' theorem. In fact, we'll give two proofs: a geometric "proof" based on the limit formula for the exterior derivative, and an algebraic proof based on the derivative formula for the exterior derivative. The geometric proof has the advantage that it makes it clear what is happening intuitively: Stokes' theorem is true because we defined $d\omega$ to be the object that makes Stokes' theorem true infinitesimally. However, the algebraic proof is easier to make rigorous.

Theorem 135. *Let ω be a k-form on $\mathbb{R}^n$ and let c be a $(k+1)$-cell in $\mathbb{R}^n$. Then*

$$\int_{\partial c} \omega = \int_c d\omega.$$

Proof. (Geometric "Proof") For simplicity, we'll just give the proof in the case $k = 1$. Let ω be a 1-form on $\mathbb{R}^n$ and let $c\colon [0,1]^2 \to \mathbb{R}^n$ be a 2-cell. Cut $[0,1]^2$ into m^2 squares S_{ij} of side length $1/m$.



Then the image $c(S_{ij})$ is approximately a parallelogram P_{ij} based at $c(\frac{i-1}{m}, \frac{j-1}{m})$ with sides $\frac{1}{m}\frac{\partial c}{\partial s}$ and $\frac{1}{m}\frac{\partial c}{\partial t}$.

We know that

$$\sum_{i=1}^{m}\sum_{j=1}^{m} d\omega(P_{ij}) \to \int_c d\omega \tag{2}$$

as $m \to \infty$. Now observe that

$$d\omega(P_{ij}) = \frac{1}{m^2} d\omega\left(\frac{\partial c}{\partial s}, \frac{\partial c}{\partial t}\right) = \frac{1}{m^2} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \int_{\partial Q_\varepsilon} \omega$$

where Q_ε is the parallelogram based at $c(\frac{i-1}{m}, \frac{j-1}{m})$ with sides $\varepsilon\frac{\partial c}{\partial s}$ and $\varepsilon\frac{\partial c}{\partial t}$. In particular, notice that $Q_{1/m}$ is P_{ij} and hence

$$\frac{1}{m^2} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \int_{\partial Q_\varepsilon} \omega \approx \frac{1}{m^2}\frac{1}{(1/m)^2} \int_{\partial Q_{1/m}} \omega = \int_{\partial P_{ij}} \omega$$

when m is large. Also we have

$$\int_{\partial P_{ij}} \omega \approx \int_{\partial(c|_{S_{ij}})} \omega$$

where $c|_{S_{ij}}$ denotes the restriction of c to S_{ij}.

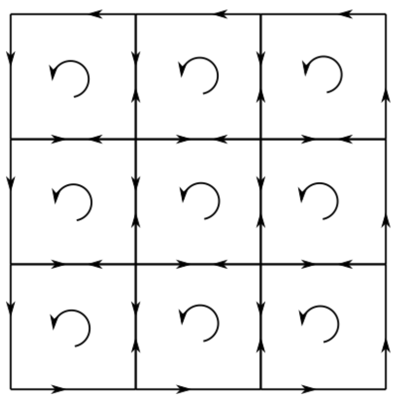
Therefore we have an approximation

$$\sum_{i=1}^{m} \sum_{j=1}^{m} d\omega(P_{ij}) \approx \sum_{i=1}^{m} \sum_{j=1}^{m} \int_{\partial(c|_{S_{ij}})} \omega, \tag{3}$$

which gets better and better as m goes to ∞. Now observe that

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \int_{\partial(c|_{S_{ij}})} \omega = \int_{\partial c} \omega \tag{4}$$

since the sum on the left integrates ω over every interior edge of our $m \times m$ grid twice with opposite orientations.



Now using equations (2), (3), (4) and letting $m \to \infty$, we get that

$$\int_{\partial c} \omega = \int_{c} d\omega,$$

as desired. $\square$

Proof. (Algebraic Proof) For simplicity, we again give the proof only in the case $k = 1$. In the general case, the notation is just more involved. So consider a 1-form $\omega = f_1\,dx_1 + \ldots + f_n\,dx_n$ on $\mathbb{R}^n$ and let $c\colon [0,1]^2 \to \mathbb{R}^n$ be an n-cell. Then

$$d\omega = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j} dx_j\,dx_i.$$

Integrating this over c yields

$$\begin{aligned}\int_c d\omega &= \sum_{i,j=1}^{n} \int_0^1\int_0^1 \frac{\partial f_i}{\partial x_j}(c(s,t))\det\begin{pmatrix} \frac{\partial c_j}{\partial s}(s,t) & \frac{\partial c_j}{\partial t}(s,t) \\ \\ \frac{\partial c_i}{\partial s}(s,t) & \frac{\partial c_i}{\partial t}(s,t)\end{pmatrix} ds\,dt \\ &= \sum_{i,j=1}^{n}\int_0^1\int_0^1 \left(\frac{\partial f_i}{\partial x_j}\circ c\right)\left(\frac{\partial c_j}{\partial s}\frac{\partial c_i}{\partial t} - \frac{\partial c_j}{\partial t}\frac{\partial c_i}{\partial s}\right) ds\,dt.\end{aligned}$$

On the other hand, let's integrate ω over the boundary of c. We have

$$\begin{aligned}\int_{\partial c}\omega = &\int_0^1 \left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial s}\right](s,0)\,ds - \int_0^1\left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial s}\right](s,1)\,ds \\ &+ \int_0^1\left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial t}\right](1,t)\,dt - \int_0^1\left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial t}\right](0,t)\,dt.\end{aligned}$$

Now by the fundamental theorem of calculus

$$\begin{aligned}&\int_0^1 \left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial s}\right](s,0) - \left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial s}\right](s,1)\,ds \\ &\quad = -\int_0^1\left(\int_0^1 \frac{d}{dt}\left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial s}\right](s,t)\,dt\right)ds \\ &\quad = -\int_0^1\int_0^1\sum_{i,j=1}^{n}\left(\frac{\partial f_i}{\partial x_j}\circ c\right)\frac{\partial c_j}{\partial t}\frac{\partial c_i}{\partial s} + (f_i\circ c)\frac{\partial^2 c_i}{\partial t\,\partial s}\,dt\,ds.\end{aligned}$$

Likewise we have

$$\begin{aligned}&\int_0^1 \left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial t}\right](1,t) - \left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial t}\right](0,t)\,dt \\ &\quad = \int_0^1\left(\int_0^1 \frac{d}{ds}\left[\sum_{i=1}^{n}(f_i\circ c)\frac{\partial c_i}{\partial t}\right](s,t)\,ds\right)dt \\ &= \int_0^1\int_0^1\sum_{i,j=1}^{n}\left(\frac{\partial f_i}{\partial x_j}\circ c\right)\frac{\partial c_j}{\partial s}\frac{\partial c_i}{\partial t} + (f_i\circ c)\frac{\partial^2 c}{\partial s\partial t}\,ds\,dt.\end{aligned}$$

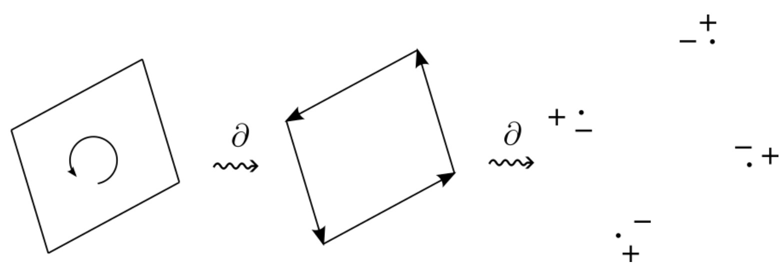
Therefore

$$\begin{aligned}\int_{\partial c}\omega &= \int_0^1\int_0^1\sum_{i,j=1}^{n}\left(\frac{\partial f_i}{\partial x_j}\circ c\right)\frac{\partial c_j}{\partial s}\frac{\partial c_i}{\partial t}+(f_i\circ c)\frac{\partial^2 c}{\partial s\partial t}\,ds\,dt\\ &\quad-\int_0^1\int_0^1\sum_{i,j=1}^{n}\left(\frac{\partial f_i}{\partial x_j}\circ c\right)\frac{\partial c_j}{\partial t}\frac{\partial c_i}{\partial s}+(f_i\circ c)\frac{\partial^2 c_i}{\partial t\,\partial s}\,dt\,ds\\ &= \int_0^1\int_0^1\sum_{i,j=1}^{n}\left(\frac{\partial f_i}{\partial x_j}\circ c\right)\left(\frac{\partial c_j}{\partial s}\frac{\partial c_i}{\partial t}-\frac{\partial c_j}{\partial t}\frac{\partial c_i}{\partial s}\right)ds\,dt.\end{aligned}$$

Hence we see that $\int_{\partial c}\omega$ and $\int_c d\omega$ are equal. $\square$

9 Closed and Exact Forms

We begin by recording some properties of the boundary and exterior derivative operators. For example, it seems intuitively clear that the boundary of a boundary should be zero. In the case of a parallelogram this is obvious. Indeed, the boundary of a parallelogram consists of four line segments.



When we take the boundary of these line segments, we see each of the four corners twice, once with a plus sign and once with a minus sign. Hence everything cancels and the boundary of the boundary is zero.

Proposition 136. *Let $c\colon [0,1]^k \to \mathbb{R}^n$ be a k-cell. Then $\partial^2 c = \partial(\partial c) = 0$.*

Proof. Writing out the formulas for ∂c and $\partial^2 c$ we get

$$\partial c = \sum_{i=1}^{k}\sum_{j=0}^{1}(-1)^{i+j}c_i^j$$

and then

$$\partial^2 c = \sum_{i=1}^{k} \sum_{j=0}^{1} \sum_{\ell=1}^{k-1} \sum_{m=0}^{1} (-1)^{i+j} (-1)^{\ell+m} (c_i^j)_\ell^m.$$

As in the case of a parallelogram, it remains to show that the terms in this sum cancel in pairs.

Consider the index set

$$\mathcal{I} = \{(i, j, \ell, m) : i \in \{1, \ldots, k\},\ \ell \in \{1, \ldots, k-1\},\ j, m \in \{0, 1\}\}.$$

We can split $\mathcal{I}$ into two pieces:

$$\begin{aligned} \mathcal{I}_1 &= \{(i, j, \ell, m) \in \mathcal{I} : i \leq \ell\}, \\ \mathcal{I}_2 &= \{(i, j, \ell, m) \in \mathcal{I} : i > \ell\}. \end{aligned}$$

Moreover, the pairing $(i, j, \ell, m) \in \mathcal{I}_1 \leftrightarrow (\ell + 1, m, i, j) \in \mathcal{I}_2$ gives a one-to-one correspondence between elements of $\mathcal{I}_1$ and elements of $\mathcal{I}_2$. Thus we can rewrite the above sum in the following way:

$$\begin{aligned} \partial^2 c &= \sum_{(i,j,\ell,m) \in \mathcal{I}} (-1)^{i+j+\ell+m} (c_i^j)_\ell^m \\ &= \sum_{(i,j,\ell,m) \in \mathcal{I}_1} \left[(-1)^{i+j+\ell+m} (c_i^j)_\ell^m + (-1)^{(\ell+1)+m+i+j} (c_{\ell+1}^m)_i^j \right]. \end{aligned}$$

Hence to show that $\partial^2 c = 0$ it's enough to check that

$$(-1)^{i+j+\ell+m} (c_i^j)_\ell^m + (-1)^{(\ell+1)+m+i+j} (c_{\ell+1}^m)_i^j = 0$$

for each $(i, j, \ell, m) \in \mathcal{I}_1$. So fix some $(i, j, \ell, m) \in \mathcal{I}_1$. Unwinding the definition of the face maps, we see that

$$\begin{aligned} (c_i^j)_\ell^m (t_1, \ldots, t_{k-2}) &= c_i^j (t_1, \ldots, t_{\ell-1}, m, t_\ell, \ldots, t_{k-2}) \\ &= c(t_1, \ldots, t_{i-1}, j, t_i, \ldots, t_{\ell-1}, m, t_\ell, \ldots, t_{k-2}), \end{aligned}$$

where we've used the fact that $i \leq \ell$ to get the second equality. Thus $(c_i^j)_\ell^m$ plugs j into the ith slot of c and plugs m into the $(\ell + 1)$st slot of c. Likewise,

$$\begin{aligned} (c_{\ell+1}^m)_i^j &= c_{\ell+1}^m (t_1, \ldots, t_{i-1}, j, t_{i+1}, \ldots, t_{k-2}) \\ &= c(t_1, \ldots, t_{i-1}, j, t_i, \ldots, t_{\ell-1}, m, t_\ell, \ldots, t_{k-2}), \end{aligned}$$

where again we've used the fact that $i \le \ell$ to get the second equality. Thus $(c^m_{\ell+1})^j_i$ also plugs j into the ith slot of c and plugs m into the $(\ell+1)$st slot of c. In other words, $(c^j_i)^m_\ell = (c^m_{\ell+1})^j_i$. Thus we have

$$(-1)^{i+j+\ell+m}(c^j_i)^m_\ell + (-1)^{(\ell+1)+m+i+j}(c^m_{\ell+1})^j_i = 0,$$

as needed. $\square$

By Stokes' theorem, there must be an analog of this property for the exterior derivative. Indeed, for any k-form ω and any $(k+2)$-cell c, Stokes' theorem says that

$$\int_c d^2\omega = \int_{\partial c} d\omega = \int_{\partial^2 c} \omega.$$

But the integral on the right is zero since $\partial^2 c = 0$. Hence the integral of $d^2\omega$ over every $(k+2)$-cell is zero and it follows that $d^2\omega = 0$. We can also verify this directly from the formula for $d^2\omega$.

Proposition 137. *For every differential form ω we have $d^2\omega = 0$.*

Proof. Let ω be a k-form. We can write $\omega = \sum_I f_I \, dx_I$. Then

$$d\omega = \sum_I (df_I)\, dx_I = \sum_{j=1}^n \sum_I \frac{\partial f_I}{\partial x_j}\, dx_j\, dx_I.$$

Now apply d again to get

$$\begin{aligned} d^2\omega &= \sum_{j=1}^n \sum_I d\left(\frac{\partial f_I}{\partial x_j}\right) dx_j\, dx_I \\ &= \sum_{\ell=1}^n \sum_{j=1}^n \sum_I \frac{\partial^2 f_I}{\partial x_\ell \partial x_j}\, dx_\ell\, dx_j\, dx_I \\ &= \sum_{1\le \ell < j \le n} \sum_I \left(\frac{\partial^2 f_I}{\partial x_\ell \partial x_j} - \frac{\partial^2 f_I}{\partial x_j \partial x_\ell}\right) dx_\ell\, dx_j\, dx_I = 0, \end{aligned}$$

where the last equality uses the fact that mixed partial derivatives are zero. $\square$

Forms which have exterior derivative zero, and forms which are exterior derivatives are important enough to warrant special names.

Definition 138. A k-form ω is called closed if $d\omega = 0$.

Definition 139. A k-form ω is called exact provided $\omega = d\alpha$ for some $(k-1)$-form α.

Notice that the previous proposition implies that every exact form is closed. Indeed if $\omega = d\alpha$ is exact, then $d\omega = d^2\alpha = 0$ and so ω is closed.

Example 140. To prove a given form is not exact, it suffices to prove the form is not closed. For example, consider the form $\omega = y\,dx + x^2\,dy$ on $\mathbb{R}^2$. We have

$$\begin{aligned} d\omega &= \left(\frac{\partial(y)}{\partial x}\,dx + \frac{\partial(y)}{\partial y}\,dy\right)dx + \left(\frac{\partial(x^2)}{\partial x}\,dx + \frac{\partial(x^2)}{\partial y}\,dy\right)dy \\ &= (2x-1)\,dx\,dy \neq 0. \end{aligned}$$

Hence ω is not closed, and it follows that ω is not exact either.

The opposite implication need not be true: there are forms which are closed but not exact.

Example 141. Consider the rotation form

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

We have

$$\begin{aligned} d\alpha_{\mathrm{rot}} &= \frac{\partial}{\partial y}\left(\frac{-y}{x^2+y^2}\right)dy\,dx + \frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right)dx\,dy \\ &= \frac{-(x^2+y^2)+2y^2}{(x^2+y^2)^2}\,dy\,dx + \frac{(x^2+y^2)-2x^2}{(x^2+y^2)^2}\,dx\,dy \\ &= \frac{y^2-x^2}{(x^2+y^2)^2}\,dy\,dx + \frac{y^2-x^2}{(x^2+y^2)^2}\,dx\,dy = 0 \end{aligned}$$

and thus α_{rot} is closed. However, α_{rot} is not exact. Indeed, by Stokes' theorem, the integral of any exact 1-form $\omega = df$ over a closed curve γ is zero:

$$\int_\gamma \omega = \int_\gamma df = f(\gamma(b)) - f(\gamma(a)) = 0.$$

But there are closed curves γ such that $\int_\gamma \alpha_{\mathrm{rot}} \neq 0$.

10 Wedge Product and Pullback

Wedge product and pullback are operations that can be performed on differential forms. Actually we've already implicitly used both operations. But now we'd like to give them names and discuss some of their properties.

10.1 Wedge Product

We'll start with the wedge product.

Definition 142. Let $\alpha = \sum_I f_I \, dx_I$ be a k-form, and let $\beta = \sum_J g_J \, dx_J$ be an ℓ-form. Then their wedge product is the $(k+\ell)$-form $\alpha \wedge \beta$ given by

$$\alpha \wedge \beta = \sum_I \sum_J f_I g_J \, dx_I \, dx_J.$$

Hence to compute the wedge product of two forms, we just symbolically multiply the forms together in the obvious way.

Example 143. Let $\alpha = y \, dx + z \, dz$ and $\beta = x^2 \, dx \, dy + xz \, dx \, dz$. Then

$$\begin{aligned}\alpha \wedge \beta &= (y \, dx + z \, dz) \wedge (x^2 \, dx \, dy + xz \, dx \, dz) \\ &= yx^2 dx \, dx \, dy + zx^2 \, dz \, dx \, dy + yxz \, dx \, dx \, dz + xz^2 \, dz \, dx \, dz \\ &= x^2 z \, dx \, dy \, dz.\end{aligned}$$

Example 144. Let $\alpha = dx_1 \, dx_3 \, dx_4$ and $\beta = x_3 \, dx_2 + x_4 \, dx_5$. Then

$$\begin{aligned}\alpha \wedge \beta &= (dx_1 \, dx_3) \wedge (x_3 \, dx_2 + x_4 \, dx_4) \\ &= x_3 \, dx_1 \, dx_3 \, dx_2 + x_4 \, dx_1 \, dx_3 \, dx_4 \\ &= -x_3 \, dx_1 \, dx_2 \, dx_3 + x_4 \, dx_1 \, dx_3 \, dx_4\end{aligned}$$

Example 145. Notice that $dx \, dy = dx \wedge dy$. Likewise

$$d\left(\sum_I f_I \, dx_I\right) = \sum_I df_I \wedge dx_I.$$

Thus we have already implicitly been using the wedge product. As with ordinary multiplication of numbers, it is common to drop the wedge symbol and just write $\alpha\beta$ for $\alpha \wedge \beta$.

Remark 146. Suppose α is a k-form and β is an ℓ-form. There is also a coordinate independent definition of $\alpha \wedge \beta$. Namely,

$$(\alpha \wedge \beta)_p(v_1, \ldots, v_{k+\ell}) = \sum (-1)^\nu \alpha_p(v_{i_1}, \ldots, v_{i_k}) \beta_p(v_{j_1}, \ldots, v_{j_\ell}).$$

Here the sum is taken over all $i_1 < \ldots < i_k$ and $j_1 < \ldots < j_\ell$ such that $(i_1, \ldots, i_k, j_1, \ldots, j_\ell)$ is a permutation of $(1, 2, \ldots, k+\ell)$. The number ν is 1 if this permutation is odd and 2 if this permutation is even.

Next we record some basic properties of the wedge product.

Proposition 147. *Let $\alpha = \sum_I f_I \, dx_I$ be a k-form and let $\beta = \sum_J g_J \, dx_J$ be an ℓ-form. Then*

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

(ii) $d(\alpha\beta) = (d\alpha)\beta + (-1)^k \alpha(d\beta)$,

(iii) if α and β are closed then $\alpha\beta$ is closed,

(iv) if α is exact and β is closed then $\alpha\beta$ is exact.

Remark 148. Often property (i) is referred to as graded commutativity. Similarly property (ii) is called the graded product rule.

Proof. (i) Observe that

$$\alpha\beta = \sum_I \sum_J f_I g_J \, dx_I \, dx_J, \quad \beta\alpha = \sum_I \sum_J f_I g_J \, dx_J \, dx_I.$$

But it takes $k\ell$ swaps to interchange the dx_I with the dx_J:

$$\begin{aligned} dx_{j_1} \ldots dx_{j_\ell} \, dx_{i_1} \ldots dx_{i_k} &= (-1)^k dx_{j_1} \ldots dx_{j_{\ell-1}} \, dx_{i_1} \ldots dx_{i_k} dx_{j_\ell} \\ &= (-1)^{2k} dx_{j_1} \ldots dx_{j_{\ell-2}} \, dx_{i_1} \ldots dx_{i_k} \, dx_{j_{\ell-1}} \, dx_{j_\ell} \\ &= \cdots \\ &= (-1)^{k\ell} dx_{i_1} \ldots dx_{i_k} \, dx_{j_1} \ldots dx_{j_k}. \end{aligned}$$

Thus $\alpha\beta = (-1)^{k\ell} \beta\alpha$.

(ii) We have

$$
\begin{aligned}
d(\alpha\beta) &= d\left(\sum_I \sum_J f_I g_J \, dx_I \, dx_J\right) \\
&= \sum_I \sum_J d(f_I g_J) \, dx_I \, dx_J \\
&= \sum_I \sum_J \sum_{m=1}^n \left(\frac{\partial f_I}{\partial x_m} g_J + \frac{\partial g_J}{\partial x_m} f_I\right) dx_m \, dx_I \, dx_J.
\end{aligned}
$$

Likewise we have

$$
\begin{aligned}
(d\alpha)\beta &= \left(\sum_I \sum_{m=1}^n \frac{\partial f_I}{\partial x_m} \, dx_m \, dx_I\right) \left(\sum_J g_J \, dx_J\right) \\
&= \sum_I \sum_J \sum_{m=1}^n \frac{\partial f_I}{\partial x_m} g_J \, dx_m \, dx_I \, dx_J
\end{aligned}
$$

and

$$
\begin{aligned}
\alpha(d\beta) &= \left(\sum_I f_I \, dx_I\right) \left(\sum_J \sum_{m=1}^n \frac{\partial g_J}{\partial x_m} \, dx_m \, dx_J\right) \\
&= \sum_I \sum_J \sum_{m=1}^n \frac{\partial g_J}{\partial x_m} f_I \, dx_I \, dx_m \, dx_J.
\end{aligned}
$$

Since it takes k swaps to interchange dx_m with dx_I, this implies that $d(\alpha\beta) = (d\alpha)\beta + (-1)^k \alpha(d\beta)$.

(iii) Assume that $d\alpha = 0$ and $d\beta = 0$. Then property (ii) implies that $d(\alpha\beta) = (d\alpha)\beta + (-1)^k \alpha(d\beta) = 0$.

(iv) Assume that $\alpha = d\omega$ and $d\beta = 0$. Then property (ii) implies that $d(\omega\beta) = (d\omega)\beta + (-1)^k \omega(d\beta) = \alpha\beta$. □

10.2 Pullback

Next we discuss pullback. Assume that $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is a smooth map and let ω be a k-form on $\mathbb{R}^n$. Let P be a parallelepiped in $\mathbb{R}^m$ based at p with sides $v_1, \ldots, v_k$. Then the image $f(P)$ is very nearly a parallelepiped Q based at $f(p)$ with sides $(Df)v_1, \ldots, (Df)v_k$. We can define a k-form on $\mathbb{R}^m$ which assigns to P the number $\omega(Q)$. This form is called the pullback of ω by f.

Definition 149. Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be a smooth map and let ω be a k-form on $\mathbb{R}^n$. The pullback $f^*\omega$ is a k-form on $\mathbb{R}^m$ given by

$$(f^*\omega)_p(v_1, \dots, v_k) = \omega_{f(p)}((Df)v_1, \dots, (Df)v_k).$$

Example 150. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(r,\theta) = (r\cos\theta, r\sin\theta)$. Call the coordinates on the target x and y and let $\omega = dx\,dy$. Then

$$Df = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \quad (Df)e_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad (Df)e_2 = \begin{pmatrix} -r\sin\theta \\ r\cos\theta \end{pmatrix}.$$

Consequently

$$\begin{aligned}
(f^*\omega)_{(r,\theta)}(e_1, e_2) &= \omega_{f(r,\theta)}((Df)e_1, (Df)e_2) \\
&= dx\,dy\left(\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \begin{pmatrix} -r\sin\theta \\ r\cos\theta \end{pmatrix}\right) \\
&= \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r.
\end{aligned}$$

Since $f^*\omega$ is a 2-form on $\mathbb{R}^2$ we know that $f^*\omega = g(r,\theta)\,dr\,d\theta$ for some function g. The above calculation shows that $g(r,\theta) = r$ and hence $f^*\omega = r\,dr\,d\theta$.

Example 151. We already used the pullback construction when we defined the integral. Indeed, suppose $c\colon [0,1]^2 \to \mathbb{R}^3$ is a 2-cell and let $\omega = f\,dx\,dy + g\,dx\,dz + h\,dy\,dz$ be a 2-form on $\mathbb{R}^3$. Then

$$\begin{aligned}
\int_c \omega &= \int_0^1 \int_0^1 (f \circ c)\begin{vmatrix} \frac{\partial c_1}{\partial s} & \frac{\partial c_1}{\partial t} \\ \frac{\partial c_2}{\partial s} & \frac{\partial c_2}{\partial t} \end{vmatrix} + (g \circ c)\begin{vmatrix} \frac{\partial c_1}{\partial s} & \frac{\partial c_1}{\partial t} \\ \frac{\partial c_3}{\partial s} & \frac{\partial c_3}{\partial t} \end{vmatrix} + (h \circ c)\begin{vmatrix} \frac{\partial c_2}{\partial s} & \frac{\partial c_2}{\partial t} \\ \frac{\partial c_3}{\partial s} & \frac{\partial c_3}{\partial t} \end{vmatrix}\, ds\,dt \\
&= \int_0^1 \int_0^1 \omega_{c(s,t)}\left(\frac{\partial c}{\partial s}, \frac{\partial c}{\partial t}\right) ds\,dt = \int_0^1 \int_0^1 (c^*\omega)_{(s,t)}(e_1, e_2)\, ds\,dt.
\end{aligned}$$

Note that if we let $K\colon [0,1]^2 \to \mathbb{R}^2$ be the identity parameterization of $[0,1]^2$, i.e., $K(s,t) = (s,t)$ then

$$\int_K c^*\omega = \int_0^1 \int_0^1 (c^*\omega)_{(s,t)}(e_1, e_2)\, ds\,dt.$$

Thus we get that

$$\int_c \omega = \int_K c^*\omega.$$

More generally, for any k-cell $c\colon [0,1]^k \to \mathbb{R}^n$ and any k-form ω on $\mathbb{R}^n$, the same reasoning shows that

$$\begin{aligned}
\int_c \omega &= \int_0^1 \cdots \int_0^1 \omega_{c(t_1,\ldots,t_k)} \left(\frac{\partial c}{\partial t_1}, \ldots, \frac{\partial c}{\partial t_k} \right) dt_1 \ldots dt_k \\
&= \int_0^1 \cdots \int_0^1 (c^*\omega)_{(t_1,\ldots,t_k)}(e_1, \ldots, e_k)\, dt_1 \ldots dt_k \\
&= \int_K c^*\omega
\end{aligned}$$

where $K\colon [0,1]^k \to \mathbb{R}^k$ is given by $K(t_1, \ldots, t_k) = (t_1, \ldots, t_k)$.

Next we show how to compute the pullback in coordinates.

Proposition 152. *Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$. Call the coordinates on the domain x_j and the coordinates on the target y_i. Let $\omega = g\, dy_{i_1} \ldots dy_{i_k}$ be a k-form on $\mathbb{R}^n$. Then $f^*\omega = (g \circ f)\, df_{i_1} \ldots df_{i_k}$.*

Before giving the proof, let's do an example illustrating how to apply the formula.

Example 153. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(s,t) = (t, s^2, s+t)$ and let $\omega = x\, dy\, dz$ be a 2-form on $\mathbb{R}^3$. Then Proposition 152 says

$$\begin{aligned}
f^*\omega = t\, df_2\, df_3 &= t \left(\frac{\partial(s^2)}{\partial s}\, ds + \frac{\partial(s^2)}{\partial t}\, dt \right) \left(\frac{\partial(s+t)}{\partial s}\, ds + \frac{\partial(s+t)}{\partial t}\, dt \right) \\
&= t(2s\, ds)(ds + dt) = 2st\, ds\, dt.
\end{aligned}$$

Let's verify this is correct using the definition of the pullback. We have

$$Df = \begin{pmatrix} 0 & 1 \\ 2s & 0 \\ 1 & 1 \end{pmatrix}, \quad (Df)e_1 = \begin{pmatrix} 0 \\ 2s \\ 1 \end{pmatrix}, \quad (Df)e_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

and so

$$(f^*\omega)_{(s,t)}(e_1, e_2) = \omega_{(t,s^2,s+t)} \left(\begin{pmatrix} 0 \\ 2s \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right) = t \begin{vmatrix} 2s & 0 \\ 1 & 1 \end{vmatrix} = 2st.$$

Hence $f^*\omega$ is indeed equal to $2st\, ds\, dt$.

We now proceed with the proof of the proposition.

Proof. (Proposition 152) We know that $f^*\omega$ is a k-form and so

$$f^*\omega = \sum_J (f^*\omega)_p(e_{j_1}, \ldots, e_{j_k})\, dx_J.$$

We need to compute these coefficients $(f^*\omega)_p(e_{j_1}, \ldots, e_{j_k})$. From the definition of pullback, we see that

$$\begin{aligned}
(f^*\omega)_p(e_{j_1}, \ldots, e_{j_k}) &= \omega_{f(p)}\left(\frac{\partial f}{\partial x_{j_1}}, \ldots, \frac{\partial f}{\partial x_{j_k}}\right) \\
&= g(f(p))\, dy_{i_1} \ldots dy_{i_k}\left(\frac{\partial f}{\partial x_{j_1}}, \ldots, \frac{\partial f}{\partial x_{j_k}}\right) \\
&= g(f(p)) \det\left(\frac{\partial f_I}{\partial x_J}\right),
\end{aligned}$$

where

$$\frac{\partial f_I}{\partial x_J} = \begin{pmatrix} \frac{\partial f_{i_1}}{\partial x_{j_1}} & \cdots & \frac{\partial f_{i_1}}{\partial x_{j_k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{i_k}}{\partial x_{j_1}} & \cdots & \frac{\partial f_{i_k}}{\partial f_{j_k}} \end{pmatrix}.$$

It follows that

$$f^*\omega = \sum_J (g \circ f) \det\left(\frac{\partial f_I}{\partial x_J}\right) dx_J.$$

On the other hand, we compute that

$$\begin{aligned}
df_{i_1} \cdots df_{i_k} &= \left(\sum_{j_1=1}^m \frac{\partial f_{i_1}}{\partial x_{j_1}}\, dx_{j_1}\right)\left(\sum_{j_2=1}^m \frac{\partial f_{i_2}}{\partial x_{j_2}}\, dx_{j_2}\right) \cdots \left(\sum_{j_k=1}^m \frac{\partial f_{i_k}}{\partial x_{j_k}}\, dx_{j_k}\right) \\
&= \sum_{j_1 \neq j_2 \neq \cdots \neq j_k} \frac{\partial f_{i_1}}{\partial x_{j_1}} \frac{\partial f_{i_2}}{\partial x_{j_2}} \cdots \frac{\partial f_{i_k}}{\partial x_{j_k}}\, dx_{j_1} \ldots dx_{j_k} \\
&= \sum_{j_1 < j_2 < \ldots < j_k} \left(\sum_{\sigma \in S_k} \operatorname{sign}(\sigma) \frac{\partial f_{i_1}}{\partial x_{j_{\sigma(1)}}} \frac{\partial f_{i_2}}{\partial x_{j_{\sigma(2)}}} \cdots \frac{\partial f_{i_k}}{\partial x_{j_{\sigma(k)}}}\right) dx_{j_1} \ldots dx_{j_k} \\
&= \sum_J \det\left(\frac{\partial f_I}{\partial x_J}\right) dx_J.
\end{aligned}$$

It follows that $f^*\omega = (g \circ f)\, df_{i_1} \ldots df_{i_k}$, as needed. $\square$

10.3 Pushforward

It is a meta principle that every operation on differential forms has a dual operation on cells. For example, the exterior derivative operator on forms is dual to the boundary operator on cells. The dual operation to pullback is called pushforward.

Definition 154. Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be a smooth map and let $c\colon [0,1]^k \to \mathbb{R}^m$ be a k-cell in $\mathbb{R}^m$. The pushforward $f_*c\colon [0,1]^k \to \mathbb{R}^n$ is the k-cell in $\mathbb{R}^n$ given by $f_*c(p) = f(c(p)) = (f \circ c)(p)$.

Remark 155. The idea here is that if c parameterizes a set S in $\mathbb{R}^m$ then f_*c parameterizes the set $f(S)$ in $\mathbb{R}^n$.

Definition 156. The pushforward of a k-chain is defined by taking the pushforward of each cell in the chain individually. Thus if $\sum a_i c_i$ is a k-chain, we define

$$f_*\left(\sum a_i c_i\right) = \sum a_i (f_* c_i)$$

The following proposition shows the duality between pushforward and pullback.

Proposition 157. *Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be smooth, let $c\colon [0,1]^k \to \mathbb{R}^m$ be a k-cell, and let ω be a k-form on $\mathbb{R}^n$. Then*

$$\int_c f^*\omega = \int_{f_*c} \omega.$$

Proof. We have

$$\begin{aligned}
\int_c f^*\omega &= \int_0^1 \cdots \int_0^1 (f^*\omega)_{c(t_1,\ldots,t_k)}((Dc)e_1,\ldots,(Dc)e_k)\,dt_1\ldots dt_k \\
&= \int_0^1 \cdots \int_0^1 \omega_{f(c(t_1,\ldots,t_k))}((Df)(Dc)e_1,\ldots,(Df)(Dc)e_k)\,dt_1\ldots,dt_k \\
&= \int_0^1 \cdots \int_0^1 \omega_{(f\circ c)(t_1,\ldots,t_k)}((D(f\circ c))e_1,\ldots,(D(f\circ c))e_k)\,dt_1\ldots,dt_k \\
&= \int_{f\circ c} \omega = \int_{f_*c} \omega,
\end{aligned}$$

where we used the chain rule to get from the second to the third line. □

10.4 Pullback and the Exterior Derivative

There is one more crucial property of pullback: $f^*(d\omega) = d(f^*\omega)$. In other words, pullback commutes with exterior derivative. There is also a dual property for pushforward: $f_*(\partial c) = \partial(f_* c)$. That is, pushforward commutes with boundary. This property of pushforward is a little easier to understand intuitively. It essentially says that "the boundary of the image is the image of the boundary."

Proposition 158. *Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be a smooth map and let $c\colon [0,1]^k \to \mathbb{R}^m$ be a k-cell. Then $\partial(f_* c) = f_*(\partial c)$.*

Proof. The faces of $f_* c$ are the maps

$$\begin{aligned}(f_* c)_i^j(t_1, \ldots, t_{k-1}) &= (f_* c)(t_1, \ldots, t_{i-1}, j, t_{i+1}, \ldots, t_{k-1}) \\ &= f(c(t_1, \ldots, t_{i-1}, j, t_i, \ldots, t_{k-1})) \\ &= f(c_i^j(t_1, \ldots, t_{k-1})) = (f_*(c_i^j))(t_1, \ldots, t_{k-1}).\end{aligned}$$

Thus $(f_* c)_i^j = f_*(c_i^j)$ and it follows that

$$\partial(f_* c) = \sum_{i=1}^{k} \sum_{j=0}^{1} (-1)^{i+j} (f_* c)_i^j = \sum_{i=1}^{k} \sum_{j=0}^{1} (-1)^{i+j} f_*(c_i^j) = f_*(\partial c).$$

This proves the proposition. □

What about the corresponding property of pullback? Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be a smooth map and let ω be a k-form on $\mathbb{R}^n$. Then

$$\begin{aligned}[d(f^*\omega)]_p(v_1, \ldots, v_{k+1}) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^{k+1}} \int_{\partial P_\varepsilon} f^*\omega \\ &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^{k+1}} \int_{f_*(\partial P_\varepsilon)} \omega = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^{k+1}} \int_{\partial(f_* P_\varepsilon)} \omega,\end{aligned}$$

where $P_\varepsilon\colon [0,1]^{k+1} \to \mathbb{R}^m$ is given by

$$P_\varepsilon(t_1, \ldots, t_{k+1}) = p + \varepsilon t_1 v_1 + \ldots + \varepsilon t_{k+1} v_{k+1}.$$

But the image of the parallelepiped P_ε under f is very nearly the parallelepiped $Q_\varepsilon\colon [0,1]^{k+1} \to \mathbb{R}^n$ given by

$$Q_\varepsilon(t_1, \ldots, t_{k+1}) = f(p) + \varepsilon t_1 (Df) v_1 + \ldots + \varepsilon t_{k+1} (Df) v_{k+1}.$$

Hence we expect that

$$
\begin{aligned}
\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^{k+1}} \int_{\partial(f_* P_\varepsilon)} \omega \text{ “ = ” } \lim_{\varepsilon\to 0} \frac{1}{\varepsilon^{k+1}} \int_{\partial(Q_\varepsilon)} \omega \\
&= d\omega_{f(p)}((Df)v_1, \dots, (Df)v_{k+1}) \\
&= [f^*(d\omega)]_p(v_1, \dots, v_{k+1}).
\end{aligned}
$$

This shows that $d(f^*\omega) = f^*(d\omega)$. Unfortunately, this is not a rigorous proof because $f_* P_\varepsilon$ is not literally equal to Q_ε. We'd have to estimate the error in this approximation to justify the equals sign in quotes. This could be done in principle, but instead we will give a rigorous proof based on the formula for pullback in coordinates.

Proposition 159. *Let $f\colon \mathbb{R}^m \to \mathbb{R}^n$ be a smooth function and let ω be a k-form on $\mathbb{R}^n$. Then $d(f^*\omega) = f^*(d\omega)$.*

Proof. Let x_j be the coordinates on $\mathbb{R}^m$ and let y_i be the coordinates on $\mathbb{R}^n$. Suppose that $\omega = g\, dy_{i_1} \dots dy_{i_k}$. Then

$$
d\omega = \sum_{\ell=1}^{n} \frac{\partial g}{\partial y_\ell}\, dy_\ell\, dy_{i_1} \dots dy_{i_k}.
$$

It follows that

$$
\begin{aligned}
f^*\omega &= (g \circ f)\, df_{i_1} \dots df_{i_k}, \\
f^*(d\omega) &= \sum_{\ell=1}^{n} \left(\frac{\partial g}{\partial y_\ell} \circ f \right) df_\ell\, df_{i_1} \dots df_{i_k}.
\end{aligned}
$$

Finally, the graded product rule says that

$$
d(f^*\omega) = \sum_{j=1}^{m} \frac{\partial}{\partial x_j}(g \circ f)\, dx_j\, df_{i_1} \dots df_{i_k} + (g \circ f)\, d(df_{i_1} \dots df_{i_k}).
$$

Since a product of closed forms is closed, we have $d(df_{i_1} \dots df_{i_k}) = 0$. Hence the above formula becomes

$$
\begin{aligned}
d(f^*\omega) &= \sum_{j=1}^{m} \sum_{\ell=1}^{n} \left(\frac{\partial g}{\partial y_\ell} \circ f \right) \frac{\partial f_\ell}{\partial x_j}\, dx_j\, df_{i_1} \dots df_{i_k} \\
&= \sum_{\ell=1}^{n} \left(\frac{\partial g}{\partial y_\ell} \circ f \right) df_\ell\, df_{i_1} \dots df_{i_k}.
\end{aligned}
$$

Thus we see that $d(f^*\omega) = f^*(d\omega)$, as needed. $\square$

11 The Poincare Lemma

We have seen that there are closed forms on $\mathbb{R}^n \setminus \{0\}$ which are not exact. It is a very special property of $\mathbb{R}^n$ that all closed forms defined on $\mathbb{R}^n$ turn out to be exact. This fact is usually referred to as the Poincare lemma.

Theorem 160 (Poincare Lemma)**.** *Every closed k-form on $\mathbb{R}^n$ is exact.*

Example 161. Let's check that every closed 1-form on $\mathbb{R}^n$ is exact. Let

$$\alpha = \sum_{i=1}^{n} f_i \, dx_i$$

be a closed 1-form on $\mathbb{R}^n$. Typically we find antiderivatives by some sort of integration process. Motivated by this, define a function $g\colon \mathbb{R}^n \to \mathbb{R}$ by setting

$$g(p) = \int_\gamma \alpha$$

where $\gamma\colon [0,1] \to \mathbb{R}^n$ is some curve starting at 0 and ending at p.

We need to check that this is well-defined, i.e., that it does not depend on the particular choice of curve connecting the origin to p. So suppose that $\gamma, \eta\colon [0,1] \to \mathbb{R}^n$ are two curves connecting the origin to p. Define $h\colon [0,1] \times [0,1] \to \mathbb{R}^n$ by

$$h(s,t) = s\eta(t) + (1-s)\gamma(t).$$

Note that $h(s,0) = 0$ and $h(s,1) = p$ for all $s \in [0,1]$. This h is a homotopy between γ and η holding the endpoints of these curves fixed. Since α is closed, Stokes' theorem says

$$\begin{aligned} 0 = \int_h d\alpha = \int_{\partial h} \alpha &= \int_{h_2^0} \alpha + \int_{h_1^1} \alpha - \int_{h_2^1} \alpha - \int_{h_1^0} \alpha \\ &= \int_\eta \alpha - \int_\gamma \alpha. \end{aligned}$$

Here the final equality follows from the fact that $h_2^0(s) = h(s,0) = 0$, $h_1^1(t) = h(1,t) = \eta(t)$, $h_2^1(s) = h(s,1) = p$, and $h_1^0(t) = \gamma(t)$. Thus

$$\int_\eta \alpha = \int_\gamma \alpha,$$

proving that g is well-defined.

We claim that $dg = \alpha$. To see this, fix a point p in $\mathbb{R}^n$, and observe that

$$\frac{\partial}{\partial x_i} g(p) = \lim_{\varepsilon \to 0} \left[\int_{\gamma_{p+\varepsilon e_i}} \alpha - \int_{\gamma_p} \alpha \right]$$

where $\gamma_{p+\varepsilon e_i}$ is some path from 0 to $p + \varepsilon e_i$, and γ_p is some path from 0 to p. Actually, we are free to choose these paths, and hence we can assume that $\gamma_{p+\varepsilon e_i}$ consists of γ_p followed by the straight line from p to $p + \varepsilon e_i$. For this choice of $\gamma_{p+\varepsilon e_i}$, we get

$$\lim_{\varepsilon \to 0} \left[\int_{\gamma_{p+\varepsilon e_i}} \alpha - \int_{\gamma_p} \alpha \right] = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\eta_\varepsilon} \alpha$$

where $\eta_\varepsilon \colon [0, \varepsilon] \to \mathbb{R}^n$ is given by $\eta_\varepsilon(t) = p + te_i$. But the fundamental theorem of calculus implies that

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\eta_\varepsilon} \alpha = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^\varepsilon f_i(p + te_i)\, dt = f_i(p).$$

Thus we get

$$dg = \sum_{i=1}^n \frac{\partial g}{\partial x_i}\, dx_i = \sum_{i=1}^n f_i\, dx_i = \alpha,$$

as needed.

To prove the general case of the Poincare lemma, we need to find a way to "integrate" k-forms to get $(k-1)$-forms. Given a k-form $\omega = \sum_I f_I\, dx_I$ on $\mathbb{R}^n$, we can split the dx_I's into two groups: those that contain dx_1 and those that don't. This gives a decomposition

$$\omega = \sum_J f_{1,J}\, dx_1\, dx_J + \sum_K f_K\, dx_K,$$

where the first sum is taken over increasing lists J of $k-1$ numbers taken from $2, \ldots, n$, and the second sum is taken over increasing lists K of k numbers taken from $2, \ldots, n$. We define the "integral"

$$I\omega = \sum_J \left(\int_0^{x_1} f_{1,J}(s, x_2, \ldots, x_n)\, ds \right) dx_J.$$

This is a $(k-1)$-form on $\mathbb{R}^n$ obtained by integrating out the dx_1 in the terms with a dx_1, and forgetting the terms without a dx_1.

Example 162. Consider the 2-form

$$\omega = x_1x_2\, dx_1\, dx_2 + x_1^2x_3\, dx_1\, dx_3 + x_2\, dx_2\, dx_3.$$

We have

$$\begin{aligned} I\omega &= \left(\int_0^{x_1} sx_2\, ds\right) dx_2 + \left(\int_0^{x_1} s^2x_3\, ds\right) dx_3 \\ &= \frac{x_1^2x_2}{2}\, dx_2 + \frac{x_1^3x_3}{3}\, dx_3. \end{aligned}$$

Thus we've integrated out the dx_1 in $dx_1\, dx_2$ and $dx_1\, dx_3$, and we've forgotten the term $dx_2\, dx_3$.

Proposition 163. *Define maps* $\iota\colon \mathbb{R}^{n-1} \to \mathbb{R}^n$ *and* $\pi\colon \mathbb{R}^n \to \mathbb{R}^{n-1}$ *by setting* $\iota(x_2, \ldots, x_n) = (0, x_2, \ldots, x_n)$ *and* $\pi(x_1, x_2, \ldots, x_n) = (x_2, \ldots, x_n)$*. Then*

$$d(I\omega) + I(d\omega) = \omega - \pi^*\iota^*\omega.$$

Proof. Let ω be a k-form and write

$$\omega = \sum_J f_{1,J}\, dx_1\, dx_J + \sum_K f_K\, dx_K.$$

Then we have

$$I\omega = \sum_J \left(\int_0^{x_1} f_{1,J}(s, \cdot)\, ds\right) dx_J,$$

where the $\cdot$ stands for the variables $x_2, \ldots, x_n$. It follows that

$$\begin{aligned} d(I\omega) &= \sum_J \left(f_{1,J}\, dx_1 + \sum_{j=2}^n \frac{\partial}{\partial x_j}\left[\int_0^{x_1} f_{1,J}(s, \cdot)\, ds\right] dx_j\right) dx_J \\ &= \sum_J \left(f_{1,J}\, dx_1 + \sum_{j=2}^n \left[\int_0^{x_1} \frac{\partial f_{1,J}}{\partial x_j}(s, \cdot)\, ds\right] dx_j\right) dx_J. \end{aligned}$$

On the other hand, we have

$$d\omega = \sum_J \sum_{j=2}^n \frac{\partial f_{1,J}}{\partial x_j}\, dx_j\, dx_1\, dx_J + \sum_K \frac{\partial f_I}{\partial x_1}\, dx_1\, dx_K + \substack{\text{terms that don't} \\ \text{have } dx_1},$$

and hence

$$
\begin{aligned}
I(d\omega) &= -\sum_J \sum_{j=2}^{n} \left(\int_0^{x_1} \frac{\partial f_{1,J}}{\partial x_j}(s,\cdot) \right) dx_j\, dx_J \\
&\qquad + \sum_K \left(\int_0^{x_1} \frac{\partial f_I}{\partial x_1}(s,\cdot)\, ds \right) dx_K \\
&= -\sum_J \sum_{j=2}^{n} \left(\int_0^{x_1} \frac{\partial f_{1,J}}{\partial x_j}(s,\cdot) \right) dx_j\, dx_J \\
&\qquad + \sum_K f_K(x_1,\cdot)\, dx_I - \sum_K f_K(0,\cdot)\, dx_K.
\end{aligned}
$$

Thus we get

$$d(I\omega) + I(d\omega) = \omega - \sum_K f_K(0,\cdot)\, dx_K.$$

To complete the proof, we need to check that

$$\sum_K f_K(0,\cdot)\, dx_K = \pi^* \iota^* \omega.$$

Note that $\iota^*(dx_i) = dx_i$ for $i \geq 2$ and $\iota^*(dx_1) = 0$. Thus

$$\iota^* \omega = \sum_K f_K(0,\cdot)\, dx_K,$$

where we are thinking of this as a form on $\mathbb{R}^{n-1}$ with coordinates $x_2, \ldots, x_n$. To get $\pi^*(\iota^*\omega)$, we simply treat this as a form on $\mathbb{R}^n$ with the same formula. $\square$

Using this we can prove the Poincare lemma.

Proof. (Poincare lemma) Fix an integer $k \geq 1$. We will prove the result by induction on the dimension n of the ambient space $\mathbb{R}^n$. For the base case, assume that $n = k$. Let ω be a closed k-form on $\mathbb{R}^k$. Then the previous proposition says

$$\omega = d(I\omega) + I(d\omega) + \pi^* \iota^* \omega.$$

But $d\omega = 0$ since ω is closed and $\iota^*\omega = 0$ since it is a k-form on $\mathbb{R}^{k-1}$. Thus $\omega = d(I\omega)$ is exact. This proves the base case.

Now fix a positive integer $n \geq k$. Assume by way of induction that every closed k-form on $\mathbb{R}^n$ is exact. To complete the inductive step, we need to show that every closed k-form on $\mathbb{R}^{n+1}$ is exact. So assume ω is a closed k-form on $\mathbb{R}^{n+1}$. Note that $d(\iota^*\omega) = \iota^*(d\omega) = 0$, and so $\iota^*\omega$ is a closed k-form on $\mathbb{R}^n$. By the inductive hypothesis, there is a form α on $\mathbb{R}^n$ such that $\iota^*\omega = d\alpha$. It follows that

$$\begin{aligned} \omega = d(I\omega) + I(d\omega) + \pi^*\iota^*\omega &= d(I\omega) + \pi^*(d\alpha) \\ &= d(I\omega) + d(\pi^*\alpha) = d(I\omega + \pi^*\alpha), \end{aligned}$$

and hence ω is exact. This finishes the inductive step, and the proof is complete. $\square$

12 Applications

Gauss's Law in physics says that the flux of the electric field through a closed surface is proportional to the net amount of charge enclosed by the surface. We can use Stokes' theorem to prove this in a couple of special cases.

12.1 Gauss's Law for Point Charges

In three dimensions, the electric field generated by a unit point charge at the origin is

$$F(x,y,z) = \frac{1}{4\pi}\frac{(x,y,z)}{\|(x,y,z)\|^3}.$$

Actually, in any dimension $n \geq 2$, there is an analogous field F which we will still refer to as an electric field.

Definition 164. Fix an integer $n \geq 2$. Let $c_{n-1} = \mathrm{Vol}_{n-1}(\text{unit sphere in } \mathbb{R}^n)$ denote the $(n-1)$-dimensional volume of the unit sphere in $\mathbb{R}^n$.

Example 165. Note that $c_1 = 2\pi$ is the perimeter of the unit circle in $\mathbb{R}^2$ and $c_2 = 4\pi$ is the surface area of the unit sphere in $\mathbb{R}^3$.

Definition 166. Fix a dimension $n \geq 2$. The electric field associated to a point charge at the origin is

$$F(\vec{x}) = \frac{1}{c_{n-1}}\frac{\vec{x}}{\|\vec{x}\|^n}.$$

Note that this field is only defined on $\mathbb{R}^n \setminus \{0\}$.

So far we've only discussed the flux of vector fields in $\mathbb{R}^3$ over surfaces in $\mathbb{R}^3$. However, there is also a notion of flux for vector fields $F\colon \mathbb{R}^n \to \mathbb{R}^n$ over $(n-1)$-dimensional objects in $\mathbb{R}^n$.

Definition 167. Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a vector field on $\mathbb{R}^n$. Associated to F is a flux form α_F. It is the $(n-1)$-form on $\mathbb{R}^n$ given by the formula

$$(\alpha_F)_p(v_1, \dots, v_{n-1}) = \det(F(p), v_1, \dots, v_{n-1}).$$

Thus α_F computes the flux of F over small $(n-1)$-parallelepipeds in $\mathbb{R}^n$.

By expanding the determinant along the first column we can also express α_F in terms of the dx_i. Indeed,

$$\det(F(p), v_1, \dots, v_{n-1}) = \sum_{i=1}^{n} (-1)^{i+1} F_i(p)(dx_1 \dots \widehat{dx_i} \dots dx_n)(v_1, \dots, v_{n-1}),$$

where the hat over dx_i indicates that this term should be omitted. Thus we can write

$$\alpha_F = \sum_{i=1}^{n} (-1)^{i+1} F_i \, dx_1 \dots \widehat{dx_i} \dots dx_n.$$

For small n this gives the formulas:

$$\begin{aligned}
&n = 2 : F_1 \, dy - F_2 \, dx \\
&n = 3 : F_1 \, dy \, dz - F_2 \, dx \, dz + F_3 \, dx \, dy \\
&n = 4 : F_1 \, dx_2 \, dx_3 \, dz_4 - F_2 \, dx_1 \, dx_3 \, dx_4 + F_3 \, dx_1 \, dx_2 \, dx_4 - F_4 \, dx_1 \, dx_2 \, dx_3.
\end{aligned}$$

Note in particular that for $n = 3$ we recover the familiar formula for the flux form of a vector field on $\mathbb{R}^3$.

Proposition 168. *Let α_F be the flux form associated to a vector field $F\colon \mathbb{R}^n \to \mathbb{R}^n$. Then*

$$d\alpha_F = \operatorname{div}(F) \, dx_1 \dots dx_n$$

where $\operatorname{div}(F) = \sum_{i=1}^{n} \frac{\partial F_i}{\partial x_i}$ *is the divergence of the vector field F.*

Proof. We have

$$\begin{aligned}
d\alpha_F &= \sum_{i=1}^{n} (-1)^{i+1} \left(\sum_{j=1}^{n} \frac{\partial F_i}{\partial x_j}\, dx_j \right) dx_1 \dots \widehat{dx_i} \dots dx_n \\
&= \sum_{i=1}^{n} (-1)^{i+1} \frac{\partial F_i}{\partial x_i}\, dx_i\, dx_1 \dots \widehat{dx_i} \dots dx_n \\
&= \sum_{i=1}^{n} \frac{\partial F_i}{\partial x_i} dx_1 \dots dx_i \dots dx_n = \operatorname{div}(F)\, dx_1 \dots dx_n,
\end{aligned}$$

as needed. $\square$

Proposition 169. *Let F be the electric field generated by a point charge at the origin in $\mathbb{R}^n$. Then $d\alpha_F = 0$ on $\mathbb{R}^n \setminus \{0\}$.*

Proof. According to the previous proposition, we need to check that $\operatorname{div}(F) = 0$. Recalling that

$$F(\vec{x}) = \frac{1}{c_{n-1}} \frac{\vec{x}}{\|\vec{x}\|^n},$$

we compute that

$$\begin{aligned}
\operatorname{div}(F) &= \frac{1}{c_{n-1}} \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(\frac{x_i}{(x_1^2 + \ldots + x_n^2)^{n/2}} \right) \\
&= \frac{1}{c_{n-1}} \sum_{i=1}^{n} \frac{(x_1^2 + \ldots + x_n^2)^{n/2} - n x_i^2 (x_1^2 + \ldots + x_n^2)^{n/2-1}}{(x_1^2 + \ldots + x_n^2)^n} = 0,
\end{aligned}$$

as desired. $\square$

We can now verify Gauss's law in the case of point charges.

Proposition 170. *Fix $n \geq 2$ and let F be the electric field associated to a point charge at the origin in $\mathbb{R}^n$. Let α_F be the flux form for F. Let Σ be a closed $(n-1)$-dimensional surface in $\mathbb{R}^n \setminus \{0\}$ and let Ω be the region enclosed by Σ. Then*

$$\int_\Sigma \alpha_F = \begin{cases} 0, & \text{if } 0 \notin \Omega \\ 1, & \text{if } 0 \in \Omega. \end{cases}$$

Proof. There are two cases to consider. First assume that Σ does not enclose the origin. Then F is defined everywhere on Ω. Moreover, since $\Sigma = \partial\Omega$, Stokes' theorem gives

$$\int_\Sigma \alpha_F = \int_\Omega d\alpha_F = \int_\Omega \operatorname{div}(F)\, dx_1 \ldots dx_n = 0.$$

This verifies Gauss's law when Σ does not enclose the origin.

Now suppose instead that Σ does enclose the origin. Fix a very small number $r > 0$ and let B_r be a ball of radius r centered at the origin. Then $\operatorname{div}(F) = 0$ everywhere on $\Omega \setminus B_r$ and so

$$0 = \int_{\Omega\setminus B_r} d\alpha_F = \int_{\partial(\Omega\setminus B_r)} \alpha_F = \int_\Sigma \alpha_F - \int_{\partial B_r} \alpha_F,$$

where both integrals on the right hand side are computed with respect to the outer normals. Thus we get

$$\int_\Sigma \alpha_F = \int_{\partial B_r} \alpha_F.$$

On the other hand, F is everywhere normal to ∂B_r and $\|F\| = \frac{1}{c_{n-1} r^{n-1}}$ on ∂B_r and so

$$\int_{\partial B_r} \alpha_F = \frac{1}{c_{n-1} r^{n-1}} \operatorname{Vol}_{n-1}(\partial B_r) = \frac{1}{c_{n-1} r^{n-1}} \left[r^{n-1} \operatorname{Vol}_{n-1}(\partial B_1) \right] = 1.$$

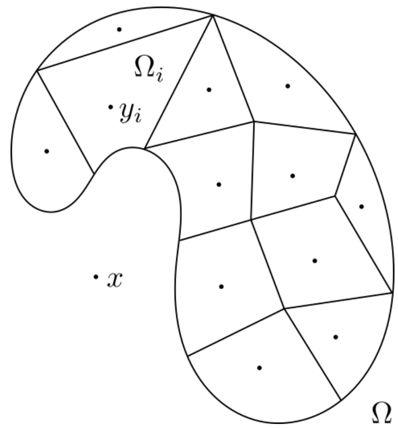
Here we've used the fact that $\operatorname{Vol}_{n-1}(\partial B_r) = r^{n-1} \operatorname{Vol}_{n-1}(\partial B_1)$ and the fact that $c_{n-1} = \operatorname{Vol}_{n-1}(\partial B_1)$. Combining everything above yields

$$\int_\Sigma \alpha_F = 1,$$

and so we've verified Gauss's law in this case as well. $\square$

12.2 Gauss's Law for Uniform Charge Distributions

Of course, there are many other possible charge distributions. We'll also address the case of uniform charge densities on regions in $\mathbb{R}^n$. Consider a nice region Ω in $\mathbb{R}^n$. What is the electric field generated by a uniform charge density on Ω?



To answer this intuitively, cut Ω into small pieces Ω_i. In each piece Ω_i pick a point y_i. Then we can approximate Ω_i by a point charge at y_i with net charge $\mathrm{Vol}(\Omega_i)$. Because of this, the force a point charge at x experiences due to Ω_i is very nearly equal to

$$\frac{1}{c_{n-1}} \frac{x - y_i}{\|x - y_i\|^n} \mathrm{Vol}(\Omega_i).$$

Summing over all the pieces Ω_i shows that the total force at x is approximately

$$\frac{1}{c_{n-1}} \sum_i \frac{x - y_i}{\|x - y_i\|^n} \mathrm{Vol}(\Omega_i).$$

This converges to

$$\frac{1}{c_{n-1}} \int_{y \in \Omega} \frac{x - y}{\|x - y\|^n} \, dV(y)$$

as the approximation gets better and better. Based on this we make the following definition.

Definition 171. Let Ω be a nice region in $\mathbb{R}^n$. The electric field associated to a uniform charge density on Ω is

$$F(x) = \frac{1}{c_{n-1}} \int_{y \in \Omega} \frac{x - y}{\|x - y\|^n} \, dV(y).$$

It turns out that this electric field is defined at every $x \in \mathbb{R}^n$. This isn't immediately obvious from the formula above since we may divide by 0 in the integrand if $x \in \Omega$. To see that the integral still converges, fix a point $x \in \Omega$. For $n = 2$, in polar coordinates centered at x, the integrand becomes

$$\frac{x-y}{\|x-y\|^2}\, dV(y) = \frac{-(r\cos\theta, r\sin\theta)}{r^2}\, r\, dr\, d\theta.$$

Thus the Jacobian factor cancels the singularity leaving something integrable. For $n = 3$, in spherical coordinates centered at x, the integrand becomes

$$\frac{x-y}{\|x-y\|^3}\, dV(y) = \frac{-(r\cos\theta\sin\phi, r\sin\theta\sin\phi, r\cos\phi)}{r^3}\, r^2\, \sin\phi\, dr\, d\theta\, d\phi.$$

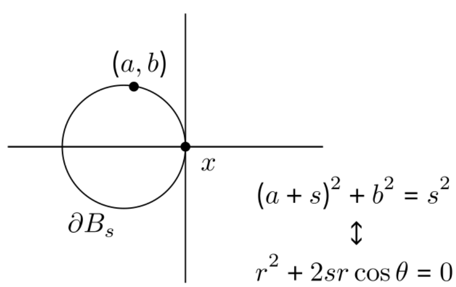
Again the Jacobian factor cancels the singularity leaving something integrable. For general n, the same computation in n-dimensional polar coordinates shows that the integral is well-defined.

Proposition 172. *Let B_s be a ball of radius s centered at the origin in $\mathbb{R}^2$. Let F be the electric field associated to a uniform charge density on B_s. Then*

$$F(\vec{x}) = \frac{\vec{x}}{2}$$

for $\vec{x} \in \partial B_s$.

Proof. By rotational symmetry, it's enough to prove this for $\vec{x} = (s, 0)$. Introduce polar coordinates centered at x. In these coordinates, ∂B_s is given by the equation $r = -2s\cos\theta$.



Therefore we get

$$\begin{aligned} F(x) &= \frac{1}{2\pi}\int_{y\in B_s} \frac{x-y}{\|x-y\|^2}\, dV(y) \\ &= \frac{1}{2\pi}\int_{\theta=\frac{\pi}{2}}^{\theta=\frac{3\pi}{2}} \int_{r=0}^{r=-2s\cos\theta} \frac{-(r\cos\theta, r\sin\theta)}{r^2}\, r\, dr\, d\theta \\ &= \frac{2s}{2\pi}\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} (\cos^2\theta, \sin\theta\cos\theta)\, d\theta = \left(\frac{s}{2}, 0\right) = \frac{x}{2}, \end{aligned}$$

as needed. $\square$

We'd like to prove Gauss's law for continuous charge distributions. The first step is to compute the divergence of the electric field associated to a uniform charge distribution on Ω.

Proposition 173. *Let Ω be a nice region in $\mathbb{R}^n$ and let F be the electric field associated to a uniform charge distribution on Ω. Then F is continuous on $\mathbb{R}^n$ and smooth away from $\partial\Omega$. Moreover, we have*

$$\operatorname{div}(F)(x) = \begin{cases} 0, & \text{if } x \text{ is outside } \Omega \\ 1, & \text{if } x \text{ is inside } \Omega. \end{cases}$$

The full proof of this is slightly beyond the scope of the class. So we'll just indicate the main ideas.

Proof. (Sketch) There are two cases to consider.

Case 1: Assume that x is a point outside of Ω. We want to compute $\operatorname{div}(F)$. So we need to find, for instance,

$$\frac{\partial}{\partial x_1} F_1 = \frac{1}{c_{n-1}} \frac{\partial}{\partial x_1} \left[\int_{y \in \Omega} \frac{x_1 - y_1}{\|x - y\|^n} \, dV(y) \right].$$

Thus we need to take the derivative of an integral. Recall that derivatives commute with sums:

$$\frac{\partial}{\partial x_1} \left(\sum_{i=1}^{\ell} f_i \right) = \sum_{i=1}^{\ell} \frac{\partial f_i}{\partial x_1}.$$

Since an integral is a sort of "infinite sum" we might therefore hope that derivatives commute with integrals. In our case, this turns out to be valid provided x does not belong to Ω.

Fact. *If x lies outside Ω then*

$$\frac{\partial}{\partial x_1} \left[\int_{y \in \Omega} \frac{x_1 - y_1}{\|x - y\|^n} \, dV(y) \right] = \int_{y \in \Omega} \frac{\partial}{\partial x_1} \left(\frac{x_1 - y_1}{\|x - y\|^n} \right) dV(y).$$

Given the fact, we can finish the proof of case 1. Indeed, we get

$$\begin{aligned} \operatorname{div}(F)(x) &= \frac{1}{c_{n-1}} \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left[\int_{y \in \Omega} \frac{x_i - y_i}{\|x - y\|^n} \, dV(y) \right] \\ &= \frac{1}{c_{n-1}} \int_{y \in \Omega} \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(\frac{x_i - y_i}{\|x - y\|^n} \right) dV(y) = 0 \end{aligned}$$

since the divergence of a point charge centered at y is zero.

Case 2: Assume that x lies inside Ω. We'd still like to use the fact. To make this work, we'll remove a small ball around x from Ω. Let $s > 0$ be a very small number and let B_s be a ball of radius s centered at x. Then we can split F into two vector fields:

$$F(p) = \underbrace{\frac{1}{c_{n-1}} \int_{y \in \Omega \setminus B_s} \frac{p-y}{\|p-y\|^n} \, dV(y)}_{G_s(p)} + \underbrace{\frac{1}{c_{n-1}} \int_{y \in B_s} \frac{p-y}{\|p-y\|^n} \, dV(y)}_{H_s(p)}.$$

Note that if $p \in B_s$, then p lies outside $\Omega \setminus B_s$, and so again we can use the fact to commute derivatives with integrals. This implies that $\operatorname{div}(G_s)(p) = 0$ for $p \in B_s$.

Since $\operatorname{div}(G_s)(p) = 0$ in B_s, Stokes' theorem gives that

$$\int_{\partial B_s} \alpha_{G_s} = \int_{B_s} d\alpha_{G_s} = 0.$$

Recall from Example 118 that we can compute the divergence of F at x as a limit of the flux of F through small surfaces centered at x. Thus we have

$$\begin{aligned} \operatorname{div}(F)(x) &= \lim_{s \to 0} \frac{1}{\operatorname{Vol}_n(B_s)} \int_{\partial B_s} \alpha_F \\ &= \lim_{s \to 0} \frac{1}{\operatorname{Vol}_n(B_s)} \left[\int_{\partial B_s} \alpha_{G_s} + \int_{\partial B_s} \alpha_{H_s} \right] \\ &= \lim_{s \to 0} \frac{1}{\operatorname{Vol}_n(B_s)} \int_{\partial B_s} \alpha_{H_s}. \end{aligned}$$

Finally, one verifies by explicit computation that

$$\frac{1}{\operatorname{Vol}_n(B_s)} \int_{\partial B_s} \alpha_{H_s} = 1,$$

and this completes the proof. □

Remark 174. Let us carry out this explicit computation for $n = 2$. In this case, H_s is the electric field associated to a uniform charge density on a ball of radius s. By Proposition 172, we know that $\|H_s\| = \frac{s}{2}$ on ∂B_s and H_s everywhere normal to ∂B_s. Thus

$$\frac{1}{\operatorname{Area}(B_s)} \int_{\partial B_s} \alpha_{H_s} = \frac{1}{\pi s^2} \left[\frac{s}{2} \cdot 2\pi s \right] = 1,$$

as needed.

Now we can verify Gauss's law for continuous charge distributions as well.

Proposition 175. *Let Ω be a nice region in $\mathbb{R}^n$ and let F be the electric field associated to a uniform charge density on Ω. Let Σ be a closed $(n-1)$-dimensional surface in $\mathbb{R}^n$ which encloses a region U. Then*

$$\int_\Sigma \alpha_F = \mathrm{Vol}(U \cap \Omega).$$

Proof. Since $U = \partial\Sigma$, Stokes' theorem gives that

$$\int_\Sigma \alpha_F = \int_U d\alpha_F = \int_U \mathrm{div}(F)\, dx_1 \ldots dx_n = \mathrm{Vol}(U \cap \Omega).$$

Here the final equality follows from the fact that $\mathrm{div}(F) = 1$ inside Ω and $\mathrm{div}(F) = 0$ outside Ω. $\square$

12.3 The Isoperimetric Inequality

We can use Gauss's law for uniform charge distributions along with Stokes' theorem to give a nice proof of the isoperimetric inequality in the plane. The isoperimetric problem in the plane asks "what is the most efficient way to enclose a given amount of area in the plane?" In other words, of all the regions in the plane with a given area, which one has the smallest perimeter? Intuitively, we expect that the answer is a circle. The following inequality confirms that this is the case.

Theorem 176 (Isoperimetric Inequality)**.** *Let Ω be region in $\mathbb{R}^2$ bounded by a smooth, closed curve. Then*

$$\mathrm{Length}(\partial\Omega) \geq 2\sqrt{\pi}\sqrt{\mathrm{Area}(\Omega)}.$$

Moreover, if equality holds then Ω is a round disc.

Remark 177. Note that the right hand side of the above inequality is the perimeter of the circle with the same area as Ω. Indeed, let r be the radius of such a circle. Then $\pi r^2 = \mathrm{Area}(\Omega)$ and so

$$r = \sqrt{\mathrm{Area}(\Omega)/\pi}.$$

It follows that the perimeter of the circle is $2\pi r = 2\sqrt{\pi}\sqrt{\mathrm{Area}(\Omega)}$. Hence the isoperimetric inequality really says that the perimeter of Ω is at least the perimeter of the circle with the same area as Ω. In particular, the isoperimetric inequality implies that the circle is the solution to the isoperimetric problem.

Proof. (Due to Gromov) Let Ω be as in the statement of the theorem. Let F be the electric field generated by a uniform charge density on Ω. Then $\mathrm{div}(F) = 1$ on Ω and so Stokes' theorem says that

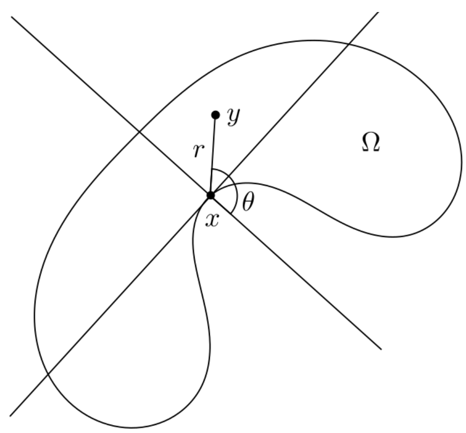
$$\mathrm{Area}(\Omega) = \int_{\Omega} d\alpha_F = \int_{\partial\Omega} \alpha_F = \int_{\partial\Omega} F \cdot \nu \, d\ell.$$

Here ν is the unit outward normal to $\partial\Omega$.

It remains to estimate the right hand side. Fix a point $x \in \partial\Omega$. Then

$$F(x) \cdot \nu = \frac{1}{2\pi} \int_{y \in \Omega} \frac{(x-y) \cdot \nu}{\|x-y\|^2} \, dV(y).$$

Introduce polar coordinates centered at x such that $r = 1$, $\theta = 0$ corresponds to ν.



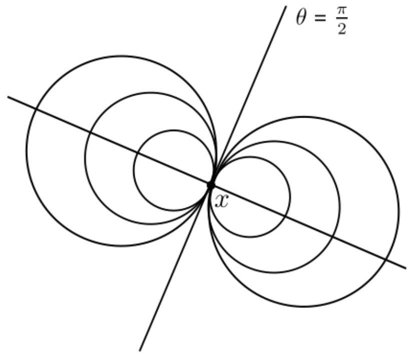
In this coordinate system the integrand becomes

$$\frac{(x-y) \cdot \nu}{\|x-y\|^2} = \frac{-r\cos\theta}{r^2} = -\frac{\cos\theta}{r}.$$

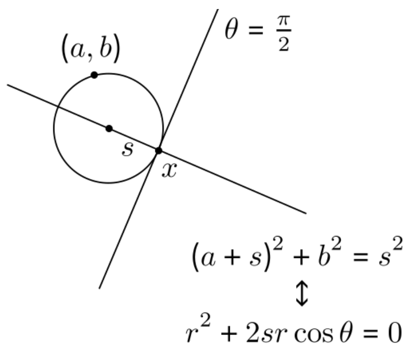
Next let's analyze the level sets of this function.

We claim that the level sets of $-\frac{\cos\theta}{r}$ are circles tangent to the line $\theta = \pi/2$

at x.



To see this, first consider a circle of radius s which lies in the region $\pi/2 \leq \theta \leq 3\pi/2$ and is tangent to $\theta = \pi/2$ at x.

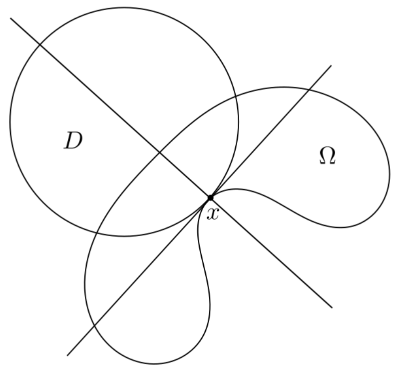


Computing, we find that $-\frac{\cos\theta}{r} = \frac{1}{2s}$ on this circle. Likewise, consider a circle of radius s which lies in the region $-\pi/2 \leq \theta \leq \pi/2$ and is tangent to $\theta = \pi/2$ at x. Then $-\frac{\cos\theta}{r} = -\frac{1}{2s}$ on this circle. Finally, notice that $-\frac{\cos\theta}{r} = 0$ on the line $\theta = \pi/2$. Thus we have a complete description of the level sets of $-\frac{\cos\theta}{r}$.

Now let D be a disc of the same area as Ω which lies in the region $\pi/2 \leq \theta \leq 3\pi/2$ and is tangent to $\theta = \pi/2$ at x. Let

$$s = \sqrt{\text{Area}(\Omega)/\pi}$$

denote the radius of D.



From the above description of the level sets of $-\frac{\cos\theta}{r}$, we see that

$$-\frac{\cos\theta}{r} \geq \frac{1}{2s} \quad \text{inside } D, \qquad -\frac{\cos\theta}{r} \leq \frac{1}{2s} \quad \text{outside } D.$$

Moreover, the fact that $\text{Area}(D) = \text{Area}(\Omega)$ implies that $\text{Area}(D \setminus \Omega) = \text{Area}(\Omega \setminus D)$. It follows that

$$\begin{aligned}\int_{y\in\Omega\setminus D} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y) &\leq \frac{1}{2s}\,\text{Area}(\Omega\setminus D)\\ &= \frac{1}{2s}\,\text{Area}(D\setminus\Omega) \leq \int_{y\in D\setminus\Omega} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y).\end{aligned}$$

Hence we can estimate

$$\begin{aligned}F(x)\cdot\nu &= \frac{1}{2\pi}\int_{y\in\Omega} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y)\\ &= \frac{1}{2\pi}\int_{y\in\Omega\cap D} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y) + \frac{1}{2\pi}\int_{y\in\Omega\setminus D} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y)\\ &\leq \frac{1}{2\pi}\int_{y\in\Omega\cap D} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y) + \frac{1}{2\pi}\int_{y\in D\setminus\Omega} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y)\\ &= \frac{1}{2\pi}\int_{y\in D} \frac{(x-y)\cdot\nu}{\|x-y\|^2}\, dV(y) = \frac{s}{2},\end{aligned}$$

where we used Proposition 172 to get the last equality.

Putting everything together, we get

$$\text{Area}(\Omega) = \int_{\partial\Omega} F\cdot\nu\, d\ell \leq \frac{s}{2}\cdot\text{Length}(\partial\Omega) = \frac{\sqrt{\text{Area}(\Omega)/\pi}}{2}\cdot\text{Length}(\partial\Omega).$$

Rearranging, this implies that

$$\text{Length}(\partial\Omega) \geq 2\sqrt{\pi}\sqrt{\text{Area}(\Omega)},$$

as desired. To complete the proof, notice that if equality holds then $F(x) \cdot \nu = s/2$ for every $x \in \partial\Omega$. From the above chain of inequalities, it's clear that the only way this can happen is if $\Omega = D$. $\square$

12.4 Brouwer's Fixed Point Theorem

We can also use Stokes' theorem to give a proof of the Brouwer fixed point theorem. This theorem has important applications in both math and other disciplines. For example, in game theory, the Brouwer fixed point theorem is used to establish the existence of Nash equilibria.

Theorem 178 (Brouwer)**.** *Let $D = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ be the closed unit disc in $\mathbb{R}^n$. Let $f\colon D \to D$ be a smooth map. Then f has a fixed point, i.e., there is a point $x \in D$ such that $f(x) = x$.*

Remark 179. (i) It is essential that D is the *closed* unit disc. The formula

$$f(x, y) = \left(\frac{x}{2} + \frac{1}{2}, \frac{y}{2}\right)$$

defines a smooth map from the open unit disc to itself. However, it does not have a fixed point in the open unit disc. Of course, the "missing" fixed point $(1, 0)$ does belong to the closed unit disc.

(ii) The assumption that f is smooth is not essential. In fact, the theorem remains true so long as f is merely continuous.

(iii) The theorem remains true if we replace the closed unit disc D by any other subset of $\mathbb{R}^n$ that "looks like" D. More precisely, suppose A is a subset of $\mathbb{R}^n$ such that there is a continuous bijection $\phi\colon A \to D$ with continuous inverse. Then Brouwer's theorem applies to continuous maps $f\colon A \to A$ as well. For example, A could be a closed cube or a closed tetrahedron. However, a spherical shell $A = \{x \in \mathbb{R}^n : 1 \leq \|x\| \leq 2\}$ is *not* allowed.

Example 180. As a silly example, suppose someone takes a cup of coffee, stirs it, and then allows the liquid to return to rest. The Brouwer fixed point theorem says that some molecule must return to the exact position at which it started.

We will deduce the Brouwer fixed point theorem as a consequence of the following lemma.

Lemma 181. *Let D be the closed unit ball in $\mathbb{R}^n$. There does not exist a smooth map $g\colon D \to \partial D$ such that $g(p) = p$ for all $p \in \partial D$.*

Remark 182. Intuitively, this says there is no way to retract the closed unit ball to its boundary without "ripping" the disc.

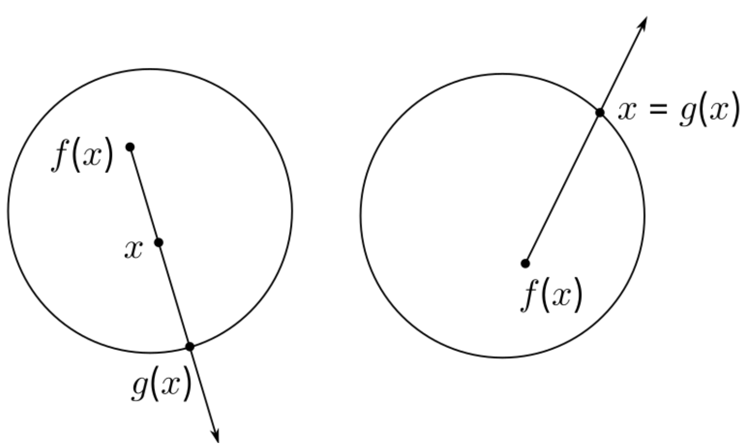
Proof. Suppose for contradiction that such a map g exists. Let F be the electric field generated by a point charge at the origin and consider the associated flux form α_F on $\mathbb{R}^n \setminus \{0\}$. Recall that $d\alpha_F = 0$. Also notice that g restricts to the identity map on ∂D and so $g_*(\partial D) = \partial D$. Hence we get

$$1 = \int_{\partial D} \alpha_F = \int_{g_*(\partial D)} \alpha_F = \int_{\partial D} g^*\alpha_F = \int_D d(g^*\alpha_F) = \int_D g^*(d\alpha_F) = 0.$$

This is clearly impossible. [Note that we are allowed to apply Stokes' theorem to $g^*\alpha_F$ since it is defined everywhere on D, whereas α_F is not.] $\square$

Now let's use the lemma to prove Brouwer's theorem.

Proof. (Brouwer Fixed Point Theorem) Suppose for contradiction there is a smooth map $f\colon D \to D$ that does not have a fixed point. For each $x \in D$, let $R(x)$ be the ray originating at $f(x)$ and passing through x. Note that $R(x)$ is well-defined since f has no fixed points. Now define $g\colon D \to \partial D$ by letting $g(x)$ be the point at which $R(x)$ intersects ∂D.



Note that g is smooth and that $g(x) = x$ for $x \in \partial D$. This is a contradiction because no such g exists by Lemma 181. $\square$

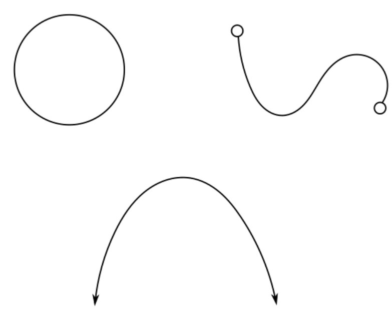
13 Manifolds

13.1 Intuition

We begin with an intuitive discussion of manifolds.

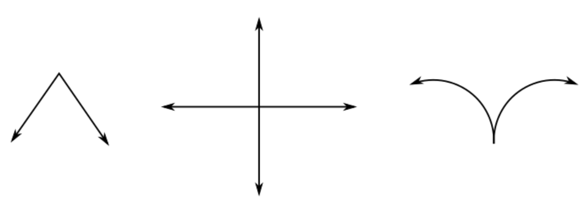
Definition 183. (Informal) An n-dimensional manifold is a set $M \subset \mathbb{R}^N$ such that M looks locally like a n-dimensional plane near each of its points.

Example 184. The following are examples of 1-dimensional manifolds in $\mathbb{R}^2$.



Notice that M can be bounded or unbounded. The important thing is that M has a well-defined tangent line at each of its points.

Example 185. The following are not examples of 1-manifolds.



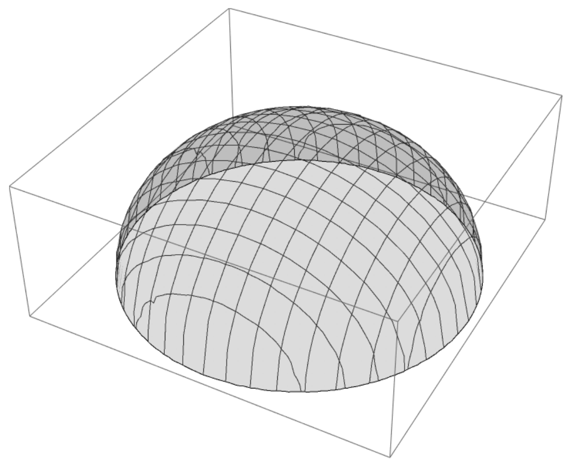
In each case there is a point where M is not locally modeled by a line.

Example 186. A ray is not a one manifold since it is locally modeled on a half-line near one of its points.

Nevertheless, this is still an important object that we would like to consider. It is an example of a manifold with boundary.

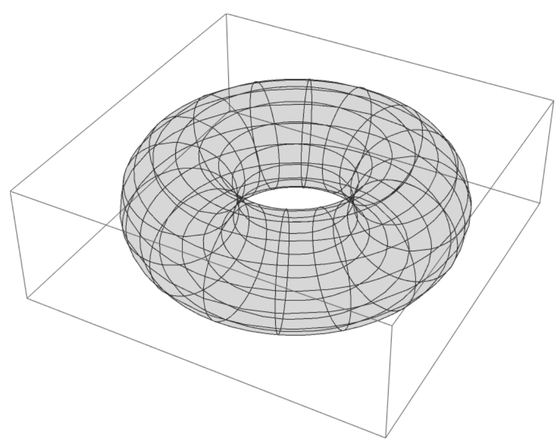
Definition 187. (Informal) A n-dimensional manifold with boundary is a set $M \subset \mathbb{R}^N$ that looks locally like either an n-dimensional plane or half of an n-dimensional plane near each of its points. The points where M looks like half of an n-dimensional plane are called the boundary points of M.

Example 188. Let M be the upper hemisphere of the unit sphere in $\mathbb{R}^3$. That is, let $M = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, \ z > 0\}$.



Then M is a 2-dimensional manifold. If we modified M to include the equator, then we would get a manifold with boundary.

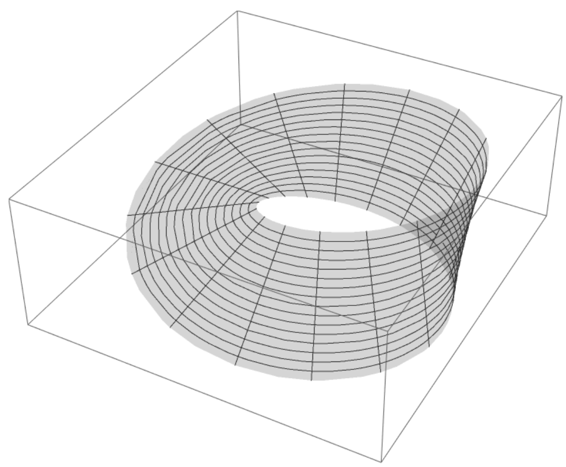
Example 189. A torus is the surface of a donut in $\mathbb{R}^3$.



It is an example of a 2-dimensional manifold. It is also possible to think about a torus as an abstract 2-dimensional manifold without giving an embedding into Euclidean space. Indeed, think about the screen in the game of Pac-Man.

When Pac-Man goes off one edge of the screen, he comes back on the other side. Thus we can think of the screen in Pac-Man as a square with the top edge glued to the bottom edge and the left edge glued to the right edge. Starting from a square, gluing the top edge to the bottom edge makes a cylinder, and then gluing the left edge to the right edge connects the two ends of the cylinder to make a torus.

Example 190. The Möbius band is a 2-dimensional manifold with boundary obtained by taking a rectangular strip, putting a single twist in it, and then gluing a pair of opposite edges together.



Unlike the previous examples we've seen, the Möbius band is one-sided. An ant that walks once around the Möbius band will come back to it's starting point on the opposite side of the band.

13.2 Formal Definitions

Now we proceed with the formal definition of a manifold. The basic building blocks of manifolds are called embeddings.

Definition 191. Let U be an open subset of $\mathbb{R}^n$. An embedding is a map $\psi\colon U \to \mathbb{R}^N$ such that

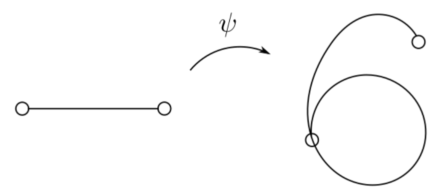
(i) ψ is injective,

(ii) $D\psi(x)$ has n linearly independent columns for every $x \in U$,

(iii) $\psi^{-1}\colon \psi(U) \to U$ is continuous.

The motivation for each of these conditions is the following. Condition (i) ensures that the image of ψ has no self-intersections. Condition (ii) ensures that $\psi(U)$ has a well-defined n-dimensional tangent plane at each point. Finally, condition (iii) prevents the image of ψ from "wrapping around to itself" as we approach the boundary of U.

Example 192. Consider the map $\psi\colon (-1,1) \to \mathbb{R}^2$ which maps the open interval into a 6-shaped figure.



This map ψ satisfies conditions (i) and (ii) for an embedding. However, it fails condition (iii). The map ψ^{-1} is not continuous from the 6-shaped figure to the open interval.

Example 193. One basic example of an embedding is the graph of a function. Let $u\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function. Then $\psi\colon \mathbb{R}^n \to \mathbb{R}^{n+1}$ given by

$$\psi(x_1,\ldots,x_n) = (x_1,\ldots,x_n,u(x_1,\ldots,x_n))$$

is an embedding that parameterizes the graph of u. To see that this satisfies conditions (i)-(iii), first observe that $\psi\colon U \to \psi(U)$ is invertible with inverse $\psi^{-1}\colon \psi(U) \to U$ given by $\psi^{-1}(x_1,\ldots,x_n,x_{n+1}) = (x_1,\ldots,x_n)$. The existence of an inverse shows that ψ is injective. Moreover, ψ^{-1} is continuous since it is the restriction of a continuous map defined on all of $\mathbb{R}^{n+1}$.

It remains to check that $D\psi(x_1,\ldots,x_n)$ has n linearly independent columns. Observe that

$$D\psi(x_1,\ldots,x_n) = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ * & * & \cdots & * \end{pmatrix}$$

where the stars stand for some numbers that don't matter. Call the columns $v_1,\ldots,v_n$ and suppose that $w = a_1v_1 + \ldots + a_nv_n = 0$. Observe that the first coordinate of w is a_1 and therefore $a_1 = 0$. The second coordinate of w is a_2 and therefore $a_2 = 0$. Likewise $a_i = 0$ for all other i. This shows that $v_1, \ldots, v_n$ are linearly independent.

Definition 194. An n-dimensional manifold is a subset M of $\mathbb{R}^N$ such that for each point $p \in M$ there exists

(i) an open set $U \subset \mathbb{R}^n$,

(ii) an open set $V \subset \mathbb{R}^N$ containing p, and

(iii) an embedding $\psi\colon U \to \mathbb{R}^N$ such that $M \cap V = \psi(U)$.

The map ψ is called a local parameterization for M near p. It's inverse $\psi^{-1}\colon \psi(U) \to U$ is called a chart for M near p. A collection of parameterizations $(\psi_i)_{i\in\mathcal{I}}$ is called an atlas for M provided M is contained in the union of the images $\psi_i(U_i)$.

Example 195. Let $u\colon \mathbb{R}^n \to \mathbb{R}$ be a smooth function and let $M \subset \mathbb{R}^{n+1}$ be the graph of u. Then M is an n-dimensional manifold. The map $\psi\colon \mathbb{R}^n \to \mathbb{R}^{n+1}$ given by $\psi(x) = (x, u(x))$ is a single parameterization that covers all of M.

Example 196. Let S^2 be the unit sphere in $\mathbb{R}^3$. Let's show that S^2 is a 2-dimensional manifold by exhibiting an explicit atlas for S^2. Let U be the open unit ball in $\mathbb{R}^2$. Note that $\{z > 0\}$ is an open set in $\mathbb{R}^3$ and $S^2 \cap \{z > 0\}$ consists of the upper hemisphere of S^2 without the equator. The map $\psi_{z>0}\colon U \to \mathbb{R}^3$ given by

$$\psi_{z>0}(x, y) = (x, y, \sqrt{1 - x^2 - y^2})$$

is an embedding (since it is the graph of a function) and $\psi(U) = S^2 \cap \{z > 0\}$. Likewise, there are five other embeddings

$$\begin{aligned}
\psi_{z<0}(x, y) &= (x, y, -\sqrt{1 - x^2 - y^2}), & \psi_{z<0}(U) &= S^2 \cap \{z < 0\} \\
\psi_{y>0}(x, z) &= (x, \sqrt{1 - x^2 - z^2}, z), & \psi_{y>0}(U) &= S^2 \cap \{y > 0\} \\
\psi_{y<0}(x, z) &= (x, -\sqrt{1 - x^2 - z^2}, z), & \psi_{y<0}(U) &= S^2 \cap \{y < 0\} \\
\psi_{x>0}(y, z) &= (\sqrt{1 - y^2 - z^2}, y, z), & \psi_{x>0}(U) &= S^2 \cap \{x > 0\} \\
\psi_{x<0}(y, z) &= (-\sqrt{1 - y^2 - z^2}, y, z), & \psi_{x<0}(U) &= S^2 \cap \{x < 0\}
\end{aligned}$$

each of which parameterizes a hemisphere of S^2. The collection of these six embeddings forms an atlas for S^2.

Example 197. Let T^2 be the torus in $\mathbb{R}^3$ obtained by taking the circle

$$(x-2)^2 + z^2 = 1$$

in the xz-plane and revolving it around the z-axis. We'd like to show that this is a 2-manifold by giving an explicit atlas. Note that

$$\begin{aligned}\psi(s,t) &= (2\cos s, 2\sin s, 0) + \cos t(\cos s, \sin s, 0) + \sin t(0,0,1) \\ &= (2\cos s + \cos s\cos t, 2\sin s + \sin s\cos t, \sin t)\end{aligned}$$

parameterizes T^2. However, ψ is not injective if we allow s and t to vary over too large a domain. Hence to obtain embeddings, define maps

$$\begin{aligned}\psi_1 &\colon \left(0, \frac{3\pi}{2}\right) \times \left(0, \frac{3\pi}{2}\right) \to \mathbb{R}^3, \\ \psi_2 &\colon \left(\pi, \frac{5\pi}{2}\right) \times \left(0, \frac{3\pi}{2}\right) \to \mathbb{R}^3, \\ \psi_3 &\colon \left(0, \frac{3\pi}{2}\right) \times \left(\pi, \frac{5\pi}{2}\right) \to \mathbb{R}^3, \\ \psi_4 &\colon \left(\pi, \frac{5\pi}{2}\right) \times \left(\pi, \frac{5\pi}{2}\right) \to \mathbb{R}^3,\end{aligned}$$

by $\psi_1(s,t) = \psi(s,t)$, $\psi_2(s,t) = \psi(s,t)$, $\psi_3(s,t) = \psi(s,t)$, and $\psi_4(s,t) = \psi(s,t)$. Together these four parameterizations form an atlas for T^2.

13.3 Tangent Spaces

Given an n-dimensional manifold in $\mathbb{R}^N$ we would like to define the tangent plane to M at p. This should be an n-dimensional plane $T_pM \subset T_p\mathbb{R}^N$.

Definition 198. Fix a point $p \in M$. Choose a local parameterization $\psi\colon U \to \mathbb{R}^N$ such that there is a point $x \in U$ with $\psi(x) = p$. The tangent space to M at p is the vector space spanned by the columns of $D\psi(x)$. This space has dimension n since the columns of $D\psi(x)$ are linearly independent. We denote the tangent space to M at p by T_pM.

Example 199. Let $u\colon \mathbb{R}^2 \to \mathbb{R}$ be a smooth function and let Σ_u be the graph of f. Then $\psi(x,y) = (x, y, u(x,y))$ is a parameterization for Σ_u. The columns of $D\psi$ are

$$\frac{\partial\psi}{\partial x} = \begin{pmatrix}1\\0\\\frac{\partial u}{\partial x}\end{pmatrix} \quad\text{and}\quad \frac{\partial\psi}{\partial y} = \begin{pmatrix}0\\1\\\frac{\partial u}{\partial y}\end{pmatrix}.$$

These span the tangent space $T_{(x,y,u(x,y))}\Sigma_u$.

14 The Regular Value Theorem

Even in the simple examples we have seen, it is already somewhat tedious to construct an atlas by hand. Fortunately, there are other more practical ways of checking that a set is a manifold. The most useful of these is the regular value theorem. This theorem describes criteria for the level set of a function to be a manifold.

14.1 The Implicit Function Theorem

Before we can prove the regular value theorem, we first need to discuss the implicit function theorem. We begin with a special case. Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ is smooth and $f(x_0, y_0) = 0$. We would like to know whether it is possible to write the level set $\{f = 0\}$ as the graph of a function of x in a small neighborhood of (x_0, y_0).

As a motivating calculation, suppose there is a function $g\colon (x_0-\varepsilon, x_0+\varepsilon) \to \mathbb{R}$ with $g(x_0) = y_0$ and $f(x, g(x)) = 0$. Then we can try to Taylor expand g using implicit differentiation:

$$0 = \frac{d}{dx}\bigg|_{x=x_0} \Big[f(x, g(x)) \Big] = \frac{\partial f}{\partial x}(x_0, y_0) + \frac{\partial f}{\partial y}(x_0, y_0) g'(x_0).$$

There are now two possibilities. If $\frac{\partial f}{\partial y}(x_0, y_0) = 0$ then this equation may have no solution. If the equation has no solution, we have arrived at a contradiction and g does not exist.

On the other hand we may have $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$. In this case we can solve for $g'(x_0)$ to get

$$g'(x_0) = -\frac{\frac{\partial f}{\partial x}(x_0, y_0)}{\frac{\partial f}{\partial y}(x_0, y_0)}$$

Now let's try to go one stage further and find $g''(x_0)$. By implicit differentiation again

$$\begin{aligned} 0 &= \frac{d}{dx}\left[\frac{\partial f}{\partial x}(x, g(x)) + \frac{\partial f}{\partial y}(x, g(x)) g'(x) \right] \\ &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x \partial y} g'(x) + \left[\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2} g'(x) \right] g'(x) + \frac{\partial f}{\partial y} g''(x). \end{aligned}$$

Since $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$, we can solve this equation to find $g''(x_0)$. Hence there is no new obstruction to finding $g''(x_0)$ beyond the first derivative being non-zero. Continuing in this fashion it is possible to compute $g'''(x_0)$, $g^{(4)}(x_0)$, etc.

While this does not constitute a proof, this strongly suggests that such a g should exist provided $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$. The implicit function theorem says that this is indeed the case.

Theorem 200 (Implicit Function Theorem, Special Case)**.** *Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ is smooth and $f(x_0, y_0) = 0$ and $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$. Then there is an open interval U centered at x_0 and an open interval V centered at y_0 and a smooth function $g\colon U \to V$ such that the solutions to $f(x, y) = 0$ in $U \times V$ are exactly the points $(x, g(x))$.*

Remark 201. Of course we can interchange the roles of x and y in the theorem. That is if $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$, then the level set $\{f = 0\}$ is locally given by the graph of a function of y near (x_0, y_0).

Example 202. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x^2 + y^2 - 1$ so that the level set $\{f = 0\}$ is the unit circle. Observe that

$$\frac{\partial f}{\partial y} = 2y$$

which is non-zero so long as $y \neq 0$. Hence the implicit function theorem guarantees that the unit circle can locally be written as the graph of the function of x about any point except for $(1, 0)$ and $(-1, 0)$. Likewise

$$\frac{\partial f}{\partial x} = 2x$$

which is non-zero so long as $x \neq 0$. Thus the implicit function theorem says the unit circle can locally be written as the graph of a function of y at any point except $(0, 1)$ and $(0, -1)$.

Now we state the general case of the implicit function theorem. First we set up the notation. Let $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be a smooth map. Write $x_1, \ldots, x_n$ for the coordinates on $\mathbb{R}^n$ and write $y_1, \ldots, y_m$ for the coordinates on $\mathbb{R}^m$. The derivative Df can be broken into two pieces:

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \quad \text{and} \quad \frac{\partial f}{\partial y} = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial f_m}{\partial y_m} \end{pmatrix}$$

Thus $\partial f/\partial x$ has m rows and n columns, while $\partial f/\partial y$ is an $m \times m$ square matrix.

Theorem 203 (Implicit Function Theorem)**.** *Let $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be a smooth map. Assume that $f(x_0, y_0) = 0$ and that*

$$\left(\frac{\partial f}{\partial y}\right)(x_0, y_0)$$

is invertible. Then there is an open set $U \subset \mathbb{R}^n$ containing x_0 and an open set V in $\mathbb{R}^m$ containing y_0 and a smooth function $g\colon U \to V$ such that the solutions to $f(x, y) = 0$ in $U \times V$ are exactly the points of the form $(x, g(x))$. In other words, the level set $\{f = 0\}$ can be written as the graph of a function of x near (x_0, y_0).

14.2 The Regular Value Theorem

The regular value theorem gives a condition for the level set of a function to be a manifold.

Theorem 204 (Regular Value Theorem)**.** *Let $f\colon \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a smooth function. Fix $c \in \mathbb{R}^n$ and assume that at each point $p \in \{f = c\}$ the derivative $(Df)(p)$ has rank m (i.e. has m linearly independent columns). Then the level set $\{f = c\}$ is an n-dimensional manifold in $\mathbb{R}^{n+m}$.*

Proof. Let f and c be as in the statement of the theorem. The idea is to use the implicit function theorem to construct charts for $\{f = c\}$. Fix a point p in the level set $\{f = c\}$. By assumption $(Df)(p)$ has m linearly independent columns. Therefore we can relabel the coordinates on $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ as $x_1, \ldots, x_n, y_1, \ldots, y_m$ in such a way that

$$\left(\frac{\partial f}{\partial y}\right)(x_0, y_0)$$

is invertible. Then by the implicit function theorem there is an open set $U \subset \mathbb{R}^n$ containing x_0 and an open set $V \subset \mathbb{R}^m$ containing y_0 and a smooth function $g\colon U \to \mathbb{R}^m$ such that $\{f = c\} \cap (U \times V)$ is the graph of g. Thus $\psi\colon U \to \mathbb{R}^{n+m}$ given by $\psi(x) = (x, g(x))$ is an embedding such that

$$\{f = c\} \cap (U \times V) = \psi(U).$$

It follows that ψ is a local parameterization of $\{f = c\}$ near p. Since p was arbitrary, this proves that $\{f = c\}$ is a manifold. $\square$

It is also possible to compute the tangent space to the level set $\{f = c\}$ in terms of Df.

Proposition 205. *Let $f\colon \mathbb{R}^{n+m} \to \mathbb{R}^m$ be a smooth function. Fix $c \in \mathbb{R}^n$ and assume that at each point $p \in \{f = c\}$ the derivative $(Df)(p)$ has rank m. Then the tangent space to $\{f = c\}$ at p is* $\ker(Df(p))$.

Proof. Let $M = \{f = c\}$ which is an n-dimensional manifold by the regular value theorem. Let $\psi\colon U \to \mathbb{R}^{n+m}$ be a local parameterization for M near p. Let x be the point in U such that $\psi(x) = p$. By definition, the tangent space T_pM is spanned by the columns of $D\psi(x)$. Note that $f(\psi(x + te_i)) = c$ for all t. Differentiating this equation in t setting $t = 0$ gives

$$Df(p)[D\psi(x)e_i] = 0.$$

Therefore the ith column of $D\psi(x)$ belongs to $\ker(Df(p))$. Since i was arbitrary, this implies that $\ker(Df(p))$ contains T_pM. On the other hand, since $Df(p)$ has rank m, it follows that $\ker(Df(p))$ is n-dimensional. Since T_pM is also n-dimensional and $\ker(Df(p)) \supseteq T_pM$, it follows that we must have equality. □

Example 206. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = x^2 - y^2$. Then $Df(x, y) = (2x, -2y)$. This has rank 1 unless $(x, y) = (0, 0)$. Therefore for every $c \neq 0$, the level set $\{f = c\}$ is a 1-dimensional manifold by the regular value theorem. In this case, we can also verify this by hand since equation $x^2 - y^2 = c$ defines a hyperbola for $c \neq 0$.

The tangent space to $\{f = 3\}$ at the point $(2, 1)$ is the kernel of $Df(2, 1)$. Now $Df(2, 1) = (4, -2)$ and the kernel of this matrix is spanned by

$$v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Therefore the tangent space to $\{f = 3\}$ at $(2, 1)$ is spanned by v.

14.3 Manifolds of Matrices

To illustrate the utility of the regular value theorem, we will now show that several natural collections of matrices form manifolds. The space of all $n \times n$ matrices is denoted $\mathbb{R}^{n\times n}$. It is naturally identified with n^2-dimensional Euclidean space. There are many interesting subsets of $\mathbb{R}^{n\times n}$. For example, consider

(i) $\mathrm{GL}(n) = \{A \in \mathbb{R}^{n\times n} : \det(A) \neq 0\}$,

(ii) $\mathrm{SL}(n) = \{A \in \mathbb{R}^{n\times n} : \det(A) = 1\}$,

(iii) $\mathrm{O}(n) = \{A \in \mathbb{R}^{n\times n} :\ A^T A = I\}$.

The general linear group $\mathrm{GL}(n)$ consists of all invertible $n \times n$ matrices. The special linear group $\mathrm{SL}(n)$ consists of all $n \times n$ matrices with determinant one. These are exactly the matrices that preserve the oriented volume of n-dimensional parallelepipeds. Finally the orthogonal group $\mathrm{O}(n)$ consists of all rotation and reflection matrices.

Proposition 207. *The general linear group* $\mathrm{GL}(n)$ *is an* n^2*-dimensional manifold in* $\mathbb{R}^{n\times n}$*.*

Proof. It is enough to check that $\mathrm{GL}(n)$ is an open subset of $\mathbb{R}^{n\times n}$. This follows from the fact that the determinant is a continuous function of the entries in a matrix. Hence if A has non-zero determinant, then any small perturbation of A will also have non-zero determinant. □

Proposition 208. *The special linear group* $\mathrm{SL}(n)$ *is an* (n^2-1)*-dimensional manifold in* $\mathbb{R}^{n\times n}$*.*

Proof. Consider the function $\det\colon \mathbb{R}^{n\times n} \to \mathbb{R}$. The special linear group is exactly the level set $\{\det = 1\}$. Therefore the regular value theorem will give the conclusion provided $(D\det)(A)$ has rank 1 at every matrix $A \in \mathrm{SL}(n)$.

To verify this, we need to compute $D\det(A)$. We begin with the special case $A = I$. In this case

$$(D\det)(I)B = \lim_{t\to 0} \frac{\det(I+tB) - \det(I)}{t}.$$

Let $b_1, \ldots, b_n$ be the columns of B. Then by the scaling and additivity properties of determinant

$$\begin{aligned}\det(I+tB) &= \det(e_1 + tb_1, \ldots, e_n + tb_n)\\ &= \det(e_1, \ldots, e_n) + t\sum_{i=1}^{n} \det(e_1, \ldots, e_{i-1}, b_i, e_{i+1}, \ldots, e_n)\\ &\qquad + \text{terms involving higher powers of } t.\end{aligned}$$

Now observe that $\det(e_1, \ldots, e_{i-1}, b_i, e_{i+1}, \ldots, e_n) = b_{ii}$ and hence

$$\det(I + tB) = 1 + t\,\mathrm{tr}(B) + \text{terms involving higher powers of } t.$$

Thus

$$\lim_{t\to 0} \frac{\det(I + tB) - \det(I)}{t} = \mathrm{tr}(B)$$

and so $(D\det)(I)B$ equals the trace of B.

Now consider an arbitrary invertible $n \times n$ matrix A. We have

$$(D\det)(A)B = \lim_{t\to 0} \frac{\det(A + tB) - \det(A)}{t}.$$

We can compute this limit by factoring out an A and using the computation we've already done at the identity:

$$\begin{aligned}\lim_{t\to 0} \frac{\det(A + tB) - \det(A)}{t} &= \det(A) \lim_{t\to 0} \frac{\det(I + tA^{-1}B) - \det(I)}{t} \\ &= \det(A)\,\mathrm{tr}(A^{-1}B).\end{aligned}$$

Thus $(D\det)(A)B = \det(A^{-1})\,\mathrm{tr}(A^{-1}B)$.

Now fix a matrix $A \in \mathrm{SL}(n)$. To show that $(D\det)(A)$ has rank 1, it is enough to show that $(D\det)(A)\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ is not the zero map. But this is obvious from the previous formula since

$$(D\det)(A)A = \det(A^{-1})\,\mathrm{tr}(A^{-1}A) = \mathrm{tr}(I) = n \neq 0.$$

Hence the regular value theorem applies and yields that $\mathrm{SL}(n) = \{\det = 1\}$ is an $(n^2 - 1)$-dimensional manifold. $\square$

Proposition 209. *The orthogonal group* $\mathrm{O}(n)$ *is an* $\frac{n(n-1)}{2}$ *dimensional manifold in* $\mathbb{R}^{n\times n}$.

Proof. Define a function $f\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ by setting $f(A) = A^T A$. Then the orthogonal group is exactly the level set $\{f = I\}$. We would like to apply the regular value theorem. However, we are not yet in a position to do this. Indeed, if we could apply the regular value theorem to this f, it would tell us that the level set $\{f = I\}$ is a manifold of dimension $n^2 - n^2 = 0$! But this is clearly not the case.

To get around this, observe that

$$(A^T A)^T = A^T A.$$

In other words, the matrix $A^T A$ is always symmetric. Let S be the set of all symmetric $n \times n$ matrices. Since a symmetric matrix is determined by its entries on and above the diagonal, the space S is naturally identified with $\mathbb{R}^{n(n+1)/2}$. Now consider f as a map $\mathbb{R}^{n\times n} \to S$. The orthogonal group is still the level set $\{f = I\}$. But now the regular value theorem, if applicable, will tell us that $\{f = I\}$ is a manifold of dimension

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

as desired.

Thus we aim to apply the regular value theorem to the map $f\colon \mathbb{R}^{n\times n} \to S$. We need to check that $(Df)(A)$ has rank $\frac{n(n+1)}{2}$ for all $A \in \mathrm{O}(n)$. First let's find a formula for $(Df)(A)B$. We have

$$\begin{aligned}(Df)(A)B &= \lim_{t\to 0} \frac{f(A+tB) - f(A)}{t} \\ &= \lim_{t\to 0} \frac{(A+tB)^T(A+tB) - A^T A}{t} = A^T B + B^T A.\end{aligned}$$

Now fix a matrix $A \in \mathrm{O}(n)$. To show that $(Df)(A)$ has rank $\frac{n(n+1)}{2}$, it is equivalent to show that $(Df)(A)$ is surjective as a map from $\mathbb{R}^{n\times n} \to S$.

So fix a matrix $C \in S$. We want to find a matrix B such that

$$A^T B + B^T A = C = \frac{C}{2} + \frac{C^T}{2}.$$

Since $B^T A = (A^T B)^T$, it is equivalent to find a matrix B such that

$$A^T B = \frac{C}{2}.$$

Since A belongs to $\mathrm{O}(n)$, we know that $AA^T = I$ and therefore $B = \frac{AC}{2}$ is a solution to the previous equation. This proves that $(Df)(A)$ is surjective for every $A \in \{f = I\}$. Therefore the regular value theorem applies and yields that $\mathrm{O}(n) = \{f = I\}$ is a manifold of dimension $\frac{n(n-1)}{2}$. $\square$

15 Differential Forms on Manifolds

15.1 Definition and First Examples

A differential form on a manifold is still a rule that assigns a number to oriented parallelepipeds. However, a form on M can only be applied to parallelepipeds that are tangent to M. As before, if ω is a form on M, we write $\omega_p(v_1, \dots, v_n)$ for the value of ω on the parallelepiped based at p and spanned by the vectors $v_1, \dots, v_n \in T_pM$. We shall again require that ω_p satisfies the scaling, additivity, and alternation properties. Moreover, we require ω_p to depend smoothly on p. (We will give the precise definition of what this means later after discussing how to write forms in coordinates.)

Example 210. Let S^2 be the unit sphere in $\mathbb{R}^3$. Then

$$\omega_p(v) = v \cdot e_3, \quad v \in T_pS^2$$

is a 1-form on S^2.

The previous example is a special case of a more general construction.

Definition 211. Let $M \subset \mathbb{R}^N$ be a manifold. A vector field F on M is a choice of vector $F(p) \in T_p\mathbb{R}^N$ for each $p \in M$. The vector $F(p)$ is required to depend smoothly on p.

Remark 212. Note that the vector $F(p)$ need not be tangent to M at p. In the case that $F(p) \in T_pM$ for all $p \in M$ we say that F is tangential.

Example 213. On $S^2 \subset \mathbb{R}^3$ let $F(p) = p$. This is the unit outward normal to the unit sphere.

Example 214. On $S^2 \subset \mathbb{R}^3$ let $F(p) = e_3$. This vector field is not tangential. We could build a tangential vector field from F by projecting each vector $F(p)$ to T_pM. Since the normal vector to T_pM is p, this projected vector field is given by the formula $F^T(p) = e_3 - (e_3 \cdot p)p$.

Given a vector field F on M there is an associated work 1-form given by $(\omega_F)_p(v) = F(p) \cdot v$. Thus the form in Example 210 is the work 1-form for the vector field $F(p) = e_3$. When defining a work form, only the tangential part of a vector field matters. Indeed, given a vector field F on M let $F = F^T + F^\perp$ be its decomposition into tangential and normal components. Then

$$(\omega_F)_p(v) = F(p) \cdot v = (F^T(p) + F^\perp(p)) \cdot v = F^T(p) \cdot v = (\omega_{F^T})_p(v)$$

since $v \cdot F^\perp(p) = 0$ for $v \in T_pM$. Hence when defining work forms, it is enough to consider only tangential vector fields.

Example 215. Consider $S^2 \subset \mathbb{R}^3$. The area form ω on S^2 is a 2-form given by $\omega_p(v, w) = \det(p, v, w)$. Since p has length one and is normal to T_pM, this is just the oriented area of the parallelogram spanned by v and w.

15.2 Forms in Coordinates

In Euclidean space, we were able to express forms in terms of the coordinates dx_1, …, dx_n. Writing forms in coordinates like this was often very convenient for doing computations. We'd like to have a way to write forms on manifolds in coordinates. This can be done by means of a parameterization.

Let α be a k-form on an n-dimensional manifold $M \subset \mathbb{R}^N$. Pick a point $p \in M$. By the definition of a manifold, there is a local parameterization $\psi\colon U \subset \mathbb{R}^n \to \mathbb{R}^N$ parameterizing a neighborhood of p in M. We can consider the pullback $\psi^*\alpha$. This is the k-form on U given by

$$(\psi^*\alpha)_x(v_1, \ldots, v_k) = \alpha_{\psi(x)}((D\psi)v_1, \ldots, (D\psi)v_k).$$

Note that the right hand side makes sense, since each $(D\psi)v_i$ is an element of T_pM. Since the left hand side is a form on an open subset of $\mathbb{R}^n$, we can express it in terms of dx_1, …, dx_n:

$$\psi^*\alpha = \sum_I f_I \, dx_I.$$

This is the formula for α in the local coordinates induced by ψ.

Remark 216. When we say that α_p depends smoothly on p, we mean that when we express α in any coordinate system

$$\psi^*\alpha = \sum_I f_I \, dx_I$$

the functions f_I are smooth functions on U.

Example 217. On S^2 consider the 1-form $\alpha_p(v) = e_3 \cdot v$. Let's write this in terms of the local coordinates given by $\psi(x,y) = (x, y, \sqrt{1-x^2-y^2})$. To do this, we need to find

$$(\psi^*\alpha)_{(x,y)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad (\psi^*\alpha)_{(x,y)} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Observe that

$$D\psi(x,y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{-x}{\sqrt{1-x^2-y^2}} & \frac{-y}{\sqrt{1-x^2-y^2}}. \end{pmatrix}$$

Therefore it follows that

$$(\psi^*\alpha)_{(x,y)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \alpha_{(x,y,\sqrt{1-x^2-y^2})} \begin{pmatrix} 1 \\ 0 \\ \frac{-x}{\sqrt{1-x^2-y^2}} \end{pmatrix} = \frac{-x}{\sqrt{1-x^2-y^2}},$$

$$(\psi^*\alpha)_{(x,y)} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \alpha_{(x,y,\sqrt{1-x^2-y^2})} \begin{pmatrix} 0 \\ 1 \\ \frac{-y}{\sqrt{1-x^2-y^2}} \end{pmatrix} = \frac{-y}{\sqrt{1-x^2-y^2}},$$

Hence

$$\psi^*\alpha = -\frac{x}{\sqrt{1-x^2-y^2}}\,dx - \frac{y}{\sqrt{1-x^2-y^2}}\,dy$$

is the expression for α in local coordinates.

Example 218. On $S^2 \subset \mathbb{R}^3$ consider the area form $\omega_p(v,w) = \det(p,v,w)$. Again let $\psi(x,y) = (x,y,\sqrt{1-x^2-y^2})$. To write ω in local coordinates, we need to find

$$(\psi^*\omega)_{(x,y)} \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right).$$

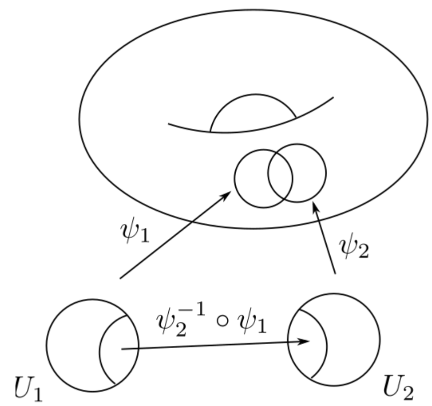
Using the expression for $D\psi$ from above, we compute that

$$\begin{aligned}
(\psi^*\omega)_{(x,y)} \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) &= \omega_{(x,y,\sqrt{1-x^2-y^2})} \left(\begin{pmatrix} 1 \\ 0 \\ \frac{-x}{\sqrt{1-x^2-y^2}} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \frac{-y}{\sqrt{1-x^2-y^2}} \end{pmatrix} \right) \\
&= \det \begin{pmatrix} x & 1 & 0 \\ y & 0 & 1 \\ \sqrt{1-x^2-y^2} & \frac{-x}{\sqrt{1-x^2-y^2}} & \frac{-y}{\sqrt{1-x^2-y^2}} \end{pmatrix} \\
&= \frac{1}{\sqrt{1-x^2-y^2}}.
\end{aligned}$$

Thus $\psi^*\omega = \frac{1}{\sqrt{1-x^2-y^2}}\,dx\,dy$ in coordinates.

15.3 Change of Coordinates Formula

Let α be a k-form on a manifold M. If two parameterizations ψ_1 and ψ_2 for M overlap, then we get two different expressions for α in local coordinates. We would like to know how to go back and forth between these different coordinate formulas.



Suppose $\psi_1\colon U_1 \to \mathbb{R}^n$ and $\psi_2\colon U_2 \to \mathbb{R}^n$ are two parameterizations such that $\psi_1(U_1)$ and $\psi_2(U_2)$ overlap. Then $\psi_2^{-1} \circ \psi_1$ is well-defined on $\psi_1^{-1}(\psi_2(U_2))$. Moreover $\psi_2 \circ (\psi_2^{-1} \circ \psi_1) = \psi_1$ and so

$$\psi_1^*\alpha = (\psi_2^{-1} \circ \psi_1)^* \psi_2^*\alpha \quad \text{on } \psi_1^{-1}(\psi_2(U_2)).$$

This is called the change of coordinates formula.

Example 219. On $S^2 \subset \mathbb{R}^3$ consider the area form ω. Consider the two parameterizations

$$\psi_1(x,y) = (x, y, \sqrt{1-x^2-y^2}),$$
$$\psi_2(y,z) = (\sqrt{1-y^2-z^2}, y, z).$$

We've already computed that

$$\psi_1^*\omega = \frac{1}{\sqrt{1-x^2-y^2}}\,dx\,dy$$

in Example 218. A very similar computation yields that

$$\psi_2^*\omega = \frac{1}{\sqrt{1-y^2-z^2}}\, dy\, dz.$$

Note that $\psi_2^{-1} \circ \psi_1$ is defined on the set $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1, x > 0\}$. Moreover,

$$\psi_2^{-1} \circ \psi_1(x, y) = \psi_2^{-1}(x, y, \sqrt{1-x^2-y^2}) = (y, \sqrt{1-x^2-y^2})$$

on its domain of definition.

Let's verify the change of coordinates formula:

$$\begin{aligned}
(\psi_2^{-1} \circ \psi_1)^*\psi_2^*\omega &= (\psi_2^{-1} \circ \psi_1)^* \frac{1}{\sqrt{1-y^2-z^2}}\, dy\, dz \\
&= \frac{1}{\sqrt{1-y^2-(\sqrt{1-x^2-y^2})^2}}\, dy\, d(\sqrt{1-x^2-y^2}) \\
&= \frac{1}{\sqrt{1-y^2-(1-x^2-y^2)}}\, dy \frac{-x}{\sqrt{1-x^2-y^2}}\, dx \\
&= \frac{1}{\sqrt{1-x^2-y^2}}\, dx\, dy.
\end{aligned}$$

Note that we used the fact that $\sqrt{x^2} = x$ since $x > 0$ to get the final equality. Thus $(\psi_2^{-1} \circ \psi_1)^*\psi_2^*\omega = \psi_1^*\omega$ where defined.

15.4 Defining Forms in Coordinates

In Euclidean space we can also define differential forms by specifying them in coordinates. If we choose a smooth function f_I for each increasing multi-index of length k, then we can define a k-form α by setting

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

We can attempt to define forms on manifolds in a similar fashion. Suppose M is an n dimensional manifold in $\mathbb{R}^N$. Assume for simplicity that M has a finite atlas $\psi_1, \ldots, \psi_m$. For each integer $i = 1, \ldots, m$ and each increasing multi-index I of length k, pick a smooth function $f^i_I\colon U_i \to \mathbb{R}$. Then we can try to define a k-form α on M by stipulating that

$$\psi_i^* \alpha = \sum_I f^i_I \, dx_I$$

on U_i. In order for this to give rise to a well-defined form on M, the various formulas must agree about what α does when two coordinate patches overlap. In other words, the change of variables formula

$$(\psi_j^{-1} \circ \psi_i)^* \psi_j^* \alpha = \psi_i^* \alpha$$

must hold on $\psi_i^{-1}(\psi_j(U_j))$. If this change of variables formula is satisfied for all i and j, then there is indeed a unique k-form α on M which is given by the specified formulas in coordinates. We summarize this in the following proposition.

Proposition 220 (Defining forms in coordinates)**.** *Let M be an n-dimensional manifold in $\mathbb{R}^N$. Let $\psi_1, \ldots, \psi_m$ be an atlas for M. For each integer $i = 1, \ldots, m$ and each increasing multi-index I of length k, pick a smooth function $f^i_I\colon U_i \to \mathbb{R}$. Define k-forms β_i on U_i by*

$$\beta_i = \sum_I f^i_I \, dx_I.$$

If the change of variables formula $(\psi_j^{-1} \circ \psi_i)^* \beta_j = \beta_i$ holds on $\psi_i^{-1}(\psi_j(U_j))$ for all i and j then there is a unique k-form α on M such that $\psi_i^* \alpha = \beta_i$ for all i.

16 Volume Forms and Orientation

We already seen how to define an area 2-form ω on S^2 that measures the oriented area of parallelograms tangent to S^2. To do this, we took a parallelogram tangent to S^2 and completed it to a parallelepiped by appending a unit normal vector. Then we took the oriented volume of this parallelepiped. This same construction works more generally for two-sided hypersurfaces.

An n-dimensional manifold M in $\mathbb{R}^{n+1}$ is called a hypersurface. A hypersurface M is called two-sided provided there exists a smooth, unit normal vector field on M. That is, M is two-sided if we can find a smooth vector field N on M such that $N(p)$ has unit length and is perpendicular to T_pM for all $p \in M$.

Example 221. The unit sphere S^n in $\mathbb{R}^{n+1}$ is a two-sided hypersurface since $N(p) = p$ is a smooth unit normal vector field. The Möbius band is an example of a hypersurface which is not two-sided.

Given a two-sided hypersurface $M \subset \mathbb{R}^{n+1}$, it is possible to define a volume form ω on M by

$$\omega_p(v_1, \ldots, v_n) = \det(N(p), v_1, \ldots, v_n).$$

This computes the oriented volume of n-dimensional parallelepipeds tangent to M.

16.1 Area of n-dimensional parallelepipeds in $\mathbb{R}^N$.

We'd like to be able to define a volume form on an n-dimensional manifold M contained in $\mathbb{R}^N$. In order to do this, we need to know how to compute the volume of an n-dimensional parallelepiped in $\mathbb{R}^N$. So suppose $a_1, \ldots, a_n$ are n vectors in $\mathbb{R}^N$:

$$a_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{Ni} \end{pmatrix}.$$

What is the volume of the parallelepiped spanned by $a_1, \ldots, a_n$?

There is one special case where the answer is easy to compute. Suppose that actually $a_1, \ldots, a_n$ all live in the $x_1 \cdots x_n$-coordinate plane. That is, suppose that each a_i has the form

$$a_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Then we can think of the parallelepiped spanned by $a_1, \ldots, a_n$ as an n-dimensional parallelepiped in $\mathbb{R}^n$ and hence its volume is given by the absolute value of the determinant:

$$\mathrm{Vol}(a_1, \ldots, a_n) = \left| \det \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \right|.$$

We would like to re-express this in terms of the original vectors a_i and not the vectors a_i with the zeros removed.

To do this, let A be the $N \times n$ matrix with columns $a_1, \dots, a_n$. Observe that

$$\begin{aligned}
\mathrm{Vol}(a_1, \dots, a_n)^2 &= \det \left[\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}^T \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \right] \\
&= \det \left[\begin{pmatrix} a_{11} & \cdots & a_{n1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & & \\ a_{1n} & \cdots & a_{nn} & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \\ 0 & & 0 \\ \vdots & & \vdots \\ 0 & & 0 \end{pmatrix} \right] \\
&= \det(A^T A).
\end{aligned}$$

Therefore $\mathrm{Vol}(a_1, \dots, a_n) = \sqrt{\det(A^T A)}$.

Now consider the general case where a_i can be any vectors in $\mathbb{R}^N$. We can reduce this case to the previous one via rotation. Indeed, we can always find an $N \times N$ rotation matrix Q such that the vectors $b_i = Q a_i$ belong to the $x_1 \cdots x_n$-plane. Since rotation doesn't change volume

$$\mathrm{Vol}(a_1, \dots, a_n) = \sqrt{\det(B^T B)}$$

where B is the $N \times n$ matrix with columns $b_1, \dots, b_n$. Since Q is a rotation matrix, it satisfies $Q^T Q = Q Q^T = I$. Therefore $a_i = Q^T b_i$ and $A = Q^T B$. It follows that

$$\det(B^T B) = \det(B^T Q Q^T B) = \det((Q^T B)^T (Q^T B)) = \det(A^T A).$$

Hence we again get that $\mathrm{Vol}(a_1, \dots, a_n) = \sqrt{\det(A^T A)}$. We have now proved the following proposition.

Proposition 222. *The volume of the n-dimensional parallelepiped in $\mathbb{R}^N$ spanned by $a_1, \dots, a_n$ is given by* $\mathrm{Vol}(a_1, \dots, a_n) = \sqrt{\det(A^T A)}$.

Remark 223. One can show that

$$\det(A^T A) = \sum_I \left[\det \begin{pmatrix} a_{i_1 1} & \cdots & a_{i_1 n} \\ \vdots & \ddots & \vdots \\ a_{i_n 1} & \cdots & a_{i_n n} \end{pmatrix} \right]^2$$

where the sum is taken over all increasing multi-indices of length n in $\mathbb{R}^N$. This is a consequence of the Cauchy-Binet formula (which we won't prove). The above formula can be thought of as a generalized Pythagorean theorem. Indeed, let P be the parallelepiped spanned by $a_1, \ldots, a_n$. For each increasing multi-index I of length n, let P_I be the parallelepiped obtained by projecting P to the $x_{i_1} \cdots x_{i_n}$-coordinate plane. Then the formula says that $\mathrm{Vol}(P)^2 = \sum_I \mathrm{Vol}(P_I)^2$.

16.2 Volume Forms

Equipped with the previous formula, we can now define volume forms on general manifolds. Let M be an n-dimensional manifold in $\mathbb{R}^N$. As a first attempt, we can try to define a volume form ω on M by

$$\omega_p(v_1, \ldots, v_n) = \sqrt{\det(V^T V)}$$

where V is the $N \times n$ matrix with columns $v_1, \ldots, v_n$. However, this is *not* a differential form since it fails to satisfy the alternating property. The problem is that ω does not take orientation into account.

As a second attempt, we can try to define a volume form by specifying it in coordinates. Let $\psi_1, \ldots, \psi_m$ be an atlas for M. For each $i = 1, \ldots, m$ define an n-form β_i on U_i by

$$\beta_i = \sqrt{\det\left[D\psi^T D\psi\right]}\, dx_1 \ldots dx_n.$$

These forms will piece together to define a form on M provided the compatibility conditions are satisfied:

$$\beta_i = (\psi_j^{-1} \circ \psi_i)^* \beta_j \quad \text{on } \psi_i^{-1}(\psi_j(U_j)).$$

Let $f = \psi_j^{-1} \circ \psi_i$. Let $x_1, \ldots, x_n$ be the coordinates on U_i and let $y_1, \ldots, y_n$ be the coordinates on U_j. Then

$$\begin{aligned}
f^*\beta_j &= \sqrt{\det\left[D\psi_j(f(x))^T D\psi_j(f(x))\right]}\, f^*(dy_1 \ldots dy_n) \\
&= \sqrt{\det\left[D\psi_j(f(x))^T D\psi_j(f(x))\right]}\, \det(Df(x))\, dx_1 \ldots dx_n \\
&= \pm\sqrt{\det\left[Df(x)^T D\psi_j(f(x))^T D\psi_j(f(x)) Df(x)\right]}\, dx_1 \ldots\, dx_n \\
&= \pm\sqrt{\det\left[D\psi_i(x)^T D\psi_i(x)\right]}\, dx_1 \ldots\, dx_n \\
&= \pm\beta_i.
\end{aligned}$$

Here we've used the fact that $\psi_j \circ f = \psi_i$ and the chain rule to get from the third to the fourth line. Also the formula has the $+$ sign if $\det(Df(x)) > 0$ and the $-$ sign if $\det(Df(x)) < 0$. Thus the compatibility conditions will be satisfied provided $\det(D(\psi_j^{-1} \circ \psi_i)) > 0$ whenever defined. This motivates the following definition.

Definition 224. An n-dimensional manifold is called orientable if there is an atlas for M with the property that $\det(D(\psi_j^{-1} \circ \psi_i)) > 0$ whenever defined. Such an atlas is called an oriented atlas for M.

Definition 225. Given an oriented n-dimensional manifold M equipped with an oriented atlas, there is a unique n-dimensional form ω on M such that

$$\psi_i^* \omega = \sqrt{\det[D\psi^T D\psi]}\, dx_1 \dots dx_n$$

on U_i for all i. This form ω is called the volume form for M.

16.3 Orientation

Let V be an n-dimensional vector space. An ordered basis for V is an ordered list of vectors $v_1, \dots, v_n$ such that $v_1, \dots, v_n$ forms a basis for V. For example, e_1, e_2, e_3 is an ordered basis for $\mathbb{R}^3$ and e_2, e_1, e_3 is another (different) ordered basis for $\mathbb{R}^3$. Given two ordered bases $v_1, \dots, v_n$ and $w_1, \dots, w_n$ for V, there is a unique change of basis matrix A satisfying $Av_i = w_i$ for $i = 1, \dots, n$.

Definition 226. Let V be an n-dimensional vector space. An orientation on V is a function

$$\mathrm{or} \colon \{\text{ordered bases for } V\} \to \{\pm 1\}$$

which satisfies the following property:

$$\mathrm{or}(v_1, \dots, v_n) = \mathrm{sign}(\det A)\mathrm{or}(w_1, \dots, w_n) \tag{5}$$

whenever $v_1, \dots, v_n$ and $w_1, \dots, w_n$ are two ordered bases for V and A is the associated change of basis matrix.

Given a vector space V and an orientation or on V we say an ordered basis $v_1, \dots, v_n$ is positively oriented if $\mathrm{or}(v_1, \dots, v_n) = 1$ and negatively oriented if $\mathrm{or}(v_1, \dots, v_n) = -1$. Note in particular that swapping two basis vectors turns a positively oriented basis into a negatively oriented basis and vice versa. In other words $\mathrm{or}(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -\mathrm{or}(v_1, \dots, v_j, \dots, v_i, \dots, v_n)$.

Example 227. The standard orientation on $\mathbb{R}^3$ is

$$\operatorname{or}(v_1, v_2, v_3) = \begin{cases} +1, & \text{if } v_1, v_2, v_3 \text{ is right handed} \\ -1, & \text{if } v_1, v_2, v_3 \text{ is left handed.} \end{cases}$$

There is also an "unstandard" orientation on $\mathbb{R}^3$ given by

$$\overline{\operatorname{or}}(v_1, v_2, v_3) = \begin{cases} +1, & \text{if } v_1, v_2, v_3 \text{ is left handed} \\ -1, & \text{if } v_1, v_2, v_3 \text{ is right handed.} \end{cases}$$

Proposition 228. *Let V be an n-dimensional vector space. Then there are exactly two orientations on V.*

Proof. Property (5) implies that an orientation function is completely determined by its value on a single ordered basis. Thus there are at most two orientation functions. On the other hand, fix an ordered basis $v_1, \ldots, v_n$. Define

$$\operatorname{or}(w_1, \ldots, w_n) = \det(A) \quad \text{and} \quad \overline{\operatorname{or}}(w_1, \ldots, w_n) = -\det(A)$$

where A is the change of basis matrix from $v_1, \ldots, v_n$ to $w_1, \ldots, w_n$. Then or and $\overline{\operatorname{or}}$ are orientation functions. □

There is an equivalent definition of oriented manifolds in terms of orientations on tangent spaces.

Definition 229. An n-dimensional manifold M in $\mathbb{R}^N$ is orientable if for each $p \in M$ we can choose an orientation or_p on T_pM in such a way that or_p varies smoothly with p. Here varying smoothly with p means that for any parameterization $\psi\colon U \to \mathbb{R}^N$ the function $\operatorname{or}_{\psi(x)}((D\psi)(x)e_1, \ldots, (D\psi)(x)e_n)$ is constant on each connected component of U.

Proposition 230. *A manifold M is orientable in the sense of Definition 224 if and only if it is orientable in the sense of Definition 229.*

Proof. (Sketch) Assume that M has an oriented atlas. Let p be a point in M and choose a parameterization $\psi\colon U \to \mathbb{R}^N$ such that $p = \psi(x)$. Let or_p be the orientation on T_pM such that

$$\operatorname{or}_p(D\psi(x)e_1, \ldots, D\psi(x)e_n) = 1.$$

If we had picked a different parameterization ϕ, we would still choose the same orientation or_p on T_pM because the change of coordinates map between ϕ and ψ is orientation preserving.

On the other hand, suppose there is a smooth choice of orientation or_p on each T_pM. Let $(\psi_i)_{i\in\mathcal{I}}$ be an atlas for M. We can assume the domains U_i are connected. If

$$\mathrm{or}_{\psi(x)}((D\psi)(x)e_1,\ldots,(D\psi)(x)e_n) = 1$$

on U_i, define $\phi_i\colon U_i \to \mathbb{R}^N$ by $\phi_i(x) = \psi_i(x)$. On the other hand, if

$$\mathrm{or}_{\psi(x)}((D\psi)(x)e_1,\ldots,(D\psi)(x)e_n) = -1$$

on U_i, then let $\tilde{U}_i$ be the reflection of U_i across the x_1 axis and define $\phi_i\colon \tilde{U}_i \to \mathbb{R}^N$ by

$$\phi_i(x_1,x_2,\ldots,x_n) = \psi_i(-x_1,x_2,\ldots,x_n).$$

Then $(\phi_i)_{i\in\mathcal{I}}$ is an oriented atlas for M. $\square$

Given an oriented n-dimensional manifold M we can equivalently define the volume form ω by

$$\omega_p(v_1,\ldots,v_n) = \mathrm{or}_p(v_1,\ldots,v_n)\sqrt{\det(V^TV)}.$$

Technically $\mathrm{or}_p(v_1,\ldots,v_n)$ isn't defined if $v_1,\ldots,v_n$ are not linearly independent in T_pM, but in this case the volume form outputs 0.

17 Manifolds with Boundary

Recall that an n-dimensional manifold with boundary is a subset M of $\mathbb{R}^N$ that looks locally like either an n-dimensional plane or half of an n-dimensional plane near each of its points. In this section we give the formal definition of a manifold with boundary.

Definition 231. The n-dimensional half-space is $\mathbb{H}^n = \{x \in \mathbb{R}^n : x_n \geq 0\}$. The boundary of $\mathbb{H}^n$ is the $(n-1)$-dimensional plane $\partial\mathbb{H}^n = \{x \in \mathbb{R}^n : x_n = 0\}$.

Definition 232. An n-dimensional manifold with boundary in $\mathbb{R}^N$ is a set $M \subset \mathbb{R}^N$ such that for each $p \in M$ there exist

(i) an open set $U \subset \mathbb{R}^n$,

(ii) an open set $V \subset \mathbb{R}^N$ containing p, and

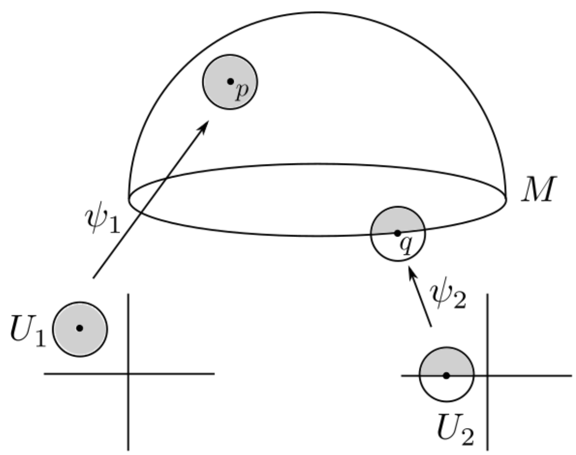
(iii) an embedding $\psi\colon U \to \mathbb{R}^N$ such that $\psi(U \cap \mathbb{H}^n) = M \cap V$.

If $p = \psi(t)$ with $t \in \partial\mathbb{H}^n$ we say that p belongs to ∂M. If $p = \psi(t)$ with $t \in \mathbb{H}^n \setminus \partial\mathbb{H}^n$, we say that p is an interior point of M.

Example 233. The closed upper hemisphere

$$M = \{(x, y, z) \in \mathbb{R}^3;\ x^2 + y^2 + z^2 = 1 \text{ and } z \geq 0\}$$

is a 2-dimensional manifold with boundary. The boundary is the equatorial circle. The following figure illustrates a parameterization for M near an interior point p and a parameterization for M near a boundary point q.



Note that $U_1 \cap \mathbb{H}^2$ is an entire disk while $U_2 \cap \mathbb{H}^2$ is half of a disk.

Example 234. The closed unit ball $B^3 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}$ is a 3-dimensional manifold with boundary in $\mathbb{R}^3$. The boundary is the unit sphere S^2.

Remark 235. If M is an n-dimensional manifold with boundary then ∂M is an $(n-1)$-dimensional manifold without boundary. Local parameterizations for ∂M can be constructed by restricting local parameterizations for M to $\partial\mathbb{H}^n$.

An n-dimensional manifold M with boundary still has an n-dimensional tangent space T_pM at each point $p \in M$. If p is an interior point, we define the tangent space in the usual fashion. If p is a boundary point, pick a local parameterization $\psi \colon U \to \mathbb{R}^N$ with $\psi(x) = p$ with $x \in \partial \mathbb{H}^n$. Then we define T_pM to be the span of the columns of $D\psi(x)$. This is still an n-dimensional subspace of $\mathbb{R}^N$. Suppose $p \in \partial M$ and consider a tangent vector $v \in T_pM$. Then by definition $v = D\psi(x)u$ for some vector $u \in \mathbb{R}^n$. We can classify v into one of three types according to whether the n-th component of u is positive, negative, or zero. We say that

(i) v points inward if $u_n > 0$

(ii) v points outward if $u_n < 0$

(iii) v is tangent to ∂M if $u_n = 0$.

The outward normal to ∂M at a point $p \in \partial M$ is the unique vector $\nu \in T_pM$ such that ν is perpendicular to $T_p\partial M$ and ν points outward.

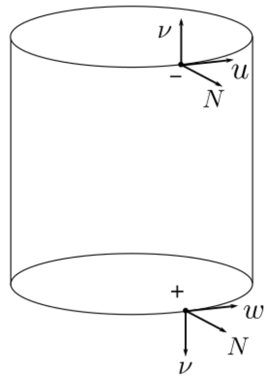
If M is oriented, then ∂M inherits a natural orientation from M. More precisely, let $v_1, \dots, v_{n-1}$ be an ordered basis for $T_p\partial M$. Then $\nu(p), v_1, \dots, v_{n-1}$ is an ordered basis for T_pM. We say that $v_1, \dots, v_{n-1}$ is positively or negatively oriented in $T_p\partial M$ according to whether $\nu(p), v_1, \dots, v_{n-1}$ is positively or negatively oriented in T_pM.

Example 236. Let M be cylindrical shell

$$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1,\ -1 \leq z \leq 1\}.$$

At each $p \in M$ let $N(p)$ be the outward pointing normal vector to T_pM. Orient M so that v, w forms a positively oriented basis for T_pM if and only if $N(p), v, w$ is right handed.

The boundary of M consists of two circles. On the upper circle, the outward normal to ∂M is given by $\nu(p) = e_3$. On the lower circle, the outward normal to ∂M is given by $\nu(p) = -e_3$. A vector $w \in T_p\partial M$ is positively oriented if and only if $\nu(p), w$ is positively oriented in T_pM. In turn, this is the case if and only if $N(p), \nu(p), w$ is right handed.



The above figure depicts a negatively oriented tangent vector u to the upper boundary circle and a positively oriented tangent vector w to the lower boundary circle.

18 Operations on Forms

We can define all the usual operations (e.g. wedge product, exterior derivative, etc.) for forms on manifolds simply by doing everything in charts.

Definition 237. Let α be a k-form on M and let $(\psi_i)_{i\in I}$ be an atlas for M. For each i define $\beta_i = d(\psi_i^*\alpha)$. This is a $(k+1)$-form on U_i. Then $d\alpha$ is the unique $(k+1)$-form on M whose such that $\psi_i^*(d\alpha) = \beta_i$ on U_i.

Remark 238. In order for this definition to make sense, we must verify that the β_i's satisfy the change of coordinates formula. This is a consequence of the fact that exterior derivative commutes with pullback:

$$(\psi_j^{-1} \circ \psi_i)^*\beta_j = (\psi_j^{-1} \circ \psi_i)^* d\psi_j^*\alpha = d(\psi_j^{-1} \circ \psi_i)^*\psi_j^*\alpha = d\psi_i^*\alpha = \beta_i$$

on $\psi_i^{-1}(\psi_j(U_j))$. Thus $d\alpha$ is well-defined.

Definition 239. Let α and β be forms on M and let $(\psi_i)_{i\in I}$ be an atlas for M. For each i let $\beta_i = \psi_i^*\alpha \wedge \psi_j^*\beta$. This is a form on U_i. Define $\alpha \wedge \beta$ to be the unique form on M such that $\psi_i^*(\alpha \wedge \beta) = \beta_i$ on U_i.

Remark 240. Again this is well-defined because wedge product commutes with pullback.

Example 241. Let α be the 1-form on S^2 given by $\alpha_p(v) = v \cdot e_3$. Let $f\colon S^2 \to \mathbb{R}$ be the function $f(x, y, z) = z$. We claim that $\alpha = df$. To check this, we need to verify that the equation holds in every chart in an atlas. Consider, for example, the local parameterization $\psi(x, y) = (x, y, \sqrt{1 - x^2 - y^2})$. We saw in Example 217 that

$$\psi^*\alpha = -\frac{x}{\sqrt{1 - x^2 - y^2}}\, dx - \frac{y}{\sqrt{1 - x^2 - y^2}}\, dy.$$

Also we have $\psi^* f = \sqrt{1 - x^2 - y^2}$. Thus

$$\begin{aligned} d(\psi^* f) &= \frac{\partial}{\partial x}(\sqrt{1 - x^2 - y^2})\, dx + \frac{\partial}{\partial y}(\sqrt{1 - x^2 - y^2})\, dy \\ &= -\frac{x}{\sqrt{1 - x^2 - y^2}}\, dx - \frac{y}{\sqrt{1 - x^2 - y^2}}\, dy = \psi^*\alpha \end{aligned}$$

and so $\alpha = df$ in the coordinates induced by ψ. Likewise, one could check that $\alpha = df$ in all the other charts in an atlas for S^2.

We would also like to define pullback for smooth maps between manifolds. In order to do this, we need to decide what we mean by a smooth map from one manifold to another. As with everything else, we say a map between manifolds is smooth if it is smooth when we look at it in charts.

Definition 242. Let $M \subset \mathbb{R}^a$ and $N \subset \mathbb{R}^b$ be manifolds and let $f\colon M \to N$ be a continuous function. Assume that $N \subset \mathbb{R}^L$. Let $(\psi_i)_{i \in I}$ be an atlas for M. Then f is smooth if for every i the function $f \circ \psi_i$ is smooth.

Remark 243. Note that $f \circ \psi_i$ sends U_i into $\mathbb{R}^b$. Thus $f \circ \psi_i$ sends Euclidean space to Euclidean space and so we know how to decide if it is smooth.

Example 244. Define a map $f\colon S^2 \to S^2$ by $f(x, y, z) = (-x, -y, -z)$. This is called the antipodal map. Let $\psi(x, y) = (x, y, \sqrt{1 - x^2 - y^2})$. Then

$$(f \circ \psi)(x, y) = (-x, -y, -\sqrt{1 - x^2 - y^2})$$

is a smooth function from the open unit ball in $\mathbb{R}^2$ into $\mathbb{R}^3$. Likewise, one could check that $f \circ \phi$ is smooth for all other charts ϕ in an atlas for S^2. Thus the antipodal map is smooth.

Associated to a smooth map $f\colon M \to N$ is a differential Df. For each $p \in M$, the differential $Df(p)$ is a linear map from T_pM to $T_{f(p)}N$. It is defined by taking the derivative of f in charts. More precisely, pick a vector $v \in T_pM$. Then we can find a local parameterization $\psi\colon U \to \mathbb{R}^a$ and a point $x \in U$ such that $p = \psi(x)$. There is a unique vector u such that $D\psi(x)u = v$. Define $Df(p)v = D(f \circ \psi)(x)u$. It is now possible to define pullback via the usual formula.

Definition 245. Let $f\colon M \to N$ be a smooth map and let α be a k-form on N. Then $f^*\alpha$ is the k-form on M given by

$$(f^*\alpha)_p(v_1, \ldots, v_k) = \alpha_{f(p)}(Df(p)v_1, \ldots, Df(p)v_k).$$

Equivalently, $f^*\alpha$ is the unique k-form on M with coordinate representations $\psi_i^*(f^*\alpha) = (f \circ \psi_i)^*\alpha$ on U_i.

19 Integrating Forms on Manifolds

If M is an orientable n-dimensional manifold and ω is an n-form on M we should be able to integrate ω over M. Like with everything else, the way to do this is via charts. Let $\psi_1, \ldots, \psi_m$ be an atlas for M. As a first attempt, we could try to set

$$\int_M \omega = \int_{U_1} \psi_1^*\alpha + \ldots + \int_{U_m} \psi_m^*\alpha.$$

However, this is **not** a good definition. It overcounts the integral of α over the regions where the coordinate charts overlap.

To get around this, we'd like to weight each of the pieces on the right hand side so that the total weight at every point is 1. The technical device for doing this is called a partition of unity.

Definition 246. Let $\psi_1, \ldots, \psi_m$ be an atlas for M. A partition of unity subordinate to (ψ_i) is a collection of smooth functions $f_1, \ldots, f_m\colon M \to \mathbb{R}$ such that

(i) $f_i \geq 0$,

(ii) $f_i = 0$ outside $\psi_i(U_i)$,

(iii) $f_1(p) + \ldots + f_m(p) = 1$ for all $p \in M$.

Think of the f_i as a weight associated to each chart.

We will need the following fact.

Fact. *Let $\psi_1, \ldots, \psi_m$ be an atlas for the manifold M. Then there is a partition of unity $f_1, \ldots, f_m$ subordinate to $\psi_1, \ldots, \psi_m$.*

Equipped with this, we can define the integral of a form over a manifold.

Definition 247. Let M be an oriented n-dimensional manifold and let ω be an n-form on M. Let $\psi_1, \ldots, \psi_m$ be an oriented atlas for M and let $f_1, \ldots, f_m$ be a subordinate partition of unity. Then we define

$$\int_M \omega = \int_{U_1} \psi_1^*(f_1 \alpha) + \ldots + \int_{U_m} \psi_m^*(f_m \alpha).$$

Remark 248. Partitions of unity are very useful as a theoretical tool. However, we would never use the above formula to actually compute an integral. In practice, to compute the integral of a form over a manifold, we can use a collection of charts with no overlap that cover M up to a set of measure zero. For example, the parameterizations $\psi(x, y) = (x, y, \sqrt{1 - x^2 - y^2})$ and $\phi(x, y) = (x, y, -\sqrt{1 - x^2 - y^2})$ have disjoint images and cover all of S^2 apart from the equator. Since the equator has zero area, we can safely neglect it when computing an integral. Thus we have

$$\int_{S^2} \alpha = \int_{\{x^2+y^2<1\}} \psi^* \alpha + \int_{\{x^2+y^2<1\}} \phi^* \alpha,$$

which avoids the need to write down an explicit partition of unity.

We can define the integral of a form over a manifold with boundary in the same way.

Definition 249. Let M be an oriented n-dimensional manifold with boundary and let ω be an n-form on M. Let $\psi_1, \ldots, \psi_m$ be an oriented atlas for M and let $f_1, \ldots, f_m$ be a subordinate partition of unity. Then we define

$$\int_M \omega = \int_{U_1 \cap \mathbb{H}^n} \psi_1^*(f_1 \omega) + \ldots + \int_{U_m \cap \mathbb{H}^n} \psi_1^*(f_m \omega)$$

Stokes' theorem continues to hold on manifolds with boundary.

Theorem 250 (Stokes' Theorem)**.** *Let M be an oriented n-dimensional manifold with boundary and let α be an $(n-1)$ form on M. Then*

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

Proof. This follows by applying the Euclidean Stokes' theorem in each chart. More explicitly, let $\psi_1, \ldots, \psi_m$ be an oriented atlas for M. Then

$$\begin{aligned}\int_M d\alpha &= \int_{U_1 \cap \mathbb{H}^n} \psi_1^*(f_1 d\alpha) + \ldots + \int_{U_1 \cap \mathbb{H}^n} \psi_m^*(f_m d\alpha) \\ &= \int_{U_1 \cap \mathbb{H}^n} \psi_1^*(d(f_1\alpha)) + \ldots + \int_{U_1 \cap \mathbb{H}^n} \psi_m^*(d(f_m\alpha)) \\ &\quad - \int_{U_1 \cap \mathbb{H}^n} \psi_1^*((df_1)\alpha) - \ldots - \int_{U_1 \cap \mathbb{H}^n} \psi_m^*((df_m)\alpha)\end{aligned}$$

Now observe that df_i is zero outside of $\psi_i(U_i)$ and so

$$\begin{aligned}\int_{U_1 \cap \mathbb{H}^n} &\psi_1^*((df_1)\alpha) + \ldots + \int_{U_1 \cap \mathbb{H}^n} \psi_m^*((df_m)\alpha) \\ &= \int_M (df_1)\alpha + \ldots + \int_M (df_m)\alpha \\ &= \int_M d(f_1 + \ldots + f_m)\alpha = 0\end{aligned}$$

since $f_1 + \ldots + f_m = 1$. Therefore

$$\begin{aligned}\int_M d\alpha &= \int_{U_1 \cap \mathbb{H}^n} \psi_1^*(d(f_1\alpha)) + \ldots + \int_{U_1 \cap \mathbb{H}^n} \psi_m^*(d(f_m\alpha)) \\ &= \int_{U_1 \cap \mathbb{H}^n} d\psi_1^*(f_1\alpha) + \ldots + \int_{U_1 \cap \mathbb{H}^n} d\psi_m^*(f_m\alpha) \\ &= \int_{U_1 \cap \partial\mathbb{H}^n} \psi_1^*(f_1\alpha) + \ldots + \int_{U_1 \cap \partial\mathbb{H}^n} \psi_m^*(f_m\alpha) = \int_{\partial M} \alpha,\end{aligned}$$

as desired. □

20 Distinguishing the Sphere and the Torus

We'll end the course by using differential forms to show that the sphere and torus are genuinely different manifolds.

Definition 251. Two manifolds M and N are called diffeomorphic if there is a smooth bijection $f\colon M \to N$ which has a smooth inverse $f^{-1}\colon N \to M$.

Intuitively two manifolds are diffeomorphic if they are intrinsically equivalent. One can think of diffeomorphic manifolds as two copies of the same abstract space that are situated differently in their ambient Euclidean spaces. For example, a sphere and an ellipsoid are diffeomorphic. Also a torus and a (smoothed out) coffee cup are diffeomorphic. However, we expect that a sphere and a torus are not diffeomorphic since the torus has a hole in it while the sphere doesn't.

Theorem 252. *The sphere S^2 and the torus T are not diffeomorphic.*

This will follow from a pair of lemmas about the structure of the differential forms on these manifolds.

Lemma 253. *There is a closed 1-form on T which is not exact.*

Lemma 254. *Every closed 1-form on S^2 is exact.*

Given the lemmas, we can prove the theorem as follows.

Proof. (Theorem 252) Assume for contradiction that the sphere and the torus are diffeomorphic. Then there is a smooth bijection $f\colon S^2 \to T$ with smooth inverse $f^{-1}\colon T \to S^2$. By Lemma 253, there is a closed 1-form α on T that is not exact. Then $f^*\alpha$ is a closed 1-form on S^2. By Lemma 254, there exists a smooth function g on S^2 such that $f^*\alpha = dg$. But this implies that $\alpha = (f^{-1})^* f^*\alpha = (f^{-1})^*(dg) = d((f^{-1})^*g)$ is exact. This is a contradiction. $\square$

It remains to prove the lemmas.

Proof. (Lemma 253) Think of T as the manifold obtained by rotating the circle $(x-2)^2 + z^2 = 1$ in the xz-plane around the z-axis. Let

$$C = \{(x, y, 0) \in \mathbb{R}^3 : x^2 + y^2 = 4\}$$

be the central circle inside T. Then there is a projection $\pi\colon T \to C$ given by

$$\pi(x, y, z) = \frac{2(x, y)}{\|(x, y)\|}.$$

Let ω be the length form on C. Then ω is closed since it is a 1-form on a 1-manifold. Thus $\alpha = \pi^*\omega$ is a closed 1-form on T^2.

We claim that α is not exact. Suppose for contradiction that α is exact. Then $\pi^*\omega = dg$ for some function g on T. Now consider the curve $\gamma\colon [0, 2\pi] \to T$ given by $\gamma(t) = (2\cos t, 2\sin t, 1)$. By Stokes' theorem we have

$$\int_\gamma \pi^*\omega = \int_\gamma dg = g(\gamma(2\pi)) - g(\gamma(0)) = 0.$$

On the other hand $\pi_*\gamma(t) = (2\cos t, 2\sin t)$ traverses the central circle C exactly once and so

$$\int_\gamma \pi^*\omega = \int_{\pi_*\gamma} \omega = \operatorname{length}(C) = 4\pi.$$

This is a contradiction. Hence α is a closed 1-form on T which is not exact. $\square$

Remark 255. Let

$$\psi(s,t) = (2\cos s, 2\sin s, 0) + \cos t(\cos s, \sin s, 0) + \sin t(0,0,1)$$

and let $\psi_1, \psi_2, \psi_3, \psi_4$ be the restrictions of ψ to $(0, \frac{3\pi}{2}) \times (0, \frac{3\pi}{2})$, $(0, \frac{3\pi}{2}) \times (\pi, \frac{5\pi}{2})$, $(\pi, \frac{5\pi}{2}) \times (0, \frac{3\pi}{2})$, $(\pi, \frac{5\pi}{2}) \times (\pi, \frac{5\pi}{2})$ respectively. The $\psi_1, \psi_2, \psi_3, \psi_4$ form an atlas for the torus.

The 1-form α essentially measures how a much curve in the torus T rotates around the z-axis. Therefore α is given in coordinates by $\psi_i^*(\alpha) = 2\,ds$. (The factor of 2 is there because ω is the length form on C and not the angle form. Because C has radius 2, length and angle are related by a factor of 2.) It is easy to see that α is closed from this expression because $d(2\,ds) = 0$.

On the other hand, α is not exact. Even though $2\,ds = d(2s)$ is exact in every coordinate chart, the functions $2s$ do not satisfy the change of coordinates formula and hence do not piece together to give a well-defined function on T. For example, ψ_1 thinks the s coordinate of $(\sqrt{2}, \sqrt{2}, 1)$ is $\pi/4$ while ψ_3 thinks the s coordinate of $(\sqrt{2}, \sqrt{2}, 1)$ is $2\pi + \pi/4$. So there is no well-defined value for the function s at the point $(\sqrt{2}, \sqrt{2}, 1)$.

To check that every closed 1-form on S^2 is exact, we first need to prove some preliminary results.

Proposition 256. *Let $\gamma\colon [0,1] \to S^2$ be a smooth curve. Then γ is not surjective.*

Proof. (Sketch) The crucial point is that

$$L = \operatorname{length}(\gamma) = \int_0^1 \|\gamma'(t)\|\, dt$$

is finite. Let k be a positive integer and pick points $p_0, p_1, \ldots, p_k$ along γ that split γ into k pieces each of length L/k. Let B_i be a ball of radius L/k centered at p_i. Then the image of γ is contained in the union of the balls B_i for $i = 0, 1, \ldots, k$. Consequently, the image of γ has area at most

$$\sum_{i=0}^{k} \operatorname{Area}(B_i) = \sum_{i=0}^{k} \frac{\pi L^2}{k^2} = \frac{\pi L^2(k+1)}{k^2}.$$

Letting $k \to \infty$, the number on the right hand side goes to 0. Therefore, the image of γ cannot cover a set of positive area. In particular, γ cannot be surjective. $\square$

Remark 257. There does exist a continuous curve $\gamma\colon [0,1] \to S^2$ which is surjective. (The interested reader can look up Hilbert's space-filling curve.)

Proposition 258. *Every smooth, closed curve $\gamma\colon [0,1] \to S^2$ is homotopic in S^2 to a point.*

Proof. Let $\gamma\colon [0,1] \to S^2$ be a smooth, closed curve. By the previous proposition, γ is not surjective. Without loss of generality, we can assume that the north pole $(0,0,1)$ is not in the image of γ. Consider the straight line homotopy from γ to the south pole $(0,0,-1)$ in $\mathbb{R}^3$. Since γ does not pass through the north pole, none of the lines in the straight line homotopy pass through the origin. Therefore, we can renormalize to get a homotopy

$$h(s,t) = \frac{(1-s)\gamma(t) + s(0,0,-1)}{\|(1-s)\gamma(t) + s(0,0,-1)\|}$$

from γ to $(0,0,-1)$ in S^2. $\square$

Proposition 259. *Let α be a closed 1-form on S^2 and let $\gamma\colon [0,1] \to S^2$ be a smooth, closed curve. Then $\int_\gamma \alpha = 0$.*

Proof. By the previous proposition, there exists a homotopy $h\colon [0,1]\times[0,1] \to S^2$ from γ to a point. Hence by Stokes' theorem

$$\int_\gamma \alpha = \int_{\partial h} \alpha = \int_h d\alpha = 0,$$

as required. □

Finally we can prove Lemma 254.

Proof. (Lemma 254) Let α be a closed 1-form on S^2. Define a function $g\colon S^2 \to \mathbb{R}$ by setting

$$g(p) = \int_{\gamma_p} \alpha$$

where γ_p is some curve from the south pole to p in S^2. Note that if γ_p^1 and γ_p^2 are two curves from the south pole to p, then $\gamma_p^1 - \gamma_p^2$ is a closed curve in S^2 and so

$$\int_{\gamma_p^1} \alpha - \int_{\gamma_p^2} \alpha = 0.$$

Consequently $g(p)$ is well-defined, i.e., it does not depend on the particular choice of curve γ_p. It remains to show that $dg = \alpha$. This follows by essentially the same argument as in Example 161. □