The intuition behind calculus on manifolds

Daniel O'Connor [1]

[1] If you have any questions, comments, or suggestions about how to clarify these notes, feel free to email me at daniel.v.oconnor@gmail.com.

Contents

1 *Introduction*

The fundamental strategy of calculus is to take a nonlinear function f (difficult) and approximate it locally by a linear function (easy):

$$f(y) \approx f(x) + f'(x)(y - x) \quad \text{when } y \text{ is close to } x. \tag{1}$$

Technically, the function on the right in (1) is called an "affine" function.

For a function $f : \mathbb{R}^n \to \mathbb{R}$, this approximation can be written as

$$f(y) \approx f(x) + \langle \nabla f(x), y - x \rangle.$$

Once we internalize this approximation, which is sometimes called Newton's approximation, we realize something delightful: everything in calculus can be derived easily. Calculus is *easy*. Shouldn't this also be true of calculus on manifolds, which is often viewed as a formidable subject? Shouldn't it be easy to derive even the generalized Stokes's theorem? After all, it's still just calculus.

$$\int_{\partial M} \omega = \int_M d\omega$$

How would you think of this?

A manifold is the natural setting in which to do calculus. Developing our theory at the appropriate level of generality should make things more beautiful and more elegant, not more difficult.

There is something strange about the way that calculus on manifolds is taught. Calculus is first taught non-rigorously, and rightly so, because a fact such as the mean value theorem is very believable and there is no benefit in getting bogged down, initially, in its rigorous proof. You can discover and understand calculus perfectly clearly, like the old masters did, without the rigorous proofs. But for some reason, calculus on manifolds is usually taught in full rigor right from the start. In that approach, the intuition is lost.

There are in fact several books which attempt to explain calculus on manifolds intuitively, such as

- Visual Differential Geometry and Forms by Needham (available June 2021)
- Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach by Hubbard and Hubbard
- Multivariable Mathematics: Linear Algebra, Multivariable Calculus, and Manifolds by Shifrin
- Second Year Calculus: From Celestial Mechanics to Special Relativity by Bressoud
- Discrete Differential Geometry: An Applied Introduction by Keenan Crane (see also his YouTube videos)

Many calculus on manifolds textbooks drop a lot of abstract-sounding definitions on the poor student without explaining the intuitive idea behind what's going on. Tensors and differential forms are pulled out of thin air, and you might have no idea how they emerge naturally and unavoidably when you chop up a manifold into tiny pieces and compute the contribution of each piece to an integral. The generalized Stokes's theorem is also pulled out of thin air, without mentioning that it can be discovered using the same simple type of argument that physicists use to derive Green's theorem and the divergence theorem. (See figure 1.) The exterior derivative is pulled out of thin air, without showing how it emerges when deriving Stokes's theorem. The goal of these notes is to explain calculus on manifolds in such a way that you can imagine how someone might have thought of it. In other words, the goal is to reveal how calculus on manifolds, like basic calculus, is almost obvious.

I will freely use vector calculus and linear algebra, though. If you already love linear algebra, like I do, then there is little standing in the way to a clear intuitive understanding of calculus on manifolds.

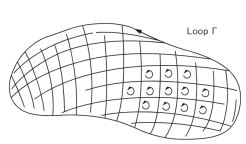

Fig. 3–9. Some surface bounded by the loop Γ is chosen. The surface is divided into a number of small areas, each approximately a square. The circulation around Γ is the sum of the circulations around the little loops.

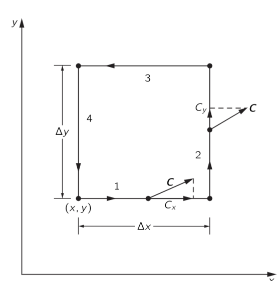

Fig. 3–10. Computing the circulation of C around a small square.

Figure 1: These diagrams taken from the Feynman Lectures on Physics show how physicists derive the classical Stokes's theorem and divergence theorem using highly intuitive arguments. **We can do the same for the generalized Stokes's theorem.**

Disclaimer: I'm not an expert on smooth manifolds. Rather, these notes represent the intuition that I only uncovered with much difficulty, after reading parts of several books on the topic and pondering the ideas for a long time. I'm sure these ideas are obvious to everyone who understands this material deeply — in fact, so obvious that they are not even worth mentioning, apparently.

2 *What is a manifold?*

In these notes the term "manifold" will always mean "smooth manifold."

A manifold is a generalization of the idea of a curve or of a surface. A smooth curve can be approximated locally by a straight line. A surface can be approximated locally by a plane. A smooth k-manifold in $\mathbb{R}^n$ can be approximated locally by a k-dimensional affine subspace of $\mathbb{R}^n$.

An "affine subspace" of $\mathbb{R}^n$ is what you get if you take a subspace of $\mathbb{R}^n$ and shift it away from the origin. Note that an affine subspace of $\mathbb{R}^n$ is not actually a subspace of $\mathbb{R}^n$. For that reason, some might prefer the term "affine manifold."

A smooth manifold is a natural setting in which to do calculus, for the following reason. The key idea of calculus is to approximate a nonlinear function locally by an affine function. But the domain of an affine function must be an affine space. So how can we approximate a function f by an affine function if the domain of f is not even an affine space? We must at least be able to locally *approximate* the domain of f by an affine space.

In order for the approximation $f(y) \approx f(x) + f'(x)(y - x)$ to make sense, we need $y - x$ to be a vector, the appropriate type of object to be plugged into a linear transformation or multiplied by a matrix. The distinctive feature of an affine space is that the difference of any two points is a vector.

Suppose that a manifold M is approximated near a point $x \in M$ by an affine subspace W. Vectors in $\mathbb{R}^n$ of the form $w - x$, where $w \in W$, are called "tangent vectors" to M at x. The set of all tangent vectors to M at x is called the **tangent space** of M at x, and is denoted $T_x(M)$.

Note that the way I have defined the tangent space of M at x, it is a subspace of $\mathbb{R}^n$. I don't visualize it as passing through the origin, though. I visualize it as being attached to M at x. When I visualize a vector v that is tangent to M at x, the arrow that I picture has its tail at x. (See figure 2.)

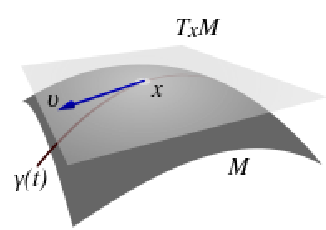

Figure 2: A tangent vector to a manifold

2.1 *Manifolds with boundary*

Some manifolds have a boundary. For example, the boundary of a ball in $\mathbb{R}^3$ is a sphere. A sphere in $\mathbb{R}^3$ has no boundary, but a hemisphere has a boundary which is a circle. (See figure 3.)

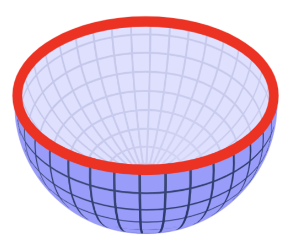

Figure 3: The boundary of a hemisphere is a circle.

The above examples suggest that if a smooth k-manifold M has a boundary, then the boundary of M is a smooth manifold of dimension $k - 1$. We denote the boundary of M by ∂M. At each point on ∂M, there is a unique outward unit normal vector. This is illustrated in figure 5. We might also conjecture that ∂M has no boundary.

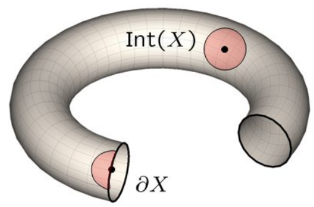

Figure 4: A point on the boundary of a manifold called X.

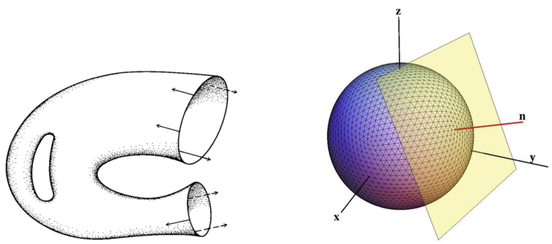

Figure 5: Outward and inward-pointing normal vectors at various points on the boundary of a 2-manifold (left) and a 3-manifold (a ball, right).

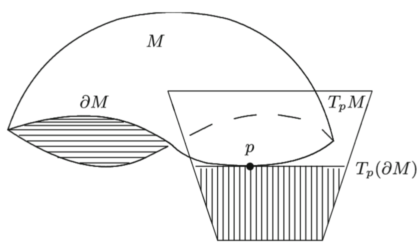

Figure 6: If $p \in \partial M$, then $T_p(\partial M)$ is a subspace of $T_p(M)$. In this example, $T_p(M)$ is two-dimensional and $T_p(\partial M)$ is one-dimensional.

Notice that if $x \in \partial M$, then $T_x(\partial M)$ is a subspace of $T_x(M)$. (See figure 6.) The dimension of $T_x(\partial M)$ is $k-1$, whereas the dimension of $T_x(M)$ is k. When $T_x(\partial M)$ is viewed as a subspace of $T_x(M)$, the orthogonal complement of $T_x(\partial M)$ has dimension 1. This explains why, at each point in ∂M, there is a unique outward unit normal vector. (See figure 5.)

3 *Integration on manifolds*

The way that integration works is that you chop the region that you're integrating over into tiny pieces, compute the contribution of each piece, then add up all those individual contributions. When we chop up a manifold M, we are going to chop it up into tiny pieces such that each piece is approximately a parallelepiped, spanned by vectors which are tangent to M. (See figure 7.)

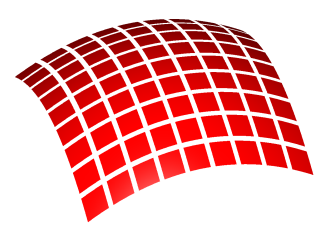

Figure 7: A manifold is chopped up into tiny parallelepipeds, each one of which is spanned by tangent vectors.

3.1 *What type of object should we integrate over a manifold?*

In vector calculus, we integrate vector fields over surfaces. At least, that is how we are taught to think of it. What kind of mathematical object is the natural thing to integrate over a manifold? Should it be a vector field or something else? (Answer: it should be something else.)

Let's think about how integration is going to work. Following the usual pattern for integration, we chop up our manifold into tiny pieces, in such a way that the ith piece is approximately a parallelepiped spanned by tangent vectors $v_1^i, \ldots, v_k^i$. The contribution of the ith piece can be viewed as being a function of these k vectors. Thus, to compute the contribution of each piece of the manifold, what we need is a gadget ω that will assign to each point p on our manifold a function $\omega(p)$ that takes as input a list of tangent vectors $v_1, \ldots, v_k$ and returns as output a real number (the contribution of the piece of the manifold that is spanned by these tangent vectors). The integral of ω over M is defined by

$$\int_M \omega \approx \sum_i \omega(p_i)(v_1^i, v_2^i, \ldots, v_k^i). \tag{2}$$

Each p_i is a point chosen arbitrarily in the ith parallelepiped. A precise definition would state that $\int_M \omega$ is in some sense a limit of "Riemann sums" like this.

You can argue that $\omega(p)$ should be alternating and multilinear, because chopping up the manifold more finely should not change the value of the integral (and because degenerate parallelepipeds should contribute 0).

For $\omega(p)$ to be alternating means that interchanging any two of its inputs should reverse the sign of the output. For $\omega(p)$ to be multilinear means that it is linear in each of its inputs.

Why should $\omega(p)$ be multilinear?

Imagine that we chop the ith piece of M in half, so that each of the two new pieces is spanned by the tangent vectors $v_1^i/2, v_2^i, \ldots, v_k^i$. The sum on the right in equation (2) should not change, right? At least, it should only change by a negligible amount. After all, the integral is supposed to be the *limit* of sums like this, so the sums are supposed to *converge* to the limiting value. Once we have chopped up M into sufficiently small pieces, we are not supposed to be able to keep changing the value of the sum by chopping ever more finely.

In order for the value of the sum to be unchanged by this extra chop, we require that

$$\omega(p_i)\left(\frac{v_1^i}{2}, v_2^i, \ldots, v_k^i\right) = \frac{1}{2}\omega(p_i)(v_1^i, \ldots, v_k^i).$$

In other words, when a piece is cut in half, its contribution to the integral is cut in half. The contribution of the original piece is the sum of the contributions from the two new pieces.

Similar reasoning suggests that $\omega(p_i)$ should have the property that

$$\omega(p_i)(v_1^i, \dots, cv_j^i, \dots, v_k^i) = c\omega(p_i)(v_1^i, \dots, v_j^i, \dots, v_k^i) \tag{3}$$

for all $c > 0$ and $j = 1, \dots, k$. And this in turn suggests that $\omega(p_i)$ should be a multilinear function.

It may seem to be a leap from (3) to the assumption that $\omega(p_i)$ must be multilinear. I will offer two additional pieces of evidence that requiring $\omega(p)$ to be multilinear for all p is the correct choice to make. First of all, if you parametrize the same manifold in two different ways, and use each parametrization to compute the integral, you should get the same answer. It can be shown that when each $\omega(p)$ is multilinear (and alternating), we do indeed get the same answer both ways. Secondly, the classical integrals of vector calculus can be expressed as special cases of the integral over a manifold defined in equation (2), with $\omega(p)$ alternating and multilinear. This is shown in sections 4.2 and 4.3.

WHY SHOULD $\omega(p)$ BE ALTERNATING? If any two of the vectors in the list $v_1^i, \dots, v_k^i$ are equal to each other, then the ith parallelopiped is degenerate and its contribution to the integral should be 0. In other words, $\omega(p_i)(v_1^i, \dots, v_k^i) = 0$. Because we are assuming that $\omega(p_i)$ is multilinear, this implies that $\omega(p_i)$ must be alternating.

Here is an explanation of this fact in the case where $k = 2$. I'm assuming that $\omega(p)$ is multilinear and that if both inputs to $\omega(p)$ are equal then the output is 0. Notice that

$$\begin{aligned} 0 &= \omega(p)(v_1 + v_2, v_1 + v_2) \\ &= \underbrace{\omega(p)(v_1, v_1)}_{0} + \omega(p)(v_1, v_2) + \omega(p)(v_2, v_1) + \underbrace{\omega(p)(v_2, v_2)}_{0} \\ &= \omega(p)(v_1, v_2) + \omega(p)(v_2, v_1). \end{aligned}$$

We conclude that

$$\omega(p)(v_2, v_1) = -\omega(p)(v_1, v_2)$$

which means that $\omega(p)$ is alternating.

LET'S MAKE A FEW DEFINITIONS, motivated by the above observations. A multilinear function $T : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{k \text{ times}} \to \mathbb{R}$ is called a k-tensor on $\mathbb{R}^n$. More generally, if V is a subspace of $\mathbb{R}^n$ (such as the tangent space to a manifold at a point p), then a multilinear function $T : \underbrace{V \times \cdots \times V}_{k \text{ times}} \to \mathbb{R}$ is called a k-tensor on V. If T is also alternating (so that interchanging any two inputs to T reverses the sign of the output), then T is called an alternating k-tensor. A function ω that assigns to each point $p \in M$ an alternating k-tensor $\omega(p)$ on the tangent space to M at p is called a **differential k-form** on M.

A differential k-form can also be called a k-form or a differential form.

A differential form is the natural type of object to integrate over a manifold. It exists in order to compute the contribution of each piece of a chopped up manifold M. Each tiny piece is a parallelepiped spanned by tangent vectors. A differential form ω looks at each tiny parallelepiped (or equivalently, it looks at the tangent vectors that span the parallelepiped) and it tells us the contribution of each tiny parallelepiped to the total integral. Add up all those individual contributions to obtain the value of the integral $\int_M \omega$.

3.2 Orientation

There is one important issue that I have not yet addressed when explaining how to integrate a differential form ω over a manifold M. Suppose we chop up M into tiny pieces and the ith piece is approximately a parallelepiped P based at a point $p \in M$ and spanned by tangent vectors $v_1, \dots, v_k$. To compute the contribution of the ith piece, we plug these tangent vectors into the k-tensor $\omega(p)$. But here is the crucial question: when we plug these tangent vectors into $\omega(p)$, *how should they be ordered?* I could have equally well said that the parallelepiped P is spanned by any permutation of the tangent vectors $v_1, \dots, v_k$. No permutation of these tangent vectors is any more or less valid than any other.

Since $\omega(p)$ is alternating, permuting the inputs to $\omega(p)$ either changes the sign of the output (if the permutation is odd) or else does not change the output at all (if the permutation is even). Somehow we must specify a consistent way of ordering the tangent vectors for each tiny piece of our chopped up manifold.

The most direct way to do this is to provide a continuous differential k-form μ on M which is "non-vanishing" in the sense that if tangent vectors $v_1, \dots, v_k$ at a point $p \in M$ are linearly independent then $\mu(p)(v_1, \dots, v_k) \neq 0$. Then, when computing the contribution of the ith piece of our chopped up manifold M, we order the tangent vectors $v_1, \dots, v_k$ so that the output of $\mu(p)$ is positive. (We need μ to be non-vanishing because otherwise it might not be possible to order the tangent vectors $v_1, \dots, v_k$ so that the output of $\mu(p)$ is positive.) Such a differential form μ is called an "orientation form" for M. An "oriented" manifold is a manifold M for which an orientation form has been specified.

What does it mean for a differential form ω on M to be "continuous" or to vary continuously over M? I can imagine a vector field varying continuously over M, but what would it mean for a tensor field to vary continuously over M?

Here is one way to think about it. If the functions $v_1 : M \to \mathbb{R}^n, \dots, v_k : M \to \mathbb{R}^n$ are continuous tangent vector fields on M, then the function $p \mapsto \omega(p)(v_1(p), \dots, v_k(p))$ should vary continuously as the point p ranges over M.

Similarly, to say that a differential form ω on M is "smooth" means that if the tangent vector fields $v_1, \dots, v_k$ on M are smooth then the function $p \mapsto \omega(p)(v_1(p), \dots, v_k(p))$ varies smoothly as p ranges over M.

For example, suppose that M is a sphere in $\mathbb{R}^3$. To orient M, we could let μ be the differential form on M defined by $\mu(p)(v_1, v_2) = (v_1 \times v_2) \cdot n_p$, where n_p is the outward unit normal vector at p. Notice that $\mu(p)(v_1, v_2)$ is positive if $v_1 \times v_2$ points outward and is negative if $v_1 \times v_2$ points inward.

A Mobius strip provides an amazing demonstration that some manifolds can't be oriented. Suppose M is a Mobius strip and μ is an orientation form for M. Let v_1 and v_2 be tangent vectors to M at a point $p \in M$, ordered so that $\mu(p)(v_1, v_2) > 0$. Now imagine sliding or "continuously morphing" v_1 and v_2 along M, without ever passing through a degenerate (linearly dependent) configuration, so that the output of μ at each location along the way remains positive. Eventually you arrive back at p, and you discover that $\mu(p)(v_2, v_1) > 0$, which is a contradiction. (See figure 8.)

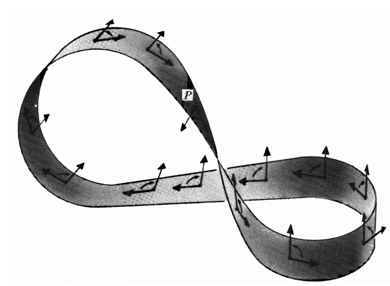

Figure 8: A Mobius strip cannot be oriented. This figure is from Calculus on Manifolds by Spivak.

3.3 An induced orientation for the boundary of M

Suppose μ is an orientation form for a k-manifold with boundary M. The boundary of M is a $(k-1)$-manifold, and we can define an orientation form $\partial\mu$ on ∂M as follows: if $p \in \partial M$, and $v_1, \ldots, v_{k-1} \in T_p(\partial M)$, then

$$\partial\mu(p)(v_1, \ldots, v_{k-1}) = \mu(p)(n_p, v_1, \ldots, v_{k-1}),$$

where n_p is the outward unit vector normal to ∂M at p. The slogan of $\partial\mu$ is "outward-pointing vector comes first."

3.4 Integrating a constant k-form over a parallelepiped

Suppose that $M \subset \mathbb{R}^n$ is a parallelepiped based at a point $p \in \mathbb{R}^n$ and spanned by vectors $v_1, \ldots, v_k \in \mathbb{R}^n$. This set M is a k-manifold.

At each point $x \in M$, the affine subspace $x + \operatorname{span}\{v_1, \ldots, v_k\}$ provides a perfect local approximation of M.

Let ω be a differential form on $\mathbb{R}^n$ which is constant on M (so that $\omega(p) = \omega(q)$ for all points $p, q \in M$). Then

$$\int_M \omega = \omega(p)(v_1, \ldots, v_k).$$

This fact might seem intuitive. It is analogous to integrating a constant function in multivariable calculus. When you chop up M into tiny identical pieces, each tiny piece contributes the same amount. However, I'll attempt a more detailed explanation.

Is it clear that M is orientable? At each point $x \in M$, the tangent space to M at x is $V = \operatorname{span}\{v_1, \ldots, v_k\}$. Let μ be the k-form on $\mathbb{R}^n$ defined so that $\mu(x)(w_1, \ldots, w_k)$ is equal to 0 if the vectors $w_1, \ldots, w_k$ are linearly dependent, and otherwise is equal to the determinant of the change of basis matrix from the ordered basis $(w_1, \ldots, w_k)$ to the ordered basis $(v_1, \ldots, v_k)$. This μ is an orientation form for M.

If we define two ordered bases of V to have the "same orientation" when the determinant of the change of basis matrix from one to the other is positive, then we can say that this k-form μ returns a positive value if and only if $(w_1, \ldots, w_k)$ has the same orientation as $(v_1, \ldots, v_k)$.

Let N be a large positive integer. Chop up M into tiny parallelepipeds, each of which is spanned by the vectors $v_1/N, \ldots, v_k/N$. The total number of pieces of M is N^k. The sum of all the contributions to $\int_M \omega$ is

$$\begin{aligned}
\sum_i \omega(p)(v_1/N, \ldots, v_k/N) &= \sum_i \frac{1}{N^k}\omega(p)(v_1, \ldots, v_k) \\
&= N^k \left(\frac{1}{N^k}\omega(p)(v_1, \ldots, v_k)\right) \\
&= \omega(p)(v_1, \ldots, v_k).
\end{aligned}$$

By a limiting argument, we conclude that $\int_M \omega = \omega(p)(v_1, \ldots, v_k)$.

4 Examples of tensors and differential forms

4.1 The differential of a smooth function $f : \mathbb{R}^n \to \mathbb{R}$

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function. We define the "differential of f", denoted df, to be the differential 1-form on $\mathbb{R}^n$ such that

$$df(p)(v) = \langle \nabla f(p), v \rangle \tag{4}$$

for all $p, v \in \mathbb{R}^n$.

You can think of df as telling you approximately how much f changes when the input to f changes from p to $p + v$. That's because Newton's approximation

$$f(p+v) \approx f(p) + \langle \nabla f(p), v \rangle$$

can be written equivalently as

$$f(p+v) \approx f(p) + df(p)(v).$$

4.2 *Forms appearing in surface integrals and line integrals*

Let M be an orientable surface in $\mathbb{R}^3$ and let F be a smooth vector field on $\mathbb{R}^3$. One side of M is declared to be the "inside" and the other side is the "outside". In vector calculus, to compute $\int_M F \cdot dA$, we chop up M into tiny pieces, each of which is approximately a parallelogram. Suppose that the ith piece is based at a point $p \in M$ and spanned by tangent vectors v_1 and v_2. These tangent vectors are ordered so that $v_1 \times v_2$ points outward. The contribution of the ith piece is $(v_1 \times v_2) \cdot F(p)$. We now recognize that

$$\int_M F \cdot dA = \int_M \omega$$

where ω is the differential 2-form on $\mathbb{R}^3$ defined by

$$\omega(p)(v_1, v_2) = (v_1 \times v_2) \cdot F(p).$$

So the surface integrals we do in vector calculus can be thought of as integrating a particular differential form over a surface.

Similarly, if C is a curve in $\mathbb{R}^3$, then the line integral $\int_C F \cdot dr$ is equal to $\int_C \omega$, where ω is the 1-form on $\mathbb{R}^3$ defined by

$$\omega(p)(v) = v \cdot F(p).$$

4.3 *Forms appearing in volume integrals*

Now let M be an open subset of $\mathbb{R}^3$ and let $F : \mathbb{R}^3 \to \mathbb{R}$ be a smooth function. In vector calculus, to compute the volume integral $\int_M F\, dV$, we chop up M into tiny parallelepipeds. Suppose that the ith parallelepiped is based at a point $p \in M$ and spanned by vectors v_1, v_2, and v_3, which are ordered according to the right-hand rule (so $(v_1 \times v_2) \cdot v_3 > 0$). The volume of the ith piece is $\det \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ and the contribution of the ith piece is $F(p) \det \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$. We recognize that $\int_M F\, dV = \int_M \omega$, where ω is the 3-form on $\mathbb{R}^3$ defined by

$$\omega(p)(v_1, v_2, v_3) = F(p) \det \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}.$$

So the volume integrals we do in vector calculus can also be interpreted as integrating a differential form.

4.4 The determinant

The determinant of an $n \times n$ matrix, viewed as a function of the columns of the matrix, is a great example of an alternating n-tensor on $\mathbb{R}^n$. Can we somehow use the determinant to construct an alternating k-tensor on $\mathbb{R}^n$, where $k < n$? Yes, and it's the simplest thing you would try to do. To be concrete, I'll show how it works when $k = 3$ and $n = 5$. Our tensor is given as input three vectors

$$v_1 = \begin{bmatrix} v_{11} \\ v_{12} \\ v_{13} \\ v_{14} \\ v_{15} \end{bmatrix}, \quad v_2 = \begin{bmatrix} v_{21} \\ v_{22} \\ v_{23} \\ v_{24} \\ v_{25} \end{bmatrix}, \quad v_3 = \begin{bmatrix} v_{31} \\ v_{32} \\ v_{33} \\ v_{34} \\ v_{35} \end{bmatrix}.$$

We choose integers i_1, i_2, i_3 with $1 \leq i_1 < i_2 < i_3 \leq 5$. To be concrete, I'll take $i_1 = 2, i_2 = 4$, and $i_3 = 5$, and I'll name our tensor φ_{245} to reflect this choice. We will keep only components $2, 4$, and 5 of each input vector and then we will compute a determinant. So, the definition of φ_{245} is

$$\varphi_{245}(v_1, v_2, v_3) = \det \begin{bmatrix} v_{12} & v_{22} & v_{32} \\ v_{14} & v_{24} & v_{34} \\ v_{15} & v_{25} & v_{35} \end{bmatrix}.$$

With this construction, we obtain one alternating k-tensor on $\mathbb{R}^n$ for each list of integers $i_1, \ldots, i_k$ with $1 \leq i_1 < \cdots < i_k \leq n$. In other words, we obtain one alternating k-tensor for each k-element subset of $\{1, \ldots, n\}$. Thus, the total number of alternating k-tensors on $\mathbb{R}^n$ that we have obtained with this construction is $\binom{n}{k}$.

Let $e_1, \ldots, e_5$ be the standard basis vectors for $\mathbb{R}^5$. Notice that

$$\varphi_{245}(e_{i_1}, e_{i_2}, e_{i_3}) = \begin{cases} \pm 1 & \text{if } \{i_1, i_2, i_3\} = \{2, 4, 5\}, \\ 0 & \text{otherwise.} \end{cases}$$

In particular, we have

$$\varphi_{245}(e_2, e_4, e_5) = 1.$$

More generally, if $i_1 < \cdots < i_k$, then

$$\varphi_{i_1 \cdots i_k}(e_{i_1}, \ldots, e_{i_k}) = 1.$$

5 A basis for the vector space of alternating k-tensors on $\mathbb{R}^n$

The set of all alternating k-tensors on $\mathbb{R}^n$ is a vector space, denoted $\Lambda^k(\mathbb{R}^n)$. It turns out that the set B of all k-tensors $\varphi_{i_1 \cdots i_k}$ with $i_1 < \cdots < i_k$ is a basis of this vector space. Let's try to understand why.

First I'll show that B spans $\Lambda^k(\mathbb{R}^n)$. Thanks to multilinearity, an alternating k-tensor α on $\mathbb{R}^n$ is determined completely by the values of $\alpha(e_{i_1}, \ldots, e_{i_k})$ where $i_1 < \cdots < i_k$. To be concrete, I'll show this explicitly in the case where $k = 2$ and $n = 4$. Notice that

$$\begin{aligned}
&\alpha(c_{11}e_1 + c_{12}e_2 + c_{13}e_3 + c_{14}e_4, c_{21}e_1 + c_{22}e_2 + c_{23}e_3 + c_{24}e_4) \\
&= c_{11}\alpha(e_1, c_{21}e_1 + c_{22}e_2 + c_{23}e_3 + c_{24}e_4) \\
&\quad + c_{12}\alpha(e_2, c_{21}e_1 + c_{22}e_2 + c_{23}e_3 + c_{24}e_4) \\
&\quad + c_{13}\alpha(e_3, c_{21}e_1 + c_{22}e_2 + c_{23}e_3 + c_{24}e_4) \\
&\quad + c_{14}\alpha(e_4, c_{21}e_1 + c_{22}e_2 + c_{23}e_3 + c_{24}e_4) \\
&= c_{11}c_{21}\cancel{\alpha(e_1, e_1)} + c_{11}c_{22}\alpha(e_1, e_2) + c_{11}c_{23}\alpha(e_1, e_3) + c_{11}c_{24}\alpha(e_1, e_4) \\
&\quad + c_{12}c_{21}\alpha(e_2, e_1) + c_{12}c_{22}\cancel{\alpha(e_2, e_2)} + c_{12}c_{23}\alpha(e_2, e_3) + c_{12}c_{24}\alpha(e_2, e_4) \\
&\quad + c_{13}c_{21}\alpha(e_3, e_1) + c_{13}c_{22}\alpha(e_3, e_2) + c_{13}c_{23}\cancel{\alpha(e_3, e_3)} + c_{13}c_{24}\alpha(e_3, e_4) \\
&\quad + c_{14}c_{21}\alpha(e_4, e_1) + c_{14}c_{22}\alpha(e_4, e_2) + c_{14}c_{23}\alpha(e_4, e_3) + c_{14}c_{24}\cancel{\alpha(e_4, e_4)} \\
&= (c_{11}c_{22} - c_{12}c_{21})\alpha(e_1, e_2) + (c_{11}c_{23} - c_{13}c_{21})\alpha(e_1, e_3) \\
&\quad + (c_{11}c_{24} - c_{14}c_{21})\alpha(e_1, e_4) + (c_{12}c_{23} - c_{13}c_{22})\alpha(e_2, e_3) \\
&\quad + (c_{12}c_{24} - c_{14}c_{22})\alpha(e_2, e_4) + (c_{13}c_{24} - c_{14}c_{23})\alpha(e_3, e_4).
\end{aligned}$$

This reveals that α is determined entirely by the values of $\alpha(e_{i_1}, e_{i_2})$ with $i_1 < i_2$. Now, we can easily construct a linear combination of the 2-tensors $\varphi_{i_1 i_2}$ which agrees with α for these particular inputs, as follows:

$$\alpha = \alpha(e_1, e_2)\varphi_{12} + \alpha(e_1, e_3)\varphi_{13} + \alpha(e_1, e_4)\varphi_{14} + \alpha(e_2, e_4)\varphi_{24} + \alpha(e_3, e_4)\varphi_{34}.$$

Try plugging in (e_1, e_2) on both sides, for example. On the left we have $\alpha(e_1, e_2)$. On the right, all terms but the first vanish, and we are left with

$$\alpha(e_1, e_2)\varphi_{12}(e_1, e_2) = \alpha(e_1, e_2).$$

More generally, we see that if $\alpha \in \Lambda^k(\mathbb{R}^n)$ then

$$\alpha = \sum_{i_1 < \cdots < i_k} \alpha(e_{i_1}, \ldots, e_{i_k})\varphi_{i_1 \cdots i_k}.$$

So, B spans $\Lambda^k(\mathbb{R}^n)$.

Next let's check that B is linearly independent. Again to be concrete I'll take $k = 2$ and $n = 4$. Suppose that

$$c_{12}\varphi_{12} + c_{13}\varphi_{13} + c_{14}\varphi_{14} + c_{23}\varphi_{23} + c_{24}\varphi_{24} + c_{34}\varphi_{34} = 0.$$

Plugging in (e_1, e_2) on both sides, all terms but the first on the left vanish, and we are left with

$$c_{12}\varphi_{12}(e_1, e_2) = c_{12} = 0.$$

Similarly, we see that all the coefficients $c_{i_1 i_2}$ are 0.

The same argument, but written in more generality, shows that if

$$\sum_{i_1 < \cdots < i_k} c_{i_1 \cdots i_k} \varphi_{i_1 \cdots i_k} = 0$$

then all the coefficients $c_{i_1 \cdots i_k}$ are 0. This shows that B is linearly independent. Thus, B is a basis for $\Lambda^k(\mathbb{R}^n)$.

If ω is a differential k-form on $\mathbb{R}^n$, then for each $p \in \mathbb{R}^n$ the k-tensor $\omega(p)$ can be written as a linear combination of the basis elements $\varphi_{i_1 \cdots i_k}$:

$$\omega(p) = \sum_{i_1 < \cdots < i_k} c_{i_1 \cdots i_k} \varphi_{i_1 \cdots i_k}. \tag{5}$$

Define $\omega_{i_1 \cdots i_k} : \mathbb{R}^n \to \mathbb{R}$ so that $\omega_{i_1 \cdots i_k}(p)$ is equal to the coefficient $c_{i_1 \cdots i_k}$ appearing in (5). Then

$$\omega = \sum_{i_1 < \cdots < i_k} \omega_{i_1 \cdots i_k} \varphi_{i_1 \cdots i_k}. \tag{6}$$

This way of representing ω will be helpful when we discover Stokes's theorem.

6 *Discovering the generalized Stokes's theorem*

Let $M \subset \mathbb{R}^n$ be an oriented k-manifold with boundary. Let μ be an orientation form that specifies the orientation for M, and let ω be a smooth differential $(k-1)$-form on $\mathbb{R}^n$. Our goal in this section is to find a formula for $\int_{\partial M} \omega$.

You might object that ω only needs to be defined on M, whereas I'm assuming ω is defined on all of $\mathbb{R}^n$. I believe we do not miss the key ideas when making this assumption. There is a related issue in vector calculus: when integrating a vector field over a curve or a surface, the vector field only needs to be defined on the curve or on the surface. However, in applications the vector field is typically an electric field or a force field or a fluid velocity field which is defined throughout all of $\mathbb{R}^3$.

6.1 *A special case*

In light of equation (6), let's simplify the problem by assuming that

$$\omega(p) = f(p)\eta \tag{7}$$

for some smooth function $f : \mathbb{R}^n \to \mathbb{R}$ and some alternating $(k-1)$-tensor η on $\mathbb{R}^n$. Once we deal with this special case, we will use equation (6) to obtain a formula for the integral of a general smooth differential form over ∂M.

Chop up M into tiny pieces such that each piece is approximately a parallelepiped. Let M_i be the ith piece of M. Notice that

$$\int_{\partial M} \omega = \sum_i \int_{\partial M_i} \omega,$$

because the sum on the right "telescopes" and wonderful cancellation occurs. This is the same type of wonderful cancellation that occurs

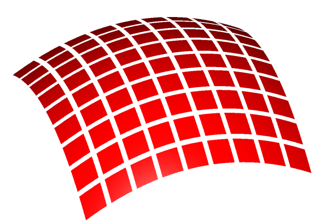

Figure 9: Chopping up a manifold.

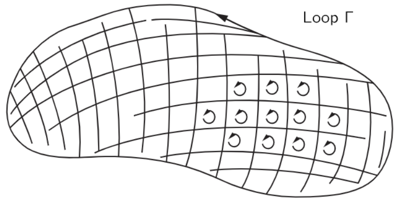

Fig. 3–9. Some surface bounded by the loop Γ is chosen. The surface is divided into a number of small areas, each approximately a square. The circulation around Γ is the sum of the circulations around the little loops.

Figure 10: This figure from the Feynman Lectures on Physics illustrates the wonderful cancellation that occurs when deriving the classical Stokes's theorem.

when physicists derive the divergence theorem or Green's theorem. (See figure 10.)

Now let's compute $\int_{\partial M_i} \omega$, the integral of ω over the boundary of the ith tiny parallelepiped. The parallelepiped M_i is based at a point $p \in M$ and spanned by tangent vectors $v_1, \ldots, v_k$, which are ordered positively, so that

$$\mu(p)(v_1, \ldots, v_k) > 0.$$

This parallelepiped has $2k$ faces, which we can think of as coming in pairs that are on opposite sides of M_i. For example, the face which is based at p and spanned by $v_2, \ldots, v_k$ is opposite from the face which is based at $p + v_1$ and spanned by $v_2, \ldots, v_k$. Let's call these faces F_1 and G_1, respectively. Likewise, the face which is based at p and spanned by $v_1, \ldots, \widehat{v_j}, \ldots, v_k$ is opposite from the face which is based at $p + v_j$ and spanned by $v_1, \ldots, \widehat{v_j}, \ldots, v_k$. Let's call these faces F_j and G_j, respectively. So

The notation $\widehat{v_j}$ means that v_j is excluded from the list.

$$\int_{\partial M_i} \omega = \sum_{j=1}^{k} \int_{F_j} \omega + \int_{G_j} \omega.$$

To evaluate $\int_{F_1} \omega$, we'll make the approximation that f is constant on F_1. In other words, f is approximately equal to $f(p)$ at all points in F_1. This yields

In section 3.4, we discussed integrating a constant differential form over a parallelepiped.

$$\int_{F_1} \omega \approx -f(p)\eta(v_2, \ldots, v_k).$$

The reason for the minus sign is that the tangent vectors $v_2, \ldots, v_k$ at the point p are not ordered correctly in the opinion of the induced boundary orientiation. The vector $-v_1$ is outward-pointing at p, and $\mu(p)(-v_1, v_2, \ldots, v_k)$ is negative.

Remember, "outward-pointing vector comes first" is the slogan of the induced boundary orientation.

To evaluate $\int_{G_1} \omega$, we make the approximation that f is approximately constant on G_1. In other words, f is approximately equal to $f(p + v_1)$ at all points in G_1. This yields

$$\int_{G_1} \omega \approx f(p + v_1)\eta(v_2, \ldots, v_k).$$

There is no minus sign this time because the tangent vectors $v_2, \ldots, v_k$ at the point $p + v_1$ are ordered correctly (in the opinion of the induced boundary orientation). The vector v_1 is outward-pointing at $p + v_1$, and $\mu(p)(v_1, v_2, \ldots, v_k)$ is positive.

Thus,

$$\begin{aligned}\int_{F_1} \omega + \int_{G_1} \omega &\approx (f(p + v_1) - f(p))\eta(v_2, \ldots, v_k) \\ &\approx \langle \nabla f(p), v_1 \rangle \eta(v_2, \ldots, v_k).\end{aligned}$$

It is no surprise that here at the crucial moment, when we are deriving the generalized Stokes's theorem, we find ourselves using Newton's approximation! Local linear approximation is indeed the key idea of calculus.

Next let's evaluate $\int_{F_2} \omega$. An argument similar to the one given above yields

$$\int_{F_2} \omega \approx f(p)\eta(v_1, v_3, \ldots, v_k).$$

There is no minus sign because the tangent vectors $v_1, v_3, \ldots, v_k$ at the point p are ordered correctly in the opinion of the induced boundary orientation. The vector $-v_2$ is outward-pointing at p, and $\mu(p)(-v_2, v_1, v_3, \ldots, v_k)$ is positive. (It is equal to $\mu(p)(v_1, v_2, v_3, \ldots, v_k)$.)

Likewise, the same reasoning also yields

$$\int_{G_2} \omega \approx -f(p + v_2)\eta(v_1, v_3, \ldots, v_k).$$

There is a minus sign because the tangent vectors $v_1, v_3, \ldots, v_k$ at the point $p + v_2$ are not ordered correctly in the opinion of the induced boundary orientation. The vector v_2 is outward-pointing at $p + v_2$, and $\mu(p)(v_2, v_1, v_3, \ldots, v_k)$ is negative.

So, we find that

$$\begin{aligned}\int_{F_2} \omega + \int_{G_2} \omega &\approx -(f(p + v_2) - f(p))\eta(v_1, v_3, \ldots, v_k) \\ &\approx -\langle \nabla f(p), v_2 \rangle \eta(v_1, v_3, \ldots, v_k).\end{aligned}$$

We can now see the pattern:

$$\int_{F_j} \omega + \int_{G_j} \omega \approx (-1)^{j+1} \langle \nabla f(p), v_j \rangle \eta(v_1, \ldots, \widehat{v_j}, \ldots, v_k)$$

for $j = 1, \ldots, k$. Thus,

$$\begin{aligned}\int_{\partial M_i} \omega &= \sum_{j=1}^{k} \int_{F_j} \omega + \int_{G_j} \omega \\ &\approx \sum_{j=1}^{k} (-1)^{j+1} \langle \nabla f(p), v_j \rangle \eta(v_1, \ldots, \widehat{v_j}, \ldots, v_k) \\ &= \sum_{j=1}^{k} (-1)^{j+1} df(p)(v_j) \eta(v_1, \ldots, \widehat{v_j}, \ldots, v_k) \qquad (8)\end{aligned}$$

where df is the differential of f, defined in equation (4).

The expression (8) reveals an interesting way to combine a 1-tensor ν with an alternating $(k-1)$-tensor η to obtain an alternating k-tensor, which we call the "wedge product" of ν and η, denoted $\nu \wedge \eta$:

$$(\nu \wedge \eta)(v_1, \ldots, v_k) = \sum_{j=1}^{k} (-1)^{j+1} \nu(v_j) \eta(v_1, \ldots, \widehat{v}_j, \ldots, v_k).$$

With this notation, we have

$$\int_{\partial M_i} \omega \approx (df(p) \wedge \eta)(v_1, \ldots, v_k).$$

So we see that

$$\begin{aligned} \int_{\partial M} \omega &= \sum_i \int_{\partial M_i} \omega \\ &\approx \sum_i (df(p) \wedge \eta)(v_1^i, \ldots, v_k^i) \qquad (9) \\ &\approx \int_M df \wedge \eta \end{aligned}$$

In the expression (9), I've introduced superscript i's to emphasize that $v_1^i, \ldots, v_k^i$ are the tangent vectors that span the ith parallelepiped M_i.

where $df \wedge \eta$ is the differential k-form on $\mathbb{R}^n$ defined by

$$(df \wedge \eta)(p) = df(p) \wedge \eta.$$

The approximations above can be made as close as we like by chopping up M into sufficiently tiny pieces. So, by a limiting argument, we conclude that

$$\int_{\partial M} \omega = \int_M df \wedge \eta.$$

6.2 *The general case*

Now let's assume that ω is any smooth differential $(k-1)$-form on $\mathbb{R}^n$. We are dropping the assumption that ω has the special form given in equation (7). From formula (6), we can express ω as

$$\omega = \sum_{i_1 < \cdots < i_{k-1}} \omega_{i_1 \cdots i_{k-1}} \varphi_{i_1 \cdots i_{k-1}}.$$

where each $\omega_{i_1 \cdots i_{k-1}} : \mathbb{R}^n \to \mathbb{R}$ is a smooth function. It follows that

$$\begin{aligned} \int_{\partial M} \omega &= \sum_{i_1 < \cdots < i_{k-1}} \int_{\partial M} \omega_{i_1 \cdots i_{k-1}} \varphi_{i_1 \cdots i_{k-1}} \\ &\approx \sum_{i_1 < \cdots < i_{k-1}} \int_M d\omega_{i_1 \cdots i_{k-1}} \wedge \varphi_{i_1 \cdots i_{k-1}} \\ &= \int_M \sum_{i_1 < \cdots < i_{k-1}} d\omega_{i_1 \cdots i_{k-1}} \wedge \varphi_{i_1 \cdots i_{k-1}} \end{aligned}$$

Again by a limiting argument, we conclude that the above approximate equality in fact holds with exact equality. If we define

$d\omega$ is called the "differential" or the "exterior derivative" of ω.

$$d\omega = \sum_{i_1 < \cdots < i_{k-1}} d\omega_{i_1 \cdots i_{k-1}} \wedge \varphi_{i_1 \cdots i_{k-1}}$$

then we obtain

$$\int_{\partial M} \omega = \int_M d\omega.$$

This is the generalized Stokes's theorem.