XII

Differential Geometry

Part XII: Contents

43 Multivariable calculus done correctly

As I have ranted about before, linear algebra is done wrong by the extensive use of matrices to obscure the structure of a linear map. Similar problems occur with multivariable calculus, so here I would like to set the record straight.

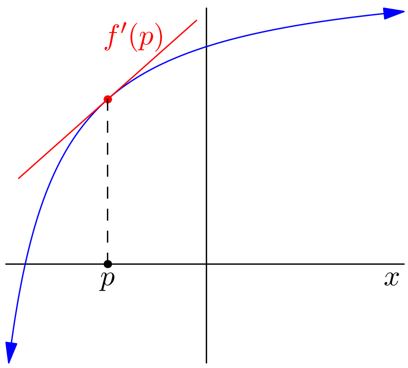
Since we are doing this chapter using morally correct linear algebra, it's imperative you're comfortable with linear maps, and in particular the dual space $V^\vee$ which we will repeatedly use.

In this chapter, all vector spaces have norms and are finite-dimensional over $\mathbb{R}$. So in particular every vector space is also a metric space (with metric given by the norm), and we can talk about open sets as usual.

§43.1 The total derivative

Prototypical example for this section: *If* $f(x,y) = x^2 + y^2$, *then* $(Df)_{(x,y)} = 2x\mathbf{e}_1^\vee + 2y\mathbf{e}_2^\vee$.

First, let $f\colon [a,b] \to \mathbb{R}$. You might recall from high school calculus that for every point $p \in \mathbb{R}$, we defined $f'(p)$ as the derivative at the point p (if it existed), which we interpreted as the *slope* of the "tangent line".



That's fine, but I claim that the "better" way to interpret the derivative at that point is as a *linear map*, that is, as a *function*. If $f'(p) = 1.5$, then the derivative tells me that if I move ε away from p then I should expect f to change by about 1.5ε. In other words,

The derivative of f at p approximates f near p by a *linear function*.

What about more generally? Suppose I have a function like $f\colon \mathbb{R}^2 \to \mathbb{R}$, say

$$f(x,y) = x^2 + y^2$$

for concreteness or something. For a point $p \in \mathbb{R}^2$, the "derivative" of f at p ought to represent a linear map that approximates f at that point p. That means I want a linear map $T\colon \mathbb{R}^2 \to \mathbb{R}$ such that

$$f(p+v) \approx f(p) + T(v)$$

for small displacements $v \in \mathbb{R}^2$.

Even more generally, if $f\colon U \to W$ with $U \subseteq V$ open (in the $\|\bullet\|_V$ metric as usual), then the derivative at $p \in U$ ought to be so that

$$f(p+v) \approx f(p) + T(v) \in W.$$

(We need U open so that for small enough v, $p+v \in U$ as well.) In fact this is exactly what we were doing earlier with $f'(p)$ in high school.

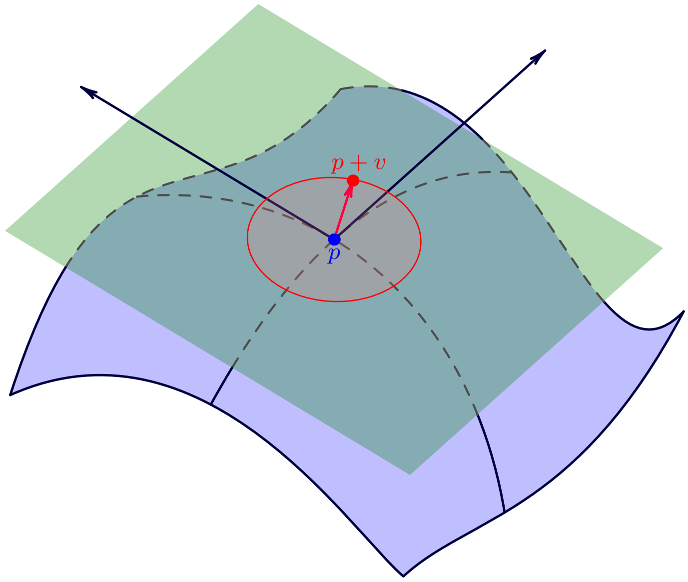


Image derived from [gk]

The only difference is that, by an unfortunate coincidence, a linear map $\mathbb{R} \to \mathbb{R}$ can be represented by just its slope. And in the unending quest to make everything a number so that it can be AP tested, we immediately forgot all about what we were trying to do in the first place and just defined the derivative of f to be a *number* instead of a *function*.

The fundamental idea of Calculus is the local approximation of functions by linear functions. The derivative does exactly this.

Jean Dieudonné as quoted in [Pu02] continues:

> In the classical teaching of Calculus, this idea is immediately obscured by the accidental fact that, on a one-dimensional vector space, there is a one-to-one correspondence between linear forms and numbers, and therefore the derivative at a point is defined as a number instead of a linear form. This **slavish subservience to the shibboleth of numerical interpretation at any cost** becomes much worse . . .

So let's do this right. The only thing that we have to do is say what "$\approx$" means, and for this we use the norm of the vector space.

Definition 43.1.1. Let $U \subseteq V$ be open. Let $f\colon U \to W$ be a continuous function, and $p \in U$. Suppose there exists a linear map $T\colon V \to W$ such that

$$\lim_{\|v\|_V \to 0} \frac{\|f(p+v) - f(p) - T(v)\|_W}{\|v\|_V} = 0.$$

Then T is the **total derivative** of f at p. We denote this by $(Df)_p$, and say f is **differentiable at p**.

If $(Df)_p$ exists at every point, we say f is **differentiable**.

Question 43.1.2. Check if that $V = W = \mathbb{R}$, this is equivalent to the single-variable definition. (What are the linear maps from V to W?)

Example 43.1.3 (Total derivative of $f(x, y) = x^2 + y^2$)
Let $V = \mathbb{R}^2$ with standard basis $\mathbf{e}_1$, $\mathbf{e}_2$ and let $W = \mathbb{R}$, and let $f(x\mathbf{e}_1 + y\mathbf{e}_2) = x^2+y^2$. Let $p = a\mathbf{e}_1 + b\mathbf{e}_2$. Then, we claim that

$$(Df)_p \colon \mathbb{R}^2 \to \mathbb{R} \quad \text{by} \quad v \mapsto 2a \cdot \mathbf{e}_1^\vee(v) + 2b \cdot \mathbf{e}_2^\vee(v).$$

Here, the notation $\mathbf{e}_1^\vee$ and $\mathbf{e}_2^\vee$ makes sense, because by definition $(Df)_p \in V^\vee$: these are functions from V to $\mathbb{R}$!

Let's check this manually with the limit definition. Set $v = xe_1 + ye_2$, and note that the norm on V is $\|(x, y)\|_V = \sqrt{x^2 + y^2}$ while the norm on W is just the absolute value $\|c\|_W = |c|$. Then we compute

$$\begin{aligned} \frac{\|f(p + v) - f(p) - T(v)\|_W}{\|v\|_V} &= \frac{\left|(a + x)^2 + (b + y)^2 - (a^2 + b^2) - (2ax + 2by)\right|}{\sqrt{x^2 + y^2}} \\ &= \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} = \sqrt{x^2 + y^2} \\ &\to 0 \end{aligned}$$

as $\|v\| \to 0$. Thus, for $p = ae_1 + be_2$ we indeed have $(Df)_p = 2a \cdot \mathbf{e}_1^\vee + 2b \cdot \mathbf{e}_2^\vee$.

Remark 43.1.4 — As usual, differentiability implies continuity.

Remark 43.1.5 — Although $U \subseteq V$, it might be helpful to think of vectors from U and V as different types of objects (in particular, note that it's possible for $0_V \notin U$). The vectors in U are "inputs" on our space while the vectors coming from V are "small displacements". For this reason, I deliberately try to use $p \in U$ and $v \in V$ when possible.

§43.2 The projection principle

You may have learned single-variable calculus as the topic of doing differentiation and integration on single-variable functions $\mathbb{R} \to \mathbb{R}$. So "multivariable calculus" ought to be calculus with functions $\mathbb{R}^n \to \mathbb{R}^m$. You might notice there are *two* upgrades happening here:

- The domain got upgraded from $\mathbb{R}$ to $\mathbb{R}^n$.
- The codomain got upgraded from $\mathbb{R}$ to $\mathbb{R}^m$.

The point of this section is that the second upgrade is is *super easy* in comparison to the first upgrade, and basically doesn't require doing anything new. All the interesting actions happens because we upgraded the domain, not the codomain. Here's why:

Theorem 43.2.1 (Projection principle)

Let U be an open subset of the vector space V. Let W be an m-dimensional real vector space with basis $w_1, \dots, w_m$. Then there is a bijection between continuous functions $f \colon U \to W$ and m-tuples of continuous $f_1, f_2, \dots, f_m \colon U \to \mathbb{R}$ by projection onto the ith basis element, i.e.

$$f(v) = f_1(v)w_1 + \dots + f_m(v)w_m.$$

Proof. Obvious. □

The theorem remains true if one replaces "continuous" by "differentiable", "smooth", "arbitrary", or most other reasonable words. Translation:

To think about a function $f \colon U \to \mathbb{R}^m$, it suffices to think about each coordinate separately.

For this reason, we'll most often be interested in functions $f \colon U \to \mathbb{R}$. That's why the dual space $V^\vee$ is so important.

§43.3 Total and partial derivatives

Prototypical example for this section: If $f(x,y) = x^2 + y^2$, then $(Df) \colon (x,y) \mapsto 2x \cdot \mathbf{e}_1^\vee + 2y \cdot \mathbf{e}_2^\vee$, and $\frac{\partial f}{\partial x} = 2x$, $\frac{\partial f}{\partial y} = 2y$.

Let $U \subseteq V$ be open and let V have a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Suppose $f \colon U \to \mathbb{R}$ is a function which is differentiable everywhere, meaning $(Df)_p \in V^\vee$ exists for every p. In that case, one can consider Df as *itself* a function:

$$\begin{aligned} Df \colon U &\to V^\vee \\ p &\mapsto (Df)_p. \end{aligned}$$

This is a little crazy: to every *point* in U we associate a *function* in $V^\vee$. We say Df is the **total derivative** of f, to reflect how much information we're dealing with. We say $(Df)_p$ is the total derivative at p.

Let's apply the projection principle now to Df. Since we picked a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of V, there is a corresponding dual basis $\mathbf{e}_1^\vee, \mathbf{e}_2^\vee, \dots, \mathbf{e}_n^\vee$. The Projection Principle tells us that Df can thus be thought of as just n functions, so we can write

$$Df = \psi_1 \mathbf{e}_1^\vee + \dots + \psi_n \mathbf{e}_n^\vee.$$

In fact, we can even describe what the ψ_i are.

Definition 43.3.1. The i^{th} **partial derivative** of $f \colon U \to \mathbb{R}$, denoted

$$\frac{\partial f}{\partial \mathbf{e}_i} \colon U \to \mathbb{R}$$

is defined by

$$\frac{\partial f}{\partial \mathbf{e}_i}(p) := \lim_{t \to 0} \frac{f(p + t\mathbf{e}_i) - f(p)}{t}.$$

You can think of it as "f' along $\mathbf{e}_i$".

Question 43.3.2. Check that if Df exists, then

$$(Df)_p(\mathbf{e}_i) = \frac{\partial f}{\partial \mathbf{e}_i}(p).$$

Remark 43.3.3 — Of course you can write down a definition of $\frac{\partial f}{\partial v}$ for any v (rather than just the $\mathbf{e}_i$).

From the above remarks, we can derive that

$$\boxed{Df = \frac{\partial f}{\partial \mathbf{e}_1} \cdot \mathbf{e}_1^\vee + \cdots + \frac{\partial f}{\partial \mathbf{e}_n} \cdot \mathbf{e}_n^\vee.}$$

and so given a basis of V, we can think of Df as just the n partials.

Remark 43.3.4 — Keep in mind that each $\frac{\partial f}{\partial \mathbf{e}_i}$ is a function from U to the *reals*. That is to say,

$$(Df)_p = \underbrace{\frac{\partial f}{\partial \mathbf{e}_1}(p)}_{\in \mathbb{R}} \cdot \mathbf{e}_1^\vee + \cdots + \underbrace{\frac{\partial f}{\partial \mathbf{e}_n}(p)}_{\in \mathbb{R}} \cdot \mathbf{e}_n^\vee \in V^\vee.$$

Example 43.3.5 (Partial derivatives of $f(x,y) = x^2 + y^2$)

Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $(x,y) \mapsto x^2 + y^2$. Then in our new language,

$$Df\colon (x,y) \mapsto 2x \cdot \mathbf{e}_1^\vee + 2y \cdot \mathbf{e}_2^\vee.$$

Thus the partials are

$$\frac{\partial f}{\partial x}\colon (x,y) \mapsto 2x \in \mathbb{R} \quad \text{and} \quad \frac{\partial f}{\partial y}\colon (x,y) \mapsto 2y \in \mathbb{R}$$

With all that said, I haven't really said much about how to find the total derivative itself. For example, if I told you

$$f(x,y) = x \sin y + x^2 y^4$$

you might want to be able to compute Df without going through that horrible limit definition I told you about earlier.

Fortunately, it turns out you already know how to compute partial derivatives, because you had to take AP Calculus at some point in your life. It turns out for most reasonable functions, this is all you'll ever need.

Theorem 43.3.6 (Continuous partials implies differentiable)

Let $U \subseteq V$ be open and pick any basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. Let $f\colon U \to \mathbb{R}$ and suppose that $\frac{\partial f}{\partial \mathbf{e}_i}$ is defined for each i and moreover is *continuous*. Then f is differentiable and Df is given by

$$Df = \sum_{i=1}^n \frac{\partial f}{\partial \mathbf{e}_i} \cdot \mathbf{e}_i^\vee.$$

Proof. Not going to write out the details, but... given $v = t_1e_1 + \cdots + t_ne_n$, the idea is to just walk from p to $p + t_1e_1$, $p + t_1e_1 + t_2e_2$, ..., up to $p + t_1e_1 + t_2e_2 + \cdots + t_ne_n = p + v$, picking up the partial derivatives on the way. Do some calculation. $\square$

Remark 43.3.7 — The continuous condition cannot be dropped. The function

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

is the classic counterexample – the total derivative Df does not exist at zero, even though both partials do.

Example 43.3.8 (Actually computing a total derivative)

Let $f(x,y) = x \sin y + x^2y^4$. Then

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \sin y + y^4 \cdot 2x \\ \frac{\partial f}{\partial y}(x,y) &= x \cos y + x^2 \cdot 4y^3. \end{aligned}$$

So Theorem 43.3.6 applies, and $Df = \frac{\partial f}{\partial x}\mathbf{e}_1^\vee + \frac{\partial f}{\partial y}\mathbf{e}_2^\vee$, which I won't bother to write out.

The example $f(x,y) = x^2 + y^2$ is the same thing. That being said, who cares about $x \sin y + x^2y^4$ anyways?

§43.4 (Optional) A word on higher derivatives

Let $U \subseteq V$ be open, and take $f \colon U \to W$, so that $Df \colon U \to \operatorname{Hom}(V,W)$.

Well, $\operatorname{Hom}(V,W)$ can also be thought of as a normed vector space in its own right: it turns out that one can define an operator norm on it by setting

$$\|T\| := \sup \left\{ \frac{\|T(v)\|_W}{\|v\|_V} \mid v \neq 0_V \right\}.$$

So $\operatorname{Hom}(V,W)$ can be thought of as a normed vector space as well. Thus it makes sense to write

$$D(Df) \colon U \to \operatorname{Hom}(V, \operatorname{Hom}(V,W))$$

which we abbreviate as D^2f. Dropping all doubt and plunging on,

$$D^3f \colon U \to \operatorname{Hom}(V, \operatorname{Hom}(V, \operatorname{Hom}(V,W))).$$

I'm sorry. As consolation, we at least know that $\operatorname{Hom}(V,W) \cong V^\vee \otimes W$ in a natural way, so we can at least condense this to

$$D^kf \colon V \to (V^\vee)^{\otimes k} \otimes W$$

rather than writing a bunch of Hom's.

Remark 43.4.1 — If $k = 2$, $W = \mathbb{R}$, then $D^2 f(v) \in (V^\vee)^{\otimes 2}$, so it can be represented as an $n \times n$ matrix, which for some reason is called a **Hessian**.

The most important property of the second derivative is that

Theorem 43.4.2 (Symmetry of $D^2 f$)
Let $f\colon U \to W$ with $U \subseteq V$ open. If $(D^2 f)_p$ exists at some $p \in U$, then it is symmetric, meaning

$$(D^2 f)_p(v_1, v_2) = (D^2 f)_p(v_2, v_1).$$

I'll just quote this without proof (see e.g. [**Pu02**, §5, theorem 16]), because double derivatives make my head spin. An important corollary of this theorem:

Corollary 43.4.3 (Clairaut's theorem: mixed partials are symmetric)
Let $f\colon U \to \mathbb{R}$ with $U \subseteq V$ open be twice differentiable. Then for any point p such that the quantities are defined,

$$\frac{\partial}{\partial \mathbf{e}_i} \frac{\partial}{\partial \mathbf{e}_j} f(p) = \frac{\partial}{\partial \mathbf{e}_j} \frac{\partial}{\partial \mathbf{e}_i} f(p).$$

§43.5 Towards differential forms

This concludes the exposition of what the derivative really is: the key idea I want to communicate in this chapter is that Df should be thought of as a map from $U \to V^\vee$.

The next natural thing to do is talk about *integration*. The correct way to do this is through a so-called *differential form*: you'll finally know what all those stupid dx's and dy's really mean. (They weren't just there for decoration!)

§43.6 A few harder problems to think about

Problem 43A* (Chain rule)**.** Let $U_1 \xrightarrow{f} U_2 \xrightarrow{g} U_3$ be differentiable maps between open sets of normed vector spaces V_i, and let $h = g \circ f$. Prove the Chain Rule: for any point $p \in U_1$, we have

$$(Dh)_p = (Dg)_{f(p)} \circ (Df)_p.$$

Problem 43B. Let $U \subseteq V$ be open, and $f\colon U \to \mathbb{R}$ be differentiable k times. Show that $(D^k f)_p$ is symmetric in its k arguments, meaning for any $v_1, \dots, v_k \in V$ and any permutation σ on $\{1, \dots, k\}$ we have

$$(D^k f)_p(v_1, \dots, v_k) = (D^k f)_p(v_{\sigma(1)}, \dots, v_{\sigma(k)}).$$

44 Differential forms

In this chapter, all vector spaces are finite-dimensional real inner product spaces. We first start by (non-rigorously) drawing pictures of all the things that we will define in this chapter. Then we re-do everything again in its proper algebraic context.

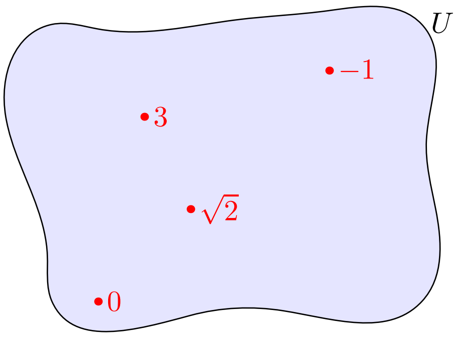
§44.1 Pictures of differential forms

Before defining a differential form, we first draw some pictures. The key thing to keep in mind is

> **"The definition of a differential form is: something you can integrate."**
> **— Joe Harris**

We'll assume that all functions are **smooth**, i.e. infinitely differentiable.

Let $U \subseteq V$ be an open set of a vector space V. Suppose that we have a function $f\colon U \to \mathbb{R}$, i.e. we assign a value to every point of U.



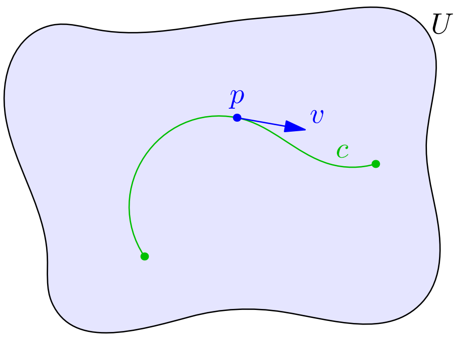
Definition 44.1.1. A **0-form** f on U is just a smooth function $f\colon U \to \mathbb{R}$.

Thus, if we specify a finite set S of points in U we can "integrate" over S by just adding up the values of the points:

$$0 + \sqrt{2} + 3 + (-1) = 2 + \sqrt{2}.$$

So, **a** 0**-form** f **lets us integrate over** 0**-dimensional "cells"**.

But this is quite boring, because as we know we like to integrate over things like curves, not single points. So, by analogy, we want a 1-form to let us integrate over 1-dimensional cells: i.e. over curves. What information would we need to do that? To answer this, let's draw a picture of a curve c, which can be thought of as a function $c\colon [0,1] \to U$.



We might think that we could get away with just specifying a number on every point of U (i.e. a 0-form f), and then somehow "add up" all the values of f along the curve. We'll use this idea in a moment, but we can in fact do something more general. Notice how when we walk along a smooth curve, at every point p we also have some extra information: a *tangent vector* v. So, we can define a 1-form α as follows. A 0-form just took a point and gave a real number, but **a 1-form will take both a point *and* a tangent vector at that point, and spit out a real number.** So a 1-form α is a smooth function on pairs (p, v), where v is a tangent vector at p, to $\mathbb{R}$. Hence

$$\alpha \colon U \times V \to \mathbb{R}.$$

Actually, for any point p, we will require that $\alpha(p, -)$ is a linear function in terms of the vectors: i.e. we want for example that $\alpha(p, 2v) = 2\alpha(p, v)$. So it is more customary to think of α as:

Definition 44.1.2. A **1-form** α is a smooth function

$$\alpha \colon U \to V^\vee.$$

Like with Df, we'll use α_p instead of $\alpha(p)$. So, at every point p, α_p is some linear functional that eats tangent vectors at p, and spits out a real number. Thus, we think of α_p as an element of $V^\vee$;

$$\alpha_p \in V^\vee.$$

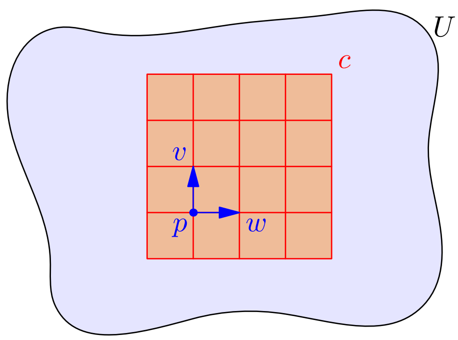
Remark 44.1.3 (Warning: arc length isn't a 1-form) — You might recall that, in high school calculus, the "arc-length element" ds can be integrated along a curve: $\int_c ds$ is the length of the curve c.

This is *not* a 1-form! More on this later. (To be brief: basically, the issue is that it's not a linear function. In some places you'll see $ds = \sqrt{dx^2 + dy^2}$ written colloquially, which should give you a sense that ds does not behave like a linear thing in dx and dy.)

Next, we draw pictures of 2-forms. This should, for example, let us integrate over a blob (a so-called 2-cell) of the form

$$c \colon [0,1] \times [0,1] \to U$$

i.e. for example, a square in U. In the previous example with 1-forms, we looked at tangent vectors to the curve c. This time, at points we will look at *pairs* of tangent vectors in U: in the same sense that lots of tangent vectors approximate the entire curve, lots of tiny squares will approximate the big square in U.



So what should a 2-form β be? As before, it should start by taking a point $p \in U$, so β_p is now a linear functional: but this time, it should be a linear map on two vectors v and w. Here v and w are not tangent so much as their span cuts out a small parallelogram. So, the right thing to do is in fact consider

$$\beta_p \in V^\vee \wedge V^\vee.$$

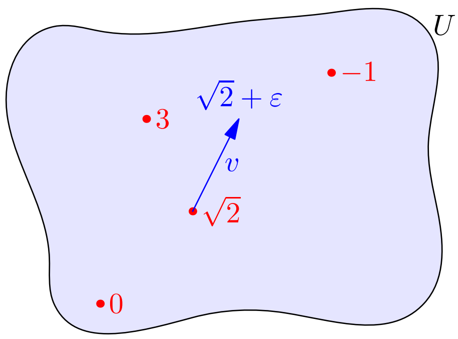
That is, to use the wedge product to get a handle on the idea that v and w span a parallelogram. Another valid choice would have been $(V \wedge V)^\vee$; in fact, the two are isomorphic,[1] but it will be more convenient to write it in the former.[2]

§44.2 Pictures of exterior derivatives

Next question:

How can we build a 1-form from a 0-form?

Let f be a 0-form on U; thus, we have a function $f : U \to \mathbb{R}$. Then in fact there is a very natural 1-form on U arising from f, appropriately called df. Namely, given a point p and a tangent vector v, the differential form $(df)_p$ returns the *change in f along v*. In other words, it's just the total derivative $(Df)_p(v)$.



Thus, df measures "the change in f".

Now, even if I haven't defined integration yet, given a curve c from a point a to b, what do you think

$$\int_c df$$

should be equal to? Remember that df is the 1-form that measures "infinitesimal change in f". So if we add up all the change in f along a path from a to b, then the answer we get should just be

$$\int_c df = f(b) - f(a).$$

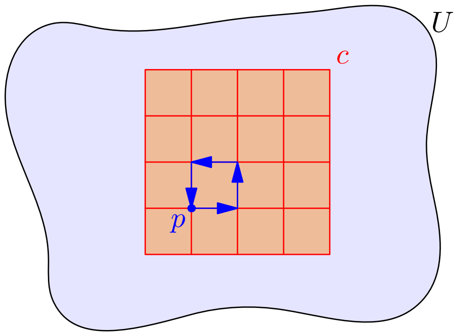
This is the first case of something we call Stokes' theorem.

Generalizing, how should we get from a 1-form to a 2-form? At each point p, the 2-form β gives a β_p which takes in a "parallelogram" and returns a real number. Now suppose we have a 1-form α. Then along each of the edges of a parallelogram, with an appropriate sign convention the 1-form α gives us a real number. So, given a 1-form α, we define $d\alpha$ to be the 2-form that takes in a parallelogram spanned by v and w, and returns **the measure of α along the boundary**.

[1] We only consider finite-dimensional V.

[2] See Section 44.5.

Now, what happens if you integrate $d\alpha$ along the entire square c? The right picture is that, if we think of each little square as making up the big square, then the adjacent boundaries cancel out, and all we are left is the main boundary. This is again just a case of the so-called Stokes' theorem.



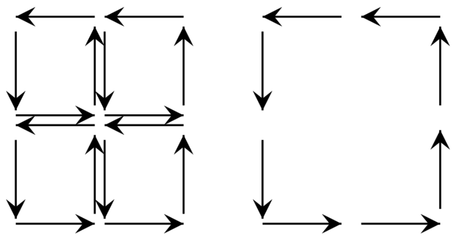


Image from [Na]

§44.3 Differential forms

Prototypical example for this section: Algebraically, something that looks like $f\mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee + \dots$, and geometrically, see the previous section.

Let's now get a handle on what dx means. Fix a real vector space V of dimension n, and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a standard basis. Let U be an open set.

Definition 44.3.1. We define a **differential k-form** α on U to be a smooth (infinitely differentiable) map $\alpha \colon U \to \bigwedge^k(V^\vee)$. (Here $\bigwedge^k(V^\vee)$ is the wedge product.)

Like with Df, we'll use α_p instead of $\alpha(p)$.

Example 44.3.2 (k-forms for $k = 0, 1$)

(a) A 0-form is just a function $U \to \mathbb{R}$.

(b) A 1-form is a function $U \to V^\vee$. For example, the total derivative Df of a function $V \to \mathbb{R}$ is a 1-form.

(c) Let $V = \mathbb{R}^3$ with standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Then a typical 2-form is given by

$$\alpha_p = f(p) \cdot \mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee + g(p) \cdot \mathbf{e}_1^\vee \wedge \mathbf{e}_3^\vee + h(p) \cdot \mathbf{e}_2^\vee \wedge \mathbf{e}_3^\vee \in \bigwedge^2(V^\vee)$$

where $f, g, h \colon V \to \mathbb{R}$ are smooth functions.

Now, by the projection principle (Theorem 43.2.1) we only have to specify a function on each of $\binom{n}{k}$ basis elements of $\bigwedge^k(V^\vee)$. So, take any basis $\{\mathbf{e}_i\}$ of V, and take the usual basis for $\bigwedge^k(V^\vee)$ of elements

$$\mathbf{e}_{i_1}^\vee \wedge \mathbf{e}_{i_2}^\vee \wedge \dots \wedge \mathbf{e}_{i_k}^\vee.$$

Thus, a general k-form takes the shape

$$\alpha_p = \sum_{1 \le i_1 < \dots < i_k \le n} f_{i_1,\dots,i_k}(p) \cdot \mathbf{e}_{i_1}^\vee \wedge \mathbf{e}_{i_2}^\vee \wedge \dots \wedge \mathbf{e}_{i_k}^\vee.$$

Since this is a huge nuisance to write, we will abbreviate this to just

$$\alpha = \sum_I f_I \cdot d\mathbf{e}_I$$

where we understand the sum runs over $I = (i_1, \dots, i_k)$, and $d\mathbf{e}_I$ represents $\mathbf{e}_{i_1}^\vee \wedge \dots \wedge \mathbf{e}_{i_k}^\vee$.

Now that we have an element $\bigwedge^k(V^\vee)$, what can it do? Well, first let me get the definition on the table, then tell you what it's doing.

Definition 44.3.3 (How to evaluate a differential form at a point)**.** For linear functions $\xi_1, \dots, \xi_k \in V^\vee$ and vectors $v_1, \dots, v_k \in V$, set

$$(\xi_1 \wedge \dots \wedge \xi_k)(v_1, \dots, v_k) := \det \begin{bmatrix} \xi_1(v_1) & \dots & \xi_1(v_k) \\ \vdots & \ddots & \vdots \\ \xi_k(v_1) & \dots & \xi_k(v_k) \end{bmatrix}.$$

You can check that this is well-defined under e.g. $v \wedge w = -w \wedge v$ and so on.

Example 44.3.4 (Evaluation of a differential form)

Set $V = \mathbb{R}^3$. Suppose that at some point p, the 2-form α returns

$$\alpha_p = 2\mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee + \mathbf{e}_1^\vee \wedge \mathbf{e}_3^\vee.$$

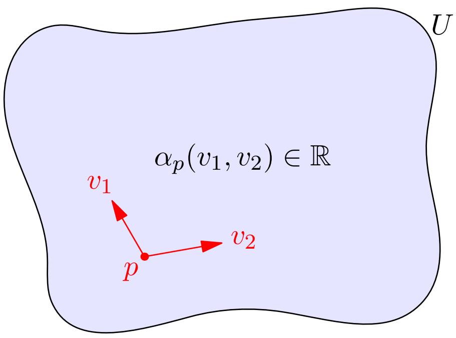
Let $v_1 = 3\mathbf{e}_1 + \mathbf{e}_2 + 4\mathbf{e}_3$ and $v_2 = 8\mathbf{e}_1 + 9\mathbf{e}_2 + 5\mathbf{e}_3$. Then

$$\alpha_p(v_1, v_2) = 2\det \begin{bmatrix} 3 & 8 \\ 1 & 9 \end{bmatrix} + \det \begin{bmatrix} 3 & 8 \\ 4 & 5 \end{bmatrix} = 21.$$

What does this definition mean? One way to say it is that

If I walk to a point $p \in U$, a k-form α will take in k vectors $v_1, \dots, v_k$ and spit out a number, which is to be interpreted as a (signed) volume.

Picture:



In other words, at every point p, we get a function α_p. Then I can feed in k vectors to α_p and get a number, which I interpret as a signed volume of the parallelepiped spanned by the $\{v_i\}$'s in some way (e.g. the flux of a force field). That's why α_p as a "function" is contrived to lie in the wedge product: this ensures that the notion of "volume" makes sense, so that for example, the equality $\alpha_p(v_1, v_2) = -\alpha_p(v_2, v_1)$ holds.

This is what makes differential forms so fit for integration.

§44.4 Exterior derivatives

Prototypical example for this section: Possibly $(dx_1)_p = \mathbf{e}_1^\vee$.

We now define the exterior derivative[3] df that we gave pictures of at the beginning of the chapter. It turns out that the exterior derivative is easy to compute given explicit coordinates to work with.

Firstly, we define the exterior derivative of a function $f \colon U \to \mathbb{R}$, as

$$df := Df = \sum_i \frac{\partial f}{\partial \mathbf{e}_i} \mathbf{e}_i^\vee$$

In particular, suppose $V = \mathbb{R}^n$ and $f(x_1, \dots, x_n) = x_1$ (i.e. $f = \mathbf{e}_1^\vee$). Then:

> **Question 44.4.1.** Show that for any $p \in U$,
> $$\left(d(\mathbf{e}_1^\vee)\right)_p = \mathbf{e}_1^\vee.$$

Abuse of Notation 44.4.2. Unfortunately, someone somewhere decided it would be a good idea to use "x_1" to denote $\mathbf{e}_1^\vee$ (because *obviously*[4] x_1 means "the function that takes $(x_1, \dots, x_n) \in \mathbb{R}^n$ to x_1") and then decided that

$$dx_1 := d(\mathbf{e}_1^\vee).$$

This notation is so entrenched that I have no choice but to grudgingly accept it. Note that it's not even right, since technically it's $(dx_1)_p = \mathbf{e}_1^\vee$; dx_1 is a 1-form.

> **Remark 44.4.3** — This is the reason why we use the notation $\frac{df}{dx}$ in calculus now: given, say, $f \colon \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$, it is indeed true that
> $$df = 2x \cdot \mathbf{e}_1^\vee = 2x \cdot dx$$
> and so by (more) abuse of notation we write $df/dx = 2x$.

More generally, we can define the exterior derivative in terms of our basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ as follows:

Definition 44.4.4. If $\alpha = \sum_I f_I d\mathbf{e}_I$ then we define the **exterior derivative** as

$$d\alpha := \sum_I df_I \wedge d\mathbf{e}_I = \sum_I \sum_j \frac{\partial f_I}{\partial \mathbf{e}_j} d\mathbf{e}_j \wedge d\mathbf{e}_I.$$

It turns out this doesn't depend on the choice of basis; we'll mention a basis-free definition at the end of this section.

> **Example 44.4.5** (Computing some exterior derivatives)
> Let $V = \mathbb{R}^3$ with standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Let $f(x, y, z) = x^4 + y^3 + 2xz$. Then we compute
> $$df = Df = (4x^3 + 2z)\, dx + 3y^2\, dy + 2x\, dz.$$

[3] The name "exterior derivative" comes from the wedge product $\wedge$ being alternatively called the exterior product.

[4] Sarcasm.

Next, we can evaluate $d(df)$ as prescribed: it is

$$\begin{aligned} d^2 f &= (12x^2\, dx + 2dz) \wedge dx + (6y\, dy) \wedge dy + 2(dx \wedge dz) \\ &= 12x^2(dx \wedge dx) + 2(dz \wedge dx) + 6y(dy \wedge dy) + 2(dx \wedge dz) \\ &= 2(dz \wedge dx) + 2(dx \wedge dz) \\ &= 0. \end{aligned}$$

So surprisingly, $d^2 f$ is the zero map. Here, we have exploited Abuse of Notation 44.4.2 for the first time, in writing dx, dy, dz.

And in fact, this is always true in general:

Theorem 44.4.6 (Exterior derivative vanishes)
Let α be any k-form. Then $d^2(\alpha) = 0$. Even more succinctly,

$$d^2 = 0.$$

The proof is left as Problem 44B.

> **Exercise 44.4.7.** Compare the statement $d^2 = 0$ to the geometric picture of a 2-form given at the beginning of this chapter. Why does this intuitively make sense?

Here are some other properties of d:

- As we just saw, $d^2 = 0$.
- For a k-form α and ℓ-form β, one can show that
$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k(\alpha \wedge d\beta).$$
- If $f \colon U \to \mathbb{R}$ is smooth, then $df = Df$.

In fact, one can show that df as defined above is the *unique* map sending k-forms to $(k+1)$-forms with these properties. So, one way to *define* df is to take as axioms the bulleted properties above and then declare d to be the unique solution to this functional equation. In any case, this tells us that our definition of d does not depend on the basis chosen.

Recall that df measures the change in boundary. In that sense, $d^2 = 0$ is saying something like "the boundary of the boundary is empty". We'll make this precise when we see Stokes' theorem in the next chapter.

§44.5 Digression: $\bigwedge^k(V^\vee)$ versus $(\bigwedge^k(V))^\vee$

Earlier on, we remarked that $\bigwedge^k(V^\vee) \cong (\bigwedge^k(V))^\vee$ canonically, but we use the former for convenience.

The former notation is indeed more convenient (wedge product of two differential form is natural), but it's not clear why Definition 44.3.3 is defined in such a way.

If we used $(\bigwedge^k(V))^\vee$ instead, it's trivial to evaluate a differential form: For $f \in (\bigwedge^k(V))^\vee$ and vectors $v_1, \dots, v_k \in V$, then

$$f(v_1, \dots, v_k) \coloneqq f(v_1 \wedge \cdots \wedge v_k).$$

This is because f naturally takes in an element of $\bigwedge^k(V)$ and returns a real number.

But now, it is not clear how we can take $f \in (\bigwedge^1(V))^\vee$ and $g \in (\bigwedge^1(V))$, and return something like $f \wedge g \in (\bigwedge^2(V))^\vee$: The natural choice $(v \wedge w \mapsto f(v)g(w))$ isn't even well-defined![5]

To figure out what to do, we have to take a step back and consider the tensor product. For a vector space V, define $T^k(V) = \underbrace{V \otimes V \otimes \cdots \otimes V}_{k \text{ times}}$.

We have the following diagram:

$$\begin{tikzcd}
T^k(V^\vee) \arrow[r, hook, "\phi"] \arrow[d, shift left, two heads, "q"] & (T^k(V))^\vee \\
\bigwedge^k(V^\vee) \arrow[u, hook, shift left, "\iota"] & (\bigwedge^k(V))^\vee \arrow[u, hook, "q^\vee"']
\end{tikzcd}$$

What is going on?

First, there is a natural map $T^k(V^\vee) \to (T^k(V))^\vee$ given by

$$\phi(\xi_1 \otimes \cdots \otimes \xi_k) = v_1 \otimes \cdots \otimes v_k \mapsto \xi_1(v_1)\xi_2(v_2)\cdots\xi_k(v_k)$$

and extends to all of $T^k(V^\vee)$ by linearity the obvious way.

Unlike the situation with the wedge product above, this map is indeed well-defined.

With some manual work, we can check ϕ is injective. Because both $T^k(V^\vee)$ and $(T^k(V))^\vee$ has dimension $(\dim V)^k$, ϕ is bijective.

Next, note that $\bigwedge^k(V)$ is just "$T^k(V)$ but with more relations imposed", there is a natural quotient map $q\colon T^k(V) \twoheadrightarrow \bigwedge^k(V)$. So, the tensors are divided into equivalence classes.

Example 44.5.1

If $V = \mathbb{R}^2$, then $T^2(V)$ would have elements such as $\mathbf{e}_1 \otimes \mathbf{e}_1$, $\mathbf{e}_1 \otimes \mathbf{e}_2$ or $-\mathbf{e}_2 \otimes \mathbf{e}_1$.

Mapping these elements to $\bigwedge^2(V)$, we get $\mathbf{e}_1 \wedge \mathbf{e}_1 = 0$, and $\mathbf{e}_1 \wedge \mathbf{e}_2 = -\mathbf{e}_2 \wedge \mathbf{e}_1$, i.e. $\mathbf{e}_1 \otimes \mathbf{e}_2$ and $-\mathbf{e}_2 \otimes \mathbf{e}_1$ belongs to the same equivalence class.

The map q induces a dual map $q^\vee\colon (\bigwedge^k(V))^\vee \to (T^k(V))^\vee$.

> **Question 44.5.2.** Convince yourself that a function $f \in (T^k(V))^\vee$ belongs to $\operatorname{im} q^\vee$ if and only if f assigns the same value for every element in each equivalence class, as defined above.

Thus, we get an isomorphism $q \circ \phi^{-1} \circ q^\vee\colon (\bigwedge^k(V))^\vee \to \bigwedge^k(V^\vee)$.[6]

To check this is indeed an isomorphism, we will construct its inverse map. As defined above, each equivalence class in $T^k(V^\vee)$ (fiber of $g \in \bigwedge^k(V^\vee)$) has multiple elements, but we can find a canonical element by the following:

Definition 44.5.3. For vector space V, and element $f = f_1 \otimes f_2 \otimes \cdots \otimes f_k \in T^k(V)$, we define the alternation of f as follows:

$$\operatorname{Alt} f = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma) \cdot (f_{\sigma(1)} \otimes f_{\sigma(2)} \otimes \cdots \otimes f_{\sigma(k)})$$

and extend it to all of $T^k(V)$.

Here, S_k is the permutation group. Notice the similarity with the definition of the determinant.

[5] You can try it with $f = \mathbf{e}_1^\vee$ and $g = \mathbf{e}_2^\vee$, evaluate it at $\mathbf{e}_1 \wedge \mathbf{e}_2$ and $-\mathbf{e}_2 \wedge \mathbf{e}_1$, which we know is equal.

[6] We're using q for both the map $T^k(V) \to \bigwedge^k(V)$ and $T^k(V^\vee) \to \bigwedge^k(V^\vee)$, by abuse of notation.

Example 44.5.4
As above, $V = \mathbb{R}^2$. Then we get:

$$\operatorname{Alt}(\mathbf{e}_1 \otimes \mathbf{e}_2) = \frac{\mathbf{e}_1 \otimes \mathbf{e}_2 - \mathbf{e}_2 \otimes \mathbf{e}_1}{2}.$$

Notice that if we swap the first and second component of $\mathbf{e}_1 \otimes \mathbf{e}_2$, we get $\mathbf{e}_2 \otimes \mathbf{e}_1$ which has little to do with the original tensor. However, if we swap the first and second component of $\frac{\mathbf{e}_1 \otimes \mathbf{e}_2 - \mathbf{e}_2 \otimes \mathbf{e}_1}{2}$, we get $\frac{\mathbf{e}_2 \otimes \mathbf{e}_1 - \mathbf{e}_1 \otimes \mathbf{e}_2}{2}$, which is exactly the negation of the original tensor!

We see that $\operatorname{Alt} f$ is a desirable representative of the equivalence class of f because:

- $\operatorname{Alt}(\operatorname{Alt} f) = \operatorname{Alt} f$;
- $q(f) = q(\operatorname{Alt} f)$ where q is the quotient map $T^k(V) \twoheadrightarrow \bigwedge^k(V)$;
- $\operatorname{Alt} f$ is an *alternating tensor* — that is, if we swap two adjacent components of $\operatorname{Alt} f$ for each pure tensor, then the whole tensor gets negated.

Thus it makes sense for us to define $\iota\colon \bigwedge^k(V^\vee) \hookrightarrow T^k(V^\vee)$ that takes each element to the alternating tensor in $T^k(V^\vee)$.

Example 44.5.5
With the same example as above, $V = \mathbb{R}^2$, then we get

$$\iota(\mathbf{e}_1 \wedge \mathbf{e}_2) = \operatorname{Alt}(\mathbf{e}_1 \otimes \mathbf{e}_2) = \frac{\mathbf{e}_2 \otimes \mathbf{e}_1 - \mathbf{e}_1 \otimes \mathbf{e}_2}{2}.$$

Finally,

Exercise 44.5.6. Show that $\operatorname{im}(\phi \circ \iota) = \operatorname{im} q^\vee$, and that $\iota^\vee \circ \phi \circ \iota$ and $q \circ \phi^{-1} \circ q^\vee$ are inverses of each other.

It is common notation that we want to define the wedge product such that $\mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee$ takes in $\mathbf{e}_1 \wedge \mathbf{e}_2$ (that is, the square formed by $\mathbf{e}_1$ and $\mathbf{e}_2$), and returns 1. However, if we define the wedge product naturally by the method above, we get

$$\iota(\mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee) = \frac{\mathbf{e}_1^\vee \otimes \mathbf{e}_2^\vee - \mathbf{e}_2^\vee \otimes \mathbf{e}_1^\vee}{2}$$

which means

$$\phi(\iota(\mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee))(\mathbf{e}_1, \mathbf{e}_2) = \frac{1 \cdot 1 - 0 \cdot 0}{2} = \frac{1}{2}.$$

So, a corrective factor $k!$ is needed.

To see how "difficult" the wedge product will be if we use the second notation, let $V = \mathbb{R}^3$, $\alpha = dx \wedge dy \in \bigwedge^2(V^\vee)$, and $\beta = dz \in \bigwedge^1(V^\vee)$.

Then, we know:

- $\alpha(\mathbf{e}_1 \wedge \mathbf{e}_2) = 1$.
- $\beta(\mathbf{e}_3) = 1$.
- We should have $(\alpha \wedge \beta)(\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3) = 1$.

The last point is obvious if we let the wedge product be the map $\wedge \colon \bigwedge^2(V^\vee) \times \bigwedge^1(V^\vee) \to \bigwedge^3(V^\vee)$.

However, if we're given α and β as elements of $(\bigwedge^2(V))^\vee$ and $(\bigwedge^1(V))^\vee$ respectively (that is, we can only evaluate α at $v \wedge w$ for $v, w \in V$; and we can only evaluate β at v for $v \in V$), then it would be much more difficult to write down what $\alpha \wedge \beta$ should be. In fact,

$$(\alpha \wedge \beta)(v_1 \wedge v_2 \wedge v_3) = \alpha(v_1 \wedge v_2)\beta(v_3) - \alpha(v_1 \wedge v_3)\beta(v_2) + \alpha(v_2 \wedge v_3)\beta(v_1).$$

You can see that this is a variant of the alternation operator (or the evaluation operation), where we compute a weighted average in order to force $\alpha \wedge \beta$ to be alternating.

§44.6 Tangential remark: Arc length ds is not a 1-form

As mentioned in a remark earlier, the arc length ds is not a 1-form.[7]

We said earlier that differential form is something you can integrate. You can certainly integrate ds, but it's not considered a 1-form!

While we can easily check against the definition that ds is not linear (Problem 45E), it still raises the question that why we would want to define differential form to exclude ds. What's going on here?

In fact, the true story is that the objects that are integrable over a smooth curve are **1-densities**. We will define this later.

For simplicity, we work over $\mathbb{R}^2$ in this section. Given a (smooth) 1-density ω that can be integrated over a smooth curve c, we would like the integral $\int_c \omega$ to have the following properties:

- It is additive: if c is the concatenation of two curves c_1 and c_2, then $\int_c \omega = \int_{c_1} \omega + \int_{c_2} \omega$.
- Because everything is smooth, we would expect that if c is a tiny line segment, then in fact $\int_{c_1} \omega \approx \int_{c_2} \omega$ if we divide c into two segments c_1 and c_2 of equal length.

  Thus, it's natural to require $\int_c \omega$ to be "approximately linear" when the length of c is small enough.

  In symbols: for $\varepsilon > 0$, let c_ε be the initial segment of the curve c with length ε, then
  $$\lim_{\varepsilon \to 0^+} \frac{\int_{c_\varepsilon} \omega}{\varepsilon} = h$$
  for some finite constant h.

We certainly can formalize a 1-density ω to be simply a function that takes smooth curves c and returns the value $\int_c \omega$ satisfying the two conditions above, but this definition is clunky.

A better way to do it is to observe that, if we know $\int_c \omega$ for tiny curves c, then we can integrate ω over any smooth curves c by chopping it up into tiny curves. But this isn't completely formal — of course, as the length of a curve tends to 0, the integral $\int_c \omega$ also tends to 0 — so instead, we consider the limit above:

$$\lim_{\varepsilon \to 0^+} \frac{\int_{c_\varepsilon} \omega}{\varepsilon}.$$

[7] https://mathoverflow.net/q/90455 has a discussion on this.

Question 44.6.1. Convince yourself that, given two curves $c\colon [0,1] \to \mathbb{R}^2$ and $c_2\colon [0,1] \to \mathbb{R}^2$ that starts at the same point $c(0) = c_2(0) = p$, and moves in the same direction $c'(0) = c_2'(0) = v$, then basic smoothness condition of $\int_c \omega$ would guarantee that the limit above is the same.

Thus,

We can define a 1-density ω to be a function that takes in a point p and the initial direction $v \in \mathbb{R}^2$, which is understood as a tangent vector of $\mathbb{R}^2$ at p, and returns the limit.

Formally:

Definition 44.6.2 (1-density)**.** A 1-density ω is a function $\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$.

We write $\omega_p(v) \in \mathbb{R}$.

Since only the direction matters, it makes sense to make ω satisfy $\omega_p(\lambda v) = \lambda \omega_p(v)$ for $\lambda \ge 0$. In particular, $\omega_p(0) = 0$.

Then, ds is the differential form $ds_p(v) = \|v\|$. While we have not rigorously defined how to integrate over a curve (we will do this next chapter), you can intuitively see how it works.

With this definition, a 1-form is just a 1-density that is in addition linear in the second argument — $\omega_p(v + w) = \omega_p(v) + \omega_p(w)$.

So, what is the special properties that differential forms enjoys? For one, if ω is a differential form, we have:

> Let $c\colon [0,1] \to \mathbb{R}^2$ be a smooth curve, then for any sequence of smooth curves c_k that converges uniformly to c, then $\int_{c_k} \omega$ converges to $\int_c \omega$.

You can easily imagine how this can fail for ds — a sequence of piecewise smooth curves that consist of only horizontal and vertical lines can approximate a circle, but the arc length of these jagged curves can never converge to the circumference of the circle.[8]

§44.7 Closed and exact forms

Let α be a k-form.

Definition 44.7.1. We say α is **closed** if $d\alpha = 0$.

Definition 44.7.2. We say α is **exact** if for some $(k-1)$-form β, $d\beta = \alpha$. If $k = 0$, α is exact only when $\alpha = 0$.

Question 44.7.3. Show that exact forms are closed.

A natural question arises: are there closed forms which are not exact? Surprisingly, the answer to this question is tied to topology. Here is one important example.

Example 44.7.4 (The angle form)

Let $U = \mathbb{R}^2 \setminus \{0\}$, and let $\theta(p)$ be the angle formed by the x-axis and the line from the origin to p.

[8] In fact, using the same argument, you can also prove that, conversely, any smooth density that satisfies the latter property must in fact be linear!

The 1-form $\alpha \colon U \to (\mathbb{R}^2)^\vee$ defined by

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

is called the **angle form**: given $p \in U$ it measures the change in angle $\theta(p)$ along a tangent vector. So intuitively, "$\alpha = d\theta$". Indeed, one can check directly that the angle form is closed.

However, α is not exact: there is no global smooth function $\theta \colon U \to \mathbb{R}$ having α as a derivative. This reflects the fact that one can actually perform a full 2π rotation around the origin, i.e. θ only makes sense mod 2π. Thus existence of the angle form α reflects the possibility of "winding" around the origin.

So the key idea is that the failure of a closed form to be exact corresponds quite well with "holes" in the space: the same information that homotopy and homology groups are trying to capture. To draw another analogy, in complex analysis Cauchy-Goursat only works when U is simply connected. The "hole" in U is being detected by the existence of a form α. The so-called de Rham cohomology will make this relation explicit.

§44.8 A few harder problems to think about

Problem 44A. Show directly that the angle form

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

is closed.

Problem 44B. Establish Theorem 44.4.6, which states that $d^2 = 0$.

45 Integrating differential forms

We now show how to integrate differential forms over cells, and state Stokes' theorem in this context. In this chapter, all vector spaces are finite-dimensional and real.

§45.1 Motivation: line integrals

Given a function $g\colon [a,b] \to \mathbb{R}$, we know by the fundamental theorem of calculus that

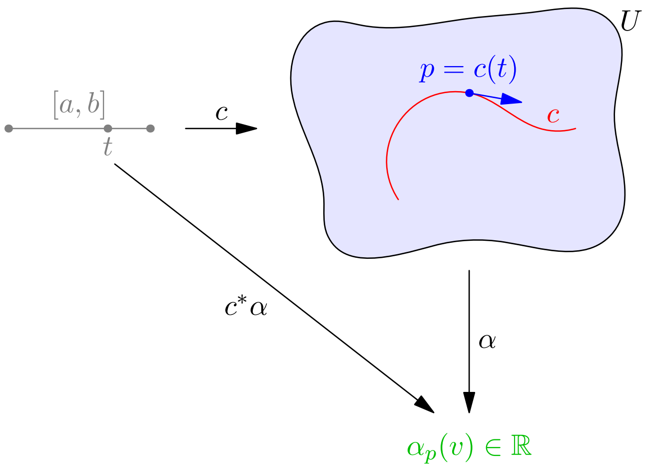
$$\int_{[a,b]} g(t)\; dt = f(b) - f(a)$$

where f is a function such that $g = df/dt$. Equivalently, for $f\colon [a,b] \to \mathbb{R}$,

$$\int_{[a,b]} g\; dt = \int_{[a,b]} df = f(b) - f(a)$$

where df is the exterior derivative we defined earlier.

Cool, so we can integrate over $[a,b]$. Now suppose more generally, we have U an open subset of our real vector space V and a 1-form $\alpha\colon U \to V^\vee$. We consider a **parametrized curve**, which is a smooth function $c\colon [a,b] \to U$. Picture:



We want to define an $\int_c \alpha$ such that:

The integral $\int_c \alpha$ should add up all the α along the curve c.

Our differential form α first takes in a point p to get $\alpha_p \in V^\vee$. Then, it eats a tangent vector $v \in V$ to the curve c to finally give a real number $\alpha_p(v) \in \mathbb{R}$. We would like to "add all these numbers up", using only the notion of an integral over $[a,b]$.

> **Exercise 45.1.1.** Try to guess what the definition of the integral should be. (By type-checking, there's only one reasonable answer.)

So, the definition we give is

$$\int_c \alpha := \int_{[a,b]} \alpha_{c(t)}\left(c'(t)\right)\; dt.$$

Here, $c'(t)$ is shorthand for $(Dc)_t(1)$. It represents the *tangent vector* to the curve c at the point $p = c(t)$, at time t. (Here we are taking advantage of the fact that $[a, b]$ is one-dimensional.)

Now that definition was a pain to write, so we will define a differential 1-form $c^*\alpha$ on $[a, b]$ to swallow that entire thing: specifically, in this case we define $c^*\alpha$ to be

$$(c^*\alpha)_t(\varepsilon) = \alpha_{c(t)} \cdot (Dc)_t(\varepsilon)$$

(here ε is some displacement in time). Thus, we can more succinctly write

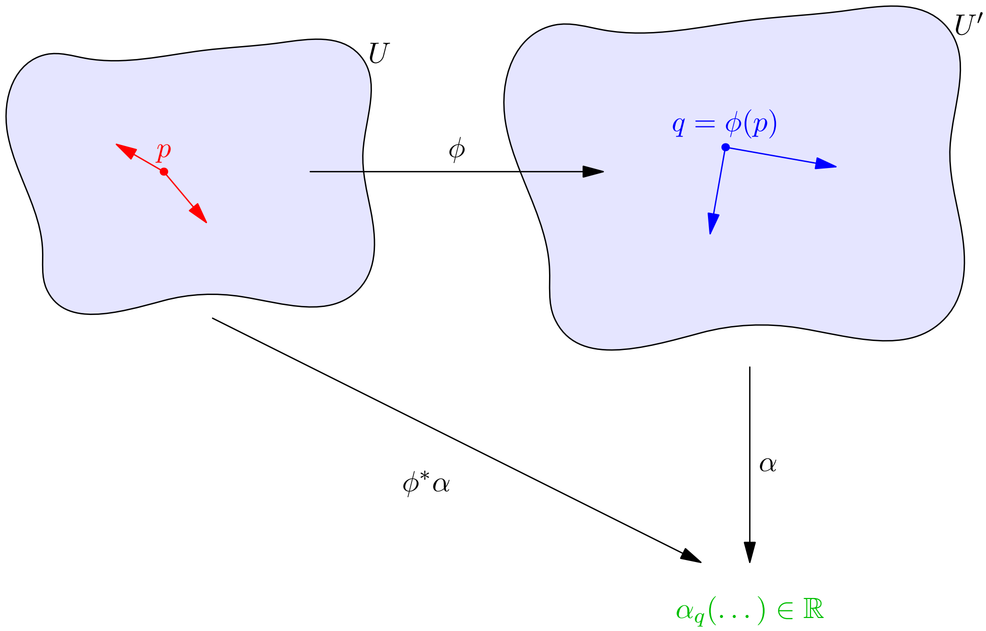
$$\int_c \alpha := \int_{[a,b]} c^*\alpha.$$

This is a special case of a *pullback*: roughly, if $\phi \colon U \to U'$ (where $U \subseteq V$, $U' \subseteq V'$), we can change any differential k-form α on U' to a k-form on U. In particular, if $U = [a, b]$,[1] we can resort to our old definition of an integral. Let's now do this in full generality.

§45.2 Pullbacks

Let V and V' be finite dimensional real vector spaces (possibly different dimensions) and suppose U and U' are open subsets of each; next, consider a k-form α on U'.

Given a map $\phi \colon U \to U'$ we now want to define a pullback in much the same way as before. Picture:



Well, there's a total of about one thing we can do. Specifically: α accepts a point in U' and k tangent vectors in V', and returns a real number. We want $\phi^*\alpha$ to accept a point in $p \in U$ and k tangent vectors $v_1, \dots, v_k$ in V, and feed the corresponding information to α.

Clearly we give the point $q = \phi(p)$. As for the tangent vectors, since we are interested in volume, we take the derivative of ϕ at p, $(D\phi)_p$, which will scale each of our vectors v_i into some vector in the target V'. To cut a long story short:

[1] OK, so $[a, b]$ isn't actually open, sorry. I ought to write $(a - \varepsilon, b + \varepsilon)$, or something.

Definition 45.2.1. Given $\phi\colon U \to U'$ and α a k-form, we define the **pullback**

$$(\phi^*\alpha)_p(v_1, \dots, v_k) := \alpha_{\phi(p)}\left((D\phi)_p(v_1), \dots, (D\phi)_p(v_k)\right).$$

There is a more concrete way to define the pullback using bases. Suppose $w_1, \dots, w_n$ is a basis of V' and $e_1, \dots, e_m$ is a basis of V. Thus, by the projection principle (Theorem 43.2.1) the map $\phi\colon V \to V'$ can be thought of as

$$\phi(v) = \phi_1(v)w_1 + \dots + \phi_n(v)w_n$$

where each ϕ_i takes in a $v \in V$ and returns a real number. We know also that α can be written concretely as

$$\alpha = \sum_{I \subseteq \{1,\dots,n\}} f_J \, dw_J.$$

Then, we define

$$\phi^*\alpha = \sum_{I \subseteq \{1,\dots,n\}} (f_I \circ \phi)(D\phi_{i_1} \wedge \dots \wedge D\phi_{i_k}).$$

A diligent reader can check these definitions are equivalent.

Example 45.2.2 (Computation of a pullback)

Let $V = \mathbb{R}^2$ with basis $\mathbf{e}_1$ and $\mathbf{e}_2$, and suppose $\phi\colon V \to V'$ is given by sending

$$\phi(a\mathbf{e}_1 + b\mathbf{e}_2) = (a^2 + b^2)w_1 + \log(a^2 + 1)w_2 + b^3 w_3$$

where w_1, w_2, w_3 is a basis for V'. Consider the form $\alpha_q = f(q) w_1^\vee \wedge w_3^\vee$, where $f\colon V' \to \mathbb{R}$. Then

$$(\phi^*\alpha)_p = f(\phi(p)) \cdot (2a\mathbf{e}_1^\vee + 2b\mathbf{e}_2^\vee) \wedge (3b^2\mathbf{e}_2^\vee) = f(\phi(p)) \cdot 6ab^2 \cdot \mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee.$$

It turns out that the pullback basically behaves nicely as possible, e.g.

- $\phi^*(c\alpha + \beta) = c\phi^*\alpha + \phi^*\beta$ (linearity)
- $\phi^*(\alpha \wedge \beta) = (\phi^*\alpha) \wedge (\phi^*\beta)$
- $\phi_1^*(\phi_2^*(\alpha)) = (\phi_2 \circ \phi_1)^*(\alpha)$ (naturality)

but I won't take the time to check these here (one can verify them all by expanding with a basis).

§45.3 Cells

Prototypical example for this section: A disk in $\mathbb{R}^2$ can be thought of as the cell $[0, R] \times [0, 2\pi] \to \mathbb{R}^2$ by $(r, \theta) \mapsto (r\cos\theta)\mathbf{e}_1 + (r\sin\theta)\mathbf{e}_2$.

Now that we have the notion of a pullback, we can define the notion of an integral for more general spaces. Specifically, to generalize the notion of integrals we had before:

Definition 45.3.1. A **k-cell** is a smooth function $c\colon [a_1, b_1] \times [a_2, b_2] \times \dots [a_k, b_k] \to V$.

Example 45.3.2 (Examples of cells)

Let $V = \mathbb{R}^2$ for convenience.

(a) A 0-cell consists of a single point.

(b) As we saw, a 1-cell is an arbitrary curve.

(c) A 2-cell corresponds to a 2-dimensional surface. For example, the map $c \colon [0,R] \times [0,2\pi] \to V$ by

$$c \colon (r,\theta) \mapsto (r\cos\theta, r\sin\theta)$$

can be thought of as a disk of radius R.

So we can now give the definition.

Definition 45.3.3 (How to integrate differential k-forms)**.** Take a differential k-form α and a k-cell $c \colon [0,1]^k \to V$. Define the integral $\int_c \alpha$ using the pullback

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

Since $c^*\alpha$ is a k-form on the k-dimensional unit box, it can be written as $f(x_1,\dots,x_n)\, dx_1 \wedge \dots \wedge dx_n$, So the above integral could also be written as

$$\int_0^1 \cdots \int_0^1 f(x_1,\dots,x_n)\, dx_1 \wedge \dots \wedge dx_n.$$

Example 45.3.4 (Area of a circle)

Consider $V = \mathbb{R}^2$ and let $c \colon (r,\theta) \mapsto (r\cos\theta)\mathbf{e}_1 + (r\sin\theta)\mathbf{e}_2$ on $[0,R] \times [0,2\pi]$ as before. Take the 2-form α which gives $\alpha_p = \mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee$ at every point p. Then

$$\begin{aligned} c^*\alpha &= (\cos\theta dr - r\sin\theta d\theta) \wedge (\sin\theta dr + r\cos\theta d\theta) \\ &= r(\cos^2\theta + \sin^2\theta)(dr \wedge d\theta) \\ &= r\; dr \wedge d\theta \end{aligned}$$

Thus,

$$\int_c \alpha = \int_0^R \int_0^{2\pi} r\; dr \wedge d\theta = \pi R^2$$

which is the area of a circle.

Here's some geometric intuition for what's happening. Given a k-cell in V, a differential k-form α accepts a point p and some tangent vectors v_1, …, v_k and spits out a number $\alpha_p(v_1,\dots,v_k)$, which as before we view as a signed hypervolume. Then the integral *adds up all these infinitesimals across the entire cell.* In particular, if $V = \mathbb{R}^k$ and we take the form $\alpha \colon p \mapsto \mathbf{e}_1^\vee \wedge \dots \wedge \mathbf{e}_k^\vee$, then what these α's give is the kth hypervolume of the cell. For this reason, this α is called the **volume form** on $\mathbb{R}^k$.

You'll notice I'm starting to play loose with the term "cell": while the cell $c \colon [0,R] \times [0,2\pi] \to \mathbb{R}^2$ is supposed to be a function I have been telling you to think of it as a unit disk (i.e. in terms of its image). In the same vein, a curve $[0,1] \to V$ should be thought of as a curve in space, rather than a function on time.

This error turns out to be benign. Let α be a k-form on U and $c\colon [a_1, b_1] \times \cdots \times [a_k, b_k] \to U$ a k-cell. Suppose $\phi\colon [a'_1, b'_1] \times \ldots [a'_k, b'_k] \to [a_1, b_1] \times \cdots \times [a_k, b_k]$; it is a **reparametrization** if ϕ is bijective and $(D\phi)_p$ is always invertible (think "change of variables"); thus

$$c \circ \phi\colon [a'_1, b'_1] \times \cdots \times [a'_k, b'_k] \to U$$

is a k-cell as well. Then it is said to **preserve orientation** if $\det(D\phi)_p > 0$ for all p and **reverse orientation** if $\det(D\phi)_p < 0$ for all p.

Exercise 45.3.5. Why is it that exactly one of these cases must occur?

Theorem 45.3.6 (Changing variables doesn't affect integrals)
Let c be a k-cell, α a k-form, and ϕ a reparametrization. Then

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

Proof. Use naturality of the pullback to reduce it to the corresponding theorem in normal calculus. $\square$

So for example, if we had parametrized the unit circle as $[0, 1] \times [0, 1] \to \mathbb{R}^2$ by $(r, t) \mapsto R\cos(2\pi t)\mathbf{e}_1 + R\sin(2\pi t)\mathbf{e}_2$, we would have arrived at the same result. So we really can think of a k-cell just in terms of the points it specifies.

§45.4 Boundaries

Prototypical example for this section: The boundary of $[a, b]$ is $\{b\} - \{a\}$. The boundary of a square goes around its edge counterclockwise.

First, I introduce a technical term that lets us consider multiple cells at once.

Definition 45.4.1. A ***k*-chain** U is a formal linear combination of k-cells over U, i.e. a sum of the form

$$c = a_1 c_1 + \cdots + a_m c_m$$

where each $a_i \in \mathbb{R}$ and c_i is a k-cell. We define $\int_c \alpha = \sum_i a_i \int c_i$.

In particular, a 0-chain consists of several points, each with a given weight.

Now, how do we define the boundary? For a 1-cell $[a, b] \to U$, as I hinted earlier we want the answer to be the 0-chain $\{c(b)\} - \{c(a)\}$. Here's how we do it in general.

Definition 45.4.2. Suppose $c\colon [0, 1]^k \to U$ is a k-cell. Then the **boundary** of c, denoted $\partial c\colon [0, 1]^{k-1} \to U$, is the $(k-1)$-chain defined as follows. For each $i = 1, \ldots, k$ define $(k-1)$-chains by

$$\begin{aligned} c_i^{\text{start}}&\colon (t_1, \ldots, t_{k-1}) \mapsto c(t_1, \ldots, t_{i-1}, 0, t_i, \ldots, t_{k-1}) \\ c_i^{\text{stop}}&\colon (t_1, \ldots, t_{k-1}) \mapsto c(t_1, \ldots, t_{i-1}, 1, t_i, \ldots, t_{k-1}). \end{aligned}$$

Then

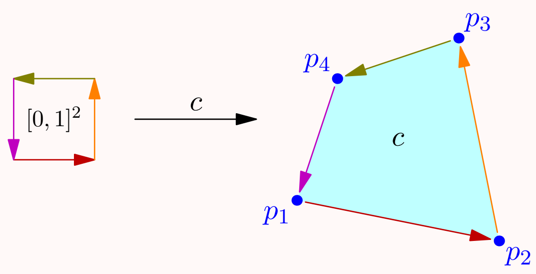
$$\partial c := \sum_{i=1}^{k} (-1)^{i+1} \left(c_i^{\text{stop}} - c_i^{\text{start}} \right).$$

Finally, the boundary of a chain is the sum of the boundaries of each cell (with the appropriate weights). That is, $\partial(\sum a_i c_i) = \sum a_i \partial c_i$.

Question 45.4.3. Satisfy yourself that one can extend this definition to a k-cell c defined on $c\colon [a_1, b_1] \times \cdots \times [a_k, b_k] \to V$ (rather than from $[0,1]^k \to V$).

Example 45.4.4 (Examples of boundaries)

Consider the 2-cell $c\colon [0,1]^2 \to \mathbb{R}^2$ shown below.



Here p_1, p_2, p_3, p_4 are the images of $(0,0)$, $(0,1)$, $(1,1)$, $(1,0)$, respectively. Formally, ∂c is given by

$$\partial c = (t \mapsto c(1,t)) - (t \mapsto c(0,t)) - (t \mapsto c(t,1)) + (t \mapsto c(t,0)).$$

I apologize for the eyesore notation caused by inline functions. Let's make amends and just write

$$\partial c = [p_2, p_3] - [p_1, p_4] - [p_4, p_3] + [p_1, p_2]$$

where each "interval" represents the 1-cell shown by the reddish arrows on the right, after accounting for the minus signs. We can take the boundary of this as well, and obtain an empty chain as

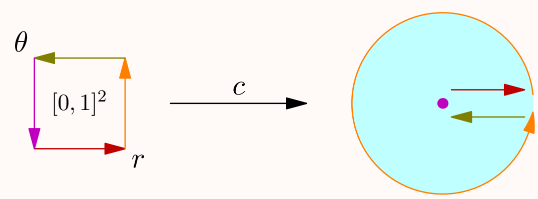
$$\partial(\partial c) = \sum_{i=1}^{4} \{p_{i+1}\} - \{p_i\} = 0.$$

Example 45.4.5 (Boundary of a unit disk)

Consider the unit disk given by

$$c\colon [0,1] \times [0,1] \to \mathbb{R}^2 \quad \text{by} \quad (r, \theta) \mapsto r\cos(2\pi\theta)\mathbf{e}_1 + r\sin(2\pi\theta)\mathbf{e}_2.$$

The four parts of the boundary are shown in the picture below:



Note that two of the arrows more or less cancel each other out when they are integrated. Moreover, we interestingly have a *degenerate* 1-cell at the center of the circle; it is a constant function $[0,1] \to \mathbb{R}^2$ which always gives the origin.

Obligatory theorem, analogous to $d^2 = 0$ and left as a problem.

Theorem 45.4.6 (The boundary of the boundary is empty)
$\partial^2 = 0$, in the sense that for any k-chain c we have $\partial^2(c) = 0$.

§45.5 Stokes' theorem

Prototypical example for this section: $\int_{[a,b]} dg = g(b) - g(a)$.

We now have all the ingredients to state Stokes' theorem for cells.

Theorem 45.5.1 (Stokes' theorem for cells)
Take $U \subseteq V$ as usual, let $c\colon [0,1]^k \to U$ be a k-cell and let $\alpha\colon U \to \bigwedge^{k-1}(V^\vee)$ be a $(k-1)$-form. Then

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

In particular, if $d\alpha = 0$ then the left-hand side vanishes.

For example, if c is the interval $[a,b]$ then $\partial c = \{b\} - \{a\}$, and thus we obtain the fundamental theorem of calculus.

§45.6 Back to Earth: A comparison to what you learned in vector calculus

Now that we've done everything abstractly, let's compare what we've learned to what you might see in $\mathbb{R}^3$ if you're doing a vector calculus course at a typical university.

In Figure 45.1 I've copied a picture I drew in fall 2024 for the 18.02 class at MIT, which is the multivariable calculus class that a lot of first-year students take. For each $0 \le d \le n \le 3$ (besides $d = n = 0$), it shows what kind of integral showed up in the class if you were doing a d-dimensional integral of a function whose domain was $\mathbb{R}^n$. Note that every integral in this picture is real-valued.

I've deliberately used the notation that was actually used at MIT, which I'll refer to as 18.02 notation, because it's similar to what you will see on Wikipedia and other places too. The goal of this section is to provide a translation system from 18.02 notation to Napkin notation. (Throughout the whole section, $\mathbb{R}^n$ is thought of as a normed vector space, so the identification $\mathbf{e}_1 \mapsto \mathbf{e}_1^\vee$ and so on is canonical.)

There is a lot going in Figure 45.1, so let's break it down piece by piece.

0-forms. A 0-form is the same as just a function, so the column of 0-D integrals should be easy to understand: it's just evaluation at point, or a sum of points.

n-forms. The case $n = d$ is also easy: we know it's possible to integrate an n-form in $\mathbb{R}^n$ and get a number. That is:

- A normal integral $\int_a^b \mathrm{d}x$ is the integral across a 1-cell $[a,b]$ across the 1-form $f \cdot \mathbf{e}_1^\vee$.
- An area integral $\int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ corresponds to integrating the 1-form $f \cdot \mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee$.
- A volume integral $\int_{a_1}^{b_1} \int_{a_2}^{b_2} \int_{a_3}^{b_3} f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ corresponds to integrating the 1-form $f \cdot \mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee \wedge \mathbf{e}_3^\vee$.

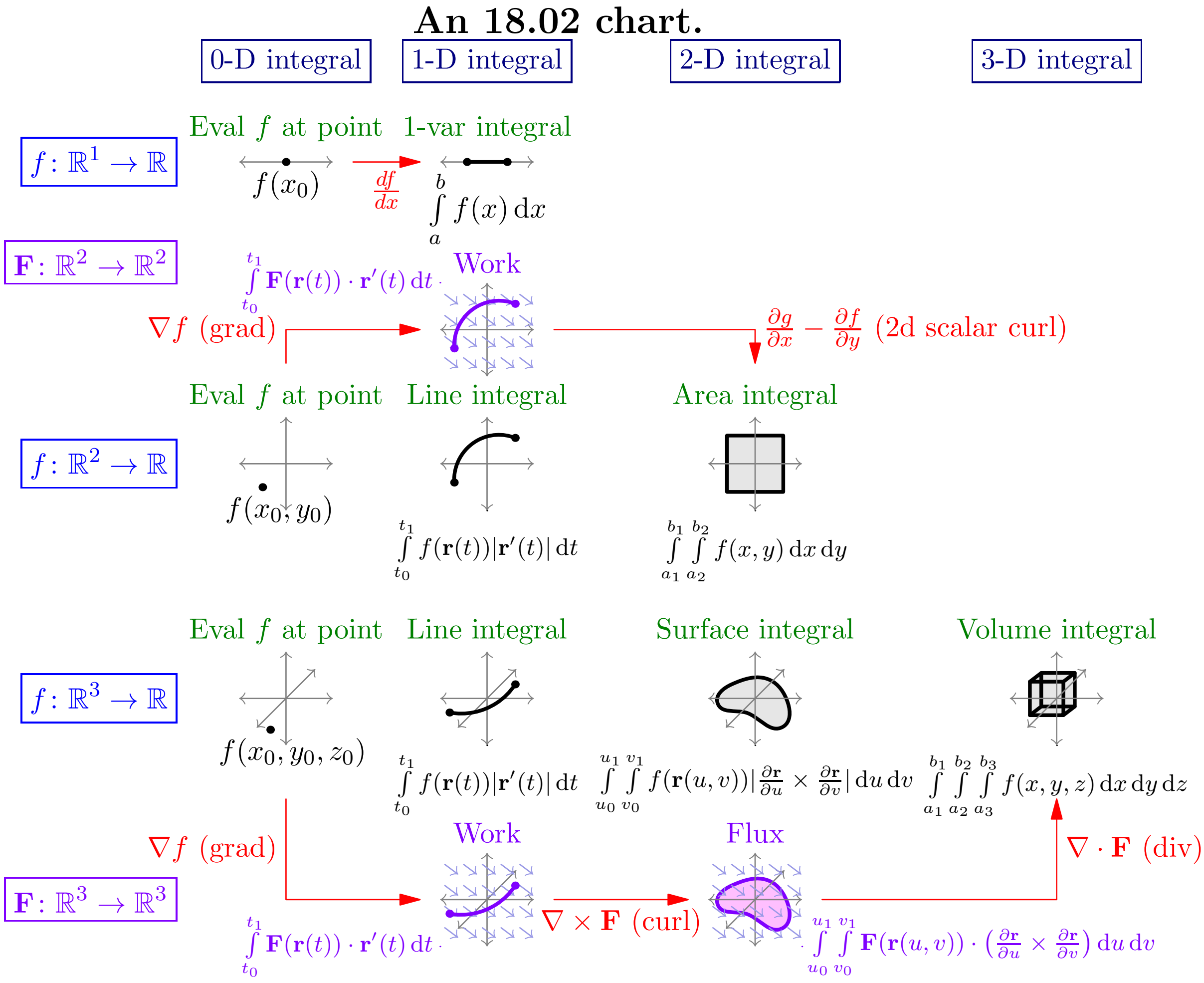


Figure 45.1: Throwback to first year of college? High-resolution version at `https://web.evanchen.cc/upload/1802/integrals-stokes.pdf`.

So this takes care of the green-labeled things on the diagonal.

We can't interpret the remaining three green pictures! The tricky part is the situations where $0 < d < n < 3$. There are three such things, the two line integrals when $d = 1$ and $n \in \{2, 3\}$ and the surface integral when $d = 2$ and $n = 3$.

In fact, these are *not* covered by our theory of differential forms! Indeed, even in the special case where $f = 1$ is a constant function, the line integrals are actually arc length, and as we mentioned in Section 44.6, that integral cannot be viewed as the integral of any differential form. Similarly, surface area isn't a differential form either.

1-**forms and** 2-**forms.** However, the three purplish integrals (over vector fields) can be viewed in our framework.

- Consider $d = 1$ and $n = 3$, i.e. the 3-D line integral. We have as input a vector-valued function $\mathbf{F}\colon \mathbb{R}^3 \to \mathbb{R}^3$. By projection principle (Theorem 43.2.1), it's the same as the data of

$$\mathbf{F}(p) = F_1(p)\mathbf{e}_1 + F_2(p)\mathbf{e}_2 + F_3(p)\mathbf{e}_3$$

for three functions $F_i\colon \mathbb{R}^3 \to \mathbb{R}$ for $i = 1, 2, 3$.

To convert the 18.02 notation $\mathbf{F}(p)$ into Napkin notation, we identify $\mathbf{F}$ with the differential form

$$\alpha_p = F_1(p)\mathbf{e}_1^\vee + F_2(p)\mathbf{e}_2^\vee + F_3(p)\mathbf{e}_3^\vee.$$

Meanwhile, the path $\mathbf{r}(t)$ parametrized by time $t \in [t_0, t_1]$ matches the concept of a 1-form $c : [t_0, t_1] \to \mathbb{R}^3$. The "work" in the integral is written as

$$\mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

but that dot product is exactly the pullback $c^*\alpha$.

- The case $d = 1$ and $n = 2$ is exactly the same, with 3 replaced by 2.
- The weirdest case is the flux integral, for $d = 2$ and $n = 3$. The parametrization $\mathbf{r}(u, v)$ is fine, and it corresponds to a 2-cell c. But $\mathbf{F}(p)$ seems to have the wrong type.

  But let's again write

  $$\mathbf{F}(p) = F_1(p)\mathbf{e}_1 + F_2(p)\mathbf{e}_2 + F_3(p)\mathbf{e}_3.$$

  There is a fairly weird hack used to convert this into Napkin notation: the form desired is

  $$\alpha_p = F_1(p)\mathbf{e}_2^\vee \wedge \mathbf{e}_3^\vee + F_2(p)\mathbf{e}_3^\vee \wedge \mathbf{e}_1^\vee + F_3(p)\mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee.$$

  Yes, that's really the identification! For this definition to be possible, we had to exploit the fact that

  $$\binom{3}{1} = \binom{3}{2}.$$

  That is the three-dimensional space $\bigwedge^2(\mathbb{R}^3)$ happens to have the same number of basis elements as $\bigwedge^1(\mathbb{R}^3) \cong \mathbb{R}^3$, so the

  $$\begin{aligned} \star \colon \bigwedge^2(\mathbb{R}^3) &\to \mathbb{R}^3 \\ \mathbf{e}_1 \wedge \mathbf{e}_2 &\mapsto \mathbf{e}_3 \\ \mathbf{e}_2 \wedge \mathbf{e}_3 &\mapsto \mathbf{e}_1 \\ \mathbf{e}_3 \wedge \mathbf{e}_1 &\mapsto \mathbf{e}_2 \end{aligned}$$

  is really an isomorphism, because it maps basis elements to basis elements. We denote this map by $\star$, because it turns out this map generalizes to the so-called Hodge star operator in higher dimensions.

  This is where I talk about cross products, which I've deliberately avoided until now. The cross product is a weird operation that takes two vectors in $\mathbb{R}^3$ and outputs a vector in $\mathbb{R}^3$. Specifically, if $\mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ and $\mathbf{w} = x'\mathbf{e}_1 + y'\mathbf{e}_2 + z'\mathbf{e}_3$, the definition of cross products taught in school is

  $$\mathbf{v} \times \mathbf{w} := (yz' - y'z)\mathbf{e}_1 + (zx' - xz')\mathbf{e}_2 + (xy' - x'y)\mathbf{e}_3.$$

  Where does this come from? The answer is that $\star(\mathbf{v} \vee \mathbf{w})$:

  $$\begin{aligned} \mathbf{v} \wedge \mathbf{w} &= (x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3) \times (x'\mathbf{e}_1 + y'\mathbf{e}_2 + z'\mathbf{e}_3) \\ &= (xy' - x'y)\mathbf{e}_1 \wedge \mathbf{e}_2 + (yz' - y'z)\mathbf{e}_2 \wedge \mathbf{e}_3 + (zx' - xz')\mathbf{e}_3 \wedge \mathbf{e}_1 \\ \star(\mathbf{v} \wedge \mathbf{w}) &\mapsto (xy' - x'y)\mathbf{e}_3 + (yz' - y'z)\mathbf{e}_1 + (zx' - xz')\mathbf{e}_2. \end{aligned}$$

With that out of the way, the weird dot-cross product

$$\mathbf{F}(\mathbf{r}(u,v)) \cdot (\mathbf{r}_u \times \mathbf{r}_v)\, du\, dv$$

is now rigged to correspond to the pullback $c^*\alpha$. So using this Hodge star, we find that flux is actually the integration of a 2-form.

Exterior derivatives Every red arrow in Figure 45.1 corresponds to the exterior derivative of the corresponding form. That is:

- The "grad" operation takes a 0-form f and outputs a vector field corresponding to the 1-form df.
- The "curl" operation takes a 1-form α and outputs a vector field corresponding to the 2-form $d\alpha$. When $n = 3$, this checks out because the space of 1-forms is $\binom{3}{1}$ dimensional, and the space of 2-forms is $\binom{3}{2}$, and thankfully $\binom{3}{1} = 3 = \binom{3}{2}$.

  The weird notation $\nabla \times \mathbf{F}$ can be checked to correspond to the exterior derivative. On the 18.02 side, if we have

  $$\mathbf{F} = F_1\mathbf{e}_1 + F_2\mathbf{e}_2 + F_3\mathbf{e}_3$$

  then the 18.02 definition of curl is that

  $$\operatorname{curl}(\mathbf{F}) \coloneqq \nabla \times \mathbf{F} \coloneqq \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\mathbf{e}_1 + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}\right)\mathbf{e}_2 + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\mathbf{e}_3$$

  The reason for the nonsensical $\nabla\times$ notation is that if you *really* abuse notation you can almost think of this as the cross product of a vector $\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle$ and the vector $\mathbf{F} = \langle F_1, F_2, F_3 \rangle$.

  Now to convert $\mathbf{F}$ into Napkin notation, remember we identified $\mathbf{F}$ with the differential form

  $$\alpha = F_1\mathbf{e}_1^\vee + F_2\mathbf{e}_2^\vee + F_3\mathbf{e}_3^\vee.$$

  If we follow our formula for exterior derivative in Definition 44.4.4, we get

  $$\begin{aligned} d\alpha &= dF_1 \wedge \mathbf{e}_1^\vee + dF_2 \wedge \mathbf{e}_2^\vee + dF_3 \wedge \mathbf{e}_3^\vee \\ &= \left(\frac{\partial F_1}{\partial x}\mathbf{e}_1^\vee + \frac{\partial F_1}{\partial y}\mathbf{e}_2^\vee + \frac{\partial F_1}{\partial z}\mathbf{e}_3^\vee\right) \wedge \mathbf{e}_1^\vee \\ &+ \left(\frac{\partial F_2}{\partial x}\mathbf{e}_1^\vee + \frac{\partial F_2}{\partial y}\mathbf{e}_2^\vee + \frac{\partial F_2}{\partial z}\mathbf{e}_3^\vee\right) \wedge \mathbf{e}_2^\vee \\ &+ \left(\frac{\partial F_3}{\partial x}\mathbf{e}_1^\vee + \frac{\partial F_3}{\partial y}\mathbf{e}_2^\vee + \frac{\partial F_3}{\partial z}\mathbf{e}_3^\vee\right) \wedge \mathbf{e}_3^\vee \\ &= \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\mathbf{e}_2^\vee \wedge \mathbf{e}_3^\vee + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}\right)\mathbf{e}_3^\vee \wedge \mathbf{e}_1^\vee + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\mathbf{e}_1^\vee \wedge \mathbf{e}_2^\vee. \end{aligned}$$

  Taking the Hodge star and then dropping all the $\vee$'s gives the same thing as $\nabla \times \mathbf{F}$, so this completes the correspondence between the 18.02 notation and the Napkin notation.
- In 18.02 terminology, the divergence div is defined by

  $$\operatorname{div}(\mathbf{F}) \coloneqq \nabla \cdot \mathbf{F} \coloneqq \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}$$

which is a scalar-valued function for input points $p \in \mathbb{R}^3$. We let you do this one in Problem 45C.

The reason for the nonsensical $\nabla \cdot$ notation is that if you *really* abuse notation you can almost think of this as the dot product of a vector $\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle$ and the vector $\mathbf{F} = \langle F_1, F_2, F_3 \rangle$.

Double derivative We know that $d^2 = 0$, which in Figure 45.1 means composing two arrows gives zero. You'll see this in 18.02 as

- The curl of a gradient is zero.
- The flux of a curl is zero.

but really they're the same theorem.

Stokes' theorem Each red arrow also gives an instance of Stokes' theorem for cells. So Stokes' theorem even for cells is really great, because we get six 18.02 theorems as special cases!

- The three arrows from 0-D integrals to 1-D integrals are all called "Fundamental Theorem of Calculus". Some authors will say "Fundamental Theorem of Calculus for line integrals" instead for $n > 1$.
- For $n = 2$, the other red arrow is called "Green's theorem"; we let you work it out as Problem 45A†.
- For $n = 3$, the arrow from work to flux is confusingly also called "Stokes' theorem"; it says the flux of a 2-D surface equals the work on the 1-D boundary.
- The rightmost red arrow for $n = 3$ is called the "divergence theorem"; it says the divergence of a 3-D volume equals the flux of the 2-D boundary surface.

§45.7 A few harder problems to think about

Problem 45A† (Green's theorem)**.** Let $f, g \colon \mathbb{R}^2 \to \mathbb{R}$ be smooth functions and c a 2-cell. Prove that

$$\int_c \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) \, dx \wedge dy = \int_{\partial c} (f \, dx + g \, dy).$$

Problem 45B. Show that $\partial^2 = 0$.

Problem 45C. Finish the correspondence of the 18.02 notation with Napkin notation. That is, let $\mathbf{F} \colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field, and let α be the 2-form corresponding to it in Napkin version. Show that the scalar-valued function defined by

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

coincides with evaluation at the 3-form $d\alpha$.

Problem 45D (Pullback and d commute)**.** Let U and U' be open sets of vector spaces V and V' and let $\phi \colon U \to U'$ be a smooth map between them. Prove that for any differential form α on U' we have

$$\phi^*(d\alpha) = d(\phi^* \alpha).$$

Problem 45E (Arc length isn't a form)**.** Show that there does *not* exist a 1-form α on $\mathbb{R}^2$ such that for a curve $c\colon [0,1] \to \mathbb{R}^2$, the integral $\int_c \alpha$ gives the arc length of c.

Problem 45F. An **exact** k-form α is one satisfying $\alpha = d\beta$ for some β. Prove that

$$\int_{C_1} \alpha = \int_{C_2} \alpha$$

where C_1 and C_2 are any concentric circles in the plane and α is some exact 1-form.

46 A bit of manifolds

Last chapter, we stated Stokes' theorem for cells. It turns out there is a much larger class of spaces, the so-called *smooth manifolds*, for which this makes sense.

Unfortunately, the definition of a smooth manifold is *complete garbage*, and so by the time I am done defining differential forms and orientations, I will be too lazy to actually define what the integral on it is, and just wave my hands and state Stokes' theorem.

§46.1 Topological manifolds

Prototypical example for this section: S^2*: "the Earth looks flat".*

Long ago, people thought the Earth was flat, i.e. homeomorphic to a plane, and in particular they thought that $\pi_2(\text{Earth}) = 0$. But in fact, as most of us know, the Earth is actually a sphere, which is not contractible and in particular $\pi_2(\text{Earth}) \cong \mathbb{Z}$. This observation underlies the definition of a manifold:

An n-manifold is a space which locally looks like $\mathbb{R}^n$.

Actually there are two ways to think about a topological manifold M:

- "Locally": at every point $p \in M$, some open neighborhood of p looks like an open set of $\mathbb{R}^n$. For example, to someone standing on the surface of the Earth, the Earth looks much like $\mathbb{R}^2$.
- "Globally": there exists an open cover of M by open sets $\{U_i\}_i$ (possibly infinite) such that each U_i is homeomorphic to some open subset of $\mathbb{R}^n$. For example, from outer space, the Earth can be covered by two hemispherical pancakes.

Question 46.1.1. Check that these are equivalent.

While the first one is the best motivation for examples, the second one is easier to use formally.

Definition 46.1.2. A **topological n-manifold** M is a Hausdorff space with an open cover $\{U_i\}$ of sets homeomorphic to subsets of $\mathbb{R}^n$, say by homeomorphisms

$$\phi_i : U_i \xrightarrow{\cong} E_i \subseteq \mathbb{R}^n$$

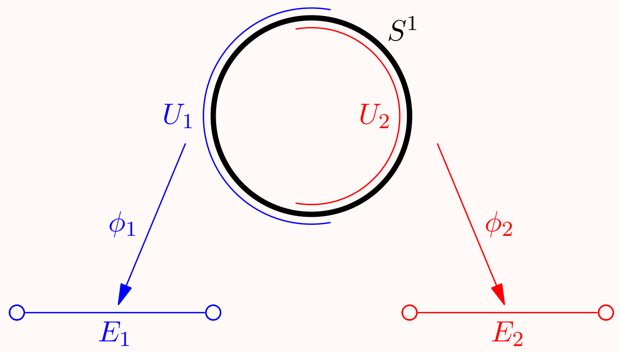
where each E_i is an open subset of $\mathbb{R}^n$. Each $\phi_i : U_i \to E_i$ is called a **chart**, and together they form a so-called **atlas**.

Remark 46.1.3 — Here "E" stands for "Euclidean". I think this notation is not standard; usually people just write $\phi_i(U_i)$ instead.

Remark 46.1.4 — This definition is nice because it doesn't depend on embeddings: a manifold is an *intrinsic* space M, rather than a subset of $\mathbb{R}^N$ for some N. Analogy: an abstract group G is an intrinsic object rather than a subgroup of S_n.

Example 46.1.5 (An atlas on S^1)

Here is a picture of an atlas for S^1, with two open sets.



Question 46.1.6. Where do you think the words "chart" and "atlas" come from?

Example 46.1.7 (Some examples of topological manifolds)

(a) As discussed at length, the sphere S^2 is a 2-manifold: every point in the sphere has a small open neighborhood that looks like D^2. One can cover the Earth with just two hemispheres, and each hemisphere is homeomorphic to a disk.

(b) The circle S^1 is a 1-manifold; every point has an open neighborhood that looks like an open interval.

(c) The torus, Klein bottle, $\mathbb{RP}^2$ are all 2-manifolds.

(d) $\mathbb{R}^n$ is trivially a manifold, as are its open sets.

All these spaces are compact except $\mathbb{R}^n$.

A non-example of a manifold is D^n, because it has a *boundary*; points on the boundary do not have open neighborhoods that look Euclidean.

§46.2 Smooth manifolds

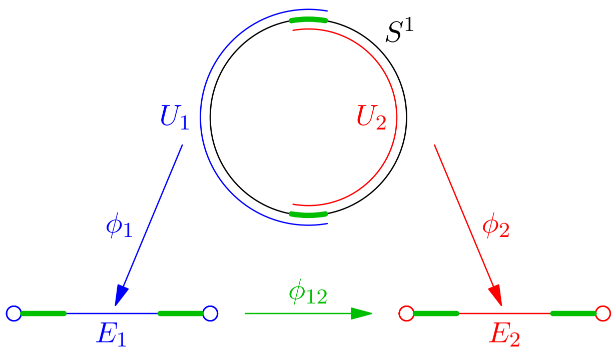
Prototypical example for this section: All the topological manifolds.

Let M be a topological n-manifold with atlas $\{U_i \xrightarrow{\phi_i} E_i\}$.

Definition 46.2.1. For any i, j such that $U_i \cap U_j \neq \varnothing$, the **transition map** ϕ_{ij} is the composed map

$$\phi_{ij} \colon E_i \cap \phi_i^{\text{img}}(U_i \cap U_j) \xrightarrow{\phi_i^{-1}} U_i \cap U_j \xrightarrow{\phi_j} E_j \cap \phi_j^{\text{img}}(U_i \cap U_j).$$

Sorry for the dense notation, let me explain. The intersection with the image $\phi_i^{\text{img}}(U_i \cap U_j)$ and the image $\phi_j^{\text{img}}(U_i \cap U_j)$ is a notational annoyance to make the map well-defined and a homeomorphism. The transition map is just the natural way to go from $E_i \to E_j$, restricted to overlaps. Picture below, where the intersections are just the green portions of each E_1 and E_2:



We want to add enough structure so that we can use differential forms.

Definition 46.2.2. We say M is a **smooth manifold** if all its transition maps are smooth.

This definition makes sense, because we know what it means for a map between two open sets of $\mathbb{R}^n$ to be differentiable.

With smooth manifolds we can try to port over definitions that we built for $\mathbb{R}^n$ onto our manifolds. So in general, all definitions involving smooth manifolds will reduce to something on each of the coordinate charts, with a compatibility condition.

As an example, here is the definition of a "smooth map":

Definition 46.2.3. (a) Let M be a smooth manifold. A continuous function $f\colon M \to \mathbb{R}$ is called **smooth** if the composition

$$E_i \xrightarrow{\phi_i^{-1}} U_i \hookrightarrow M \xrightarrow{f} \mathbb{R}$$

is smooth as a function $E_i \to \mathbb{R}$.

(b) Let M and N be smooth with atlases $\{U_i^M \xrightarrow{\phi_i} E_i^M\}_i$ and $\{U_j^N \xrightarrow{\phi_j} E_j^N\}_j$. A map $f\colon M \to N$ is **smooth** if for every i and j, the composed map

$$E_i \xrightarrow{\phi_i^{-1}} U_i \hookrightarrow M \xrightarrow{f} N \twoheadrightarrow U_j \xrightarrow{\phi_j} E_j$$

is smooth, as a function $E_i \to E_j$.

§46.3 Regular value theorem

Prototypical example for this section: $x^2 + y^2 = 1$ is a circle!

Despite all that I've written about general manifolds, it would be sort of mean if I left you here because I have not really told you how to actually construct manifolds in practice, even though we know the circle $x^2 + y^2 = 1$ is a great example of a one-dimensional manifold embedded in $\mathbb{R}^2$.

Theorem 46.3.1 (Regular value theorem)
Let V be an n-dimensional real normed vector space, let $U \subseteq V$ be open and let $f_1, \dots, f_m : U \to \mathbb{R}$ be smooth functions. Let M be the set of points $p \in U$ such that $f_1(p) = \dots = f_m(p) = 0$.

Assume M is nonempty and that the map

$$V \to \mathbb{R}^m \quad \text{by} \quad v \mapsto ((Df_1)_p(v), \dots, (Df_m)_p(v))$$

has rank m, for every point $p \in M$. Then M is a manifold of dimension $n - m$.

For a proof, see [Sj05, Theorem 6.3].

One very common special case is to take $m = 1$ above.

Corollary 46.3.2 (Level hypersurfaces)
Let V be a finite-dimensional real normed vector space, let $U \subseteq V$ be open and let $f : U \to \mathbb{R}$ be smooth. Let M be the set of points $p \in U$ such that $f(p) = 0$. If $M \neq \varnothing$ and $(Df)_p$ is not the zero map for any $p \in M$, then M is a manifold of dimension $\dim V - 1$.

Example 46.3.3 (The circle $x^2 + y^2 - c = 0$)
Let $f(x, y) = x^2 + y^2 - c$, $f : \mathbb{R}^2 \to \mathbb{R}$, where c is a positive real number. Note that

$$Df = 2x \cdot dx + 2y \cdot dy$$

which in particular is nonzero as long as $(x, y) \neq (0, 0)$, i.e. as long as $c \neq 0$. Thus:

- When $c > 0$, the resulting curve — a circle with radius $\sqrt{c}$ — is a one-dimensional manifold, as we knew.
- When $c = 0$, the result fails. Indeed, M is a single point, which is actually a zero-dimensional manifold!

We won't give further examples since I'm only mentioning this in passing in order to increase your capacity to write real concrete examples. (But [Sj05, Chapter 6.2] has some more examples, beautifully illustrated.)

§46.4 Differential forms on manifolds

We already know what a differential form is on an open set $U \subseteq \mathbb{R}^n$. So, we naturally try to port over the definition of differentiable form on each subset, plus a compatibility condition.

Let M be a smooth manifold with atlas $\{U_i \xrightarrow{\phi_i} E_i\}_i$.

Definition 46.4.1. A **differential k-form** α on a smooth manifold M is a collection $\{\alpha_i\}_i$ of differential k-forms on each E_i, such that for any j and i we have that

$$\alpha_j = \phi_{ij}^*(\alpha_i).$$

In English: we specify a differential form on each chart, which is compatible under pullbacks of the transition maps.

§46.5 Orientations

Prototypical example for this section: Left versus right, clockwise vs. counterclockwise.

This still isn't enough to integrate on manifolds. We need one more definition: that of an orientation.

The main issue is the observation from standard calculus that

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

Consider then a space M which is homeomorphic to an interval. If we have a 1-form α, how do we integrate it over M? Since M is just a topological space (rather than a subset of $\mathbb{R}$), there is no default "left" or "right" that we can pick. As another example, if $M = S^1$ is a circle, there is no default "clockwise" or "counterclockwise" unless we decide to embed M into $\mathbb{R}^2$.

To work around this we have to actually have to make additional assumptions about our manifold.

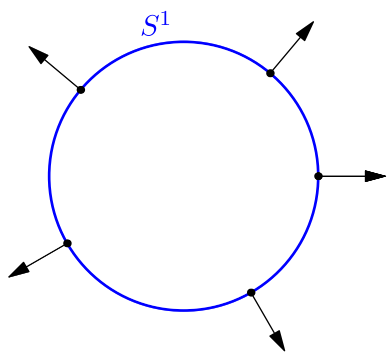
Definition 46.5.1. A smooth n-manifold is **orientable** if there exists a differential n-form ω on M such that for every $p \in M$,

$$\omega_p \neq 0.$$

Recall here that ω_p is an element of $\bigwedge^n(V^\vee)$. In that case we say ω is a **volume form** of M.

How do we picture this definition? If we recall that an differential form is supposed to take tangent vectors of M and return real numbers. To this end, we can think of each point $p \in M$ as having a **tangent plane** $T_p(M)$ which is n-dimensional. Now since the volume form ω is n-dimensional, it takes an entire basis of the $T_p(M)$ and gives a real number. So a manifold is orientable if there exists a consistent choice of sign for the basis of tangent vectors at every point of the manifold.

For "embedded manifolds", this just amounts to being able to pick a nonzero field of normal vectors to each point $p \in M$. For example, S^1 is orientable in this way.



Similarly, one can orient a sphere S^2 by having a field of vectors pointing away (or towards) the center. This is all non-rigorous, because I haven't defined the tangent plane $T_p(M)$; since M is in general an intrinsic object one has to be quite roundabout to define $T_p(M)$ (although I do so in an optional section later). In any event, the point is that guesses about the orientability of spaces are likely to be correct.

Example 46.5.2 (Orientable surfaces)

(a) Spheres S^n, planes, and the torus $S^1 \times S^1$ are orientable.

(b) The Möbius strip and Klein bottle are *not* orientable: they are "one-sided".

(c) $\mathbb{CP}^n$ is orientable for any n.

(d) $\mathbb{RP}^n$ is orientable only for odd n.

§46.6 Stokes' theorem for manifolds

Stokes' theorem in the general case is based on the idea of a **manifold with boundary** M, which I won't define, other than to say its boundary ∂M is an $n-1$ dimensional manifold, and that it is oriented if M is oriented. An example is $M = D^2$, which has boundary $\partial M = S^1$.

Next,

Definition 46.6.1. The **support** of a differential form α on M is the closure of the set

$$\{p \in M \mid \alpha_p \neq 0\}.$$

If this support is compact as a topological space, we say α is **compactly supported**.

Remark 46.6.2 — For example, volume forms are supported on all of M.

Now, one can define integration on oriented manifolds, but I won't define this because the definition is truly awful. Then Stokes' theorem says

Theorem 46.6.3 (Stokes' theorem for manifolds)

Let M be a smooth oriented n-manifold with boundary and let α be a compactly supported $(n-1)$-form. Then

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

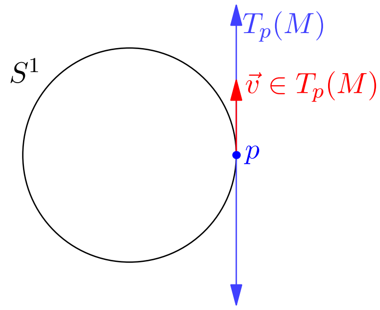
All the omitted details are developed in full in [Sj05].

§46.7 (Optional) The tangent and cotangent space

Prototypical example for this section: Draw a line tangent to a circle, or a plane tangent to a sphere.

Let M be a smooth manifold and $p \in M$ a point. I omitted the definition of $T_p(M)$ earlier, but want to actually define it now.

As I said, geometrically we know what this *should* look like for our usual examples. For example, if $M = S^1$ is a circle embedded in $\mathbb{R}^2$, then the tangent vector at a point p should just look like a vector running off tangent to the circle. Similarly, given a sphere $M = S^2$, the tangent space at a point p along the sphere would look like plane tangent to M at p.



However, one of the points of all this manifold stuff is that we really want to see the manifold as an *intrinsic object*, in its own right, rather than as embedded in $\mathbb{R}^n$.[1] So, we would like our notion of a tangent vector to not refer to an ambient space, but only to intrinsic properties of the manifold M in question.

§46.7.i Tangent space

To motivate this construction, let us start with an embedded case for which we know the answer already: a sphere.

Suppose $f : S^2 \to \mathbb{R}$ is a function on a sphere, and take a point p. Near the point p, f looks like a function on some open neighborhood of the origin. Thus we can think of taking a *directional derivative* along a vector $\vec{v}$ in the imagined tangent plane (i.e. some partial derivative). For a fixed $\vec{v}$ this partial derivative is a linear map

$$D_{\vec{v}} : C^\infty(M) \to \mathbb{R}.$$

It turns out this goes the other way: if you know what $D_{\vec{v}}$ does to every smooth function, then you can recover v. This is the trick we use in order to create the tangent space. Rather than trying to specify a vector $\vec{v}$ directly (which we can't do because we don't have an ambient space),

The vectors *are* partial-derivative-like maps.

More formally, we have the following.

Definition 46.7.1. A **derivation** D at p is a linear map $D : C^\infty(M) \to \mathbb{R}$ (i.e. assigning a real number to every smooth f) satisfying the following Leibniz rule: for any f, g we have the equality

$$D(fg) = f(p) \cdot D(g) + g(p) \cdot D(f) \in \mathbb{R}.$$

This is just a "product rule". Then the tangent space is easy to define:

Definition 46.7.2. A **tangent vector** is just a derivation at p, and the **tangent space** $T_p(M)$ is simply the set of all these tangent vectors.

In this way we have constructed the tangent space.

[1] This can be thought of as analogous to the way that we think of a group as an abstract object in its own right, even though Cayley's Theorem tells us that any group is a subgroup of the permutation group.

Note this wasn't always the case! During the 19th century, a group was literally defined as a subset of $\mathrm{GL}(n)$ or of S_n. In fact Sylow developed his theorems without the word "group". Only much later did the abstract definition of a group was given, an abstract set G which was independent of any *embedding* into S_n, and an object in its own right.

§46.7.ii The cotangent space

In fact, one can show that the product rule for D is equivalent to the following three conditions:

1. D is linear, meaning $D(af + bg) = aD(f) + bD(g)$.

2. $D(1_M) = 0$, where 1_M is the constant function on M.

3. $D(fg) = 0$ whenever $f(p) = g(p) = 0$. Intuitively, this means that if a function $h = fg$ vanishes to second order at p, then its derivative along D should be zero.

This suggests a third equivalent definition: suppose we define

$$\mathfrak{m}_p := \{f \in C^\infty M \mid f(p) = 0\}$$

to be the set of functions which vanish at p (this is called the *maximal ideal* at p). In that case,

$$\mathfrak{m}_p^2 = \left\{\sum_i f_i \cdot g_i \mid f_i(p) = g_i(p) = 0\right\}$$

is the set of functions vanishing to second order at p. Thus, a tangent vector is really just a linear map

$$\mathfrak{m}_p/\mathfrak{m}_p^2 \to \mathbb{R}.$$

In other words, the tangent space is actually the dual space of $\mathfrak{m}_p/\mathfrak{m}_p^2$; for this reason, the space $\mathfrak{m}_p/\mathfrak{m}_p^2$ is defined as the **cotangent space** (the dual of the tangent space). This definition is even more abstract than the one with derivations above, but has some nice properties:

- it is coordinate-free, and

- it's defined only in terms of the smooth functions $M \to \mathbb{R}$, which will be really helpful later on in algebraic geometry when we have varieties or schemes and can repeat this definition.

§46.7.iii Sanity check

With all these equivalent definitions, the last thing I should do is check that this definition of tangent space actually gives a vector space of dimension n. To do this it suffices to show verify this for open subsets of $\mathbb{R}^n$, which will imply the result for general manifolds M (which are locally open subsets of $\mathbb{R}^n$). Using some real analysis, one can prove the following result:

Theorem 46.7.3

Suppose $M \subset \mathbb{R}^n$ is open and $0 \in M$. Then

$$\begin{aligned} \mathfrak{m}_0 &= \{\text{smooth functions } f \mid f(0) = 0\} \\ \mathfrak{m}_0^2 &= \{\text{smooth functions } f \mid f(0) = 0, (\nabla f)_0 = 0\}. \end{aligned}$$

In other words $\mathfrak{m}_0^2$ is the set of functions which vanish at 0 and such that all first derivatives of f vanish at zero.

Thus, it follows that there is an isomorphism

$$\mathfrak{m}_0/\mathfrak{m}_0^2 \cong \mathbb{R}^n \quad \text{by} \quad f \mapsto \left[\frac{\partial f}{\partial x_1}(0), \dots, \frac{\partial f}{\partial x_n}(0)\right]$$

and so the cotangent space, hence tangent space, indeed has dimension n.

§46.8 A few harder problems to think about

Problem 46A. Show that a differential 0-form on a smooth manifold M is the same thing as a smooth function $M \to \mathbb{R}$.

some applications of regular value theorem here