On the Visualization of Differential Forms

Duarte Maia

September 20, 2025

1 Introduction

In 2021, I set up a Wordpress blog on my undergraduate webpage with the intent to post math content there. I didn't write more than three or four posts. One of these posts, however, I considered to be quite good content, even if it was blatantly unfinished. Meanwhile, some software update or incompatibility has taken the original blog down. I've been able to obtain a copy of my post from the Internet Archive. For safekeeping, this document is a direct transcription of this post (and a prerequisite post) to LaTeX form, with all the blemishes and issues of the original format. It is my hope to one day make a more refined document, but odds are I'll put it off for a few more years yet.

2 On the Visualization of Dual Space

August 8, 2021

This post talks about visualization of the dual space V^*. To read it, you need to know what a vector space is. If you don't, this post is not for you.

This post is loosely based on the excellent book Geometrical Vectors by Gabriel Weinreich, which is about visualization of vectors and covectors, as well as On the Visualisation of Differential Forms by Dan Piponi, which is a random PDF I found online but thought was great.

Without further ado,

The Dual Space and Weird Vectors

In all that follows, V refers to a finite-dimensional real vector space.

2.1 Motivation

In our experience, a vector is an arrow, pointing from the origin to some point. Of course, the fact that it starts at the origin is just convention of notation on our part: we could "place the vector anywhere", as long as it points in the same direction and has the same length. If we transform our space in some way, the vector is transformed in the same manner.

Let me elaborate on that last sentence for a moment. Let us imagine that I have a map of my bedroom, in which I have outlined the position of the chair I am writing this on, as well as my bed pillow, represented below as A and B. Let us suppose that the scale of the map is in meters (this is important!) Represented below is also the vector going from A to B, which we will henceforth call $\vec{v}$.

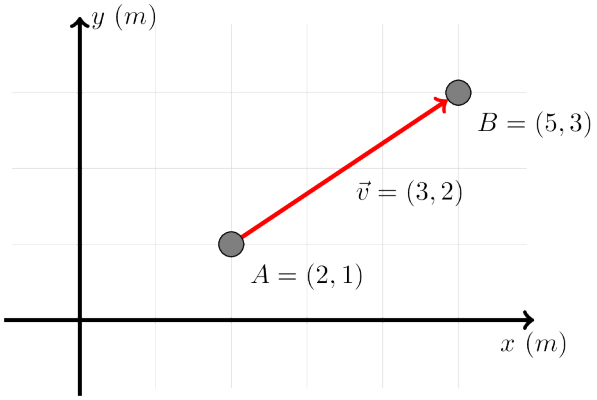


Figure 1: Schematic representation of my bedroom. Each grid square is 1m on the side.

Now, let us suppose that I decide to draw this map, but change the scale of the map. For example, suppose I scale the x axis by a half, so that x is now measured in $2 \times$ m. Then, everything in the map gets scaled, including the vector:

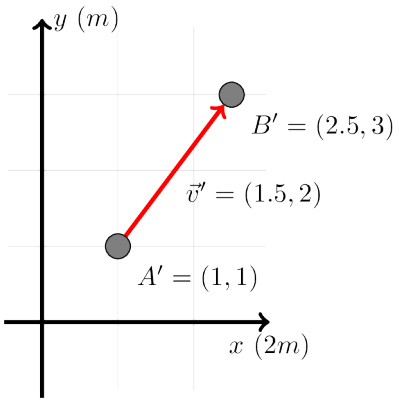


Figure 2: Still my bedroom. Now, each tick corresponds to "1" on the scale. A y tick is 1m, and an x tick is 2m.

Note: Pay attention to the difference between the map and the territory. I only rescaled the map: my actual bedroom has not been compressed. The coordinates of the points A and B, or A' and B' correspond to coordinates on the page, not in real life.

Now, to be perfectly rigorous, every time I make a rescaled map I should be giving a new name to the points. However, I would end up with something like A'''' in no time, so I'll just stop worrying about it. Just keep in mind that in each map I'm considering different points.

Anyhow, I could keep doing transformations on this map, but nothing would be surprising. If I rotate the axes, or pick a different basis, the vector $\vec{v}$ will always point from A to B, as it should. The trickiness starts when I consider some kind of vector that does not represent a displacement between two points.

One of the simplest possible examples is a slope vector. Suppose I set up a straight ramp, so that I can roll marbles from my head down to my pillow. I can represent the slope of this ramp by a vector $\vec{s}$, which points in the "downhill" direction, and whose magnitude is the slope ($\Delta z/\Delta \ell$, where z is the height and ℓ is how far along the ramp I am). For example, if my ramp decreased a total of $3^2 + 2^2 = 13$ meters (I have a very tall room), then $\vec{s}$ would be pointing in the AB direction and have magnitude $\sqrt{3^2 + 2^2}$. Schematically, we get a familiar looking picture.

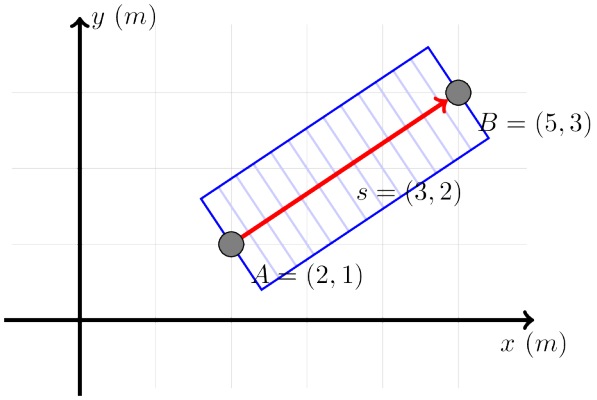


Figure 3: In blue: My ramp, as seen from above. The light blue lines are level sets: one for every meter descended.

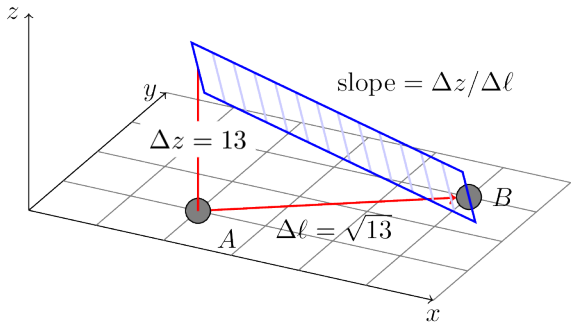


Figure 4: A 3D render of the previous figure.

[If you're thinking about s as minus the gradient of $h(x, y)$, you're not too far off, but hold your horses for the moment.]

What can I do, and what do I want to do, with this s vector? I can use it to calculate the force a ball feels when placed on the ramp, or the vector (in $\mathbb{R}^3$) orthogonal to the ramp, but both of these are specific cases of something more general: we want to recover the shape of the ramp from the s vector.

To be more specific, we want to solve the following problem: Given the s vector, which encodes the direction and slope of the ramp, we want to recover the function $h(x, y)$, which given a point (x, y) returns the height of the ramp at that point.

Note: Since the s vector only gives us information about the slope, we cannot expect to fully determine h. Therefore, we will establish that at A the ramp is at a known height h_0. Furthermore, in order not to worry about the domain of h, let us assume that the ramp has been extended to cover the whole space.

The solution to our problem is not particularly difficult. The graph of h is a plane, so h must be of the form

$$h(x, y) = ax + by + c.$$

Since $h(A)$ is known to be h_0, we can write this in a more convenient manner

$$h(p) = (a, b) \cdot (p - A) + h_0.$$

To find (a, b), recall first that the vector (a, b) is orthogonal to level sets. Looking at the figure above, we find that s is also orthogonal to level sets (this is in fact what we meant when we said "s points in the downhill direction"), so that $(a, b) = \alpha s$ for some real number α.

To discover α, we can use the fact that $h(B)$ is known, plugging $h(B) = h_0 - 13$ to get

$$\alpha s \cdot (B - A) = -13.$$

Since $s = B - A$ in our coordinate system and $\|B - A\| = \sqrt{13}$, we conclude $\alpha = -1$, hence

$$h(p) = h_0 - s \cdot (p - A).$$

Slightly more generally, it is a trivial exercise to show that $h(p) - h(q) = -s \cdot (p - q)$, so in some sense, the s vector serves as a ruler which which one can measure height differences. This is, in fact, an apt way to think about s.

Our definition of h, and the way we are thinking about s, is very closely related to the euclidean inner product. The inner product has an important property: while it behaves well under isometries, it gets distorted under other changes of coordinates. For example, if we rescale the whole map by a factor of $1/2$, and s along with it, the above formula for $h(p)$ ceases to work. For example, $h(B)$ (recall that in rigour I should be writing B') is meant to be equal to $h_0 - 13$, but if s' is the rescaled vector we get

$$h_0 - (B - A) \cdot s' = h_0 - (1.5, 1) \cdot (1.5, 1) = h_0 - \frac{13}{4} \neq h_0 - 13.$$

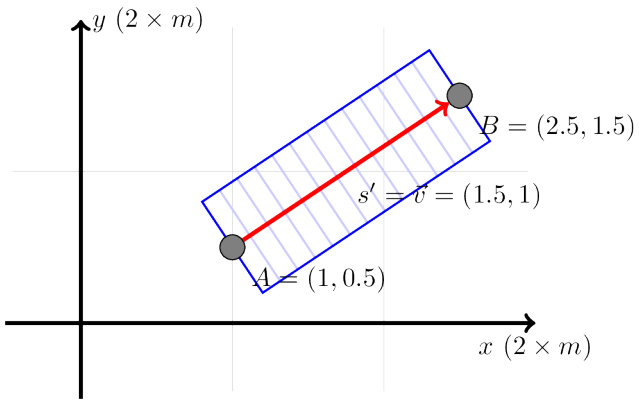


Figure 5: Rescaled map of my ramp.

As we can see, this yields the wrong result. As it should, of course: if we scaled both $(B - A)$ and s by a factor of $\frac{1}{2}$, the term $(B - A) \cdot s$ will have gotten smaller by $\frac{1}{4}$. The obvious way to fix the problem would, in fact, be to scale s up by a factor of 2, while everything else got scaled down by $\frac{1}{2}$. We can see this same phenomenon in other changes of variable. For example, suppose we scaled only the x axis in the original map, so the x axis is measured in $2 \times m$ and the y axis in m. We represent below the "naïvely scaled" vector s', as well as a vector $\hat{s}$ which has been scaled the opposite way: that is, its x coordinate was doubled instead of halved.

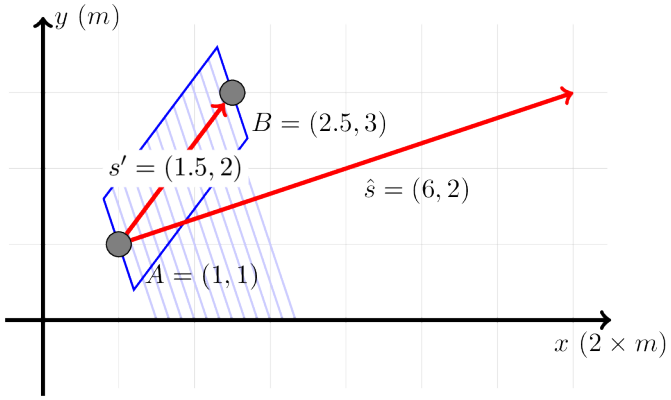


Figure 6: My ramp. s' is the vector s scaled as the referential, $\hat{s}$ is the vector scaled inversely. Level sets of the ramp extended for clarity.

Looking at the image, it is not immediately obvious why $\hat{s}$ is "the way we want s to scale". After all, it's pointing in a completely different direction. However, it is "the right vector" to ensure that the definition of h doesn't change: otherwise, we'd need to change the inner product.

This notion of "scaling differently" is at the heart of the idea of a covector. It is kind of like a vector, but it behaves very differently under changes of coordinates. And this change is made in order to retain a very specific aspect: if the vectors of the vector space V are scaled in some manner, the covectors are scaled "inversely" in order to preserve inner products between vectors and covectors!

In this manner, the role of covectors is to "be inner product'd with vectors". In some sense, the "most natural setting" for the scalar product of two vectors is actually between a vector and a covector, because it becomes independent of our coordinate system. The way we scale covectors relative to vectors will ensure that the scalar product is independent of the chosen basis.

But then, the notion of scalar product becomes somewhat superfluous. If all we're going to do with covectors is to use them to obtain numbers from vectors (and add covectors to each other and stuff), we might as well define a covector as a thing that turns vectors into numbers. But this is precisely the notion of dual vector: that is, an element of the dual space.

Recall the definition of dual space: V^* is the set of linear functions from V to $\mathbb{R}$ (or whatever field the vector space is over).

In conclusion: if V is a vector space, we use the name vector for elements of V, and covectors for elements of V^*.

2.2 How it fits together

Let us now recall elementary linear algebra to figure out how to make sense of the relation between this notion of covector and the scaling properties mentioned before.

First of all, how to represent a covector λ in coordinates? Since a covector is a different kind of object from a vector, fixing a basis of vectors $e_1, \ldots, e_n$ in V does not allow us to represent λ in as a linear combination of the e_i.

There are two distinct but equivalent ways to handle the subject matter.

2.2.1 The Dual Basis

This is the simplest way. Given a basis $e_1, \ldots, e_n$, we define the so-called dual basis $\omega^1, \ldots, \omega^n$ as:

$$\omega^i(e_j) = \delta_{ij}, \text{ or equivalently } \omega^i(\sum_j x_j e_j) = x_i.$$

Then, any covector λ can be written as a linear combination of the ω^i, as

$$\lambda(x) = \lambda\left(\sum x_i e_i\right) = \lambda\left(\sum \omega^i(x) e_i\right) = \sum \lambda(e_i)\omega^i(x),$$

whence $\lambda = \sum \lambda(e_i)\omega^i$. Therefore, we can consider the i-th coordinate of λ in some basis as λ applied to the i-th vector of this basis.

Bonus: Compare the following three formulas to write a vector or a covector in coordinates. The first is the usual formula for the coordinates of a vector in orthonormal coordinates (assuming V has an inner product), while the other two write a vector in the basis e_i or a covector in the dual basis ω^i.

$$\begin{aligned} v &= \sum \langle v, e_i \rangle \, e_i, \\ v &= \sum \omega^i(v) e_i, \\ \lambda &= \sum \lambda(e_i)\omega^i. \end{aligned}$$

2.2.2 Matrix Form

This form is slightly more enlightening, because as well as telling us what the coordinates are, it also tells us how they should be organized.

Recall from linear algebra that when we record the coordinates of a vector in some basis we do so in column-matrices.

Recall also that given a linear transformation $T\colon V \to W$, a basis of V and a basis of W, we record the "coordinates" of T in a matrix. This matrix is of size $\dim W \times \dim V$.

As it turns out, the convention of writing vectors as column-matrices can be seen as a particular case of this process: to specify an element of W is the same as to specify a linear transformation $\mathbb{R} \to W$, yielding a matrix of size $\dim W \times 1$. It stands to reason that a dual vector, which is a linear transformation $V \to \mathbb{R}$, should be represented as a vector of size $1 \times \dim V$; in other words, a row-vector!

What should the coordinates be? In this, we find the same answer as before (which is why the two points of view are equivalent). Recall from linear algebra that in order to represent T as a matrix, one must first specify a basis $\{v_k\}$ of V and a basis $\{w_k\}$ of W, and the ij-th element of said matrix is the j-th coordinate (in W) of Tv_i. In our case, W is $\mathbb{R}$, and the most natural basis is the single vector 1, and so we conclude that (given a basis $e_1, \ldots, e_n$ of V) it is natural to represent a covector λ in coordinates as the row-matrix

$$\begin{bmatrix} \lambda(e_1) & \ldots & \lambda(e_n) \end{bmatrix}.$$

This representation has two benefits. One, it reminds us that vectors and covectors are fundamentally different objects, which could be lost on us if we represented both as mere sequences of numbers. Two, it leads to a nice identity for application of covectors in local coordinates. Given a vector space V, some basis, a vector v and a covector λ, let $\hat{v}$ and $\hat{\lambda}$ be their coordinate representations. Then, the following identity holds:

$$\lambda(v) = \hat{\lambda}\hat{v}.$$

2.2.3 Coordinate Changes

To conclude this section and move on to more visual stuff, let us discuss how covectors (or rather, their representations) change under changes of basis.

Let $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ be two different bases of V. Given $v \in V$ or $\lambda \in V^*$, denote by v_e the coordinates of v in the e basis, and likewise for λ or f.

Let B be the matrix whose j-th column is the coordinates of f_j in the e basis. Then, linear algebra relates the representation of a vector in both bases by

$$v_e = Bv_f.$$

How does it relate the representation of λ_e and λ_f? From the identities we have already deduced, we conclude

$$\lambda_f = \begin{bmatrix}\lambda(f_1) & \dots & \lambda(f_n)\end{bmatrix} = \begin{bmatrix}\lambda_e(f_1)_e & \dots & \lambda_e(f_n)_e\end{bmatrix},$$

and from a little bit of matrix block multiplication we reach the conclusion

$$\lambda_f = \lambda_e \begin{bmatrix}(f_1)_e & \dots & (f_n)_e\end{bmatrix} = \lambda_e B.$$

Compare the two formulas:

$$v_e = Bv_f, \quad \lambda_e = \lambda_f B^{-1}.$$

The formulas are identical, except that B has been sent to the other side and inverted. This is as it should be, as it ensures that the formula $\lambda(v) = \lambda_e v_e$ holds in any basis, as

$$\lambda_e v_e = \lambda_f B^{-1} B v_f = \lambda_f v_f.$$

This last argument can be used as a mnemonic in a pinch to deduce how coordinates of covectors ought to change in order to "preserve the dot product".

Recall that we are representing the coordinates of the covector as a row-matrix. Note how, if we represented them in a column-matrix, the formulas would become slightly more awkward. The coordinates represented in a column-matrix are given by $(\lambda_e)^\intercal$ and $(\lambda_f)^\intercal$, and so we conclude

$$(\lambda_e)^\intercal = (B^{-1})^\intercal (\lambda_f)^\intercal,$$

in other words, if the vector coordinates have changed by B then the covector coordinates have changed by $(B^{-1})^\intercal$.

2.2.4 Visual Representationss

All this discussion is fine and all, but it doesn't help us "see" covectors. I don't know about you, but I have a hard time imagining covectors as arrows which "just scale differently". For one, using equal-looking arrows to represent two distinct kinds of vectors is inconvenient, and visualizing how they change upon distortion of the space is troublesome. Fortunately, this is only because arrows are the wrong tool for the job: there's another, more convenient way to represent covectors visually.

Let's look at the covector s from our initial discussion again, and how it changes under a coordinate change.

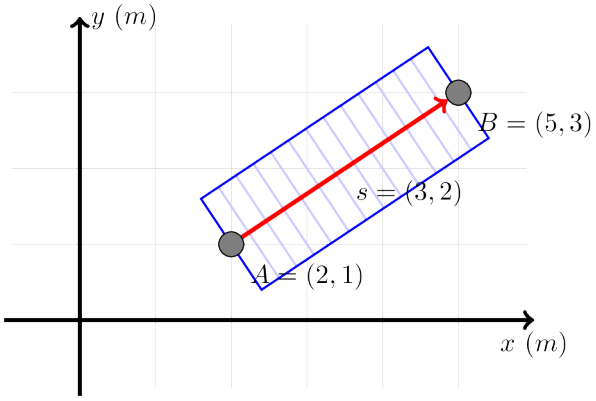


Figure 7: s in the original coordinates

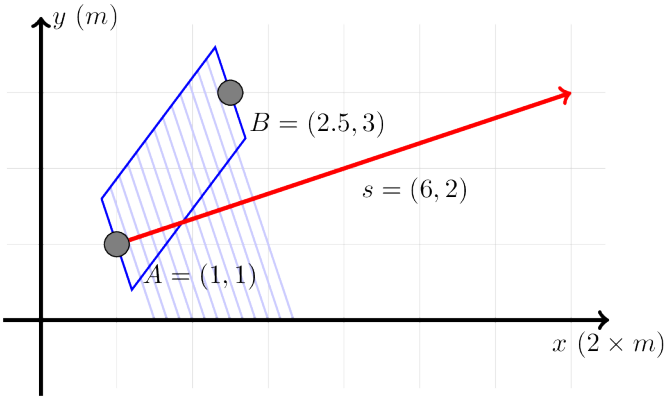


Figure 8: s when x is scaled down by $\frac{1}{2}$

To visualize how our covector changes we need to find its new direction and the new magnitude. Let's focus on direction first.

In the original coordinates, s pointed straight from A to B, but upon coordinate change that stopped happening, so "direction" in the usual sense is not reliable. However, if you'll look closely, you'll find something that hasn't changed when we changed coordinates: s is orthogonal to the level sets.

This suggests a relation between s (or rather, its representation) and these level sets. In retrospect, this shouldn't be surprising: recall that we originally defined s as a vector that represents the slope of my ramp (or, more abstractly, of a plane).

The connection goes a little bit deeper, however. Recall: What do we do with covectors? The answer: apply them to a vector to get a number. And while the inner product of vectors has an okay geometrical interpretation (project one vector onto the other and take the product of lengths), the product of a vector with a covector can be even more visual. For example, let us create a point C about a third of the way down the ramp, and let us calculate $h(C)$ using the formula $h(p) = h_0 - s \cdot (p - A)$, or, as we should be writing it if we're seeing s as a covector,

$$h(p) = h_0 - s(p - A).$$

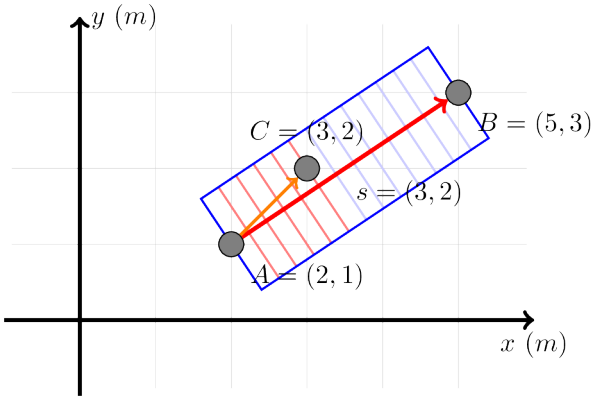


Figure 9: Orange: the vector $\overrightarrow{AC} = (1, 1)$; The level sets which intersect this vector are highlighted.

There are two ways to calculate $h(C)$. One of them is simply calculating:

$$h(C) = h_0 - (3, 2) \cdot (1, 1) = h_0 - 5.$$

Another is to look at the level sets. The vector $\overrightarrow{AC}$ intersects five level sets (not counting the one at A), and so we conclude that h "has decreased by 5 steps on the way from A to C". In other words, given a vector v we can calculate $s(v)$ by inspecting how many level sets v passes through. This shouldn't be surprising: after all, by definition these level sets measure the change in s.

This suggests a change in perspective: instead of visualizing a covector as an arrow, it is best to visualize it as the collection of its level sets. In our two-dimensional case, we visualize s as a collection of parallel lines. In n-dimensional space, we can imagine a covector λ as a collection of hyperplanes. To evaluate $\lambda(v)$, one counts how many of these hyperplanes of λ the arrow v crosses.

There are a few loose ends which still need to be tied up. First of all, to specify the level sets isn't enough: one needs to specify in which direction the numbers get bigger. But that's just a matter of specifying an orientation.

Second: The interpretation of "applying a covector to a vector is just counting intersections" is pretty cool, but comes with a lot of asterisks. If you only draw the level sets corresponding to the integers, you don't have enough resolution for non-integral results; you can't have, say, 3.5 intersections between a vector and a collection of hyperplanes.

To fix this last problem we change our perspective. Instead of drawing a discrete collection of hyperplanes, imagine infinitely many, with an infinitesimal distance δ between each of them. In order to evaluate the covector on a vector, we count how many planes the vector crosses, with the caveat that each plane counts only ε, where this is another infinitesimal number. This isn't a rigorous visualization, of course; there's no such thing as infinitesimal (real) numbers. But if we play a bit loose with rules, it works.

Here's how the visualization works under transformations. If we distort space, every hyperplane still counts for ε, but the hyperplanes themselves are rotated or squished together or torn apart. Since they represent level sets of the covector, they will always remain a faithful representation of it, and we will always be able to imagine evaluating the covector on a vector as counting the number of planes it intersects (times ε).

Finally, in practical terms, we will only draw a few of these hyperplanes, leaving the job of picturing the infinity of them with tiny spaces between them to the reader.

2.3 Summary

We have determined that, when measuring slopes in a coordinate-independent way, the notion of covector is more robust with regard to coordinate changes than a "regular" vector.

A covector can be defined as a linear transformation from V to $\mathbb{R}$. This makes sense intuitively (because to measure a slope is to assign a height to each number) and formally (because the coordinates scale as we want them to).

We can visualize such covectors, not as arrows, but as stacked (oriented) hyperplanes, an infinity of them an infinitesimal distance from each other. Under this picture, evaluation of a covector on a vector amounts to counting how many hyperplanes the vector crosses.

2.4 Closing Remark

For another example of where covectors are useful, see p. 7 of Geometrical Vectors, where the author discusses an application of covectors to electromagnetism.

While the presented method of visualizing covectors has its benefits, one place it fails is in visualization of vector space operations: you can't "add" hyperplanes.

There are ways around this, though they aren't as simple as addition of vectors "just put the tail of one on the head of the other". To learn more, take a look at Geometrical Vectors, mentioned in the introduction.

Vectors and covectors aren't the only type of vector out there. Geometrical Vectors talks about plenty more. It refers to vectors as arrows and covectors as stacks, and includes thumbtacks, sheaves, and more.

3 On the Visualization of Differential Forms

September 8, 2021

This post talks about visualization of differential forms on differential manifolds. It is aimed at a reader who is in the process of learning about differential forms (or who already has), but who has not seen any visual intuition for what they are.

This post has as a prerequisite [the above] post on the visualization of the dual of a vector space, which (hopefully) motivates a very simple particular case of what will be described in this post.

The references therein mentioned still apply, but mostly On the Visualisation of Differential Forms by Dan Piponi, on which most of this post is based.

This post is intended for purely illustrative purposes. We will not yet examine how this "formalism" is equivalent to the standard notion of differential form, nor will we turn our "proofs" and "definitions" into rigorous arguments. That is the content of a future post,[1] in which we will show how the visualizations we will build here tie into the symbolic garbage usually called "tensor algebra".[2]

3.1 Differential Forms

In all that follows, M denotes an n-dimensional smooth manifold, for some non-negative integer n.

3.1.1 Motivation (n-form)

Before defining the notion of differential form, it is first useful to consider what problem we intend to solve.

The problem at hand is that of integration. We already know how to differentiate maps between differential manifolds, and now we wish to find out how to integrate them. Unfortunately, by itself this is not possible.

For example, consider the constant function equal to one on the unit sphere S^2. Its integral really ought to be the surface area of the sphere, 4π. On the other hand, the unit sphere is diffeomorphic, and therefore indistinguishable as a smooth manifold, from the radius two sphere, which we will call $2S^2$. But integrating 1 on the latter yields a different result: 16π. What gives?

The obvious answer is that we shouldn't be integrating the 1 function, but rather the function $\frac{1}{4}$, because the diffeomorphism between S^2 and $2S^2$ "spreads out the function". And while this is legitimate intuition, that's just not how functions "go across diffeomorphisms". We need to consider some other kind of object.

Let's inspect the "physicist's way" of approaching integration. First, one divides the domain of discourse into many small rectangles, whose area we know how to calculate. Then, for each square, one multiplies the value of on that rectangle by its area . Finally, one sums over all the tiny squares, which we represent by a tall S in a familiar way as $\int f \, \mathrm{d}A$.

How does this "approach" fail on smooth manifolds? The problem is that we don't know how to measure the areas of the tiny rectangles. Indeed, the notion of area is not conserved when we change charts, so we cannot simply "measure the area of the rectangle in coordinates", because this area depends on the coordinates chosen.

Consequently, in order to define integration, we need to add additional structure. For each point x in our manifold M, we need to know how to measure areas of tiny rectangles in a coordinate-free way. The apparatus that allows us to do so is precisely the differential form: it receives an "infinitesimal parallelogram" (represented by its edges given as tangent vectors) and returns its area (or rather, its area "scaled up" as not to be infinitesimal).

[1] Author's note: This never happened.

[2] Author's note: I was feeling particularly sour at the subject after my geometry teacher at the time urged me to stop trying to visualize differential forms. "Not everything can be visualized", he said. "I spent a lot of time visualizing all Lie groups I saw, and it wasted a lot of my time and didn't end up being useful", he said. Yes, I am still feeling a little sour. But the point is, at the present time I would no longer describe tensor algebra as "symbolic garbage".

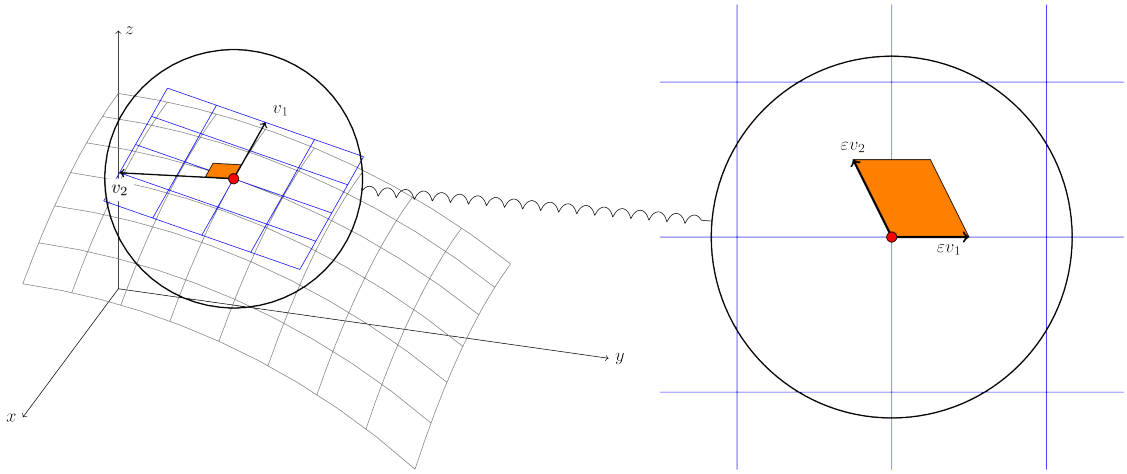


Figure 10: The relation between a parallelogram (orange) and the tangent vectors that represent it. Since our parallelogram is infinitesimal, we work not with its edge-vectors, but rather we rescale these up by a factor "$1/\varepsilon$", where ε is a previously fixed infinitesimal number. In other words, the vectors v_1 and v_2 represent a parallelogram with sides of order ε and area of order ε^2.

This, by itself, is enough motivation to define a differential form as "something that takes n tangent vectors and returns a number", and it is very natural to require multilinearity and skew-symmetry. However, we have not yet reached a useful visualization.

To proceed, let us imagine that our manifold is some kind of physical object, composed of many tiny pieces like "atoms". Then, there is an unambiguous way to measure some kind of area on the manifold: given some shape, one can count how many atoms lie in this shape. Then, assuming the atoms are uniformly distributed in some sense, one must only multiply the atom count by some scaling factor to obtain "the area" of the shape.

This thought experiment shows that a "manifold made of atoms" has a canonical way of measuring area (modulo a scaling factor). Therefore, it might be worth toying with the idea of representing a differential form by "a cover of the manifold by atoms".

We then make the following "definition": a differential form on M is a cover of M by an infinitely dense mesh of points. To calculate the area of a set $A \subseteq M$, we count the number of atoms in A, and multiply by an appropriate infinitesimal number.

The biggest benefit of this point of view is that it behaves naturally with regard to rescalings. In other words, if we apply a diffeomorphism to the manifold that scales everything up in some manner, we are also scaling up the spaces between the atoms (but not the weight of the atoms themselves!). Therefore, area is conserved.

Equivalently, this point of view allows for coordinate-independent measurement of area, as any distortion between two charts of M will imply an equal distortion of the "atoms", and again area is conserved.

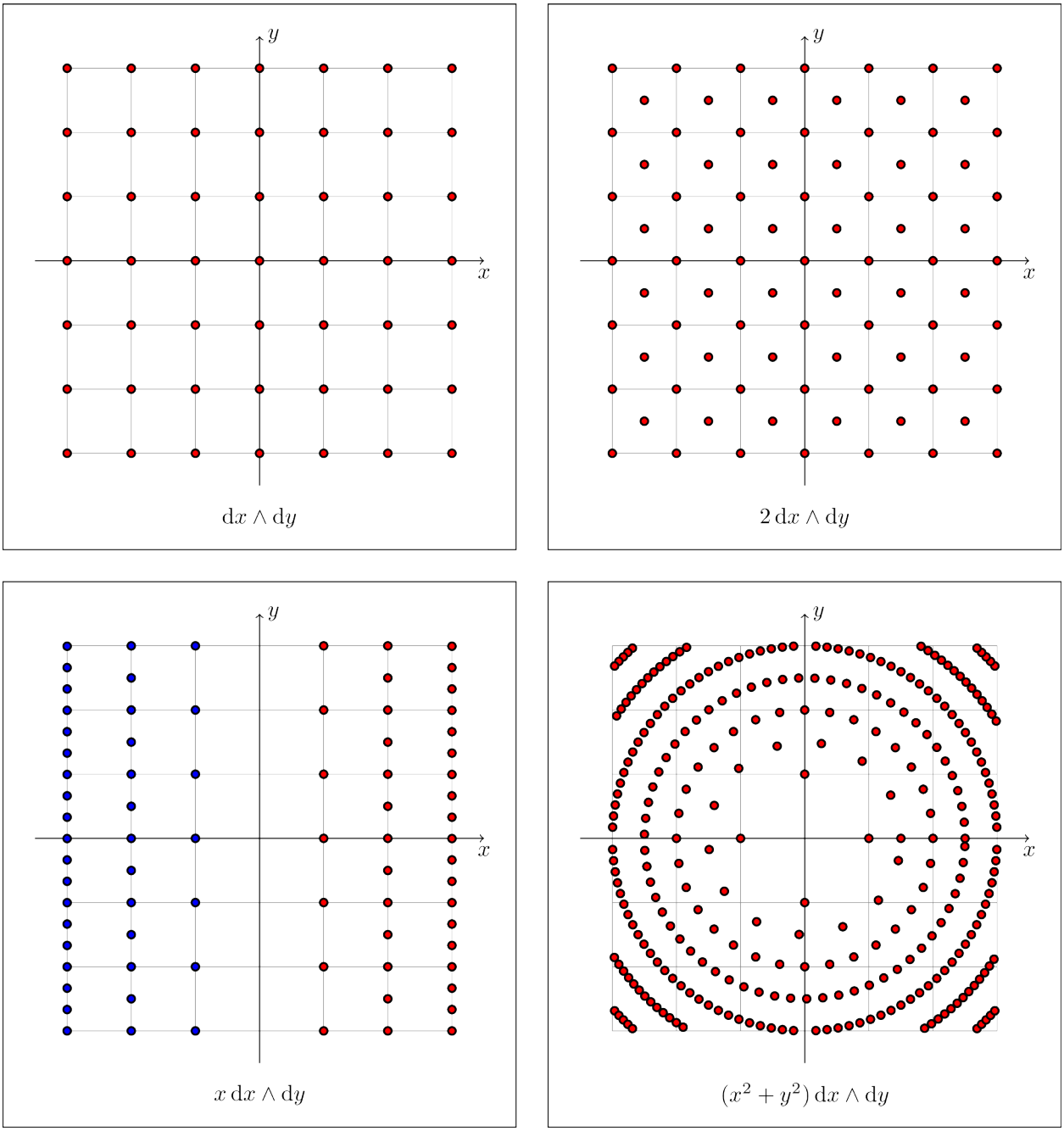


Figure 11: Visual representation of some 2-forms in $\mathbb{R}^2$. The blue points represent "negative atoms", which count negatively towards the atom count. To (approximately) measure the integral of each of these forms over a set U, simply count the amount of atoms in U.

3.1.2 Other Order Forms

Now that we have solved the problem of area measuring, what comes next? Why are other dimension forms necessary?

At the most elementary level, forms of dimensions other than n appear in the statement of the generalized

Stokes theorem (we will henceforth drop the "generalized" from the name). This is an important theorem, having as particular cases the divergence and Green's theorem, which are very useful tools of calculation. Therefore, I hope the reader will agree that the study and proof ("proof") of this theorem is a worthwhile endeavor.

If the reader is not yet familiar with the Stokes theorem, they may have some trouble understanding the following few paragraphs. However, I hope that the subject matter will become clearer over time. In any case, I suggest that such a reader would continue linearly until we state and "prove" the Stokes theorem, and then return here and reread this part with more context.

The statement of the Stokes theorem relates the integral of a form ω on the border of a set U with the integral of its "derivative" $\mathrm{d}\omega$ on the interior of U, with the usual statement being cleanly written as

$$\int_{\partial U} \omega = \int_U \mathrm{d}\omega.$$

Setting aside for now the "derivative" part of the statement, there is something we first have to touch. What kind of object does ω have to be for us to integrate it on ∂U? It cannot be a differential form as we have described them, because if we fill the space with a very fine grid of atoms, it is very unlikely that any of them lie exactly on the border of U.

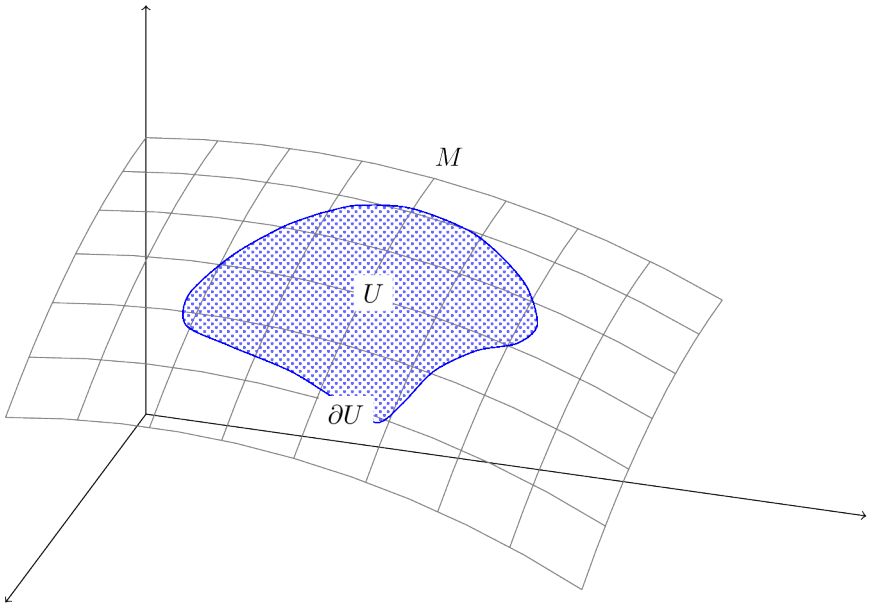


Figure 12: The set U and its border, ∂U, the integrals over which the Stokes theorem relates.

In other words, we seek to solve the following problem. We have already concluded that the kind of object used to integrate over n-dimensional sets and measure n-dimensional area is "covers of the space with points". However, this is not adequate to integrate over k-dimensional sets, with $k < n$, i.e. k-dimensional submanifolds of M. Therefore, we wish to discover what kind of object will allow us to naturally define integrals on k-dimensional submanifolds of M.

From now on, let N be a k-dimensional submanifold of the n-dimensional manifold M.

The algebraic point of view presents itself rather naturally. Recall that the algebraic definition of a differential form is a thing that takes n tangent vectors to the same point and returns a number. We want to define an object on M which can be "restricted to N", becoming an object which takes k vectors tangent to N and returns a number. It is then natural to define a "k-form on M" as an object which takes k tangent vectors at the same point of M, and the restriction to N becomes simply the restriction of ω to the collection of vectors tangent to N.

Let us now go from the algebraic point of view to the pictorial one. A k-form receives k vectors, forming the edges of an infinitesimal parallelogram of dimension k. For example, if $n = 3$, this means that we are considering an infinitesimal line segment or rectangle, instead of a cube.

If you've seen my other post on the dual space, you should already be familiar with a way to visualize a covector, a thing that receives a vector and returns a number, as a collection of hyperplanes. This carries over to the notion of 1-form, which is also a thing that recieves a vector and returns a number. Indeed, we can visualize a 1-form as a collection of hypersurfaces (surfaces of dimension $n - 1$) on our manifold, and evaluating η on a vector consists of counting how many of these hypersurfaces intersect the given vector. Note the similarities:

n-form ω	**1-form η**
Collection of points (0-dimensional)	Collection of hypersurfaces ($(n-1)$-dimensional)
To evaluate on a collection of vectors, count how many points are in the parallelogram spanned by them	To evaluate on a vector, count how many hypersurfaces intersect it

Now, it may not be obvious with only two data points, but I propose the following visualization of the intermediate steps: *Visualize a k-form as a collection of $(n - k)$-planes.*

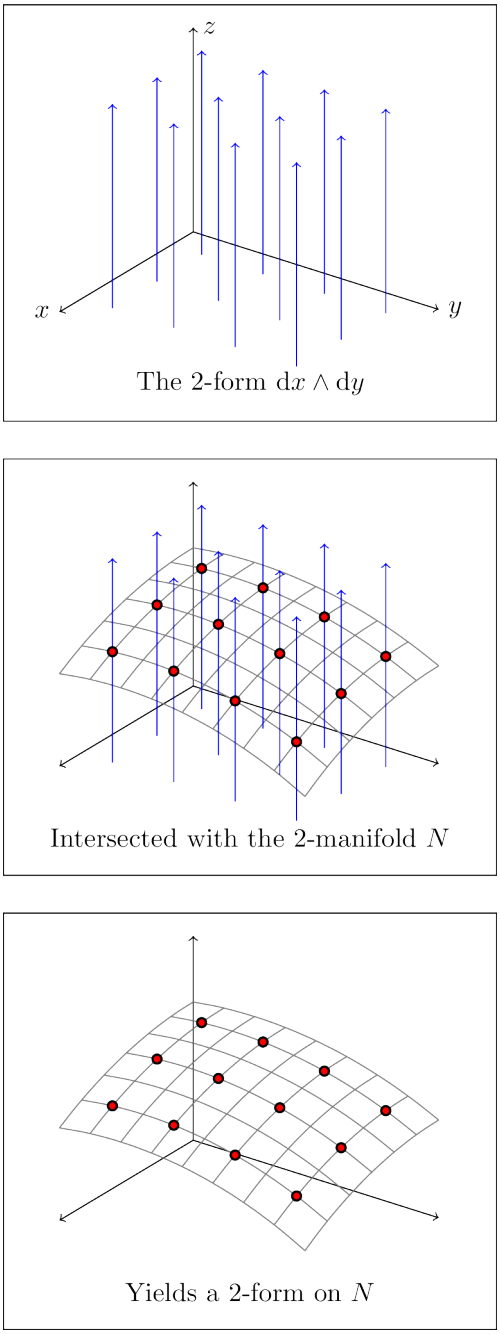


Figure 13: A visualization of the restriction process of a k-form to a k-submanifold

Under this visualization, to evaluate such a k-form ω on a collection of vectors is to count how many of these $(n-k)$-planes intersect the k-parallelogram spanned by these vectors. The dimensions check out: in general, the intersection between two objects of dimension a and b is an object of dimension $a+b-n$, and if $a = n-k$ and $b = k$, this means that these intersections will often be 0-dimensional points.

As a bonus, our original goal is also met. Indeed, if the k-form ω consists of covering the space in $(n-k)$-planes, the restriction of ω to a submanifold N consists simply of taking the intersection between these planes and N. As in the previous paragraphs, these intersections will often be points, and so we recover

a differential form (of dimension k) on N.

3.2 The Bad and the Ugly

There are two big details we have been neglecting, and will keep neglecting, over the course of this post. We will now take a slight detour, to shed some light on them before we shove them under the carpet again.

Readers who are not terribly interested are encouraged to skip ahead to "The Exterior Derivative".

3.2.1 Transversality and Perturbations

Let's go back to this paragraph: *"in general, the intersection between two objects of dimension a and b is an object of dimension $a + b - n$, and if $a = n - k$ and $b = k$, this means that these intersections will often be 0-dimensional points."* Why would this statement be true, and what do we mean by "in general"?

The answer is related to a concept known in differential geometry as transversality. The following discussion comes with a lot of asterisks, and is even more handwavey than the rest of this post.

Two submanifolds A and B of a common manifold M are said to intersect transversely at a point p if T_pA and T_pB span T_pM. Intuitively, this corresponds to the intersection being, well, transversal, as opposed to tangent. For example, two curves in the plane intersect transversely at a point if they form an angle which is not 0 or 180 degrees. Two curves in space never intersect transversely.

The most important fact about transversality, aside from the fact that transverse intersections are The Nice Kind, is that it is robust and very common. By this we mean that if two manifolds intersect transversely, then small enough perturbations of the manifolds will not mess that up: given small enough perturbations, they will still intersect, and the intersection will still be transverse.

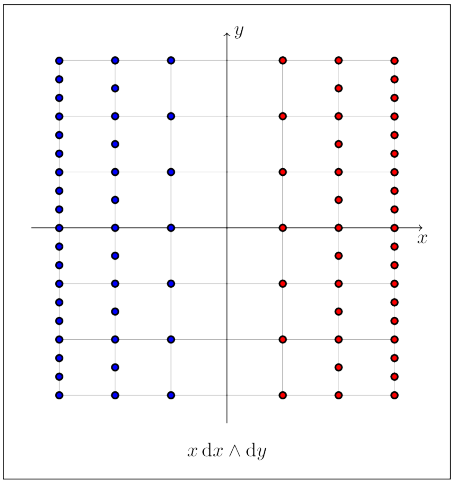
Furthermore, non-transversality is sensitive and rare. If two manifolds intersect non-transversely, then almost no perturbations will preserve this state of affairs. For example, consider two lines in space. If you perturb either of them randomly, it is almost certain that they will stop intersecting. As another example, consider in $\mathbb{R}^2$ the x axis and the graph of $f(x) = x^3$. These curves intersect non-transversely at the origin, but any tiny nudge in either of these curves in the up or down direction will make the intersection transverse.

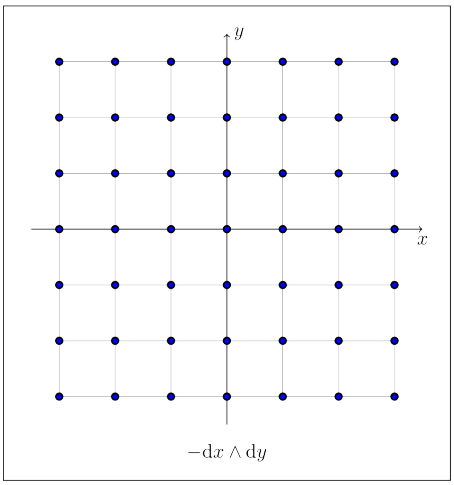
Finally, transverse intersections satisfy the dimensionality property referred at the start. That is, if two manifolds A and B, of dimension a and b respectively, intersect transversely, their intersection is a manifold of dimension $a + b - n$.

As a consequence, we will henceforth (as we have thus far) ignore the possibility of nontransverse intersections.

3.2.2 Orientations

Another important detail to be observed is the orientation of the surfaces "composing" a form, and more specifically the orientation of the intersections.





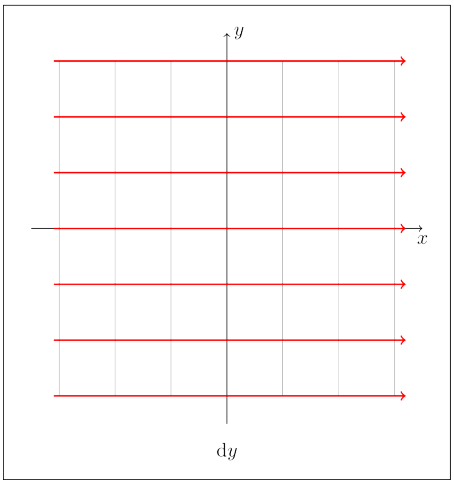


Figure 14: Some standard forms in $\mathbb{R}^2$.

For example, consider the example representation of the form $x\,\mathrm{d}x \wedge \mathrm{d}y$ pictured above, or something even simpler such as $-\mathrm{d}x \wedge \mathrm{d}y$, both in $\mathbb{R}^2$. Then, the "atom" representation requires that some or all of the

atoms composing these representations count negatively.

As another example, consider, still in $\mathbb{R}^2$, the 1-form $\mathrm{d}y$, which is represented as a uniform stack of horizontal lines. The mnemonic "to calculate $\mathrm{d}y(v)$ is to count how many horizontal lines v crosses" works if v is pointing upwards, but requires negating the result if v points downwards.

The easiest solution is to play with orientability. Suppose, for the moment, that M is oriented, and let ω be a k-form. Then, we make the following slight addendum to our visualization of ω: we represent it visually by a mesh of *oriented* $(n-k)$-submanifolds.

Here's what this means for visualization purposes:

- Points will be represented, as we have so far, as either red or blue depending on whether they are positively or negatively oriented. (Yes, a zero dimensional manifold has orientation, either positive or negative at each point!)
- Curves will be represented with an arrow representing the positive direction. The sign convention for evaluation will be given by the right-hand rule, so for example $\mathrm{d}y$ in $\mathbb{R}^2$ would be given by arrows pointing to the right, and $\mathrm{d}x \wedge \mathrm{d}y$ in $\mathbb{R}^3$ would be given by arrows pointing upwards.
- When representing 1-forms in $\mathbb{R}^3$ via surfaces, it should be understood which normal direction is positive, either via arrows indicating the positive direction or through written indication. However, this case probably won't happen a lot.
- Since I'm not going to be making pictures in 4d, that basically exhausts all possible cases, except for the case of 0-forms, which will be mentioned at the end.[3]

Again, note that all of this makes sense for orientable M. In one hand, this foreshadows an interesting fact: the theory of forms is more well-behaved for orientable manifolds than non-orientable ones. For example, a Riemannian metric on M induces a volume form if and only if M is orientable; in fact, a differentiable manifold is orientable if and only if it admits an n-form which is never null. Finally, and most problematic, the definition of integral of an n-form on a manifold requires that the manifold be orientable, so without orientability everything we intend to do with forms breaks down. [Note: There is a way around this last problem called "densities", which is a sort of generalization of n-forms. Unfortunately, I know nothing about it.]

On the other hand, I would like to point out that despite these problems, a lot of what we will do works equally well in non-orientable surfaces. Indeed, every manifold is locally orientable, so our pictures work locally, and non-orientability will not invalidate our discussions on the exterior derivative, wedges or Lie derivatives. As a rule of thumb, local or pointwise statements will work in every manifold, while global statements require orientability.

Author's note: I don't know why I didn't write about this at the time, but I certainly thought about it. The reason that I focus on orientations is because they are more elementary and easier to describe, but really, the true definition of "atom" should be: A ***co-oriented*** *proper submanifold with boundary.*

3.3 The Exterior Derivative

Recall the form $\mathrm{d}y$ on $\mathbb{R}^2$, which can be represented as a collection of horizontal lines. A trivial modification to this form would be considering $y\,\mathrm{d}y$, in which the lines are getting denser the further up you go.

[3] Author's note: I never got around to this, so let me summarize it in this footnote. As is known, zero-forms coincide with functions on the manifold. The function f corresponds, in this visualization, to the collection of its sub-level sets. This works out because, in this visualization, a zero-form is roughly a collection of open sets. At a place where $f(x) = 5$, say, we'd have roughly $5\varepsilon^{-1}$ surfaces covering that point. Negative values correspond to "open sets that count negatively".

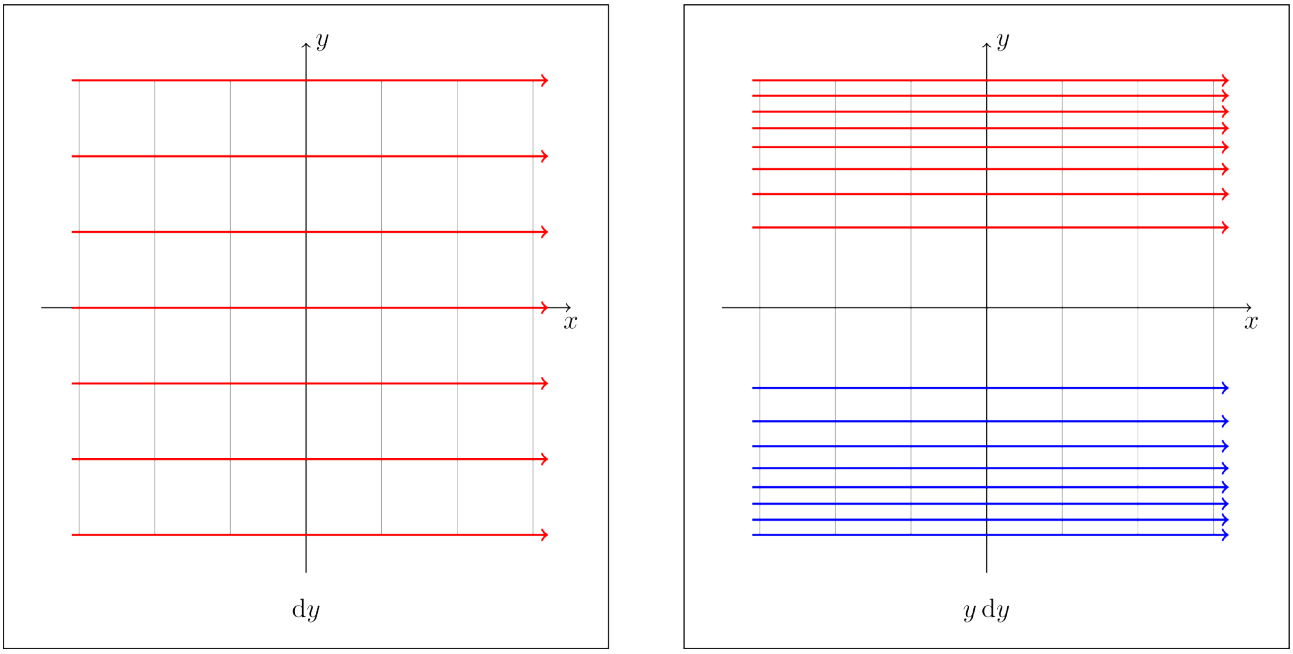


Figure 15: Representation of the two forms $\mathrm{d}y$ and $y\,\mathrm{d}y$. The blue lines represent "negative lines", and intersections with them count as -1.

A slightly less trivial modification would be the form $x\,\mathrm{d}y$. In this case, the lines are getting denser the further to the *right* you go, yet the lines must be horizontal, in order for horizontal vectors to intersect none of them. Consequently, you get a new-looking picture:

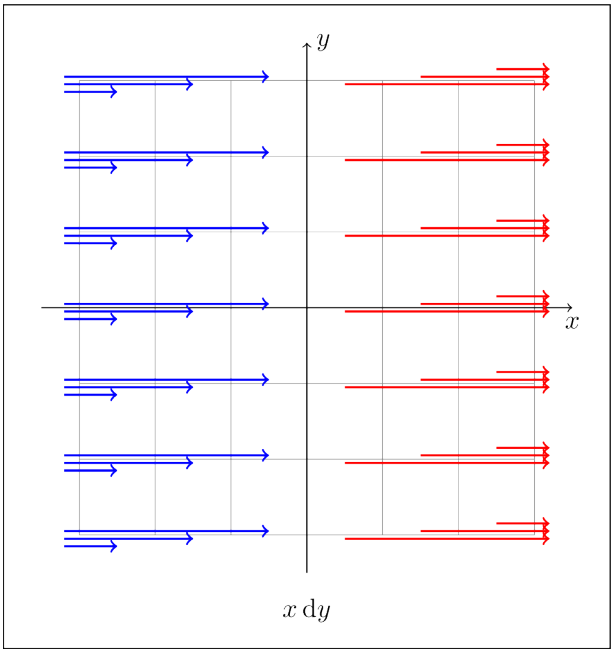


Figure 16: The form $x\,\mathrm{d}y$

The new phenomenon exhibited here is the "creation of lines". This phenomenon is meaningless for n-forms, but for forms of other dimensions it is the distinction between the "atoms" being $(n-k)$-manifolds, versus being $(n-k)$-manifolds *with boundary*.

All this to say, our formalism does not require that the surfaces representing forms be boundaryless. Indeed, some forms, like $x\,\mathrm{d}y$, require that the surfaces representing them have boundary. As a consequence, it is natural to ask, given a form, whats its boundary is. You can probably see where this is going: if the k-form ω is represented by a collection of $(n-k)$-surfaces, their boundaries (which are $(n-k-1)$-forms) make up a $(k+1)$-form which we usually denote $\mathrm{d}\omega$.

Let's see what this looks like in the above examples. The forms $\mathrm{d}y$ and $y\,\mathrm{d}y$ are represented without boundary, so their exterior derivatives are null, which agrees with the symbolic computation. On the other hand, if we look at the borders of $x\,\mathrm{d}y$, we get a uniform mesh of points, also known as the familiar $\mathrm{d}x \wedge \mathrm{d}y$.

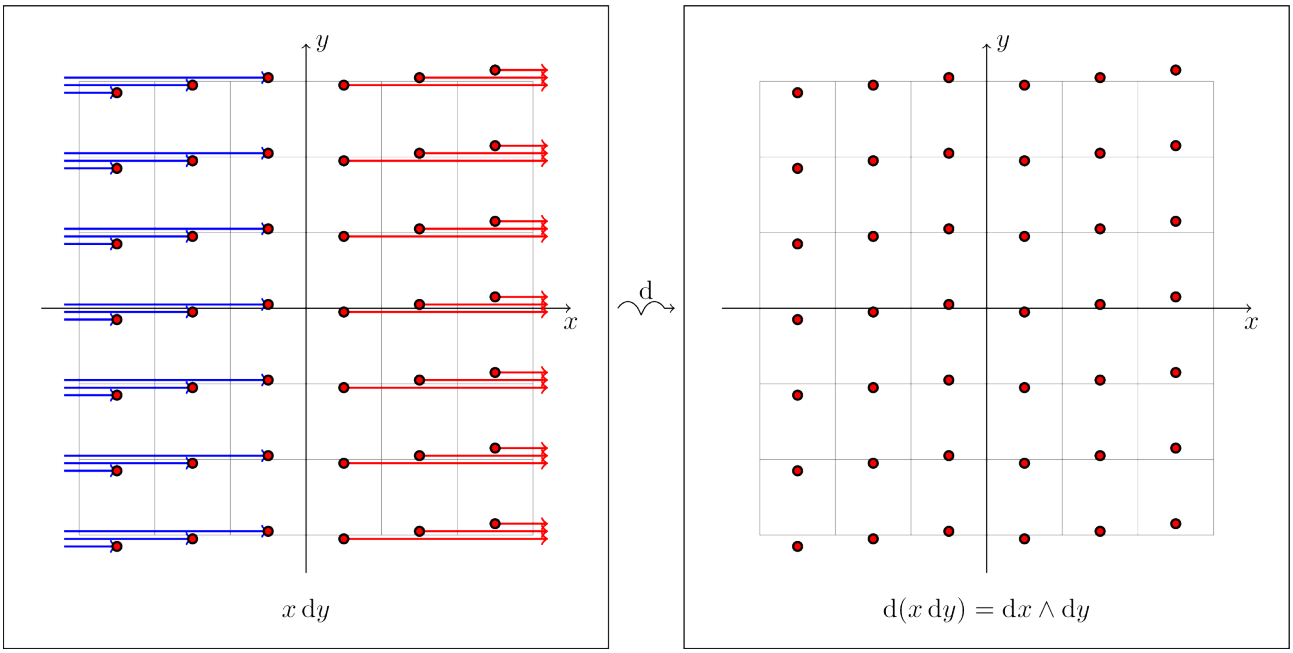


Figure 17: Visual representation of the computation of $\mathrm{d}(x\,\mathrm{d}y)$. Even though the visual representation on the right is not exactly a uniform grid, it is supposed to be: the vertical offset is only an artifact of the visualization of $x\,\mathrm{d}y$.

Let's look at another, more complicated example, now in three-dimensional space.

Consider the 1-form $xy\,\mathrm{d}z$. As in the $x\,\mathrm{d}y$ example, this form must be represented by horizontal planes. Their density depends on x and y, so that near the axes there are less planes, and far from the axes there are more. Furthermore, the amount of planes at a point is proportional to xy, so it makes sense to represent this form by a collection of planes with hyperbolic borders:

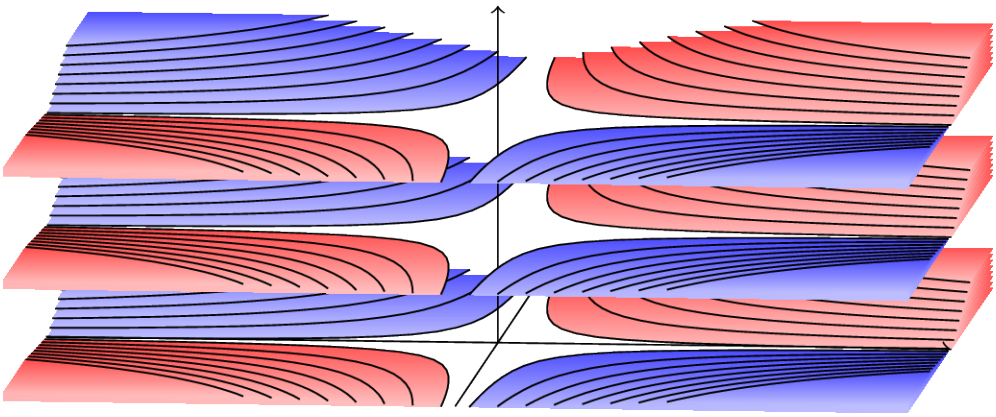


Figure 18: Visualization of $xy\,\mathrm{d}z$. The red planes are oriented such that an arrow going "up through" them counts positively, while such an arrow would count negatively going through a blue plane.

Let us now compute visually the 2-form $\mathrm{d}(xy\,\mathrm{d}z)$. The picture is as follows.

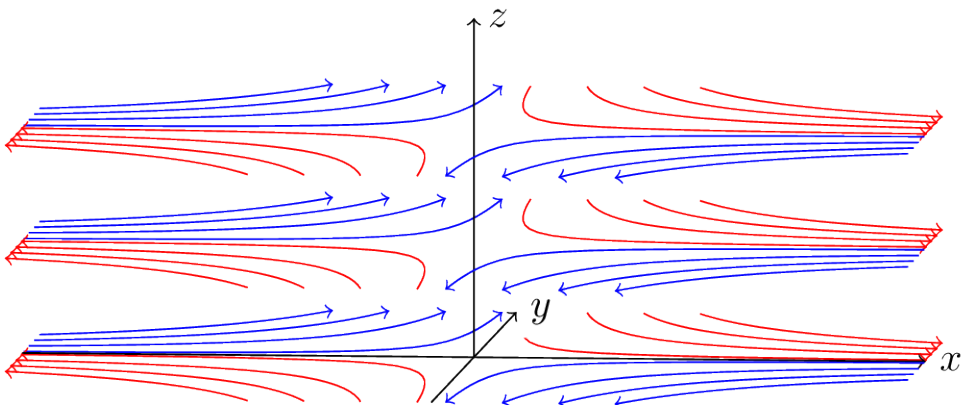


Figure 19: Visual representation of $\mathrm{d}(xy\,\mathrm{d}z)$. Note that the borders of planes are oriented with the right-hand rule.

The result is a collection of curves satisfying the equation $xy = c$, with constant z. Let us label these curves by c and z. We will calculate the form $\mathrm{d}(xy\,\mathrm{d}z)$, let us call it ω from now on, at a point $p = (x_0, y_0, z_0)$.

We'll revisit this example again when we look at the wedge product,[4] but until then we will sketch the answer by calculating it for a well-chosen basis of the tangent space at p. Consider the following three vectors:

[4]Author's note: I did not end up doing this, but the intent was to note that, by differential calculus, $\mathrm{d}(xy\,\mathrm{d}z) = \mathrm{d}(xy) \wedge \mathrm{d}z$, and one can recover Figure 19 using the facts: (1) Wedge corresponds to intersection, and (2) If f is a function, the 1-form $\mathrm{d}f$ has as atoms the level sets of f at equally-spaced infinitesimal intervals.

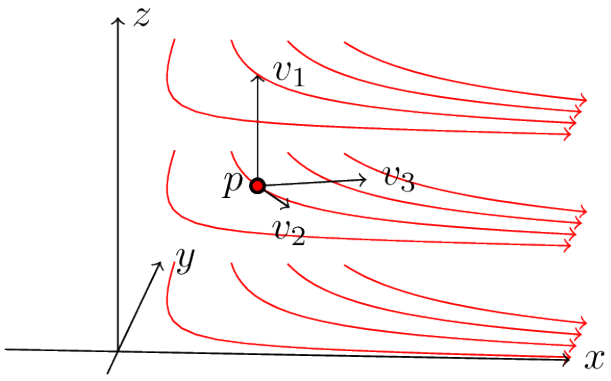


Figure 20: The basis we are constructing for $T_p\mathbb{R}^3$.

The vector v_1 is simply $(0,0,1)$. The vector v_2 is chosen to be tangent to the curves, and v_3 is meant to point in a direction of "increasing c". We'll look at $\omega(v_1, v_3)$ shortly, but until then it should be clear that $\omega(v_1, v_2) = \omega(v_2, v_3) = 0$. In general, whenever a vector is tangent to the curves defining a form, any parallelogram with it as an edge will have no (or negligible) intersection with these curves.

With only this, we are in position of finding a partial expression for ω. Indeed, if a_1, a_2, a_3, b_1, b_2, and b_3 are real numbers,

$$\omega(\sum a_i v_i, \sum b_i v_i) = \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} \omega(v_1, v_3).$$

It remains to compute a_1 and a_3 for a vector of the form $a = (x, y, z) = \sum a_i v_i$.

The a_1 component is trivial: it is simply the z component of a. To compute a_3, it is convenient to calculate v_3. If v_1, v_2, and v_3 are chosen to make an orthonormal basis, a_3 can be easily written as $a_3 = a \cdot v_3$.

First, we compute v_2. Since it is tangent to horizontal curves, it is of the form $(v_{2x}, v_{2y}, 0)$. Since it is tangent to the curve $xy = c$, the value of c for $p + \varepsilon v_2$ ought to be equal to $x_0 y_0$ to first order. This value is $(x_0 + \varepsilon v_{2x})(y_0 + \varepsilon v_{2y}) \approx x_0 y_0 + \varepsilon(x_0 v_{2y} + y_0 v_{2x})$. Therefore,we conclude

$$x_0 v_{2y} + y_0 x_{2x} = 0.$$

The unit norm solution to this equation with, say, $v_{2x} > 0$, is

$$v_2 = \frac{1}{r}(x_0, -y_0, 0), \quad \text{with } r = \sqrt{x_0^2 + y_0^2}.$$

Finally, we calculate v_3 as a unit vector, pointing away from the axes, orthogonal to v_1 and v_2.

Orthogonality to v_1 means simply its last coordinate is null, so we get $v_3 = (v_{3x}, v_{3y}, 0)$. Orthogonality to v_2 means $x_0 v_{3x} - y_0 v_{3y} = 0$, with obvious solution

$$v_3 = \frac{1}{r}(y_0, x_0, 0).$$

In conclusion, we may write

$$\omega(a, b) = \begin{vmatrix} a_z & b_z \\ \frac{1}{r}(y_0 a_x + x_0 a_y) & \frac{1}{r}(y_0 b_x + x_0 b_y) \end{vmatrix} \omega(v_1, v_3),$$

or, in more familiar wedge notation,

$$\omega = \omega(v_1, v_3)\, \mathrm{d}z \wedge \left(\frac{1}{r}(y_0\, \mathrm{d}x + x_0\, \mathrm{d}y)\right) = \frac{1}{r}\omega(v_3, v_1)(y_0\, \mathrm{d}x \wedge \mathrm{d}z + x_0\, \mathrm{d}y \wedge \mathrm{d}z).$$

(Note the double sign change in the second step! $\omega(v_1, v_3) \to \omega(v_3, v_1)$ and $\mathrm{d}z \wedge \mathrm{d}x \to \mathrm{d}x \wedge \mathrm{d}z$.)

Finally, we calculate $\omega(v_3, v_1)$. The order change may appear confusing (we put v_3 before v_1), but it is actually for the best: this way, the vectors v_3, v_1, and the red arrows satisfy the right-hand rule, so that the result is positive.

Since v_1 is a vertical unit vector and the system is symmetric with regard to vertical translations, we reduce the computation to finding out how many curves the vector v_3 passes through in the $z = z_0$ plane. In other words, we wish to find out how much c varies between p and $p + \varepsilon v_3$. The answer is trivial to calculate:

$$c_{\text{final}} = (x_0 + \varepsilon \frac{1}{r} y_0)(y_0 + \varepsilon \tfrac{1}{r} x_0) \approx c_{\text{start}} + \varepsilon \frac{1}{r}(x_0^2 + y_0^2) = c_{\text{start}} + \varepsilon r.$$

Therefore, we conclude that the rectangle spanned by εv_3 and εv_1 passes through roughly $\varepsilon^2 r$ curves, whence $\omega(v_3, v_1)$ should be r, and so our sketches yield

$$\omega = y\,\mathrm{d}x \wedge \mathrm{d}z + x\,\mathrm{d}y \wedge \mathrm{d}z.$$

That was a boatload of work. Good thing they invented vector calculus, with which we may verify our computation:

$$\omega = \mathrm{d}(xy\,\mathrm{d}z) = \mathrm{d}(xy) \wedge \mathrm{d}z = (y\,\mathrm{d}x + x\mathrm{d}y) \wedge \mathrm{d}z = y\,\mathrm{d}x \wedge \mathrm{d}z + x\,\mathrm{d}y \wedge \mathrm{d}z.$$

3.4 The Stokes Theorem

In this framework, the Stokes theorem is quite a triviality. (If you consider something like the Jordan curve theorem a triviality, at least...)

Intuitively, let M be an n-dimensional manifold and Q a manifold which splits M into "an inside" and "an outside". The "inside" must be small in the sense that a line cannot "go on infinitely inside". For example, the plane $z = 0$ would not split $\mathbb{R}^3$ into an inside and an outside, because neither half-space is small in this sense: you have infinite line segments living in each.

An example of a set we wish to consider is $Q = S^2$, which splits $\mathbb{R}^3$ into an inside and an outside. The inside is "small" because the only way to have an infinite line inside Q is to have it wind up infinitely. Now, it is hard to express properly why such infinite winding up is disallowed, but if we assume that it is, then the inside of S^2 counts as small. We are interested in sets whose inside satisfies this property.

Of course, all this intuitive handwaving is intended to convey a concept which mathematicians have figured out long ago: We want Q to be the border of a manifold with boundary, say N. To say that N ("the inside") is small is to say that N is compact.

Now, we intend to integrate some $(n-1)$-form ω on Q. Ordinarily, there would be no reason to believe that there would be a way to simplify this computation, but the hypothesis that Q splits the space in two implies a very important fact: *any line which enters Q must either exit it again or stop inside Q*. Therefore, if we see integrating an $(n-1)$-form on Q as counting the lines entering Q, we conclude that solving the problem of integrating ω on Q is the same as to count the number of line ends in N. The line ends are counted by $\mathrm{d}\omega$, and so we conclude the familiar formula

$$\int_{\partial N} \omega = \int_N \mathrm{d}\omega.$$

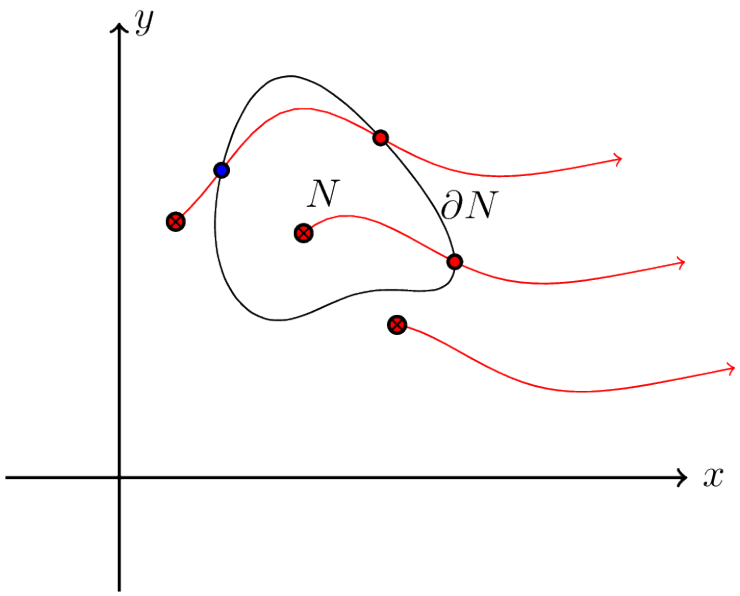


Figure 21: Visualization of the Stokes theorem. The integral $\int_{\partial N} \omega$ counts the number of intersections (red minus blue, because orientations). The integral $\int_N \mathrm{d}\omega$ counts the number of crossed dots inside N.

3.5 What Comes Next

Let's recap what we've done thus far. We have defined an n-form as an area/volume/measure measuring device. In order to measure areas of lower dimension (say, on submanifolds) we have defined the notion of k-form for $k < n$. These lower-dimensional forms displayed a phenomenon that could not be seen in the n-forms, which is the possibility that the "atoms" were manifolds with boundary. Hence, we defined the notion of exterior derivative as that which recovers these boundaries, and turns a k-form into a $(k+1)$-form. Finally, we exploited the notion of "closed surface" to simplify the computation of integrals of $(n-1)$-forms, by noticing that counting outbound lines is the same as counting line ends inside the surface. This is the celebrated Stokes Theorem.

The Stokes Theorem is certainly a very important milestone in vector calculus, but it is only the beginning. I cannot hope to describe every phenomenon and definition, but in the remainder of this post we will look at four important concepts: Wedge Products, Pullbacks, Lie Derivatives and Inner Products (aka Contraction aka that thing where you put a vector field into the form). The discussion will culminate in another celebrated formula of tensor calculus: Cartan's Magical Formula

$$L_X(\omega) = \mathrm{d}(X \lrcorner\, \omega) + X \lrcorner\, \mathrm{d}\omega$$

3.6 The Wedge Product

3.6.1 Introduction: Wedges as Symbolic Glue

One of the biggest advantages of tensor calculus is its practicality. With only a small set of rules, it is possible to do all kinds of manipulations and calculations in coordinates, which is part of what makes tensors so useful and applicable in engineering and physics.

The reason behind this practicality is that, as the reader may or may not already know, tensors are decomposible into very elementary parts. Indeed, any tensor can be written using only functions and one-forms, for example something of the sort

$$\omega = f\,\mathrm{d}x \wedge \mathrm{d}y + g\,\mathrm{d}z \wedge \mathrm{d}w.$$

The symbol between the one-forms is called the wedge product, and it is the glue that allows us to compose small-dimensional forms in order to make higher-dimensional forms.

Therefore, when we define some new type of transformation (like, say, an exterior derivative or a Lie derivative), it suffices to find how it behaves on these elementary parts, and how it behaves under the wedge product. This then gives us a recipe for computing that transformation: simply decompose, apply to the elementary parts, and tie it all back again.

[Remark: The reader might (understandably) believe that there is not one type of "symbolic glue" in play, but two: the wedge product, and the product by scalar functions. However, as it turns out, the product by a scalar function actually corresponds to the wedge product by a zero-form; see appendix A for details. *Author's note: This was never written. It is left as an exercise to the reader.*]

3.6.2 Wedges as Intersections

Our current goal is to decompose a k-form into forms of lower dimension. From the perspective of our framework, in which a k-form is represented by $(n-k)$-dimensional surfaces, we wish to write a manifold in terms of manifolds of higher dimension.

This may seem abstract, but it is very easy to motivate if we think in terms of cartesian equations. For example, consider a one-dimensional curve in $\mathbb{R}^3$ such as the red circle in the following figure:

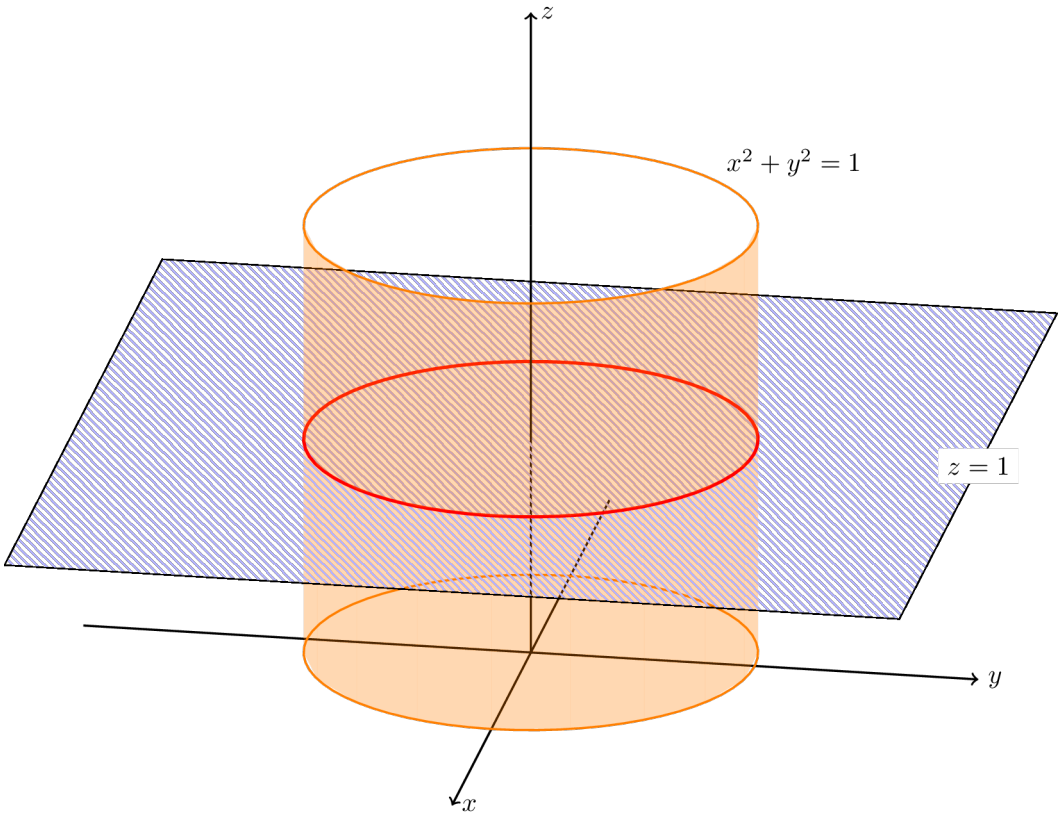


Figure 22: The red circle is the intersection of the orange cylinder and blue plane.

This curve can be written in terms of the cartesian equations

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

Now there is a trivial way to decompose this one-dimensional object as the intersection of two 2-dimensional objects: a horizontal plane $z = 1$ and a cylinder $x^2 + y^2 = 1$.

This is a more or less general phenomenon. We can define a d-dimensional surface in n-dimensional space through cartesian equations. Each of these equations separately defines an $(n-1)$-dimensional surface, and their intersection is obviously the surface we started with. In other words, we get a kind of general principle:

A surface S of dimension d contained in -dimensional space can be written as the intersection of $n - d$ hypersurfaces.

The next step is to apply to this to our framework for forms. A k-form ω is represented through $(n-k)$-surfaces, so we should expect to write the surfaces composing ω as the intersection of k hypersurfaces; that is, to write ω in terms of k one-forms. Sound familiar?

This, in turn, presents a way to "glue forms": take the intersection of the corresponding surfaces. Indeed, the notation of wedge product is even suggestive: ever notice how $\omega \wedge \eta$ kind of looks like $\omega \cap \eta$? I have no idea if this is a coincidence, but if so it is a very fortunate one.

Let ω be a k-form and η be an ℓ-form. Define the wedge product of ω and η as the $(k + \ell)$-form whose atoms are the pairwise intersections of atoms from ω and atoms from η.

For a simple example, let us show that the wedge product $\mathrm{d}x \wedge \mathrm{d}y$ corresponds to the uniform mesh of points we have come to associate with the standard area form on $\mathbb{R}^2$. Indeed, recall that $\mathrm{d}x$ is represented by uniformly spaced vertical lines, and $\mathrm{d}y$ is represented by uniformly spaced horizontal lines. Then, it is visually obvious that the collection of pairwise intersections is a square grid of points.

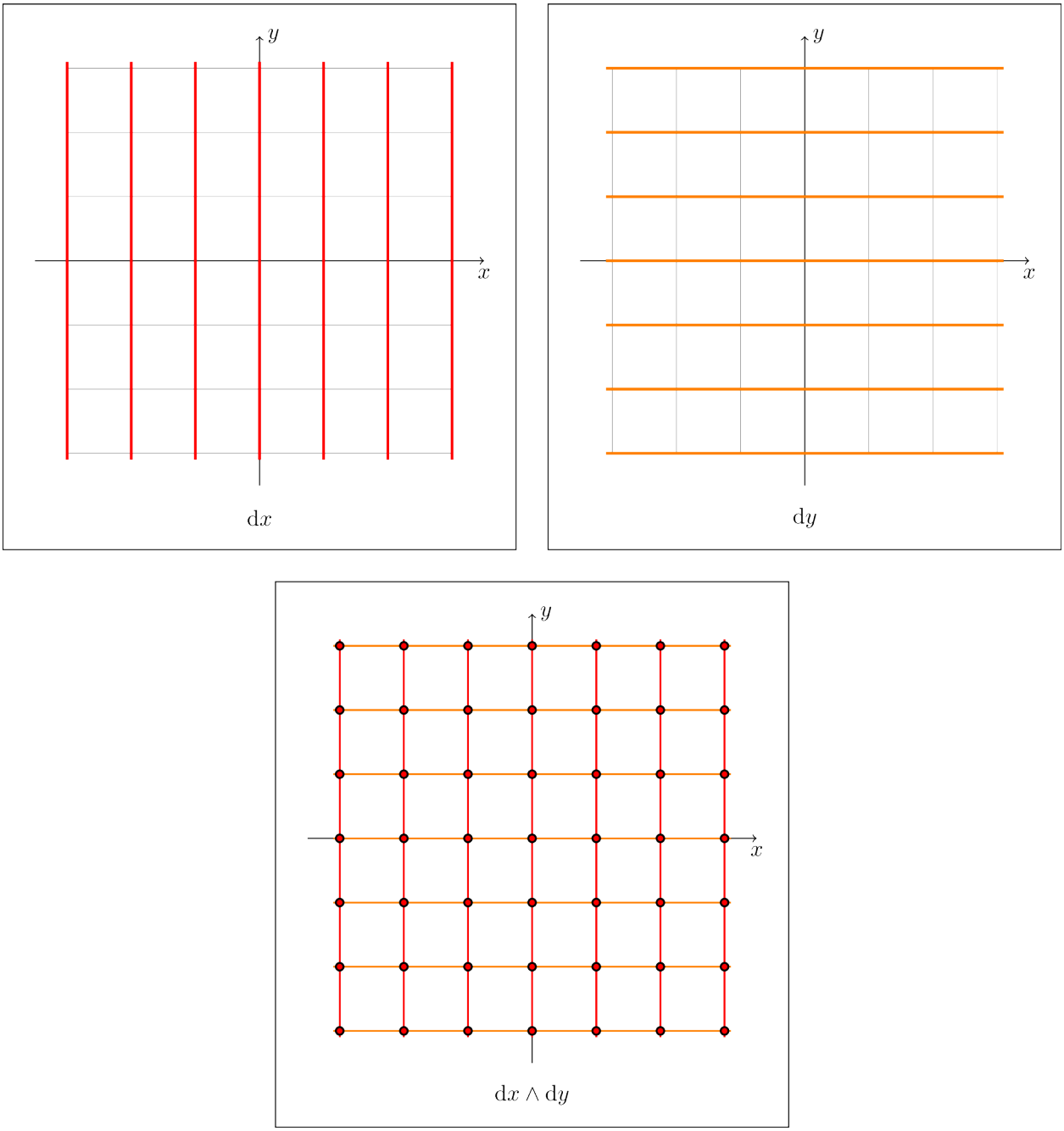


Figure 23: Writing the atoms of as intersections of horizontal and vertical lines.

3.6.3 Dimensions, Orientations and Properties

Appealing though this "definition" of wedge product may be, it remains to show (or at least, plausibly convince) that the wedge product has the properties we have come to expect it to have. For example, is it even obvious that the dimension of the wedge is the sum of the dimensions of the arguments? In this section, we sketch arguments to justify the most important properties of the wedge product.

- Dimension: First, we recall the discussion on transversality above. In particular, recall that we

will always be assuming that all intersections are transverse, which bodes very well for dimensional considerations. In particular, the math checks out: if ω is k-dimensional and η is ℓ-dimensional, they are represented respectively by $(n-k)$ and $(n-\ell)$-surfaces, whose intersection has dimension $(n-k)+(n-\ell)-n = n-(k+\ell)$, which corresponds, as desired, to a $(k+\ell)$-form.

- $\mathrm{d}x \wedge \mathrm{d}x = 0$: The calculation of such a wedge requires a bit of thought. At first glance, it would seem that $\mathrm{d}x \wedge \mathrm{d}x = \mathrm{d}x$, because we are intersecting a collection of hyperplanes with themselves. However, the computation cannot be done in this way, because a hyperplane is not traverse to itself!

  The way around that is actually to work with two distinct representations of $\mathrm{d}x$. Indeed, suppose the first is represented by the hyperplanes $x = 0, \varepsilon, 2\varepsilon, \ldots$ Then, instead of representing the second $\mathrm{d}x$ by the same collection of hyperplanes, represent it through $x = \frac{1}{2}\varepsilon, (1+\frac{1}{2})\varepsilon, (2+\frac{1}{2})\varepsilon, \ldots$ Then, we have two collections of hyperplanes which never intersect. Therefore, their intersections are transverse: indeed, they are empty. (Yes, an empty intersection is a Good Intersection.) In conclusion, $\mathrm{d}x \wedge \mathrm{d}x = 0$, as it ought to.

- $\omega \wedge \eta = (-1)^{k\ell} \eta \wedge \omega$: To talk about this property, we would need to go back to the issue of orientations. However, without getting into it, it is a priori obvious that $\omega \wedge \eta = \pm \eta \wedge \omega$, as the intersection of two surfaces does not depend on the order in which the intersection is taken. However, the sign issues arise upon consideration of the *orientation of the intersection.*

  Unfortunately, the subject of orientation is too complex to be discussed in a mere bullet point, so the term must remain mysterious for the time being.

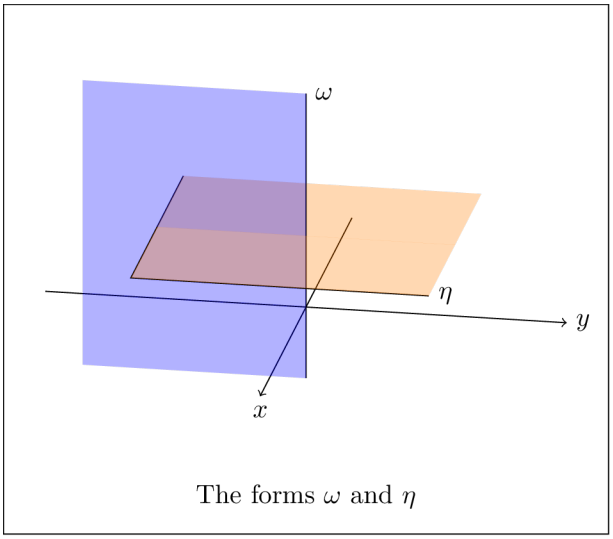
  Author's note: I never wrote down the solution, but I promise it exists. In summary, recall that (as said in a previous author note) what is actually demanded of the "atoms" is a co-orientation, and the issue is: If A and B are co-oriented transversal submanifolds, what is the natural co-orientation on $A \cap B$ vs. $B \cap A$? Indeed, one takes positively oriented bases of the normal budles of A and B, and juxtaposes them to make a (by definition, positively oriented) basis of $A \cap B$ or $B \cap A$, depending on the order. These two definitions don't generally agree: To turn one basis into the other, one needs to perform $\operatorname{codim} A \operatorname{codim} B$ *swaps in the basis, which leads to a difference of orientation of sign* $(-1)^{\operatorname{codim} A \operatorname{codim} B}$*. This is where the $k\ell$ term comes from.*

- Associativity: Like in the previous bullet point, it is obvious that the wedge product would be associative up to sign. To show that the orientations of $\alpha \wedge (\beta \wedge \gamma)$ and $(\alpha \wedge \beta) \wedge \gamma$ match up would require a deeper inspection of the orientation conventions, but it does check out.

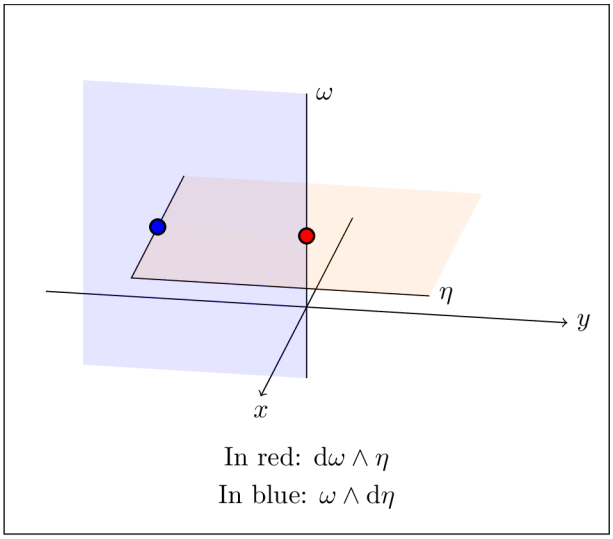
[Note to future self: Write another post discussing orientations in more detail?]

3.6.4 Exterior Derivative of a Wedge Product

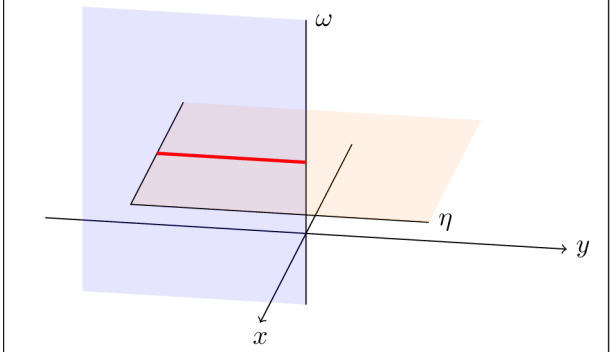
The following picture should speak for itself, modulo orientations.



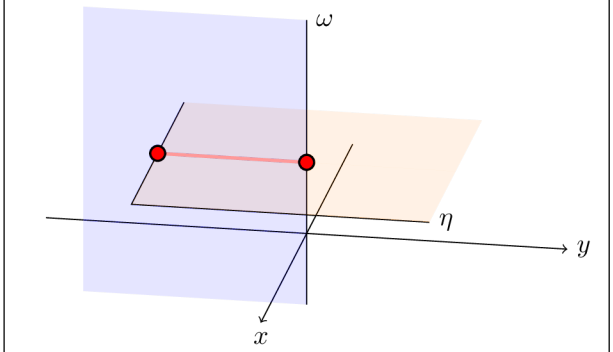
The forms ω and η



In red: $\mathrm{d}\omega \wedge \eta$
In blue: $\omega \wedge \mathrm{d}\eta$



Red line: $\omega \wedge \eta$



The border of the red line, i.e. $\mathrm{d}(\omega \wedge \eta)$

Figure 24: A visual argument for $\mathrm{d}(\omega \wedge \eta) = \pm \mathrm{d}\omega \wedge \eta \pm \omega \wedge \mathrm{d}\eta$.

Again, to figure out the signs one would need a more careful look at orientations, which is beyond the scope of this post, but the formula is given by

$$\mathrm{d}(\omega \wedge \eta) = \mathrm{d}\omega \wedge \eta + (-1)^k \omega \wedge \mathrm{d}\eta, \text{ for } \omega \text{ of dimension } k.$$

3.7 Pullbacks and Lie Derivatives

3.7.1 Pullbacks

Let M and N be manifolds, and $f\colon M \to N$ a smooth function. Then, the pullback of a form ω on N is usually defined as follows:

$$f^*\omega(v_1, \ldots, v_k) := \omega((\mathrm{d}f)v_1, \ldots, (\mathrm{d}f)v_k).$$

Symbolically, this is a very natural operation to define, but it also finds a nice interpretation in our framework.

For simplicity, let us consider a one-form ω on N, though the following argument applies to higher-dimensional forms as well. Recall that we represent ω through a collection of hypersurfaces, and to evaluate $\omega(v)$ on a vector v tangent to N *we count how many of the hypersurfaces the vector v crosses.*

Consequently, to evaluate $f^*\omega$on a vector w tangent to M, what one does is push the vector w forwards and counts how many surfaces of ω it crosses. However, a bit of mental gymnastics should convince you of the following alternative definition: *To evaluate $f^*\omega(w)$, pull the surfaces composing ω back into M (taking their preimage in f), and count how many of those w crosses.* In other words, we can represent $f^*\omega$ using the preimages of the surfaces that represent ω.

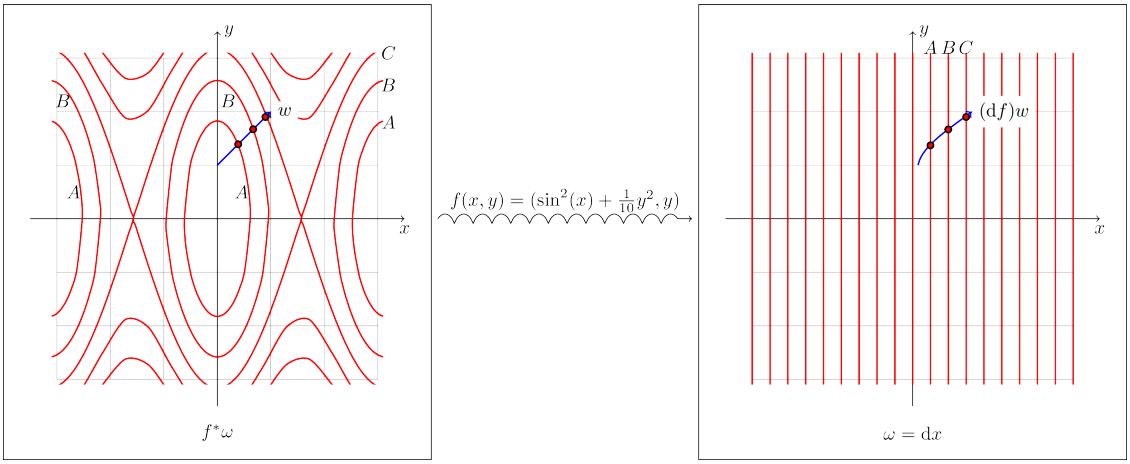


Figure 25: Left side: the manifold $N = \mathbb{R}^2$, with the form $f^*\omega$ in red and a vector w in blue. Right side: the manifold $M = \mathbb{R}^2$, with the form $\omega = \mathrm{d}x$ in red and the vector $(\mathrm{d}f)w$ in blue. Crossings between the vector and the form are highlighted as to show the equivalence in modes of computation. Some artistic liberties were taken: the atoms were drawn more densely than usual to aid in visualization, and the vector $(\mathrm{d}f)w$ is drawn slightly distorted. Indeed, what is drawn is the pointwise image of the curve which goes in a straight line from $(0, 1)$ to $(1, 2)$.

Like many others so far, this definition comes with a few of asterisks, because it is not obvious that taking the preimage under f of the atoms of ω will yield a collection of atoms of the right dimension. Less obvious yet is that preimages behave well under borders, i.e. that $\partial f^{-1}(A) = f^{-1}(\partial A)$, where A is one of the atoms of ω and ∂ is the border operator. I myself do not understand the underlying mechanics particularly well, though this post will be updated if this unfortunate state of affairs happens to change. *Author's note: As of 2025, this state of affairs has not yet changed, but in my defense I went on to do PhD in computability theory. I'm still interested in this, however, so if you have thoughts on this subject please feel free to contact me.*

However, shady though it may seem, it appears to be the case that in general the border of the preimage coincides with the preimage of the border, and so we reach the classical fact:

$$\mathrm{d}(f^*\omega) = f^*(\mathrm{d}\omega).$$

3.8 Lie Derivatives

First, let us recall the usual definition of Lie derivative.

Let ω be a form on M, and X a vector field. Let ϕ_t be the time-t flow of X. Of course, there are domain issues when considering flows, but we will sweep it all under the rug.

We define the Lie Derivative of ω in the direction of X by the formula

$$L_X\omega = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_0 \phi_t^*\omega.$$

In order to inspect this definition, let us begin by considering the pullback of a form by a flow. As we have seen, the pullback of a form can be visualized by taking the preimage of the form's atoms, so we can visualize $\phi_t^*\omega$ by "flowing the atoms of ω backwards with X".

This kind of moving visualization is harder to make drawings of, so I ask the reader to use their imagination to visualize the following example. Consider the form $\omega = \mathrm{d}x \wedge \mathrm{d}y$ in $\mathbb{R}^2$, visualized as a uniform grid of points. Let be X the radial vector field $X_{(x,y)} = x\partial_x + y\partial_y$.

We will visualize $\phi_t^*\omega$ as t varies by visualizing the motion of the atoms. Of course, for $t = 0$ there has been no motion, so we begin by seeing a uniform grid of points in $\mathbb{R}^2$.

Then, we begin flowing for positive time. The vector field in question points outwards from the origin, so since we are taking the preimage under the flow we move the atoms inwards. By solving the ODE, it is easy to check that an atom starting at (x, y) will, after time t, be at the point $\mathrm{e}^{-t}(x, y)$. Therefore, we conclude that our uniform mesh of points is being pulled towards the origin in a way that preserves uniformity: after time t, distances between points have been uniformly compressed by a factor of e^{-t}.

This even allows us to compute explicitly $\phi_t^*\omega$. Indeed, a uniform mesh of points represents a form of the type $C\,\mathrm{d}x \wedge \mathrm{d}y$, so all that remains is to find C. A heuristic argument is as follows: since in every direction, points are denser by a factor of e^t, the amount of points in a given parallelogram has been increased by $(\mathrm{e}^t)^2$. In other words, we claim that $\phi_t^*\omega = \mathrm{e}^{2t}\,\mathrm{d}x \wedge \mathrm{d}y$.

A slightly less heuristic argument is to start with the unit square at time $t = 0$, and flow it *backwards* for time t. This will result in a smaller square, of side e^{-t}, which contains exactly the same atoms, whence we conclude

$$\phi_t^*\omega(\mathrm{e}^{-t}\partial_x, \mathrm{e}^{-t}\partial_y) = \omega(\partial_x, \partial_y) = 1.$$

A few algebraic manipulations will yield again the result $C = \mathrm{e}^{2t}$.

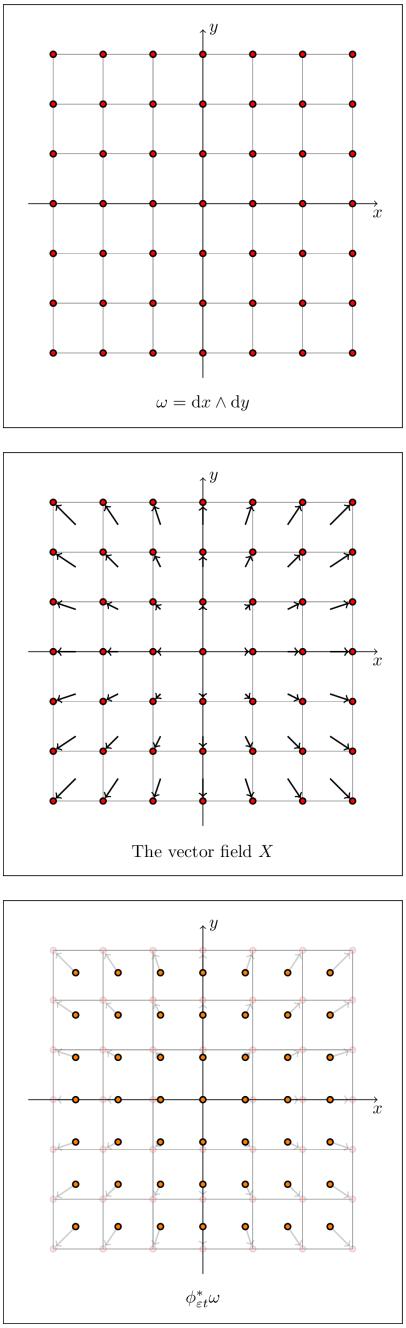


Figure 26: The construction of $\phi^*_{\varepsilon t}(\mathrm{d}x \wedge \mathrm{d}y)$ for the radial vector field, for some small value of ε. The Lie derivative will measure the difference between the red form and the orange form.

Now that we are familiar with the pullback of a form under a flow, all that remains is to differentiate. Unfortunately, the only visualization I am able to provide is very lackluster, and I encourage any reader who has a better idea to contact me.

Let us visualize differentiation using infinitesimals. In what follows, let ε be a very small positive number. Then, $\frac{\mathrm{d}}{\mathrm{d}t}\big|_0 \phi_t^*\omega = \frac{1}{\varepsilon}(\phi_\varepsilon^*\omega - \omega)$. In other words, we begin by taking ω, flowing it backwards through X for infinitesimal time, and subtracting ω again. Now, this is zero to zeroth order, so we divide by ε in order to

get a real-valued thing.

Let's elaborate a bit on the flow part for a moment. To first order, taking the flow in infinitesimal time is just calculating the velocity vector at the point and moving in the direction of the tip. In other words, (visualizing vectors as infinitesimal arrows) we may visualize $\phi_\varepsilon^*\omega$ by, for each point p of each atom A, evaluating X_p at that point and placing the arrow so that the *tip* is at p. Then, we replace p by the tail of the arrow X_p. Doing this process for each point p, we obtain a new surface B, which is the atom of $\phi_\varepsilon^*\omega$ which corresponds to the atom A in ω. Repeating this process for every atom A we obtain the atoms of $\phi_\varepsilon^*\omega$.

See Figure 26 for the example with the volume form ω on $\mathbb{R}^2$, and the radial field. Given a red atom at (x, y), the corresponding orange atom is given by $(x, y) - \varepsilon X_{(x,y)}$, that is,

$$(x, y) - \varepsilon X_{(x,y)} = (1 - \varepsilon)(x, y).$$

Therefore, we have a grid of uniformly spaced atoms, with a unit area square with sides $(1 - \varepsilon)\partial_x$ and $(1 - \varepsilon)\partial_y$. Therefore, an argument similar to above will show that the orange atoms represent the form

$$\phi_\varepsilon^*\omega = (1 - \varepsilon)^{-2}\,\mathrm{d}x \wedge \mathrm{d}y$$

and therefore we may calculate the Lie derivative of ω in the direction of X:

$$L_X\omega = \frac{1}{\varepsilon}(\phi_\varepsilon^*\omega - \omega) = \frac{1}{\varepsilon}((1 - \varepsilon)^{-2} - 1)\,\mathrm{d}x \wedge \mathrm{d}y,$$

and a bit of elementary calculus shows that, taking the limit $\varepsilon \to 0$,

$$L_X\omega = \frac{(1 - \varepsilon)^{-2} - 1}{\varepsilon}\,\mathrm{d}x \wedge \mathrm{d}y = 2\,\mathrm{d}x \wedge \mathrm{d}y.$$

3.9 Contraction/Inner Product

At first, it seems difficult to visualize the contraction, because it is explicitly defined in terms of evaluation. In case you've forgotten or have never heard of it, given a k-form ω and a vector field X, we define the contraction $X \lrcorner\, \omega$ as

$$(X \lrcorner\, \omega)(v_1, \ldots, v_{k-1}) := \omega(X, v_1, \ldots, v_{k-1}).$$

To see an easy example of how such a form can be drawn, consider the simple example $M = \mathbb{R}^2$, $X = \partial_x$, and $\omega = \mathrm{d}x \wedge \mathrm{d}y$.

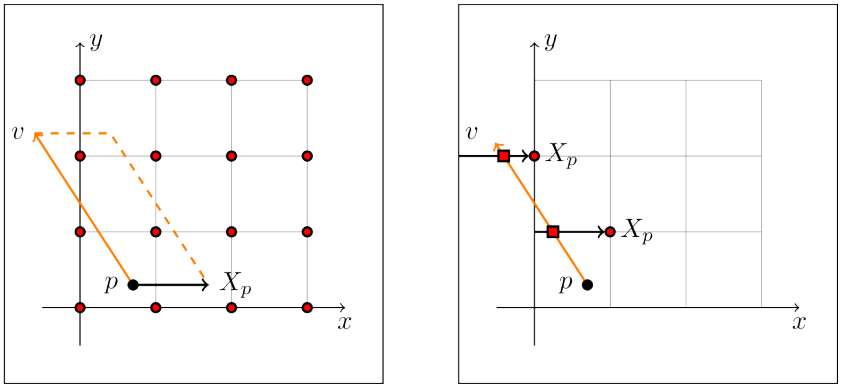


Figure 27: A visualization of calculating $(X \lrcorner\, \omega)(v)$ for some v. On the left is the trivial calculation, in which you draw the parallelogram and count how many atoms are inside it. On the right is a different perspective, in which we put a copy of X on each node and count how many of these copies of X the vector v crosses.

The idea, as seen in the image above, is to count the amount of atoms inside the parallelogram spanned by v and X by counting *intersections*. Indeed, roughly speaking, a point q is inside the parallelogram spanned by v and X if and only if there exist $\alpha, \beta \in [0, 1]$ such that

$$q = p + \alpha v + \beta X.$$

Rearranging terms, one concludes that this happens if and only if there exist such α and β such that

$$p + \alpha v = q - \beta x,$$

that is, the line segment that goes from p to $(p + v)$ intersects the line segment that goes from $(q - X)$ to q.

A simple calculation will yield a similar result for higher dimensions and for forms with non-point atoms. The general rule is as follows:

To represent $X \lrcorner \omega$ as a form, we construct its atoms out of the atoms of ω as follows. If A is an atom of ω, i.e. an $(n-k)$-manifold, at each of its points a we calculate X_a, and draw this vector with the arrow tip on a. We then take the union of all the points on all of these vectors, yielding an $(n-k+1)$-manifold, which we call $X \lrcorner A$. The atoms of $X \lrcorner \omega$ are all those of the form $X \lrcorner A$.

Alternate definition of $X \lrcorner A$: The points of the form $a - tX_a$, for $a \in A$ and $t \in [0, 1]$.

Let's look at the example above again, and calculate $X \lrcorner \omega$. This case is particularly easy, but we will see that sometimes the visualization of $X \lrcorner \omega$ requires some mental gymnastics.

In this case, we begin by drawing a uniform grid of points, and drawing X at each point. Then, we just consider the collection of resulting line segments.

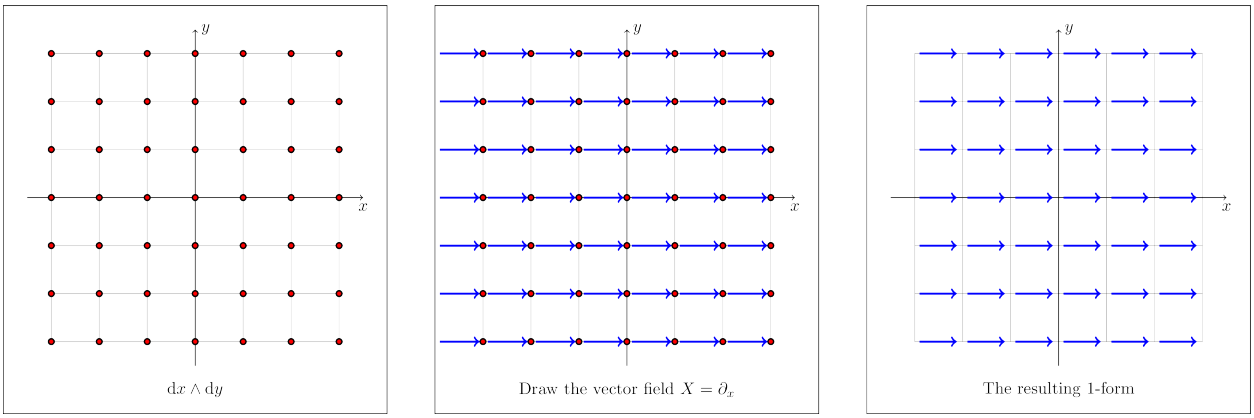


Figure 28: The construction of the 1-form given by $\partial_x \lrcorner (\mathrm{d}x \wedge \mathrm{d}y)$.

As you can see, this is a collection of line segments which are organized in rows, where each row is composed of infinitely many (length one) line segments. However, the way that the form is evaluated on a vector would not change if we glued these line segments together. So we glue them, and lo and behold, we get a uniform collection of horizontal lines, also known as $\mathrm{d}y$, so we conclude

$$\partial_x \lrcorner (\mathrm{d}x \wedge \mathrm{d}y) = \mathrm{d}y.$$

3.10 Cartan's Magical Formulas

We are now ready to look at Cartan's magic formula for the Lie derivative, as we have already seen the three necessary ingredients: the Lie derivative, contraction, and the exterior derivative.

Let us begin by noting the relation between these three concepts. Given a vector field X and a form ω, both the Lie derivative and the contraction are related to the notion of "gluing the tip of X into the atoms of ω". Indeed, if ω is a top form (i.e. its atoms are points) a useful identity appears:

$$L_X\omega = \mathrm{d}(X \lrcorner\, \omega).$$

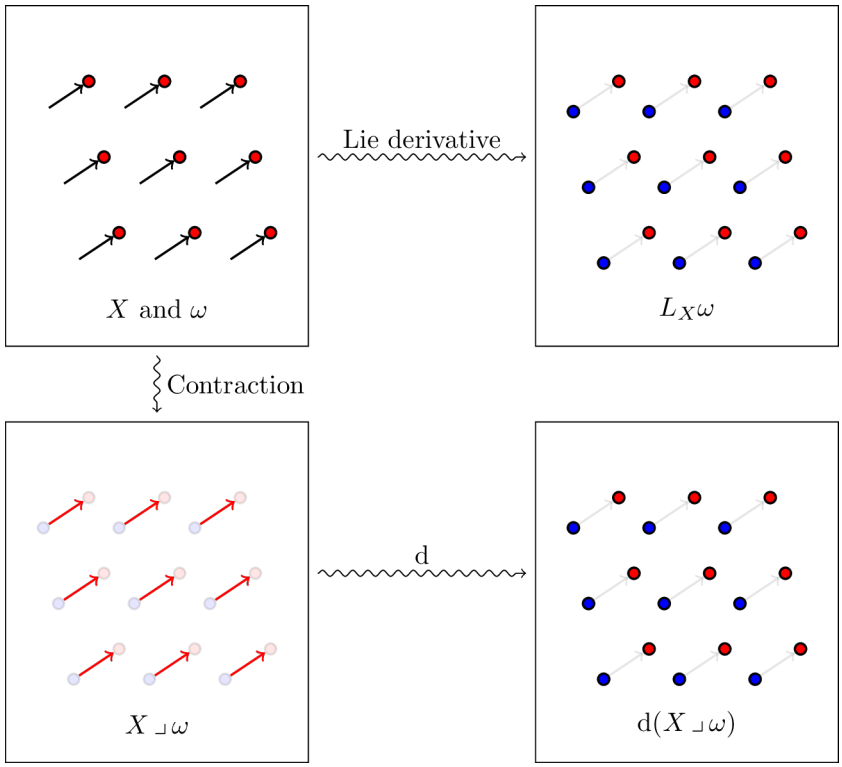


Figure 29: A visual proof that, for a top-level form ω, $L_X\omega = \mathrm{d}(X \lrcorner\, \omega)$.

This identity is the first version of Cartan's magic formula, and it actually works whenever ω is a closed form. However, when ω has atoms with nonempty border, a new term appears. To visualize why, suppose that ω is not a closed form, and has an atom A with border ∂A. Assume for the sake of terminology that A is one-dimensional, so that we may call it a curve with one or two extrema. Then, $X \lrcorner\, A$ is a strip. One of its sides is exactly A. Let us call the opposite side B, and the one or two other sides will be called C. Then, on the one hand, the atom corresponding to $L_X\omega$ is simply $A + B$. On the other hand, $\mathrm{d}(X \lrcorner\, \omega)$ includes A and B, but also C, so we need to remove that term. A little bit of thought shows that C is an atom from $X \lrcorner\, \mathrm{d}\omega$, and so, modulo signs, we reach the famous formula

$$L_X\omega = \mathrm{d}(X \lrcorner\, \omega) + X \lrcorner\, \mathrm{d}\omega.$$

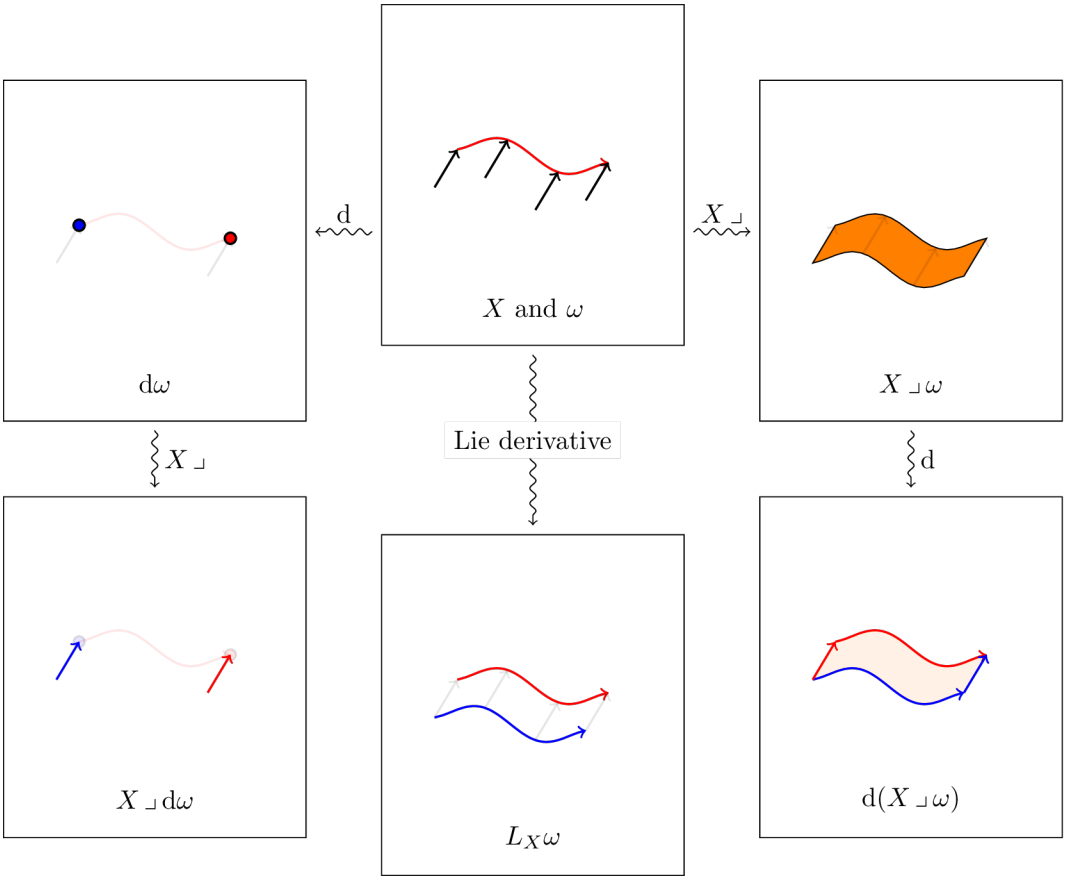


Figure 30: A visual proof of Cartan's magic formula. Note that, while orientations are usually opaque, they are reasonably understandable here. The most difficult orientations here are those of $\mathrm{d}(X \lrcorner \omega)$. To understand them, simply note that the "positive direction" on the border is the one induced by ω, and any arrow in that direction counts as positive, and in the opposite direction counts as negative.