On Tangent Spaces, Differential Forms, and Integration

Shrivathsa Pandelu

September 27, 2020

Contents

1 Prerequisites and the problem

In what follows it is assumed that the reader is familiar with basic Topology and Analysis. The purpose of this article is to seek a deeper meaning of what exactly the tangent space, differential forms are and why it is that it works. It is to take apart the framework of modern day differential geometry and see what makes it tick. First, I'll try to answer what a smooth function is. From there, we'll move to vector fields, the Lie Bracket of two vector fields and so on in a manner which I think is quite a natural progression of ideas.

Most of the text is non rigourous and there is quite a bit of hand waving. The purpose of these notes is not to derive things from the first principles, but rather to discuss why some definitions are natural. I have taken the liberty of intriducing my own notations, most of these are perhaps non standard.

2 Smooth functions

2.1 Directional derivative

Once we have a topology on a set, we know how to talk about continuous functions for that is how topology is defined. However, continuous functions have a lot of freedom and are not "nice" enough. They can have sharp corners, pinches etc. To talk of smoothness, or continuity in general, at a point, we need our functions to be defined on open neighbourhood of the point. So, let us consider an open set $U \subseteq \mathbb{R}^n$, and a point $p \in U$, and a function $f \colon U \to \mathbb{R}^m$.

As a first case, assume $n = 1$, and U is an open interval. A measure of smooth change is to look at the rate of change, i.e., how fast is f changing near a point, which is given by

$$\lim_{t \to 0} \frac{f(p+t) - f(p)}{t}.$$

When this limit exists and is finite, we say that f is differentiable at p. Notice that the limit makes sense only when f is defined on a neighbourhood of p. This is the usual derivative at p.

If we take an arbitrary n, then we can talk of the rate of change along any direction. Of course, we need to give a meaning to "direction". Let us say that the open set is embedded in some way in $\mathbb{R}^n$. Now, this $\mathbb{R}^n$ is a vector space, and here we call each vector a direction.

Given a vector v, we can talk of the limit

$$\lim_{t \to 0+} \frac{f(p+tv) - f(p)}{t}.$$

I take a one sided limit because I want to approach p, along the direction of v, not opposite to v. Notice that if we replace v by some cv for a positive c, then the limit exists and is c times the original limit. So, if the limit exists for one v, then it exists for the ray determined by v. Also, note that even if we have a bounded U, we can still talk of $f(p+tv)$ when t is very small, because U is open. Next, we obviously want f to have a finite rate of change along every direction, and if we assume this, then f gives rise to a function, which we call df_p, from $\mathbb{R}^n$ which is the space of all directions, to $\mathbb{R}^m$.

Now, I demand that df_p be a linear map. Think of a line passing through p, if df_p doesn't send this to another line, then it breaks into two rays at an angle different from $180°$. One can imagine that f is in some sense pinching U at p, and twisting it about p. We don't want a smooth function to do this, so it is reasonable to demand that it send lines to lines, basically that $df_p(cv) = cdf_p(v)$ not just for $c > 0$, but also $c < 0$(when $c = 0$, the equality trivially holds). Similarly, we want it to be additive, just to avoid "pinch points". Now, obviously this is not a very rigourous way to do things, but this is only a means for intuition. Think of lifting a cloth by pinching it at a point, the cloth then describes a function over the ground below, but at the point where your fingers hold the cloth, it is severly pinched, twisted and deformed. Those are the points that we want to avoid, because there are rapid changes occuring near that point.

So, we demand that df_p be linear. As an aside, look at the function on $\mathbb{R}^3 \to \mathbb{R}$ that sends $(x, y, z) \mapsto (xyz)^{1/3}$. This function scales properly but is not additive.

So, given f, we have a linear map $df_p \colon \mathbb{R}^n \to \mathbb{R}^m$, that assigns to each direction another vector. This map is very much dependent on p, for it takes a vector and does a specific operation on f. We can look at this operation in its own right. Fix a vector/direction v, and let denote its directional derivative by $d_v|_p$. We will come back to this operation a little later.

What we have is a set of directions $\mathbb{R}^n$, and for each point $p \in U$, the function f assigns a linear map $df_p \colon \mathbb{R}^n \to \mathbb{R}^m$. Fixing bases, df_p is a matrix, and gives rise to a map $df \colon U \to R^{m \times n}$ provided we assume that f is differentiable on U. We can again ask if this is differentiable. This way, we have the class of continuous, differentiable, $\mathcal{C}^1, \mathcal{C}^2, \ldots, \mathcal{C}^\infty$ functions. A function is smooth if all orders of derivatives exist. Note that even though the entries of the matrix depend on the basis, a change in basis doesn't change the fact that the entries are smooth, so the notion is well defined.

2.2 Tangent space

Take $U \subseteq \mathbb{R}^n$, and $p \in U$. For each vector in $\mathbb{R}^n$, and a function f on U, we have the limiting process that gives us a directional derivative. In this limiting process, since we need $p + tv$, it is essential that U is a subset of the vector space $\mathbb{R}^n$, and the topological space $\mathbb{R}^n$, that is, the limiting process makes sense when we put both structures on the same $\mathbb{R}^n$.

In the limiting process, we are only interested in the values f takes near p. I guess that the philosophy is that the rate of change only depends on the local behaviour. It doesn't matter what values f takes far from p, what matters is how f behaves as we get close to p. Thus, if we restrict f to a smaller neighbourhood of p, we will still get the same limit. Similarly, if another function g agrees with f near p, i.e., in a neighbourhood of p, then the limit for f, g agree. By design of the limiting process, it is a local operator. For a vector v, as in the previous subsection, we have the operator $d_v = d_v|_p$. To this operator, two functions defined near p are equal when they agree near p.

Write $C_p(U) = \{(V, f) | V \text{ open in } U, f \text{ differentiable } \}/\sim$ where $\sim$ is the following realtion:

$$(V_1, f_1) \sim (V_2, f_2) \text{ iff } \exists\, W \overset{\text{open}}{\subseteq} V_1 \cap V_2 \text{ such that } f_1|_W = f_2|_W.$$

It is easy to see that $\sim$ is an equivalence relation and therefore we can pass to the quotient. The resulting space $C_p(U)$ is called the space of germs of differentiable functions at p. It is common to take smooth, i.e., $\mathcal{C}^\infty$ functions, but for now, we will stick to just differentiable functions.

So, d_v acts on this set, and is a map $d_v \colon C_p(U) \to \mathbb{R}$.

Now, if we fix an open set V, then given two functions we can talk of their sum and products. These operations are carried to $C_p(U)$ with the difference that functions need not be defined on the same open set. Had we not passed to the quotient, then to add or multiply, we need our functions to be defined on the same open set. However, since we are focussing on neighbourhoods of p, we can restrict functions to smaller neighbourhoods and add/multiply them.

One can check that these operations are well defined and give a ring structure to $C_p(U)$. Of course, the functions don't need to be differentiable, the quotient is defined for any nice class of functions. For example, we could have just plain functions, or continuous functions etc. But the point is that $C_p(U)$ has a ring structure which is ultimately borrowed from that on $\mathbb{R}$. Actually, it is an $\mathbb{R}$-algebra, where $\mathbb{R}$ is embedded in $C_p(U)$ as constant functions.

Now, $d_v \colon C_p(U) \to \mathbb{R}$ is a map, which is $\mathbb{R}$-linear, but not multiplicative. In fact, d_v adheres to a Leibnitz rule, and we have $d_v(fg) = d_v(f)g(p) + f(p)d_v(g)$. Such a map from an $\mathbb{R}$-algebra to $\mathbb{R}$ that is $\mathbb{R}$-linear and satisfies Leibnitz rule is called a derivation.

Thus, to each $v \in \mathbb{R}^n$, we have a derivation on $C_p(U)$. The space of derivations at p is the set of all derivations $D \colon C_p(U) \to \mathbb{R}$. One can check that this latter space is an $\mathbb{R}$-vector space with the obvious operations, which we temporarily denote by $Der(p)$. By design, we said that f is differentiable if the map df_p is linear. With this in mind observe that the map

$$\begin{aligned} \mathbb{R}^n &\to Der(p) \\ v &\mapsto d_v \end{aligned}$$

is a vector space homomorphism.

Any text on differential geometry will give a proof that this map is in fact an isomorphism in the smooth category.

The space of derivations at p is called the tangent space. Intuitively, it is attaching a direction at each point via the map above.

2.3 A small note

Here is the flow of ideas discussed above.

- Embed U in $\mathbb{R}^n$, and fix a basis for this $\mathbb{R}^n$.
- For $p \in U$, each vector defines a limit on any function f defined near p. When this limit exists, f is said to have the corresponding directional derivative.
- We look at functions that have a directional derivative along every vector of $\mathbb{R}^n$, and that this map is linear. Then, f is said to be differentiable at p.
- Construct the space of differentiable germs at p, and the space of derivations at p. Use the isomorphism above.

We can talk of germs of differentiable functions, and the tangent space only after we have defined what differentiability is. We defined smoothness inductively in terms of the smoothness of the entries of the first derivative matrix. This matrix requires a choice of bases, which comes from the $\mathbb{R}^n$ in which U is embedded and not from individual tangent spaces, for tangent spaces in the smooth category is defined AFTER defining smooth functions. Tangent vectors at a point are defined to be derivations on the germs of smooth functions.

This is a subtle point that can be completely ignored.

3 Setting up determinants and forms

We will digress a little to introduce the concept of alternating linear forms. An important part of geometry is measurement. It may be measurement of angles, or lengths or volumes.

A first formal approach to such meausrements is the Riemann integral. Suppose we have a rectangle R and a real valued function f on this rectangle. If we want to compute the area under the function, i.e, the volume of the object $\{R \times tf(R) | t \in [0,1]\}$, then what one does in Riemannian integration is to divide the rectangle R into smaller rectangles, approximate f on these smaller rectangles by some constant (maximum, minimum or an arbitrary value taken by f) to get a cube, add up the volumes of the cubes and take a limit as the base rectangles become vanishingly small.

Of course, when I say rectangle, I mean a set of the type $[a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$, and by cube I mean a rectangle of one higher, i.e. $n+1$, dimension in this specific context. This limiting process was used by Archimedes, then by Cavalieri as the method of indivisibles, and then by Newton while developing integral calculus.

The question is, how do we measure the volume of these rectangles, or more generally parallelepipdes? A k-parallelepiped in $\mathbb{R}^n$ spanned by linearly independent vectors $v_1, \ldots, v_k$ in $\mathbb{R}^n$ is the set $\{a_1 v_1 + \cdots + a_k v_k | 0 \leq a_i \leq 1\}$. If they are not linearly independent, then the span is of some lower dimension. Observe that the parallelepiped so defined always has the origin as one vertex. This is fine because of the requirement that volume/area is preserved under translations. Formally we are trying to find a way to assign areas to subsets of $\mathbb{R}^n$ that are preserved under translations, i.e. a Haar measure (introduced by Alfred Haar in 1933) on $\mathbb{R}^n$, although at the time we are focused on parallelepipeds and other similar "regular" sets (these are examples of Borel sets).

So, we have a k parallelepiped spanned by $v_1, \ldots, v_k$ and we want to assign to it a real number denoting its volume. When $k = 1$, this volume would be the length of the vector. Therefore, if at all we are going to assign volumes to subsets of $\mathbb{R}^n$, we need a notion of length of a vector. However, simply assigning a length to each vector is not sufficient, we would like to restrict all possible lengths. There are some natural restrictions, one of which is invariance under translation (of the parallelepiped)

mentioned earlier and another is that the 0 vector has 0 length. We may also like to have rotational invariance (more on this later).

Here is another property that we may want our length function to satisfy. Suppose $L\colon \mathbb{R}^n \to \mathbb{R}$ is our length function. Now, if $v \in \mathbb{R}^n$ has length l, then it is quite natural that $2v = v + v$ has length $2l$. However, here we have two ways to proceed. We may require that length is always positive, so that $L(\alpha v) = |\alpha| L(v), \alpha \in \mathbb{R}$ or we may insist $L(\alpha v) = \alpha L(v)$ so that the algebra is neater. Both of these approaches are valid, the first is more measure theoretic while the second more algebraic.

Definition. *Let V be a vector space over a field K which is either $\mathbb{R}$ or $\mathbb{C}$, a norm on V is a map $|\cdot|\colon V \to K$ such that*

- $|v| \geq 0 \, \forall \, v \in V$ (*non negativity*)
- $|\alpha v| = |\alpha||v| \, \forall \, \alpha \in K, v \in V$ (*absolutely scalable*)
- $|v + w| \leq |v| + |w|$ (*triangle inequality*)
- $|v| = 0 \Leftrightarrow v = 0$ (*positive definite*)

A norm defines a metric d on V by $d(x, y) = |x - y|$ and positive definiteness makes V a Hausdorff space. The norm is a length function and note that if $|\cdot|$ is a norm, then so is $c|\cdot|, c > 0$.

For the moment, we let go of mathematical rigour and appeal to intuition. Suppose we know what rotation and rotational invariance means and take $V = \mathbb{R}^n, K = \mathbb{R}$. If we require a norm to be rotationally invariant, then it is uniquely defined by its value on any non zero vector because every other vector is obtained by a rotation followed by scaling. Therefore, the norm on $\mathbb{R}^n$ is uniquely determined upto a multiplicative constant. Since on $\mathbb{R}^n$ the Pythagorean norm is already rotationally invariant (again appealing to intuition because the Pythagorean norm defines spheres very nicely), it is the norm (upto a constant) on $\mathbb{R}^n$. However, there are norms that are not rotationally invariant - the p norms for example.

3.1 Angles and lengths

Let us look more closely at what we mean by lengths and angles. For convenience, we will restrict ourselves to the plane and what we discuss next easily applies to higher dimensions because angles and lines only need to dimensions to work (or so our intuition says). If we go to Euclid, then there are discussions about length and angles, but these concepts are not defined from scratch and Euclid appeals to our intuition. In the *Elements*, a line always has length (incidentally, to Euclid a line need not be straight); he defines a line as "length without breadth". Similarly, the concept of angle is not defined, instead he says "a plane angle is an inclination of the lines to one another, when the two lines in a plane meet one another, and are not lying in a straight-line" [2]. Although Euclid doesn't exactly define what length, the number, is, he doesn't need it as long as one can talk about such a property as length of a line. Throughout the *Elements* lengths are only compared or manipulated in an abstract sense without concrete numbers.

Furthermore, Euclid allows one to move figures in the plane - for example, in proving the $S - A - S$ congruence of triangles, he "applies" one triangle to another and argues that they must coincide perfectly, and therefore must be congruent. How is one allowed to move figures in the plane? Today, we would say that there is a group of transformations of the plane and that there is one transformation which takes the first triangle to the second (this is the *Erlangen program*[3]), but if we apply this reasoning to Euclid's treatment, then we are supposed to assume the existence of an infinite group of transformations. Incidentally, Hilbert takes the$S - A - S$ congruence as an axiom (see [4]), which is far fewer axioms than the existence of rigid motions.

However, even in the *Elements*, we can measure, add, subtract multiply and divide lengths assuming we have a unit length. If we were to fix an arbitrary segment as the unit length, then using a straightedge and compass we can create an ordered field (the constructible numbers). We can also create circles and a family of regular polygons of given sizes (measured in proportion of the fixed unit length). This is sufficient for most purposes. Also, given a line segment, having fixed the unit

length, we may define its length as the number of copies of the unit length required to "tile" the given segment. When talking about areas too, Euclid uses this method of tiling. However, since we cannot always tile two line segments using a fixed shorter segment, pairs of lines (similarly areas) were commensurable or incommensurable.

To measure angles between two unit segments having one common end, we may draw a circle and see what proportion of the circumference is contained between the two segments, or more directly what length of the circumference is contained between them. Note that the moment we have the idea of length of straight lines, we have the idea or notion of length for arbitrary curves (recall that for Euclid every 1- dimensional curve was a line). Once we are able to determine the angle between two unit-length segments, we can determine the angle between arbitrary segments (having one common end) because we would expect angle to be scale invariant. However, this needs us to assign lengths to circular arcs which requires a stronger theory (or a good rope to wind around the circle as in the ancient times).

Despite the difficulty in calculating exact angles, one can still talk meaningfully about angles (having defined them first in some way). For example, we can talk about angles in a triangle or polygon. One stricking feature of Euclid is that he doesn't speak of thos objects which he cannot construct with a straightedge and compass ([4]), so "irregular" angles are out of the picture. For example he defines a right angle as that angle made by a line standing on a striaght line such that the adjacent angles are equal (presumable can be overlapped perfectly), in this case the first line is perpendicular and the angles are called right angles.

Regarding areas too, we don't find convincing (according to modern standards) definitions. However, we can define the area, once we have lengths, of a rectangle as the product of the lengths of its two sides, where a rectangle is a figure with 4 right angles (and we have to prove that the opposite sides have same lengths). More generally, parallelograms have area *base*$\times$*height* and this is proved using the dissection-rearrangement method. This further generalizes to parallelepipeds in 3 dimensions, where this time we rearrange in 3 D.

Incidentally, we have the following intuitive observation : suppose we have two polygonal figures P, Q in the plane and that P can be dissected and rearraged into Q, then they have the same area. In the plane, the converse is also true and this is the Wallace-Bolyai-Gerwien theorem [5]. However, the converse is not true in 3 dimensions and this is the substance of Hilbert's third problem answered by Max Dehn using the Dehn invariant [6].

3.2 Sine and Cosine formulas

Having defined angles, one can talk of the usual trigonometric ratios. Here, we consider the unit circle, and draw a fixed radius with respect to which we measure all our angles in the clockwise sense, say. Then any angle corrsponds to a specific radius. Althought we cannot measure the angle exactly, for we cannot measure the arc lengths exactly, we can still measure straight lengths. Thus, to every angle we assign certain trigonometric ratios - $\sin, \cos, \tan$ etc depending on the lengths of perpendiculars and bases when we draw them in the unit circle. The advantage is that these numbers, which can be calculated (unless we run into the problem of measuring irrational lengths, which happens for most angles), determine the angle uniquely. However, we should still ensure that the ratios are independent of the triangle, i.e. if instead of taking the triangle from the unit circle, we take an arbitrary right triangle with the given angle.

With the Pythagoras theorem, the trigonometric ratios satisfy some relations among themselves, for example $\sin^2 + \cos^2 = 1$. Similarly, with the Pythagoras theorem, we have the sine and cosine rule:

Given a triangle with sides lengths a, b, c and corrsponding opposite angles α, β, γ, we have

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma); a^2 = b^2 + c^2 - 2bc\cos(\alpha); b^2 = a^2 + c^2 - 2ac\cos(\beta)$$

and

$$\frac{a}{\sin(\alpha)} = \frac{b}{\sin(\beta)} = \frac{c}{\sin(\gamma)}.$$

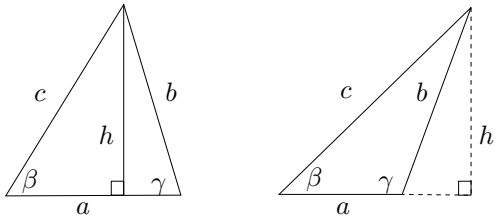

Figure 1: Sine and Cosine rules

To prove them, observe that in the figure above, the height h is given both by $b\sin(\gamma)$ and $c\sin(\beta)$. Using other altidutes, we have the sine rule. Similarly, in both cases, the base length is given by $a - b\cos(\gamma)$ taking care that cosine of an obtuse angle is negative. Using the Pythagoras theorem, we have the cosine rule :

$$c^2 = h^2 + (a - b\cos(\gamma))^2 = a^2 + b^2 - 2ab\cos(\gamma)$$

where we have used the identity $\sin^2 + \cos^2 = 1$.

The Cartesian plane is a model of the Euclidean plane, it is constructed with the Euclidean geometry in mind. We may define $\mathbb{R}^n$ as all n-tuples of real numbers, and assign the norm $\sqrt{x_1^2 + \cdots + x_n^2}$ to the point $(x_1, \ldots, x_n)$ keeping the Pythagoras theorem in mind. With this norm, in a sense, the coordinate axes are forced to be "perpeniduclar". Moreover, there is a duality of seeing n-tuples as points and as vectors - line segments from the origin. With this, the addition of points and the joining of lines are the same.

If we take three points - origin, $(x_1, \ldots, x_n)$ and $(x_1 + y_1, \ldots, x_n + y_n)$, then these points determine a triangle lying in a plane and we may use the cosine rule (because the plane $\mathbb{R}^2$ is a model for the Euclidean plane) to get

$$\sum_{i=1}^{n}(x_i + y_i)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 + 2\sqrt{(\sum_{i=1}^{n} x_i^2)}\sqrt{(\sum_{i=1}^{n} y_i^2)}\cos(\theta).$$

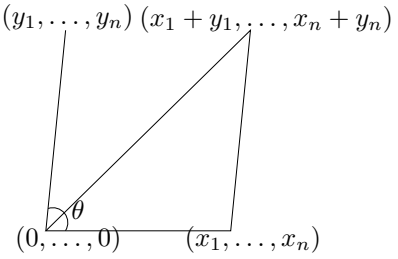

Figure 2: Cosine rule with coordinates

We set $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$ and use $\|x\|$ to denote the sum of squares norm of x and $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$. Then we get

$$\cos(\theta) = \frac{\langle x, y\rangle}{\|x\|\|y\|}.$$

This provides us a way to define the angle between two vectors. The map $\langle \cdot, \cdot \rangle$ is called the inner product (see below) and can be defined for any vector space. So, using the cosine "rule", we define the angle between two vectors. However, this requires us to know what $\cos$ is beforehand, so as to

take $\cos^{-1}$. This can be independently defined in a number of ways. One way to do it is to start with the series expansion of $\cos$ and $\sin$ and verify that certain properties hold.

Another way is to define angle as the proportion of circumference covered in the unit circle. However, this requires us to have a theory of integration first so that we can talk of the length of arcs in a meaningful way.

3.3 Linear forms

We return to the question about the k-volume of k-parallelepipeds in $\mathbb{R}^n$. We are given k linearly independent vectors $v_1, \ldots, v_k$ in $\mathbb{R}^n$ and we want to assign a volume to the parallelepiped they span. We have the following reasonable requirements for k-volume

- Translational invariance : Translation of a parallelepiped (which is another parallelepiped, though not necessarily having the origin as a vertex) should have the same volume as the original
- If $v_1, \ldots, v_k$ are dependent, then the volume should be zero, simply because it is meaningless to talk of the k-volume.
- Multilinearity : Volume should be multilinear as explained below.

The k-volume is then a function taking k - linearly independent vectors to a real number. Let us $k = n$ and we have a parallelepiped spanned by $v_1, \ldots, v_n$ which are linearly independent. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$, then there is an invertible matrix M corresponding to a linear isomorphism sending e_i to v_i. Since M is invertible, by Gaussian elimination, we may write $M = E_1 \ldots E_r$ where E_i are elementary matrices which are of the following three types:

- Swapping : It sends $e_i \mapsto e_j, e_j \mapsto e_i$ for $i \neq j$ while keeping other $e_l, l \neq i, j$ fixed.
- Scaling : It sends $e_i \mapsto ce_i$ for some $c \in \mathbb{R}$ while keeping other $e_j, j \neq i$ fixed.
- Adding : It sends $e_i \mapsto e_i + e_j$ for some $j \neq i$ while keeping other $e_l, l \neq i$ fixed.

We analyse what happens to the volume when any of these operations is applied to the parallelepiped P determined by some $w_1, \ldots, w_n$ independent. Under swapping, it is reasonable to say that the volume doesn't change because the parallelepiped doesn't change. If we scale w_i to cw_i, then we would like the volume to scale by $|c|$, or by c if we want to keep the algebra neat as we shall see below. One way to see this is that if $c \in \mathbb{Z}$, then we are making copies of the intial parallelepiped, and a similar thing holds for rational multipliers (this time we shrink or subdivide the original and add enough copies). This relies on translational invariance. In some ways, the same multiplication holds in the limit when $c \notin \mathbb{Q}$.

At last we come to the third type of elementary operation. We see what happens when we replace $w_1, \ldots, w_n$ with $w_1 + w_2, w_2, \ldots, w_n$ (this is a linearly independent set). Let the new parallelepiped be denoted by P'. By linear independence, we decompose P' into the two parts

$$\begin{aligned} P' = &\{a_1 w_1 + (a_1 + a_2)w_2 + \cdots + a_n w_n | 0 \leq a_i \leq 1, a_1 + a_2 \leq 1\} \\ &\sqcup \{a_1 w_1 + (a_1 + a_2)w_2 + \cdots + a_n w_n | 0 \leq a_i \leq 1, a_1 + a_2 > 1\}. \end{aligned}$$

If we translate the second part by $(0, -1, 0, \ldots, 0)$, then the two parts combine to form the parallelepiped P. Therefore, we can say that the volume doesn't change. Of course, we are assuming that it is meaningful to talk about the volume of the two parts, however, this we allow because the two parts are nice enough, they are the two halfs of P' when bisected by a plane.

So, the volume shouldn't change under the third operation. If we are to obtain the total change in volume, then it is scaled by $|c_1 \ldots c_l|$ where $c_1, \ldots, c_l$ are the scaling factors appearing among the elementary matrices. Therefore, once we fix the volume of the standard unit cube spanned by $e_1, \ldots, e_n$, we know the volumes of all other n-parallelepipeds by first obtaining the appropriate invertible matrix.

We know that the volume doesn't change under the third type of elementary operation, however more generally if we take $w = d_1w_1 + \cdots + d_nw_n$, and replace w_1 by $w_1 + w$, then it is easy to see that the volume is multiplied by $|1 + d_1|$. Let us say that the resulting parallelepiped is P'', and we denote the volume of a parallelepiped P by $\mu(P)$, then

$$\mu(P'') = |1 + d_1|\mu(P), \mu(w, w_2, \ldots, w_n) = |d_1|\mu(P).$$

We are quite close to obtaining additivity, and this is why we need to allow for negative volumes. If we allow for negative volumes, then upon type 2 elementary transformations, the volume scales by c instead of $|c|$. If we allow this, then in the situation above, additivity follows.

What we can conclude is that the volume we are looking for is additive in each entry, and moreover it scales appropriately. In other words, we want the n-volume function to be multilinear.

Definition. *Let V be a vector space over some field K, a multilinear form of degree k (or k-multlinear form) on V is a map $f\colon V^k \to K$ that is K-linear in each coordinate. When $k = 2$, such a map is called bilinear and so on.*

However, there is a further restriction on the volume function. Intuitively we want the n-volume of a parallelepiped to be zero if the parallelepiped is not n dimensional. By multilinearity, this boils down to the volume function being zero whenever two entries are the same.

Definition. *Let V be a vector space over some field K, an alternating k-form is a k-multilinear map $f\colon V^k \to K$ such that whenever two arguments of f are equal, f is zero.*

Lemma 1. *Let $f\colon V^k \to V$ be a k-multilinear map where V is a vector space over a field of characteristic $\neq 2$. Then f is alternating if and only if it is skew-symmetric, i.e., for every $\sigma \in S_k, v \in V^k$ we have*

$$f(\sigma(v)) = sgn(\sigma)f(v)$$

where $\sigma(v)$ is the k-tuple $(v_{\sigma(1)}, \ldots, v_{\sigma(k)})$ when $v = (v_1, \ldots, v_k)$.

Proof. Suppose f is skew-symmetric. Given $v_1, \ldots, v_k \in V$, if some $v_i = v_j$, then consider the transposition $\sigma = (ij) \in S_k$. We then have $f(v_1, \ldots, v_k) = -f(v_1, \ldots, v_k)$ and because $\text{char}(K) \neq 2, f(v_1, \ldots, v_k) = 0$ and f is alternating.

Conversely, suppose f is alternating and let $\sigma = (ij) \in S_k$ be a transposition, then using linearity and ignoring other coordinates,

$$f(v_i + v_j, v_i + v_j) = f(v_i, v_i) + f(v_i, v_j) + f(v_j, v_i) + f(v_j, v_j).$$

Because f is alternating, we get $f(\sigma(v)) = -f(v) = \text{sgn}(\sigma)f(v)$. Finally, because S_k is generated by transpositions and permution of coordinates can be done in steps and $\text{sgn}\colon S_k \to \{\pm 1\}$ is a group homomorphism, we conclude that f is skew-symmetric. □

Note that we needed the assumption on characteristic in one direction only. Therefore, our k volume form is then an alternating k form. There is a notion of symmetric forms where there is no $\text{sgn}(\sigma)$ factor. The collection of alternating, multilinear, symmetric forms each form an K-vector space.

Definition. *Let V be a vector space over K, the space of all alternating k-forms on V is a vector space denoted $\Lambda^k(V)$. The space of all symmetric k-forms is also a vector space denoted $Sym^k(V)$.*

3.4 Quadratic forms and inner products

Let V be a vector space over a field K. If f is a k-multilinear form on V, then to know the value of f on any k-tuple, it is sufficient to know the value of f on any k-tuple of basis vectors of V.

Recall that the angle is a measure between two vectors in $\mathbb{R}^n$. There are of course two choices for the angle, but let us say we have fixed one in some consistent manner. Now if v, w are two vectors, then the angle between them remains constant when v, w are scaled by constants of the same sign. If

however, we consider a function that is like the angle, but scales properly, then we are quite close to a bilinear form.

Suppose f is a bilinear form on V, then $f(\alpha v, \alpha v) = \alpha^2 f(v) \forall\, \alpha \in K, v \in V$. Then, $Q_f(v) = f(v, v)$ is very much like a norm-squared function. Note that

$$Q_f(v + w) = f(v + w, v + w) = Q_f(v) + Q_f(w) + f(v, w) + f(w, v).$$

If f is symmetric, then we can obtain $f(v, w)$ from the values of Q_f provided $\text{char}(K) \neq 2$. Let $e_1, \ldots, e_n$ be a basis of V. Then for any $v = x_1 e_1 + \cdots + x_n e_n \in V$, we have

$$Q_f(v) = f(v, v) = \sum_{1 \leq i,j \leq n} x_i x_j f(e_i, e_j).$$

Form the matrx $A = (f(e_i, e_j))_{1 \leq i,j \leq n}$, then

$$Q_f(v) = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \begin{bmatrix} f(e_1, e_1) & f(e_1, e_2) & \ldots & f(e_1, e_n) \\ f(e_2, e_1) & f(e_2, e_2) & \ldots & f(e_2, e_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(e_n, e_1) & f(e_n, e_2) & \ldots & f(e_n, e_n) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

In other words, Q_f is a homogeneous polynomial (over K) of degree 2 in the coordinates of v. When f is symmetric, so is A.

Definition. *With $K = \mathbb{R}$, let $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$ be a bilinear form. It is called*

- *positive (resp. negative) definite if $\langle x, x \rangle > 0$ (resp. < 0) for all $x \neq 0$.*
- *positive (resp. negative) semidefinite if $\langle x, x \rangle \geq 0$ (resp. ≤ 0) for all x.*

An inner product on V is a symmetric positive definite bilinear form.

If we are dealing with complex vector spaces, then it is more natural to require the inner product to be linear in one variable, conjugate linear in the other variable and conjugate symmetric, i.e.

$$\langle x, y \rangle = \overline{\langle y, x \rangle} \,\forall\, x, y \in V$$

while being linear in the first variable. This is because, as we shall see, there is a strong relation between inner products and norms. In the complex plane, given $z \in \mathbb{C}, |z|^2 = z\bar{z}$, so in order to extend this notion, we need the inner product to be conjugate symmetric.

Definition. *Given a field K, a quadratic form on n variables is a homogenours degree 2 polynomial in n variables. Given a quadratic form Q in n variables over K, we have the evaluation map $Q \colon K^n \to K$ satisfying $Q(\alpha v) = \alpha^2 Q(v) \,\forall\, \alpha \in K, v \in K^n$.*

Note that over $\mathbb{R}^n$, a quadratic form is determined by its values on the unit sphere. In general if we have a function satisfying $f(\alpha v) = \alpha^2 f(v)$ for $\alpha \in \mathbb{R}, v \in \mathbb{R}^n$, it is not true that f is given by a polynomial, for obviously we may set arbitrary values on antipodal points of the sphere and extend to $\mathbb{R}^n$.

Above, we have obtained a quadratic form Q_f from a given bilinear form f. Conversely, suppose we are given a quadratic form Q, we can construct

$$\begin{aligned} f_Q \colon V \times V &\to K \\ (v, w) &\mapsto \frac{1}{2}(Q(v + w) - Q(v) - Q(w)) \end{aligned}$$

Any quadratic form on $K^n = V$ is given by a polynomial $Q(x) = \sum_{i=1}^n a_i x_i^2 + \sum_{1 \leq i < j \leq n} 2 b_{ij} x_i x_j$. We form the matrix A having a_i on the diagonals and b_{ij} at (i, j)th entry and (j, i)th entry for

$1 \leq i < j \leq n$. Then A is symmetric and $Q(x) = x^T Ax \,\forall\, x \in K^n$. Then, we have $f_Q(x, y) = x^T Ay$ which is easily seen to be a symmetric bilinear form.

Because we are dealing with symmetric matrices, we have a bijective correspondence (over any field with characteristic $\neq 2$)

$$\text{symmetric bilinear forms} \leftrightarrow \text{symmetric matrices} \leftrightarrow \text{quadratic forms.}$$

Similar to bilinear forms, we may define positive (negative) definite and semidefinite quadratic forms. Then the above correspondence gives a bijection between inner products and quadratic forms. The familiar sum of squares is a quadratic form and the associated inner product is the familiar dot product. The standard dot product lets us define angle between two vectors, $x^T y = \|x\|\|y\| \cos(\theta)$.

3.5 Rotation

What do we mean by rotation of $\mathbb{R}^n$? Intuitively it should be a reversible transformation of $\mathbb{R}^n$ that fixes some line, the axis. Moreover, not only does it fix the axis, it sends every line passing through the origin to another line, i.e. it is a linear transformation. Furthermore, we don't expect it to change lengths or angles, two lines starting perpendicular to each other should stay perpendicular. Of course, intuitively we are looking at the standard dot product and pythagorean norm (which is the square root of the quadratic form associated to the dot product).

In the matrix notation, given $x, y \in \mathbb{R}^n$, $\langle x, y\rangle = x^T y$. If R is a rotation matrix, then we need $(Rx)^T(Ry) = x^T R^T Ry = x^T y \,\forall\, x, y \in \mathbb{R}^n$. It is easy to see from here that $R^T R = I$. Such a matrix is said to be *orthogonal*. More generally, suppose there is a bilinear form $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ given by a matrix A. Then the linear transformations fixing this inner product satisfies $R^T AR = A$.

For complex vector spaces, rotational matrices would satisfy $R^* R = I$ where R^* is the matrix obtained by taking the transpose and conjugating each entry. Such a matrix is called *unitary*. Rotational matrices are those that preserve inner products. There are a few other types of linear transformations that arise similarly:

- A matrix is called *Hermitian* if $A^* = A$
- A matrix is called *normal* if $AA^* = A^* A$

In the real case, $A^* = A^T$.

4 Alternating forms

4.1 Determinants

We take the n-volume to be an alternating form, if we need positivity we can take the absolute value after doing all the calculations. Take an n-parallelepiped spanned by $v_1, \ldots, v_n$ in $\mathbb{R}^n$. Because the volume is an alternating n-form, using multilinearity, the form is determined by its value on $(e_1, \ldots, e_n)$. Suppose $v_i = \sum_{j=1}^n a_{ij} e_j$, then

$$f(v_1, \ldots, v_n) = f(e_1, \ldots, e_n) \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma(1)} \ldots a_{n\sigma(n)}.$$

Therefore, then there is, upto a multiplying factor, a unique n-volume. Note that the indices in the sum can be interchanged, i.e we may sum $\text{sgn}(\sigma) a_{\sigma(1)1} \ldots a_{\sigma(n)n}$ instead because S_n is a group and sgn is the same on σ and σ^{-1}. In other words, the sum is the same when expanded along the rows or the columns of (a_{ij}).

The collection $v_1, \ldots, v_n$ can be arranged in an $n \times n$ matrix over $\mathbb{R}$ where the ith column contains the coordinates of v_i (with respect to the previously chosen basis $e_1, \ldots, e_n$). Then the n-volume is an alternating form on the columns of this matrix. Therefore, there is, upto a multiplicative factor, a unique map $f \colon M_n(\mathbb{R}) \to \mathbb{R}$ which is an alternating function of the columns. Because the transpose

of a matrix interchanges rows and columns, we can say that there is a unique map which is an alternating function of the rows.

Fix a matrix $A \in M_n(\mathbb{R})$, then the map $B \mapsto f(AB)$, for a given alternating form f on the columns, is also alternating in the columns (this follows from how matrix multiplication is defined). By uniqueness, we have $f(AB) = cf(B)$ for some constant c, specifically $f(A) = cf(I)$ where I is the $n \times n$ identity matrix. If we set $f(I) = 1$, then we see that f is multiplicative.

With $f(I) = 1$, the alternating form is called the determinant, denoted $\det$. The arguments above show that $\det$ is multiplicative. By the multiplicative nature, $\det$ is a group homomorphism $\det\colon GL_n(\mathbb{R}) \to \mathbb{R}^*$. Since $\det(A^T)$ is linear in the rows, and $\det(I^T) = \det(I) = 1$, we must have $\det(A^T) = \det(A)$.

Observe that in $\mathbb{R}^2$, under this definition of volume, squares and rectangles have volume given by the *base×height* formula. All we required was that the volume be translationally invariant and finitely additive (when we cut the parallelogram in two and translated them when discussing type 3 elementary operations). We did not require any notion of perpendicularity of the sides. However, we are still using a choice of the basis. Depending on which basis and what volume for that basis we chose, the volume changes, but ultimately it does not depend on the notion of perpendicularity of vectors.

4.2 Forms at a point

For $k < n$, we do not have multilinearity, or at least we do not want multilinearity. For example, when $k = 1$, we do not want the length to be additive. However, multilinearity still holds as long as we confine to a k-dimensional subspace, the arguments are exactly similar to the earlier arguments.

Let V be a k dimensional subspace of $\mathbb{R}^n$ with a basis $w_1, \ldots, w_k$. If f is any alternating k-form on V, then for any $v_1, \ldots, v_k \in V$,

$$f(v_1, \ldots, v_k) = f(w_1, \ldots, w_k) \sum_{\sigma \in S_k} a_{1\sigma(1)} \ldots a_{k\sigma(k)} \tag{1}$$

where a_{ij} are coefficients such that $v_i = \sum_{j=1}^{k} a_{ij} w_j$. Such a_{ij} exist because w_is is a bases of V. Set A to be the matrix (a_{ij}) and M_1 to be the $k \times n$ matrix with v_i in the ith row and M_2 to be the $k \times n$ matrix with w_i in the ith row. Thus, when we expand v_is and w_is into coordinates with respect to $e_1, \ldots, e_n$, then we get the matrix equation $M_1 = AM_2$ (if we want v_i, w_i to be the columns, take the transpose, it doesn't change determinant).

Suppose we have the ascending k-tuple $I = (i_1, \ldots, i_k)$ with $1 \leq i_1 < \cdots < i_k \leq n$, denote by M_I the $k \times k$ minor determined by columns $i_1, \ldots, i_k$ in a $k \times n$ matrix M. We then have for all ascending k tuples in $\{1, \ldots, n\}$, $\det M_{1I} = \det A \det M_{2I}$. This gives us

$$\det A = \frac{\sum_I \det M_{1I} \det M_{2I}}{\sum_I \det M_{2I}^2} \tag{2}$$

where the sum is over all ascending k tuples from $\{1, \ldots, n\}$. Note that because $w_1, \ldots, w_k$ are independent, some $k \times k$ minor in M_2 is invertible, therefore the denominator is non zero. So, the moment we fix a basis $\{w_1, \ldots, w_k\}$ for V and a value $f(w_1, \ldots, w_k)$, the value of f on other parallelepipeds in V are determined by the above formula.

Now, let us take $k = 1$, what are the alternating 1-forms on $\mathbb{R}^n$? Because there is only one argument, these are going to be exactly all the linear functionals $f\colon \mathbb{R}^n \to \mathbb{R}$. From linear algebra, it is easy to see that the resulting space is the dual vector space spanned by $dx_1, \ldots, dx_n$ where dx_i is the projection onto the ith coordinate. For $k \geq 1$, given $v_1, \ldots, v_k$ and an ascending k-tuple $(i_1, \ldots, i_k)$, we set

$$\begin{aligned} dx_I\colon (\mathbb{R}^n)^k &\to \mathbb{R} \\ M &\to \det M_I \end{aligned}$$

where $(\mathbb{R}^n)^k$ is thought of as the space of $k \times n$ matrices. This is an alternating k-form by the definition of the determinant. By the alternating nature, it is easy to see that these dx_I form a basis of $\Lambda^k(\mathbb{R}^n)$. We can write

$$\det A(v_1, \ldots, v_k) = \frac{\sum_I \det M_{2I} dx_I(v_1, \ldots, v_k)}{\sum_I \det M_{2I}^2}$$

where $v_1, \ldots, v_k \in V$ and M_2 is as defined above.

This helps if we are restricting ourselves to V, however we are looking for a volume function on all k-parallelepipeds. When $k = n$, there is only one n-dimensional subspace, but when $k < n$, we need a way to relate different k-dimensional subspaces and this is where we need to use rotations.

4.3 Volumes of parallelepipeds

We now give a series of results, for a proof of the above results, see Chapter 5 of [7]. In order to use rotations, we need to use the standard inner product.

Theorem 1. *Let V be a k-dimensional subspace of $\mathbb{R}^n$. Then there is a basis $w_1, \ldots, w_n$ of $\mathbb{R}^n$ such that $w_1, \ldots, w_k$ is a basis of V and for each $1 \leq i, j \leq n$, we have $w_i^T w_j = \delta_{ij}$ (Kronecker delta symbol). In other words there is an orthonormal (with respect to standard inner product) basis of V which extends to an orthonormal basis of $\mathbb{R}^n$.*

Theorem 2. *Given two orthonormal bases of $\mathbb{R}^n$, there is a rotation, i.e., a linear transformation, that sends one basis to another.*

We know that rotations fix lengths and angles, so we expect it to fix volumes as well. We have the following result.

Theorem 3. *There is a unique function V that assigns to each k-tuple $(x_1, \ldots, x_k)$ of elements of $\mathbb{R}^n$ a non negative number such that:*

1. *If $h \colon \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal transformation, then $V(h(x_1), \ldots, h(x_n)) = V(x_1, \ldots, x_n)$.*
2. *If $y_1, \ldots, y_k$ belong to the subspace $\mathbb{R}^k \times 0$ of $\mathbb{R}^n$, so that $y_i = (z_i, 0)$ for $z_i \in \mathbb{R}^k$, then $V(y_1, \ldots, y_k) = |\det \begin{bmatrix} z_1 & \ldots & z_k \end{bmatrix}|$.*

The function V vanishes if and only if the vectors $x_1, \ldots, x_k$ are dependent. It satisfies $V(x_1, \ldots, x_k) = \det(XX^T)^{1/2}$ where X is the matrix with x_i in the ith row.

Theorem 4. *For a $k \times n$ matrix M, we have*

$$\det(MM^T) = \sum_I \det(M_I)^2$$

where the sum is over all ascending k-tuples from $\{1, \ldots, n\}$.

Therefore, given a k-subspace V we can chose an orthonormal basis and the denominator of 2 becomes 1. Since the orthonormal basis is given by a rotation of the standard basis, and the volume doesn't change under rotations, its k-volume is 1. Thus, continuing the notation of the previous subsection, the volume of a k-parallelepiped in V is given by $|\det A|$. At the same time, we have a formula for $\det A$ involving k-forms on V.

When $n = 1$, the alternating forms are $dx_1, \ldots, dx_n$ which are projections onto the standard axes. So, the form dx_I gives the volume of the projection of a given k parallelepiped onto the k-subspace determined by the ascending k-tuple I. We have therefore made the following observations. Firstly, the unsinged volume of paralleopipeds is given by a Pythagorean theorem, the sum of squares of the volumes of various projections.

Secondly, to each k plane V, we can assign, a k-form on $\mathbb{R}^n$ that gives us, upto a sign, the volume of parallelepipeds in V. Of course, these forms can be applied to other k-parallelepipeds that don't reside in V. Thus, the signed k-volume function corresponding to a subspace V behaves like a dot

product. Just like the standard inner product which gives the length of the projection of one vector on another, the k-forms above give the volume of the projection of a k-parallelepiped on a k-plane. For example, when $k = 1$, the volume form looks like $a_1dx_1 + \cdots + a_ndx_n$. When evaluated on a vector v, this gives the length of the projection of v onto the line spanned by $(a_1, \ldots, a_n)$.

Lastly, note that there is an assignment of k-subspaces, with a choice of basis, to k-forms on $\mathbb{R}^n$. If we change the basis, then the form is scaled by some nonzero factor relating to the change of basis matrix. Thus, to remove the dependece on the basis, we go to the projective plane. What we so obtain is the Plucker embedding of the space of k-subspaces in $\mathbb{RP}^{\binom{n}{k}-1}$. It can be shown that this map is "nice" from an algebro-geometric perspective.

4.4 Wedge products

We have the notion of area in the plane as *base*×*height*. In higher dimensions this would correspond to obtaining the volume of a parallelepiped by looking at the volumes of its various faces. In other words, we would like to obtain the volume form for $k + l$ paralellopipeds from given k, l forms. We can do this, and the result is called the wedge product.

More generally, suppose V is a vector space over K, and we have a k-form f, and an l-form g. We wish to obtain a $k + l$-form by "multiplying" f, g. First we consider the case where f, g are just multilinear forms. In this case, we have the multilinear form $f \otimes g \colon V^{k+l} \to K$ given by

$$(v_1, \ldots, v_{k+l}) \mapsto f(v_1, \ldots, v_k)g(v_{k+1}, \ldots, v_{k+l}).$$

Of course, this is not the only such multilinear forms. Since all we are doing is taking the first k and last l vectors separately, we may as well permute the $k + l$ vectors first and then separate them Thus, we have for every $\sigma \in S_{k+l}$ a form $\sigma(f \otimes g)$ that first permutes the vectors according to σ and then applies $f \otimes g$. If we denote by $\mathcal{T}^k(V)$ all k-multilinear forms (which is a vector space), then we have a product $\otimes \colon \mathcal{T}^k(V) \times \mathcal{T}^l(V) \to \mathcal{T}^{k+l}(V)$. This tensor product (different from tensor product of vector spaces) is associative in that given multilinear forms f, g, h of degree k, l, m we have $(f \otimes g) \otimes h = f \otimes (g \otimes h)$.

To get an alternating form from two given alternating forms f, g, we take a sort of average over all the $\sigma(f \otimes g)$. Note that all alternating forms are multilinear. We take

$$f \wedge g = \frac{1}{k!}\frac{1}{l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn}(\sigma)\sigma(f \otimes g).$$

It is easy to see that $f \wedge g$ is indeed alternating. The factor $1/k!l!$ is there to avoid repetitions that arise from the permutations that permute the first k and last l elements. Basically, there is a subgroup of size $k!l!$ in S_{k+l} that permutes $\{1, \ldots, k\}$ and $\{k + 1, \ldots, k + l\}$ among themselves and these permutations multiply the terms in the summation by the same amount. Moreover, removing this multiplying factor makes $\wedge$ associative, i.e., $(f \wedge g) \wedge h = f \wedge (g \wedge h)$. The proof of this is left to the reader, or one can look at [7]. If we leave the $\operatorname{sgn}(\sigma)$ part, then we can, given two symmetric forms, obtain another symmetric form, but we won't go into that.

Thus, the wedge product is a product $\wedge \colon \Lambda^k(V) \times \Lambda^l(V) \to \Lambda^{k+l}(V)$. One can prove that the wedge product is associative and satisfies the following anticommutative law : $f \wedge g = (-1)^{kl} g \wedge f$ where f is a k-form and g an l-form. Furthermore, if dx_I is the k-form above on V^k where $I = (i_1, \ldots, i_k)$ is an ascending k-tuple, then it is true that $dx_I = dx_{i_1} \wedge \cdots \wedge dx_{i_k}$.

To summarize this section, we sought to find a way to calculate the volume of parallelepipeds. Given an n-parallelepiped in $\mathbb{R}^n$, we argued that the volume function should be an alternating form and setting the volume of the unit cube to 1, we get the standard determinant. Next, to get the k-volume of other parallelepipeds, it was necessary to use rotations and we get the formula $\det(XX^T)^{1/2}$ as the volume. Next, we looked at what happens if we study a multilinear volume function on k-subspaces V and we arrived at alternating forms on V. These two notions of volumes agree up to a sign, because both are related to calculating $\det A$ where A is the matrix as described previously. Lastly, we extended the notion of *base*×*height* to the wedge product.

5 Integration

5.1 An overview of integration

The motivation for considering integration is to determine area or volumes under curves. If we consider Riemannian integration, then given a rectangle $R \subset \mathbb{R}^n$, where rectangles are sets of the form $[a_1, b_1] \times \cdots \times [a_n, b_n]$, and a function $f\colon R \to \mathbb{R}$, then the integral of f over R is defined using partitions and step functions.

A partition of R is a subdivision of R into smaller rectangles and we can give such a partition a size, namely the maximum side lengths of the subrectangles. Given a function f, we approximate it using step functions and the integral of the step functions is a weighted sum of the volumes of the subrectangles where the weights come from the value of the step function on each rectangle.

In this simplistic picture, we can directly deal with the absolute value of volume and there is no need for alternating forms. There is especially no need for alternating k-forms for $k < n$. However, as we shall see, if we are to transfer the notion of integrals to manifolds, then we shall need to use differential forms.

So, we know how to integrate functions on rectangles. The next step is to integrate over arbitrary sets. First, we consider a bounded set S and a function f on S. To integrate f, we cover S by a rectangle R, so that $S \subseteq R$ and extend f to R by 0. We then say that $\int_S f = \int_R f$. There are certain conditions to be met for this integral to be zero, namely discontinuities of the extended f must form a measure zero set. One can show that this definition is independent of the rectangle R. Once we know how to integrate a given function over S, it is natural to ask if we can assign a volume to S, i.e., integrate $f = 1$. This is possible provided S has a measure zero boundary. Such sets are called *rectifiable* sets.

The next step is when S is not necessarily bounded. In this case, we approximate the integral using compact subsets. Given S, we take $\int_S f = \sup_D \int_D f$ where D ranges over all compact subsets of S. Following [7], we call this integral the extended integral, whereas the integrals on bounded sets are called ordinary integrals. Now, on bounded sets, we have two notions of integration. It turns out that the extended integral might exists while the ordinary one doesn't, however if the ordinary integral exists, then the extended integral exists and the two are equal.

Once we have defined integrals over arbitrary sets, we have to verify that these definitions satisfy some of the properties we expect an integral to have, eg., monotonicity over domain, monotonicity over functions, linearity etc. and these do hold true.

5.2 Integration over parametrized manifolds

Now that we know how to integrate over subsets of $\mathbb{R}^n$, we consider integrating functions over parametrized surfaces. Suppose we have a parametrization (for the moment, this is just a map, not necessarily smooth) $\alpha\colon U \to \mathbb{R}^m$ where U is open in $\mathbb{R}^n$ and that f is a function on the image of U. We would like to integrate f over $\alpha(U)$. Consider for example a curve in $\mathbb{R}^2$, unless pathological, this curve has measure zero and we cannot assign it a length or integrate functions over it if we are to look at it as a subset of $\mathbb{R}^2$. However, it is not a subset of $\mathbb{R}$, so we cannot apply the integration theory on $\mathbb{R}$ to the curve, therefore we have to resort to using parametrizations.

On the one hand, it seems obvious that we can just take the pullback of f, i.e., $f \circ \alpha$ to U and integrate that over U. However, that doesn't work, because it won't be independent of U. For example, consider the curve $(0,1) \times \{0\}$ and the two parametrizations

$$\begin{aligned} \alpha\colon (0,1) &\to (0,1) \times \{0\} \\ t &\mapsto (t, 0) \end{aligned} \qquad\qquad \begin{aligned} \beta\colon (0,1/2) &\to (0,1) \times \{0\} \\ t &\mapsto (2t, 0) \end{aligned}$$

Both are "very nice" parametrizations but the integral of $f = 1$ is 1 with respect to α and $1/2$ with respect to β. We need to consider how α stretches and shrinks U because those change the volume. Two equal area elements in U are stretched differentily by α, and therefore should be scaled accordingly along with the pullback of f, i.e., we must integrate $(f \circ \alpha)(x)c(x)$ over U where c is a function taking care of infinitesimal stretches of U.

From a measure theoretic perspective, we have the Lebesgue measure on U, which we can transfer to $\alpha(U)$ via α, where the measure on a subset $S \subseteq \alpha(U)$ is given by the measure of $\alpha^{-1}(S)$. This applies to general measurable maps but it is particularly simple when α is differentiable.

Suppose that α is differentiable, then at every point of U we have a linear map given by the total derivative of α. Recall that in Riemannian integration, we took partitions of rectangles and took the integral to be the weighted sum of the volumes of subrectangles. Let us look more closely at these volumes. One way to think about it is that we have a finite collection of points in a rectangle R and at each point we have the vectors given by the sides of the subrectangle that the point belongs to. The volume of the subrectangles are given by the evaluation of the form $dx_1 \wedge \cdots \wedge dx_n$ at one of the corner points and taking the absolute value.

When looking at $\alpha(U)$, we'll have to scale the volumes by $\det \alpha$ because that will be the resulting volume form under α. Therefore, we define $\int_{\alpha(U)} f = \int_U f \circ \alpha V(D\alpha)$ where $V(D\alpha) = \sqrt{\det D\alpha^T D\alpha}$ where $D\alpha$ is the $m \times n$ derivative matrix. Note that the transpose comes first because the vectors are now arranged as columns. We can next prove the change of variables theorem and that the integral so defined is independent of the parametrization etc.

Actually, there are two ways to approach this. One way is to define the integral as such, another way is to prove that the formula above is the push forward of the Lebesgue measure. One can prove that if ϕ is a diffeomorphism between open subsets of $\mathbb{R}^n$, then the pushforward of the measure is given by $|\det \phi^{-1}| d\lambda$ where λ is the Lebesgue measure on $\mathbb{R}^n$, this is the change of variables theorem.

5.3 General manifolds

While for parametrized manifolds we have simple formulas for the pushforward of measures which allows us to integrate functions, the same does not apply to general manifolds.

Firstly, manifolds need not be covered by a single parametrization, for example the 2-sphere cannot be covered as such for topological reasons (removing 1-point from the sphere keeps it contractible, but for open sets in the plane it creates a hole). For the same reason, it is not, in general, possible to push the Lebesgue measure forward from all the different charts, although if the manifold is equipped with additional structure, like a Riemannian metric, it is possible to construct a Riemannian measure.

Furthermore, even if it was covered by a single chart, the derivative is not given by a matrix for the simple reason that we do not have global coordinates on the manifold (unlike the parametrized manifold in the previous subsection which was contained in $\mathbb{R}^m$). Technically, we do have global coordinates given by Whitney embedding theorem (which allows us to embed an n dimensional manifold in a Euclidean space of dimension at most $2n + 1$), however these coordinates are clumsy or difficult to use (think of the projective space or other quotient spaces).

Recall that in Riemann integration we assigned to each point the form $dx_1 \wedge \cdots \wedge dx_n$, i.e., the standard determinant, or more accurately, we have the absolute value of the determinant. So, in general to capture how the volume changes, we can, as a first step, assign forms at each point. So, the object that we integrate are no longer functions where we had to force the volume form to be $dx_1 \wedge \cdots \wedge dx_n$ but forms that take care of how the volume element varies. Thus, our objects of integration are going to be maps $\omega \colon M \to \Lambda^k(TM)$ where TM is the cotangent bundle such that $\omega(p) \in \Lambda^k(T_pM)$. Keep in mind that forms are not substitutes to volume elements, i.e., not substitutes to dV, rather they are substitutes to fdV all at once.

Once we have forms varying over M, given a chart $\alpha \colon U \to M$, we can speak of the pull back $\alpha^*(\omega)$ of ω given by the form that applies ω to the pushforward of tangent vectors, i.e., given $v_1, \ldots, v_k \in T_pU$,

$$\alpha^*(\omega)(p)(v_1, \ldots, v_k) = \omega(\alpha(p))(D\alpha v_1, \ldots, D\alpha v_k).$$

Note that I am not being rigorous enough, however this is well defined because $D\alpha$, the derivative of α, sends tangent vector $v \in T_p(U)$ to that vector $w \in T_{\alpha(p)}M$ such that for a smooth germ f near $\alpha(p)$, $w(f) = v(f \circ \alpha)$. The integral of $g dx_1 dx_2 \ldots dx_n$ on an open set U of $\mathbb{R}^n$ is defined as the integral of g over U. Note how this is sensitive to the order of $dx_1, \ldots, dx_n$, it is defined upto a sign when changing parametrizations of U, i.e., under change of variables.

So, on charts we can borrow the forms from the manifold and integrate them. The form tells us how the area elements stretch and what weights to apply simultaneously. However, we cannot integrate any k-form over M. In order that the pullback gives us actual n-volumes as opposed to k-volumes for $k < n$, we must integrate n-forms over n-dimensional manifolds, although we can integrate given 1-forms, say, over curves contained in the manifold.

However, being able to integrate forms on charts doesn't tell us how to integrate over the entire manifold. Moreover, changing the integrand from functions to forms is not completely satisfactory, from a measure theoretic point of view, because we have lost nonnegativity. We shall come to these later. For now, we look at integrating forms. In order to guarantee integrability, we need certain conditions. If we had a measure, we could say, for example, the set of discontinuities should be measure zero however, we instead resort to making the forms continuous or smooth. Since we are going to be integrating the pullback of forms, this suffices for integrability.

6 Vector Fields and Forms

Given a smooth function on open sets of $\mathbb{R}^n$, we have other smooth functions given by the partial derivatives. How do we extend this notion to arbitrary manifolds? We don't have global coordinates on manifolds, so we had to resort to a different definition of tangent vectors. Consider, for example, the operation $\partial/\partial x_1$ on an open set U of $\mathbb{R}^n$. Given a smooth function f, it gives another smooth function $\partial f/\partial x_1$. Essentially, to each point of U, we have assigned a tangent vector, namely $\partial/\partial x_1|_x$, and the partial derivative function is obtained by acting on f, at each point x, by the tangent vector assigned to x, namely $\partial/\partial x_1|_x$.

The extension of this concept to arbitrary manifolds is clear. To each point p of a manifold M, we need to assign a tangent vector $X(p) \in T_p(M)$, such an assignment is called a vector field on M. However, simply assigning a tangent vector is not enough, we need it to take smooth functions to smooth functions.

Definition. *A smooth vector field on a smooth manifold M is a map $X\colon M \to TM$ such that $X(p) \in T_p(M) \forall\, p \in M$ such that for every smooth function f on M, the function $p \mapsto X(p)(f)$ is also smooth.*

Putting a chart around $p \in M$ gives a set of local coordinates, say $x_1, \ldots, x_n$. Then the tangent vectors at points q in the chart are spanned by $\partial/\partial x_1|q, \ldots, \partial/\partial x_n|q$. Therefore, X can be written locally as

$$X(q) = a_1(q)\frac{\partial}{\partial x_1}\bigg|_q + \cdots + a_n(q)\frac{\partial}{\partial x_n}\bigg|_q.$$

It turns out that a vector field X on M is smooth if and only if the coefficients $a_1, \ldots, a_n$ are smooth on each chart (the coefficients depend on the chart).

On open sets of $\mathbb{R}^n$, we also have the notion of double derivatives. There is an obvious extension of this to manifolds. Given two vector fields X, Y on M, we can define a function sending smooth function f to $X(Y(f))$. Note that because Y is smooth, $Y(f)$ is a smooth function. Suppose we go to a chart around $p \in M$ and X, Y have the form

$$X = \sum a_i \frac{\partial}{\partial x_i}, Y = \sum b_i \frac{\partial}{\partial x_i}.$$

Then, we have

$$\begin{aligned} XY &= \sum_i a_i \frac{\partial}{\partial x_i}(\sum_j b_j \frac{\partial}{\partial x_j}) \\ &= \sum_i \sum_j a_i(\frac{\partial b_j}{\partial x_i}\frac{\partial}{\partial x_j} + b_j \frac{\partial^2}{\partial x_i x_j}) \\ &= \sum_i (\sum_j a_j \frac{\partial b_i}{\partial x_j})\frac{\partial}{\partial x_i} + \sum_i \sum_j a_i b_i \frac{\partial^2}{\partial x_i x_j} \end{aligned}$$

There is an obvious assymetry between the a_is and b_js. On open sets U of $\mathbb{R}^n$, the various double derivatives commute, however this is not true over arbitrary manifolds. To see if vector fields commute, we look at the object called the Lie bracket or Jacobi-Lie braket or simply bracket of two vector fields. Given vector fields X, Y on M we define their bracket as

$$[X, Y] = XY - YX.$$

Since double derivatives commute on open sets of $\mathbb{R}^n$, from the calculation we have, on charts, the formula

$$[X, Y] = \sum_i \sum_j (a_j \frac{\partial b_i}{\partial x_j} - b_j \frac{\partial a_i}{\partial x_j}) \frac{\partial}{\partial x_i}$$

which is a vector field. The bracket operation satisfies certain nice properties. Firstly, it is $\mathbb{R}$-linear in each variable

- (Anti-symmetry) $[X, Y] = -[Y, X]$, in particular $[X, X] = 0$.
- (Jacobi identity) $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$.
- (Product rule) For smooth functions f, $[X, fY] = X(f)Y + f[X, Y]$.

6.1 Lie Derivatives

Recall how partial derivatives were defined on open sets of $\mathbb{R}^n$. We take the quotient $f(p+tv) - f(p)/t$ and then take the limit where p is a point and v is a direction at p. Let us focus on the numerator $f(p + tv) - f(p)$. Here we are using the vector space structure of $\mathbb{R}^n$. Alternatively, we may think of this as the value of f if we go a little way along the direction indicated by v. In other words, suppose we assign the direction v to every point near p (i.e., in a neighbourhood of p) and we go an infinitesimal distance from p following the directions given by v and arrive at $p + tv$. We then take the difference of the two values.

This way of thinking about the derivative allows us to extend the notion of directional derivative. It is no longer a derivative along a direction, but a derivative along a vector field. So, we may take a vector field X which assigns a direction to each point. Next, we need to move an infinitesimal distance along X. If we are to move in such a manner, we describe a curve whose tangent vectors at every point are given by X.

Therefore, given X, we must obtain a curve $c\colon (-\epsilon, \epsilon) \to M$ that passes at 0 through p and has tangent vectors given by X. We need to take the curve on both sides of 0 so that t above can be negative. This is a problem about differential equations, and there are suitable existence and uniqueness results, so such a curve can always be found.

Above we have fixed the point p, but just like how directional derivatives on open sets of $\mathbb{R}^n$ can be defined for all points, we may want p above to be variable. So, let us say we want p to vary on an open set. In this case, we want a family of curves that at $t = 0$ pass through this open set and have have tangents given by X, i.e., a smooth map $c\colon (-\epsilon, \epsilon) \times U \to M$ where U is some suitable neighbourhood of p, and for every $q \in U$, the curve $c(t, q)$ passes through q at $t = 0$ and has tangents given by X. Again, this is a question about solutions to differential equations and again, there are suitable existence and uniqueness results.

Once we have such curves, then it is simple to differentiate functions as in the case of Euclidean space. The curves reduce the problem to one dimension and we may similarly take the difference quotient. However, it is not just functions we can differentiate, with this tool we can now differentiate other vector fields and differential forms. Essentially, these curves allow us to transport data from one point to another along given vector fields. These derivatives we get in such a manner are called Lie Derivatives.

6.2 Smooth forms

Next, we know that in order to integrate, we must assign alternating forms to each point of the manifolds. However, in order to guarantee integrability, we must restrict all possible differential

forms. A differential k-form ω is attaching to each point of M an alternating k-form on $T_p(M)$, i.e., $\omega(p) \in \Lambda^k(T_p(M))$. Therefore, we can act differential k-forms on k-tuple of vector fields. Just like for vector fields, we define

Definition. *A differential k form ω is smooth if for every k-tuple of smooth vector fields $X_1, \ldots, X_k$ on M, the map $p \mapsto \omega(X_1, \ldots, X_k)(p)$ is smooth on M.*

The collection of smooth k forms on a manifold M is denoted by $\Omega^k(M)$. Just like vector fields, smoothness of forms translates to smoothness of coordinates when restricted to charts, so if a form is smooth, then we can compute its integral.

Now, suppose f is a smooth function on M, then given a smooth vector field X, we know that Xf is a smooth function. Thus, to every $f \in \mathcal{C}^\infty(M)$, we have the smooth 1-form df on M which assignes to each point, the 1-form given by evaluation at the tangent vector. Thus, if $(x_1, \ldots, x_n)$ is a chart about p, then $df(p) = \sum \partial f/\partial x_i dx_i$. Naturally, the next question is which smooth 1-forms are given as df for some smooth function f.

Because the double partial derivatives commute, there are certain necessary conditions that a 1-form must satisfy in order to be of the form df. Specifically, if a 1-form looks like $\omega = \sum a_i dx_i$ on some chart, then we must have $\partial a_i/\partial x_j = \partial a_j/\partial x_i$ for every pair i, j if ω is to be df for some f.

7 Stokes' Theorem

Suppose we have a smooth form on a manifold and we want to integrate it. If we have a single parametrization, then this will be the integral of the pullback, however when we don't have a single parametrization, we need to break the integral into simpler parts. We do this by considering a family of charts and then integrating the form over each of these charts and adding them up. However, we need to make sure that the intersections do not matter, and this is why we need partitions of unity.

Partitions of unity allow us to break the manifold into meaningful chunks so that the integral over the whole is the sum of integral over the parts. One can show that under suitable conditions, namely paracompactness and Hausdorfness, given a countable open cover U_i, we can obtain a countable collection of smooth functions ϕ_i such that $\sum \phi_i = 1$, and the support of each ϕ_i, i.e., the set $\overline{\{x|\phi(x) \neq 0\}}$, is contained in U_i. Taking U_i to be charts allows us to integrate functions. One then shows that the integral so obtained is independent of the charts and the collection ϕ_i.

We will consider the question of integrating forms later, for now we focus on the question raised in the previous subsection, namely, when is a 1-form df for some smooth function f. We have already outlined a necessary condition. Let us consider the situation in the plane, say we have a one form $\omega = pdx + qdy$ (note that because we have global coordinates, there is a canonical way to identify all the different tangent spaces with $\mathbb{R}^2$ and this gives the basis dx, dy for the cotangent space at each point). We need to solve the equations

$$\frac{\partial f}{\partial x} = p \text{ and } \frac{\partial f}{\partial y} = q.$$

One way to solve this is to integrate ω along the axes to arrive at f, this is basically an application of the fundamental theorem of calculus. So, for example set the value of f to be zero at the origin (changing f by a constant doesn't change its derivatives) and define

$$f((x_0, y_0)) = \int_0^{x_0} p(x, 0) + \int_0^{y_0} q(x_0, y).$$

Note that the pullback of ω restricted to these paths is indeed as above because in the horizontal path the dy part is pulled back to the zero form and dx to the one form etc.

In order that this f works, i.e., its partial derivatives be p, q, we need to ensure that the integral so defined do not depend on the choice of the path. For example, we may instead want to go vertically first, then horizontally, or we may want to go radially.

Another way to look at it is that if ω is indeed df for some f, then by the fundamental theorem of calculus, the integral of ω over closed paths should be zero. The simplest closed paths being

rectangles, we first look at the integrals over rectangles. For simplicity, let us consider the integral over the unit square. Note that when considering loops there are two orientations, we'll go anticlockwise along the square starting with the horizontal line.

There are four line integrals,

$$\int_{\partial[0,1]^2}(pdx+qdy) = \int_0^1 (p(t,0)-p(t,1))dt + \int_0^1 (q(1,t)-q(0,t))dt.$$

Here $\partial[0,1]^2$ is the topological boundary, i.e., the edges of the square. If p, q are differentiable, then we may apply the fundamental theorem of calculus to the integrands, to get

$$\begin{aligned}\int_{\partial[0,1]^2}(pdx+qdy) &= -\int_0^1 (\int_0^1 \frac{\partial p}{\partial y}(x,y)dy)dx + \int_0^1 (\int_0^1 \frac{\partial q}{\partial x}dx)dy \\ &= \int_{[0,1]^2}(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y})dxdy\end{aligned}$$

where we have used Fubini's theorem in the second step, assuming it is applicable. This is Green's theorem. Note that we actually obtain a scalar function to be integrated over the square, but we can interpret it as the integral of a two form (and $dxdy$ is shorthand for $dx \wedge dy$). Note that had we gone clockwise, then the integrand would have been negative.

Essentially, we have written the difference of integrals on "opposite" sides of the square as an integral through the square. In this special case, the necessary condition outlined previously makes the integrand zero. One can next show that the partial derivatives of f are indeed p, q.

Of course, we have proved path independence only for sides of rectangles, but one can prove Green's theorem more generally. Note crucially that we should be able to apply the fundamental theorem of calculus and this requires a path through the square. Intuitively then, it doesn't depend on the space being exactly the square, if there is some shape whose boundary is the boundary of the square and we are able to take paths through the shape connecting different boundary points, then the same proof should apply. This tells us that the topology of the space plays an important role and indeed consider $\omega(x,y) = (xdy - ydx)/(x^2+y^2)$ on the punctured plane, which satisfies the necessary condition but is not df for any f since the path integral around the origin is not zero.

7.1 Exterior derivative

Now we consider what happens to a 1-form in three space, let $\omega = pdx + qdy + rdz$. If we are to obtain path independence here, then by first restricting to the xy-plane, we see that we will be integrating the two form $(\partial q/\partial dx - \partial p/\partial y)dxdy$. Similarly, if we consider the other planes, we will be left with

$$d\omega = (\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y})dxdy + (\frac{\partial r}{\partial y} - \frac{\partial q}{\partial z})dydz + (\frac{\partial p}{\partial z} - \frac{\partial r}{\partial x})dzdx.$$

Note that I have not justified why we get this, but this two form restricts to the correct two forms on each of xy, yz, zx-planes. Note also that, as mentioned previously, the orientation on each plane is important to determine the signs. One can see that the same method applies to one forms in arbitrary dimensions, the hope is that path independence is guaranteed if this form is zero. If this form is zero, then it is clear that on each hyperplane (obtained by setting some $x_i = 0$) the integral of the one form is independent of the order in which we integrate along the axes (eg, whether we go horizontal first or vertical).

So, now we have a way to obtain a two-form from a given 1-form and this plays the same role df plays to f, namely in the fundamental theorem of calculus. There, if we integrate df along a curve, then we get the difference of f at the ends (with the sign dependeing on the orientation of the curve, i.e., whether we go from left to right or opposite), here if we integrate $d\omega$, we get (or rather hope) that we get the integral of ω along the boundary curve. We have proved this in the special case above, where integrating $d\omega$ on the square gave the integral of ω along the sides.

Now, we can ask the same question we had for one forms, given a two form η, when is it $d\omega$ of some one form ω. Let us first consider $n = 2$, then a two form looks like $\eta = gdxdy$. If it is $d\omega$ for $\omega = pdx + qdy$, then $g = \partial q/\partial x - \partial p/\partial y$. However, this one equation doesn't really restric anything because we can simply solve for $\partial q/\partial x = g$ and set $p = 0$. So, this case is not very interesting.

When $n = 3$, the two forms look like $\eta = fdxdy + gdydz + hdzdx$. Now if this was $d\omega$, then from the equation above, we have

$$\frac{\partial f}{\partial z} + \frac{\partial g}{\partial x} + \frac{\partial r}{\partial y} = 0.$$

Just as in the case of 1-forms, we look at the integral of a two form over the unit cube. Recall that to integrate one forms along the sides of the square, we needed to assign an orientation to each side of the square, and we put opposite signs on opposite sides, i.e., we went from left to right on the bottom side and reverse on the top etc. However, there is no such natural "anticlockwise"-ness to go along the faces of the cube.

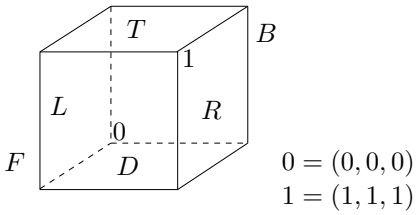

Figure 3: Integrating along the faces of a cube

However, if we are to mimick the same process we had for one forms, then we look at

$$\int_{\partial[0,1]^3} \eta = \int_F \eta - \int_B \eta + \int_R \eta - \int_L \eta + \int_U \eta - \int_D \eta$$

where $D, B, \ldots$ are the faces of the cube as in the diagram. The signs are taken from whether the face goes to the negative side or positive when going *away* from the cube. Again, we can group like faces together, so for example front and back faces give

$$\int_{[0,1]^2} (f(x, y, 1) - f(x, y, 0)) = \int_{[0,1]^3} \frac{\partial f}{\partial z}.$$

Similarly considering the other faces, we get

$$\int_{\partial[0,1]^2} \eta = \int_{[0,1]^3} (\frac{\partial f}{\partial z} + \frac{\partial g}{\partial x} + \frac{\partial r}{\partial y})dxdydz.$$

One can see that the necessary condition now ensures that the integral of η over closed surfaces is zero. Again, the topology played an important role and this time we have gone from a two form to a three form.

Essentially, there is one "operation", namely writing the difference of the integral of a form on two parallel faces as an integral through the space in between. Let us consider a k-form ω, for simplicity, say $\omega = fdx_1 \ldots dx_k$. Now, if we take a k-parallelolgram A_1 in $\mathbb{R}^n$, say parallel to the $x_1 \ldots x_k$-plane, and translate it parallel to itself to A_2 in some other plane, say $x_{k+1} = 1$, then

$$\int_{A_2} \omega - \int_{A_1} \omega = \int_{A_1} (\int_0^1 \frac{\partial f}{\partial x_{k+1}} dx_{k+1})$$

where coordinates $x_j, j \geq k+2$ are kept zero in all integrals. So, the $k+1$ form $\frac{\partial f}{\partial x_{k+1}} dx_{k+1} dx_1 \ldots dx_k$ can be thought of as the derivative of ω in the x_{k+1} direction.

What happens if we translate A_1 in the same plane, say to $x_1 = 1$? In this case, the space between the two planes is still k dimensional and we cannot talk about forms anymore. While we can still talk about the difference as an integral, we will be integrating with respect to x_1 twice. In the previous case, we had the integral of $f dx_1 \dots dx_k$ over the two (signed) faces as an integral through the space between. To generalize this, we are to take the resulting "space" as a $k+1$ dimensional object, then we must have the integral to be zero, which means we must take $\partial f / \partial x_1 dx_1 dx_1 dx_2 \dots dx_k = 0$.

In other words, the two cases are different and we are interested in writing the integral over the faces as an integral through the space bounded by them. If we observe closely, there is a definite way to go from $f dx_1 \dots dx_k$ to the $k+1$ form that contains all the various "directional derivatives" of ω and this is to consider the form obtained by $df \wedge dx_1 \dots dx_k$. We define

$$d\omega = df \wedge dx_1 \dots dx_k = \frac{\partial f}{\partial x_1} dx_1 \wedge dx_1 \wedge \dots \wedge dx_k + \dots + \frac{\partial f}{\partial x_n} dx_n \wedge dx_1 \wedge \dots \wedge dx_k.$$

Observe that this is suitable because it automatically rejects repeated dx_i (actually, the wedge product was designed to do this), and from the arguments above, when restricted to move in the $k+1$th direction, say, the irrelevant parts become zero.

We extend this operation to arbitrary k-forms, thus if $\omega = \sum_I f_/ I dx_I$ is a k-form where I runs over all ascending k-tuples, then $d\omega$ is the $k+1$ form $\sum_I df_I \wedge dx_I$. It defines a map $d \colon \Omega^k(\mathbb{R}^n) \to \Omega^{k+1}(\mathbb{R}^n)$. It is easy to check that d is $\mathbb{R}$-linear. Moreover, by the properties of wedge product, we have $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta$ where ω, η are k, l-forms respectively and $d(d\omega) = 0$. A zero form is a scalar function, so according to the definition, df is the form as defined earlier for zero forms f.

It is not surprising that when we apply d to one forms or two forms, we get back what we started with. Out original question about when a one form is df for some function f had the necessary condition that $d^\omega = 0$ and this necessity is clear from the properties of d, namely $d^2 = 0$. Historically, the d operator was called *grad, curl* and *divergence* depending on the context. Note that there is an abuse of notation in that we actually have a different d_k for each k acting on $\Omega^k(\mathbb{R}^n)$, but we are labelling all of them by d.

The d operator is called the exterior derivative. Let us go back to the flow of ideas. We started out with the question of when a one form is df for some f and this involved solving a system of partial differential equations. However, a necessary condition is that the integral of the form be zero on closed paths. We looked at the integral over the boundary of a square, being the simples curves, and using the fundamental theorem of calculus, were able to write the integral over the boundary as an integral of a specific two form over the bulk of the square. The orientation, which we shall get to soon, and topology played an important role.

The next natural question is to find out when a two form is $d\omega$ for a one form ω. We haven't explicitly found a solution (unlike the 1-form situation, where we can find a solution using path integrals), but we again looked at the integral over the faces of a cube, which is the natural extension of the previous case. Again, we were able to get hold of a triple integral over the cube (this is called the Gauss divergence theorem) using the fundamental theorem of calculus. So, now we shift our problem to integrating k-forms over boundaries of certain surfaces.

Once again, intuitively, we can reduce the integral over the boundary to an integral over the bulk space. This involves certain terms that look like derivatives of the k-form, which ultimately led us to the exterior derivative. In this sense, the exterior derivative is desgined to make Stokes' Theorem (the general version of Gauss divergence theorem and Green's theorem) true.

7.2 Boundary

Throughout in the previous sections we have talked about integrating forms on the boundary of spaces. Previously we were concerned with squares and cubes and we had an intuitive notion of boundary, namely the edges and faces. Here we formalize this notion to arbitrary manifolds. Secondly, recall that we spoke about the orientation of the curves and signs for faces, both of these come under the name of orientation.

By boundary we do not mean the topological boundary, although they agreed in the cases of squares and cubes. The notion of topological boundary requires us to embed the manifold in a larger space, so the sides of a square are its boundary only in reference to the larger Euclidean space it resides in. However, we still need a notion of boundary intrinsic to the square because, we are going to relate the integral over the edges to the integral over the bulk and secondly, it is intuitively clear that the points in the edges are different from points inside the square simply because some of the directions are not available along the edges.

Our "model" for the boundary are going to be the faces of a cube or edges of a square. These are going to be similar to the upper half plane $\mathbb{H}^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n | x_n \geq 0\}$. Note that corners of cubes are not that different from points on the faces when considered as part of the cube as a whole.

Using the upper half planes we can define manifolds with boundaries, these are topological spaces that are locally homeomorphic to either open sets in $\mathbb{R}^n$ or open sets in $\mathbb{H}^n$, and those points which are mapped to the $x_n = 0$ plane under coordinate charts are said to constitute the boundary of the manifold.

It can be shown that the boundary is well defined. When the transition maps between different charts are smooth, we have a smooth manifold with boundary. Here the notion of smoothness is extended, so that a map f on some subset S of $\mathbb{R}^n$ is smooth at a point $p \in S$ if there is a smooth function g on an open set U containing p which restricts to f on $U \cap S$.

With these definitions, if M is a n manifold, then its boundary ∂M is an $n-1$ manifold. Observe that because the $x_n = 0$ plane doesn't have a boundary when considered as an $n-1$ manifold, the boundary of a manifold has empty boundary. Manifolds without boundary are said to be closed manifolds.

Even on manifolds with boundary, we can similarly define germs of smooth functions at a point and the space of derivations, which is the tangent space. We can then define the cotangent bundle as usual and similarly forms.

7.3 Orientation

Having defined the boundary, we come to orientations. Recall that in the case of a closed curve, while integrating a one form we made a choice to go anticlockwise, similarly we had to assign signs to the faces of the unit cube to get a version of Gauss divergence theorem.

Classically, the signs appeared as the dot product of the vector field (associated with a two form) with a continuously varying outward normal on the cube or sphere. This is how people calculated the flux of vector fields, and this is how Gauss divergence theorem is classically stated.

While this picture helps to visualise things on 2 manifolds in 3 space, it doesn't generalize to other situations, nor is it uniformly applicable to other situations. Suppose instead of a surface, we had a curve in three space. In this case, we need to make a choice of going clockwise or anticlockwise, and not a choice of normal, in fact there are two normals available (although if the curve is smooth, we can define a normal using the double derivative of the curve).

Moreover, the choice of a normal vector requires our manifolds to live in an ambient $\mathbb{R}^n$. We cannot talk about a normal to the sphere for example, if there was nothing but the sphere. Similarly, in the case of curves, we'll have to replace the notion of a smoothly varying tangent lines with vector fields.

Let us go back a few steps. How do we integrate a form on a manifold? We have spoken a little about this at the start of this section. Suppose we have an n-form ω on an n manifold M. We already know how to integrate over parametrized manifolds, we simply pull the form to some open set in $\mathbb{R}^n$ and integrate using standard methods. So, one way to integrate ω over M would be to discretize it into different parametrizations (which are provided by the charts on M). Consider, for example, a sphere. If we were to integrate a form, we can integrate it over the northern and southern hemispheres and then add the result, with the equator adding zero because it is "measure zero".

Suppose we have a countable cover of M by open sets $\{U_i\}$. Then it would appear that we can simply restrict ω to each U_i and then add up the results. However, this would lead to multiple countings, and a potentially divergent/non convergent sum. One could hope that as in the sphere's

case, these open sets can be chosen so that the intersections have zero measure, however that requires us to have a concept of measure in the first place.

Assume for the moment that we know how to integrate forms over manifolds, then what would the above procedure be? The result we calculate would be

$$\int_M \omega = \sum_i \int_{U_i} \omega|_{U_i} = \sum_i \int_M \omega \chi_{U_i}$$

where χ_{U_i} is the indicator function for U_i. Supposing additivity, we get

$$\int_M \omega = \int_M \sum_i \omega \chi_{U_i}.$$

There are several problems with this approach of course, firstly even if we know how to integrate, the form $\omega\chi_{U_i}$ may be badly behaved, because χ_{U_i} need not be continuous, let alone smooth. However, one can still hope that it is at least reasonable well behaved that it is integrable, whatever our concept of integral might be. Secondly, and this is a more important problem, the functions χ_i need not add to 1. Note that in the last equation, ω is arbitrary, so a reasonable guess is that χ_{U_i} should add to 1, but this is quite impossible.

There is one fix for both the problems above and this is to approximate the characteristic functions by smooth functions and to choose these functions so that they add to 1.

Definition. *A $\mathcal{C}^\infty$ partition of unity on a manifold M is a collection of nonnegative $\mathcal{C}^\infty$ functions $\{\rho_\alpha \colon M \to \mathbb{R}\}_{\alpha \in A}$ such that*

- *the collection of supports $\{supp\rho_\alpha\}_{\alpha \in A}$ is locally finite,*
- $\sum \rho_\alpha = 1$

Given an open cover $\{U_\alpha\}_{\alpha \in A}$ of M, we say that $\{\rho_\alpha\}_{\alpha \in A}$ is subordinate to the open cover $\{U_\alpha\}$ if $supp\rho_\alpha \subset U_\alpha$ for every $\alpha \in A$.

Here the support of ρ is the set $\overline{\{x : \rho(x) \neq 0\}}$ and it is said to be locally finite if around every $p \in M$ there is a neighbourhood that intersects finitely many supports. What condition i) means is that only finitely many ρ_α are nonzero at a given point p, therefore, when evaluating $\sum \rho_\alpha$ at p, we are evaluating a finite sum.

Now we may cover M with some suitable open sets (say charts) $\{U_\alpha\}$ and obtain a partition of unity $\{\rho_\alpha\}$ subordinate to the given cover (such partitions always exist, see [8] or [7]), so that $\text{supp}\rho_\alpha \subset U_\alpha$. Then, we consider $\omega\rho_\alpha$. This form is non zero outside U and smooth, so we "integrate" it over M by integrating its restriction to U which is a parametrized. After calculating all these integrals, we can add them up.

However, there are still a few problems. Firstly, the resulting sum may be infinite, even uncountable. We can remedy this by either choosing a countable cover and hoping for a convergent sum or assuming ω itself has compact cover so that only finitely many summands are nonzero. The more important question is that of well-definedness. How do we know that the integral so obtained is independent of the charts or the partition of unity?

Before answering that question, we must ask about the well definedness of the integral over a parametrized manifold. Recall that the integral of $gdx_1 \dots dx_n$ on an open set U of $\mathbb{R}^n$ is defined as the integral of g over U. So, if $\alpha \colon U \to \mathbb{R}^n$ is a chart on M, then the integral of ω on U, is defined as the integral of $\alpha^*\omega$ on $\alpha(U)$. By the change of variables theorem, this is defined upto a sign, i.e., if we take a different parametrization $\beta \colon V \to \mathbb{R}^n$, then the integral of ω over $U \cap V$ is defined upto a sign. More directly, if we swap the coordinates of α, then clearly the integral changes by the sign of such a permutation.

Our hope for the partition of unity is that we can discretize the form and add up the integrals. However, suppose, for the sake of illustration, at some point p, the parts of unity are $1/2, 1/2$ coming from two neighbourhoods U, V. The Jacobian matrix of the transition map between U, V is the thing

that controls how the integrals differ, and since we are integrating forms, the integrals over U, V differ by the sign of the Jacobian. Thus, if the Jacobian is negative, then the parts of ω at p cancel out instead of adding.

Hence, in order that the integrals add up as we expect them to, it is necessary that the Jacobian is positive. This leads to the concept of orientability. We need to chose our initial cover in such a way that all possible Jacobians have positive sign. First we deal with manifolds without boundary.

Definition. *An atlas on M is said to be oriented if for any two overlapping charts, the transition map has positive Jacobian everywhere on the intersection. If a manifold has an oriented atlas, then it is called oriented.*

Once we have orientations, it is possible to show that the integral of forms is well defined (see [8]). Not all manifolds are orientable, a classic example is the Mobius strip. However, our definition is rather clumsy to use, it requires us to construct an orientable atlas which is not completely obvious. There are a couple of other things to check, firstly, how does orientation relate to the clockwise/anticlockwise orientation associated to a curve, and secondly, since our formulation requires us to integrate forms over a manifold and its boundary, how are the orientations of a manifold and its boundary related?

Observe that if we have two charts $(U, x_1, \ldots, x_n), (V, y_1, \ldots, y_n)$ around a point p, then the change of basis matrix is an invertible linear transformation of the tangent spaces. So, at p, we have a change of basis matrix M acting on T_pM. This transformation sends bases of T_pM to other bases. The collection of ordered bases of T_pM are divided into two equivalent classes determined by the determinant of the change of basis matrix, i.e., two ordered bases B_1, B_2 are related if the change of basis matrix between them has positive determinant.

From U we have a basis for T_pM given by $\partial/\partial x_1|_p, \ldots, \partial/\partial x_n|_p$ and from V we have a similar basis. If the Jacobian has a positive determinant, then the two bases obtained from the charts are in the same equivalent class.

So, on the vector space $\mathbb{R}^n$ we define two orientations given by the two equivalence classes, this is a generalisation of the right hand system. Recall that while defining the determinant, we set $\det I = 1$, however had we changed it to some c, then the new determinant (defined as an alternating n-form on the n-tuples of vectors) will be c times the old determinant, so the equivalence classes do not change. Then, on a manifold M we define a pointwise orientation as a choice of equivalence classes at each point.

A pointwise orientation μ on M is continuous at $p \in M$ if there is a neighbourhood U of p on which there is an n-tuple $(Y_1, \ldots, Y_n)$ of continuous vectorfields such that at each $q \in U$, the ordered collection $(Y_1(q), \ldots, Y_n(q))$ is an ordered basis of T_qM in the orientation $\mu(q)$ chosen at q.

One can then show (see [8]) that a continuous orientation exists on M if and only if there is a smooth nowhere vanishing n-form on M if and only if M has an oriented atlas. These equivalent definitions of orientability are easier to verify. One can also show that a connected manifold has precisely two orientations.

The notion of clockwise/anticlockwise orientation of a closed curve only makes sense in the plane, in general what we mean is that we have a parametrization of the curve from $[0, 1]$ and the sense of orientation is determined by whether we go from 0 to 1 or the other way. Note that on $[0, 1]$ we have a smooth nowhere vanishing 1-form, namely dt. The pushforward of this form gives an orientation of the curve.

Finally we come to orientations of a manifold with boundary. Recall that while motivating the gauss divergence theorem we used the notion of an outward normal, we next formalise this notion. As remarked in the previous subsection we can define the tangent space and differential forms on such a manifold. We can similarly define an oriented atlas. However, as in [8], in order to obtain equivalent definitions of orientability, for $n = 1$, our local model for 1-manifolds with boundaries should include the left half line $(-\infty, 0]$ along with the right half line.

Since the boundary sits inside the manifold, the tangent space $T_p(\partial M)$ is a subspace of T_pM for $p \in \partial M$. We can then define inward pointing vectors and outward pointing vectors. here one needs an important fact that if $v \in T_pM$, then there is a curve on M passing through p whose derivative at p is v. With this fact, an inward pointing vector $v \in T_pM$ is one such that $v \notin T_p(\partial M$ and there is

a half curve $[0, \epsilon) \to M$ passing through p at 0 whose derivative at 0 is v. A vector is outward if its negative is inward.

On each chart around $p \in \partial M$ coming from M we can select a smooth outward pointing vector; using partitions of unity, we can join them to get a smooth outward pointing vector field on the whole of ∂M (see [8]). Now, an orientation on M corresponds to a nowhere vanishing n-form on M. To obtain an orientation on ∂M, it suffices to obtain a nowhere vanishing $n-1$ form on ∂M.

There is an operation called the interior multiplication that, given an alternating n-form ω and a vector v, gives us the alternating $n-1$-form $\iota_v\omega$ that sends the vectors $v_1, \ldots, v_{n-1}$ to $\omega(v, v_1, \ldots, v_{n-1})$. One can extend this notion to differential forms and smooth vector fields. Using the smooth outward pointing vector field on ∂M and a nowhere vanishing n-form on M, we can, by interior multiplication, obtain a smooth $n-1$ form on ∂M.

This resulting form is well defined and gives an orientation of ∂M. For the proofs of these results, the reader is reffered to [8].

Lastly, having defined the orientation on manifolds with boundary, one can prove the Stokes' theorem in its anticipated form. For an orientable n-manifold M, smooth $n-1$ form ω on M, we have

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

One can also show that upon reversing the orientation, both signs change, therefore the equality (not the integrals) is independent of the choice of orientation.

7.4 A last note

We have looked at why forms are natural objects to integrate, but they lacked nonnegativity. Moreover, we require our manifolds to be orientable if we are to integrate forms. However, forms are interesting objects on their own. A form ω is called closed if $d\omega = 0$ and exact if $\omega = d\eta$ for some η. We have touched on the question of when a closed one form is exact, and when a closed two form is exact, etc. In trying to answer those questions, we had to extend the fundamental theorem of calculus to Stokes' theorem. This extension requires us to define the notion of boundary and orientation and in order to integrate forms we also need partitions of unity.

However, integrating forms over manifolds did not take us closer to the answer of when a closed form is exact. This question ultimately rests on the topology of the manifold. Trying to answer these questions led Poincare, Cartan and others to the concept of homotopy, homology etc. Forms are important for another reason. If two manifolds are diffeomorphic, then one can push the forms on one manifold to another. A knowledge of which closed forms are exact helps in classifying manifolds.

Next, we hoped that forms would capture how volumes changed. Forms do that job, but only to an extent. The problem is that if we are to treat integration as assigning measures to the size of sets, then forms are not the objects we should look for. However, this problem is partly remedied by a "weakness" of forms in that they are integrable only over orientable manifolds. Thus, in an orientable manifolds, upto a sign, the integral is well behaved and the nowhere vanishing n-form functions like a volume form.

In general we need an object that captures the volumes of infinitesimal parallelepipeds, but is also nonnegative. This leads to the concept of densities on manifolds. These are objects that can be integrated on any manifold, orientable or not, and behaves just as we expect any measure to behave. To some extent they are like the absolute values of n-forms.

8 Appendix - Angles

How does one define angles in a logically coherent manner? On way is to first define $\mathbb{R}$ and its topology, and the products $\mathbb{R}^n$. Then we develop a theory of integration, the Riemann integration or Jordan measure will suffice because we are going to deal with very nice sets like arcs of a circle. This theory of integration will require some preliminary notions - volumes of different parallelepipeds.

With these preliminary notions, we can assign a measure to any reasonable set using upper and lower bounds.

With the theory of integration, there is the change of variables theorem which also requires the notion of determinants. Now, in $\mathbb{R}^2$, given a line we can define its angle with respect to the x-axis as the ratio of the arc of a circle between it and the x-axis to the circumference of said circle. That this is independent of the chosen circle requires either the change of variables theorem (using a diagonal linear transformation as the diffeomorphism) or direct line integrals.

Having defined the angle as such, we can assign the trigonometric functions to the angle. We should allow for "looping" the circle, so that we get angles below 0 or above 2π. Then, one can prove various properties of trigonometric functions.

Another way is to start with power series expansions of $\sin$ and $\cos$. Then one can show that these infinite series are well defined, converge uniformly etc, and show that these series are smooth functions, periodic etc and obtain inverse functions. Again, we can prove different properties of these functions using the series expansions.

A third way is to use complex numbers. Since the angle is heuristically determined by its sine and cosine, we may take the angle to be a pair of numbers. Then we can put a structure on these numbers so that we have the familiar addition of angles etc. We also have a special angle 2π that functions as the 0 angle along with the zero angle. Basically this is a description of the circle group. Of course, to use complex numbers one first needs a definition of e^z for $z \in \mathbb{C}$. With that, we can define the $\cos(x), \sin(x)$ as the real and imaginary parts of e^{ix} respectively for $x \in \mathbb{R}$, or more properly for $x \in \mathbb{R}/2\pi\mathbb{Z}$, the circle group.

9 Appendix - (Unorganised) History of Stokes' Theorem

Keep in mind that, to early mathematicians, the meaning of $\int f dx$ was a function whose partial derivative with respect to x is f and a definite integral was the difference of this function between the limits.

On the divergence theorem, from [9]:

Divergence theorem in Cartesian form (T a volume with boundary S):

$$\int\int\int_T (\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z})dV = \int\int_S (P\cos\alpha + Q\cos\beta + R\cos\gamma)dA$$

where α, β, γ are the angles the normal makes to the x, y and z axes.

1828 - Green proves

$$\int\int\int_T [u\Delta v + \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} + \frac{\partial u}{\partial z}\frac{\partial v}{\partial z}dV = \int\int_S u\frac{\partial v}{\partial n}dA,$$

where Δ is the Laplacian operator and $\partial/\partial n$ is the directional derivative in the normal direction. This is a consequence of the divergence formula (setting $P = u\partial v/\partial x$ etc.) as is known today, but it appears that Green wasn't aware of the full divergence formula.

1813 - Gauss's paper contains identities relating surface and volume integrals, but again not in the general Cartesian form, however it contains all the major ideas involved.

1831 - Ostrogradskii publishes a paper containing the Cartesian form, however this was present in an unpublished paper from 1826-27. Method of proof was quite similaar to that of Gauss'.

1828 - Poisson also publishes a similar result in a memoir.

1893 - Heaviside presents a proof in vectorial form, but the proof relies on physical intuition.

On change of variables, from [10]:

1769 - To Euler, a double integral $\int\int Z dxdy$ was a function which when first differentiated with x, then with y gave Z. To evaluate the integral, he integrated one variable at a time, i.e. he used today's Fubini theorem. He was interested in finding out the change of variables formula, and he thought of changing the variables one at a time. With simple computations along these lines, he

arrived at a formula for the change in the area element, and verified its correctness using increasingly complex examples.

1776 - Lagrange discovered a change of variables formula for triple integrals. He was looking into the attraction of an ellipsoid on any point placed on or inside it. Again, his integration was done one variable at a time, but needed new coordinates so that the integral was easier to evaluate. He too tried obtaining the change of volume element by changing the coordinates on at a time. Legendre and Laplace used similar methods to evaluate complicated integrals.

1813 - Gauss was developing the idea of surface integrals. He started by parametrizing a surface with three function x, y, z of variables p, q. For small changes in the p, q space, he obtains a parallelogram area, which I denote by P, element in 3-space. Projecting P onto the coordinate planes, he gets 3 parallelograms with areas A_x, A_y, A_z. He then takes the area of P to be $\sqrt{A_x^2 + A_y^2 + A_z^2}$. He was able to compute surface integrals (for ellipsoids) using a special case of the divergence theorem. Setting the z-coordinate to be zero, he gets a change of variables theorem in 2 dimensions. Note that Gauss's method is considerably different from Euler's and Lagrange's.

1836 - Ostrogradskii provided the generalization of the change of variables theorem to n dimensions and extended Lagrange's proof. Given $X, Y, Z, \ldots$ as functions of $x, y, z, \ldots$, he obtains the formulas for $dX, dY, \ldots$ and then keeping $Y, Z, \ldots$ fixed obtains a formula for dX in terms of $dx, dy, \ldots$, and so on. He was able to reduce integrals on hypersurfaces in $n + 1$ dimensions to ordinary n dimensional integrals using this type of change of variables.

1841 - Carl Jacobi and Eugene Catlan publish papers giving clearly the generalization of the change of variables theorem. Catlan's proof was similar to Lagrange's. Jacobi's paper was the culmination of a series of articles about the theorem and included additional results such as the chain rule. It is due to this paper that the determinant appearing in the theorem is reffered to as the Jacobian.

1838 - Ostrogradskii published a proof of the theorem in 2 variables using the same idea as Gauss and critisizes the proof of Euler and Lagrange (and by implication his earlier proof). This time he stayed in 2 dimensions and calculated the area of an infinitesimal parallelogram element, where the sides where given by keeping one variable fixed, i.e. if the transformation is given by $(x(u, v), y(u, v))$, then the sides of the parallelogram are given by keeping u or v fixed.

1842 - DeMorgran publishes his textbook *Differential and Integral Calculus*. In it he considers integrals over regions in the plane where Fubini's theorem (which was the method to evaluate double integrals) does not apply because the region need not be bounded by nice curves. His method involved subdividing the region into those where u or v was constant, similar to Ostrogradskii's method. He also attempted to reason that infinitesimals of degree 3 (eg. du^3, du^2dv) and higher did not matter.

mid-1890s - Elie Cartan develops his theory of differential forms using Grassman rules of exterior algebras. He was able to readily prove the change of variables formula among other things.

On differential forms, from [11]:

In the mid-18th century,mathematicians considered one forms in two variables, i.e $Adx + Bdy$ where A, B are functions in two-space. For them, $Adx + Bdy = 0$ was simply the differential equation $dy/dx = -A/B$ and a solution is that function, if it exists, $f(x, y)$ such that $\partial f/\partial x = A, \partial f/\partial y = B$. If such an f exists, then $Adx + Bdy$ is the differential of f.

One forms:

1739, 1740 - Clairaut publishes two papers where he proves that the necessary and sufficient condition for $Adx + Bdy$ to be a differential is that $dA/dy = dB/dx$ (the notation of partial derivatives dates from 1840s). He also notes that the idea occured to him at about the same time as Alexis Fontaine (who never seems to have published it) and Euler.

To Clairaut, a function of two variables was a (possibly infinite) series $ax^m y^n + bx^p y^q + \ldots$. So, he computes the neccessity condition by an explicit computation and use of induction. for the sufficiency, assuming $dA/dy = dB/dx$, he asserts that the required function is of the form $\int Adx + p(y)$ where p is a function of y alone. For this f to be a solution, one needs

$$\frac{\partial f}{\partial x} = A; \frac{\partial f}{\partial y} = y.$$

The first is forced to be true, while the second requires

$$\frac{\partial f}{\partial y} = \int \frac{dA}{dy}dx + dp = \int \frac{dB}{dx}dx + dp = B + q(y) + dp$$

where $q(y)$ is another function in y. Recall that the integral was just the antiderivative. Therefore, it suffices to find p, q so that $q = -dp$. In other words, the problem reduces to one variable, which Clairaut assumed to be solvable. Note that

$$f = \int Adx + \int Bdy - \int dy[\int \frac{dB}{dx}dx].$$

In his 1740 paper, Clairaut extended the result to one forms in three variables. Again the neccesity condition is by computation. For the sufficiency, he starts by taking $f = \int Mdx$ and then uses the methods from the two variable case. He then stated the result for n variables. In modern notation, he proved that given an n-form $\omega, d\omega = 0$ if and only if $\omega = df$. This is not always true and in 1768 d'Alembert observed that it doesn't hold for $\omega = (xdy - ydx)/(x^2 + y^2)$.

1760 - Lagrange observed that 1 forms can be interpreted as line integrals, he observed that $ds = (dx^2 + dy^2 + dz^2)^{1/2}$ is a line element in a curve.

1811 - Lagrange came closer to the modern formulation of the line integral. Gauss writes in a letter to Bessel about integrating complex functions over curves.

1823 - Cauchy replaces the indefinite integrals in Clairaut's method with definite integrals along the edges of a rectangle with one corner fixed at some (x_0, y_0).

1825 - Cauchy carefully defined the integral of complex functions over curves.

1846 - Cauchy writes about about line integrals in n dimensions where the curve Γ is the boundary of some surface S and proves that if the integrand k is an exact differential, then the integral doesn't change if the curve changes slightly. Furthermore, he proves that if there are points $P', P'', \ldots$ where the integrand k fails to be continuous and if $\alpha, \beta, \ldots$ are curves surrounding $P', P'', \ldots$ respectively, then

$$\int_\Gamma kds = \int_\alpha kds + \int_\beta kds + \ldots.$$

He did not provide proofs of these results stating that they can be proved using the definite integrals along the edges of rectangles.

In the same year Cauchy also proved that integrating around singularities adds a specific constant (which he called *index of periodicity*) for each loop around the singularity. Riemann proved this rigourously.

1857 - Riemann develops the notion of multiply connected regions and the idea of cuts. He defines a simply connected region as one where every closed curve contains a regionBorrowing from [11], "A surface R said to be $n+1$-ply connected when n closed curves $A_1, A_2, \ldots, A_n$ can be drawn on it which neither individually nor in combination bound a region of R, while if augmented by any other closed curve A_{n+1}, the set bounds some region of R". He also observed that R is changed to an n-ply connected region by a cut (line is drawn from one boundary point to another passing through the interior) which does not disconnect it.

Using the idea of cuts, Riemann generalized the notion of Cauchy's index of periodicity.

1858 - Helmholtz extended the notion of multiply connected region to three dimensions and observed that some theorems in fluid dynamics and electromagnetism failed because the line integrals could not be considered as single valued functions on the regions (because of path dependency). This paper was translated into English in 1868.

1869 - William Thomson, who was investigating the motion of an ideal fluid inside a region with rigid boundary, used Helmholtz's definition and defined numbers similar to Riemann's constants and Cauchy's indices.

1873 - Maxwell generalized Thomson's idea to curves that looped around holes more than once.

Two forms 1811 - Lagrange defines the notion of an integral over a general surface. Earlier in 1788 he wrote down the integral over surfaces of the form $z = f(x, y)$. He was studying teh equilibrium

of fluids surrounding solids and was able to convert integrals of two forms into surface integrals (over surfaces in three-space).

1813 - Gauss whos was interested in attraction of ellipsoids obsereved that the integral depends on the angle of the normal to the axis and may change sign. He also used a version of divergence theorem, using geometric arguments.

1820s - Proofs of divergence theorem using surface integrals. However, explicit integrals of two forms were not in sight yet.

1850s - Stokes, Cauchy and Thomson study integrals of two forms. Jacobi observed that $d\omega = 0$ when ω is df for some 1-form f. Stokes proved the converse using solutions of Laplace's equation $\nabla U = f$. Thomson sketched a simpler proof in 1851 (see [11]) in a similar vein to Clairaut's proof for 1-forms.

Thomson needed the result in a discussion involving lamellar and solenoidal distributions of magnetism.

Generalizations of Ostrogradskii and Betti

1836 - Ostrogradskii generalized the divergence theorem to n dimensions, introduced the hyper-surface area elements and integrated $n-1$ forms. For some reason his work was largely ignored, and rediscovered by others.

1871 - Betti generalized Ostrogradskii's results and Stokes' Theorem to multiply connected spaces. He was strongly influenced by Riemann's work and had also worked in complex analysis. He also published a comprehensive definition of connectivity for an n dimensional object R. Betti compared n-fold integrals with $n-1$-fold integrals and generalising Stoke's result, he compared integrals of one forms with integrals of two forms. Using his definition of connectivity, he was also able to generalize the results of Thomson and Maxwell about integrals adding specific constants when the surfaces loop around "holes".

Generalizations of Poincare and Volterra

1887 - The condition that the integral of a one form along a curve depend only on the end points of the curve was called *integrability conditions* by Poincare. He sought to find similar conditions to surface integrals in higher dimensions, his motivations being to generalize the results of Cauchy, namely the notions of index, residues and the Cauchy's integral theorem.

He introduced a function of two complex variable and wanted to show that when the Cauchy-Riemann equations are satisfied, the integral over a closed surface vanishes and conversely. To show this, he turned to integrals of two forms in n dimensions and observed that skew symmetry is essential for that's how change of variables worked and he needed to parametrise surfaces so as to reduce the integrals into ordinary double integrals.

For $n = 3$, Poincare's result is the same as Betti's but while Betti compared n dimensional integral to $n-1$ dimensional integrals, Poincare used a calculus of variations method and needed only the two dimensional integral. Poincare was immediately abole to obtain similar conditions for the integral of n forms to depend only on the boundary of the region of integration.

Poincare, like Cauchy in 1846, considered connectivity of the domain only in the sense that the surface or its boundary shouldn't have singularities.

1889 - Volterra obtained the same conditions (namely $d\omega = 0$) while developing a theorey of "functions of lines". Poincare had shown that the integral of a function of two complex variables over a surface depends only on the "lines" bounding the surface.

In a series of papers, Volterra arrived at the modern Poincare's lemma and its converse. He also arrived at the generalized Stokes' Theorem. In essence he proved that the integral of an n-form is depends only on the boundary if and only if its differential is zero. To prove the necessity, we only need a direct calculation. Volterra proved the sufficiency by reducing the number of variables in steps - like Clairaut's proof.

Again, just like Clairaut, Volterra assumed certain differential equations can always be solved which is true locally, but not globally as d'Alembert had observed.

1895 - Poincare's *Analysis, situs*. Poincare refined the notion of connectivity of a domain and introduced the notions of homology and Betti numbers, which he further clarified in 1899.

On Stokes' Theorem, from [12]:

1813 - Gauss uses the divergence theorem in a limited sense to compute the volume of solids in three space.

1831 - Ostrogradskii publishes his general results for the Gauss divergence theorem in 3 variables. His proof involved taking cylinders parallel to the x, y and z-axes and then using fundamental theorem of calculus.

1828 - Poisson publishes a version, but he cites Ostrogradskii. Frederic Sarrus published a similar result, but his notation and ideas are not as clear as those of Ostrogradskii or Poisson. Same year, George Green also published the result in a different form using essentially the same method as Ostrogradskii. in a private publication

Each had a different motivation for the theorem:

- Gauss - Theory of magnetic attraction
- Ostrogradskii - Theory of heat
- Poisson - Theory of elasticity
- Green - Electricity and magnetism
- Sarrus - Floating bodies

1836 - Ostrogradskii generalized the divergence theorem to n dimensions with essentially the same proof as the three dimensional case.

1846 - Cauchy presents Greens theorem in two dimensions in the context of integrating complex functions over closed curves in the complex plane. He apparently never published a proof.

1851 - Riemann proves Cauchy's theorem and presents several related versions.

1854 - Stokes' Theorem appears in print in an exam paper. But Stokes had written to Kelvin about the result in 1850.

1861 - First published proof of Stokes' Theorem in a monograph by Henkel who didn't give anyone credit, only a reference to Riemann's work on the two dimensional version.

1870 - In the mid ninteenth century Hamilton, and later Tait, was developing quaternions. In 1870, Tait used quaternions to write the divergence theorem and Stokes' theorem in the vector form. Writing $\nabla = i\partial/\partial x + j\partial/\partial y + k\partial/\partial z$, given a vector function $\sigma = iX + jY + kZ$, we write the product

$$\nabla\sigma = (\frac{\partial X}{\partial x} + \frac{\partial Y}{\partial y} + \frac{\partial Z}{\partial z}) + i(\frac{\partial Z}{\partial y} - \frac{\partial Y}{\partial z}) + j(\frac{\partial X}{\partial z} - \frac{\partial Z}{\partial x}) + k(\frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y}).$$

Denote the scalar part of a quaternion p by Sp and the vector part by Vp. Then the divergence theorem becomes

$$\int\int\int S\nabla\sigma d\zeta = \int\int S(\sigma Uv)ds$$

where $d\zeta$ is an element of volume, ds surface and Uv the unit normal vector to the surface. Stokes' Theorem takse the form

$$\int S\sigma d\rho = \int\int S(V(\nabla\sigma Uv))d\zeta$$

where $d\rho$ is an element of length.

1873 - In his treatise Maxwell proposes to call $S\nabla\sigma$ as the convergence (negative of divergence in modern terminology) of σ and $V\nabla\sigma$ as its curl. Moreover, $S(pq)$ is the negative dot product of vectors p, q.

1889 - Volterra was teh first to include all three results under one general result. Apparently he did not provide a proof, but only said that it "is obtained without difficulty".

1899 - Poincare also presents the general result in a compact form. Like others from this period, Poincare was interested in integrability conditions, i.e., when is a form exact.

1899 - Ellie Cartan introduces differential forms and employs Grassman algebra. Cartan also defined the "derived expression" of a one form which is today's d operator.

1922 - Cartan extends his work on differential forms. He introduced the notions of exterior differential forms and exterior derivatives.

On Deahna's theorem, from [13]:

1740s - Euler, Caliraut and Fontaine found neccessary and sufficient conditions for 1-forms in 2 and 3 dimensions to be exact.

1814-1815 - Pfaff introduced the idea of finding integrable manifolds and submanifolds of arbitrary dimensions where a gven 1-form $a_1dx_1 + \cdots + a_ndx_n$ vanishes where a_i are smooth functions of $x_1, \ldots, x_n$. This became known as Pfaff's prblem. His answer was that one can find (locally) a system of functions $(f_1, \ldots, f_r, g_1, \ldots, g_r)$ or $(f, f_1, \ldots, f_r, g_1, \ldots, g_r)$ with independent differentials such that the given 1-form can be written as $\sum_{i=1}^{r} f_i dg_i$ or $df + \sum_{i=1}^{r} f_i dg_i$. The number r is called hte class of the 1-form.

Pfaff's work was continued in the direction of Frobenius theorem. In the language of vector fields, the question is: given vector fields $X_1, \ldots, X_r$, under what conditions can one find $n - r$ functions $f_1, \ldots, f_{n-r}$ such that the X_i nullify the f_j.

Jacobi treated the case where the vector fields commute, i.e. $[X, Y] = 0$, and A. Clebsch attempted in reducing the general case to Jacobi's case. The work was continued by many people and later Frobenius wrote a long paper reviewing what had been done before and gave his own proof, introducing the concept of exterior derivative of a 1-form. In this paper, Forbenius observes that F/ Deahna had given a proof earlier and Frobenius simplified it.

Later Ellie Cartan took up Pfaff's problem again in the language of differential forms. He also investigated a more general problem, namely that of finding integrals for a system of several Pfaffian forms.

References

[1] `https://en.wikipedia.org/wiki/History_of_calculus`

[2] *Elements* by Euclid

[3] `https://en.wikipedia.org/wiki/Erlangen_program`

[4] *Geometry: Euclid and beyond* R. Hartshorne

[5] `https://en.wikipedia.org/wiki/Wallace%E2%80%93Bolyai%E2%80%93Gerwien_theorem`

[6] `https://en.wikipedia.org/wiki/Dehn_invariant`

[7] *Analysis on Manifolds* J.R. Munkres

[8] *Introduction to Manifolds* Loring Tu

[9] Charles H. Stolze (1978), A History of the Divergence Theorem, Historia Mathematica, Vol. 5 Issue 4, pp. 437-442

[10] Victor J. Katz (1982), Change of variable in multiple integrals: Euler to Cartan, Mathematics Magazine, Vol. 55 No. 1, pp. 3-11

[11] Victor J. Katz (1981), The history of differential forms from Clairaut to Poincare, Historica Mathematica, Vol. 8 Issue 2, pp. 161-188

[12] Victor J. Katz (1979), The history of Stokes' Theorem, Mathematics Magazine, Vol. 52 No. 3, pp. 146-156

[13] Hans Samelson (2001), Differential Forms, the Early Days; or the Stories of Deahna's Theorem and of Volterra's Theorem, The American Mathematical Monthly, Vol. 108 No. 6, pp. 522-530