A Practical Introduction to Differential Forms

Alexia E. Schulz

and

William C. Schulz

March 17, 2016

Transgalactic Publishing Company
Flagstaff, Vienna, Cosmopolis

Every creator painfully experiences the chasm between his inner vision and its ultimate expression. Isaac Bashevis Singer

Dedicated to our parents, children, and cats

Contents

Chapter 1

Introduction and Basic Applications

1.1 Introduction

This book began life as an introduction to differential forms for a mathematical physics class and it still retains some of that flavor. Thus the material is introduced in a formal manner and the mathematical complexities are put off to later sections. We have tried to write so that those whose primary interest is in the applications of differential forms can avoid the theoretical material provided they are willing to accept the formulas that are derived in the mathematical sections, which are clearly marked as such. Those who wish may read the mathematical sections as they occur, or later, or indeed may put them off to a more convenient time, perhaps in a future life, without loss to the continuity of the applied thread. Anyway, such is our hope. But we want to also emphasize that those who wish will find all the mathematical details available, at a level of rigor common to the better mathematical physics books. The treatment is mostly local, and what little manifold theory is needed is quietly developed as we go. We have tried to introduce abstract material in circumstances where it is useful to do so and we have also tried to avoid introducing a lot of abstract mathematical material all at one time.

The two areas most completely addressed in this book, besides the foundational material, are coordinate changes and Maxwell's equations since we feel that these illustrate the power of differential forms quite well. We treat Maxwell's equations in both three and four dimensions in separate and independent sections. We will also look at a number of geometric applications.

Notation has been carefully chosen to be consistent with standard tensor notation to facilitate comparison with such treatments, and to facilitate learning basic differential geometry.

The treatment of Maxwell's equations requires the derivation of the potential equations. Although not strictly necessary, we have introduced the codifferential δ and the Laplace operator $\triangle = d\delta + \delta d$ since this is the natural route using modern mathematics. For example we point out that the condition of Lorenz can be expressed instantly and easily in terms of the codifferntial in four dimensions. And as long as we have it available we look at a couple of other applications of the Laplace operator on forms.

A justified criticism of this book might be that many things are done twice, which is not efficient. We have sacrificed efficiency for convenience to the reader who may wish to deal with only one particular thing, and so would like a relatively complete treatment in the section without having to read five others. Similarly, many formulas are repeated at the beginning of sections where they are used, rather than referred to in previous sections. The increase in paper is rather small, and for those reading electronically there is no waste at all. It is difficult for a mathematician to resist the call of generality but since one of us is a physicist the brakes have been applied, and we hope that the product is a reasonable compromise between the song of lovely mathematical sirens and the needs of practical physics and engineering.

1.2 Some Conventions

Here we will introduce some conventions that will be used throughout these notes. The letter A will be used for a region of 2-dimensional space, for example the unit disk consisting of points whose distance from the origin is less than or equal to 1. It's boundary would be the unit circle consisting of points whose distance from the origin is exactly 1. We will use the symbol ∂ to indicate the boundary. Thus if A is the unit disk $A = \{x \in \mathbb{R}^2 \,|\, |x| \leq 1\}$ then the boundary of A is $\partial A = \{x \in \mathbb{R}^2 \,|\, |x| = 1\}$ which is the unit circle. Notice carefully the difference between the terms DISK and CIRCLE. (DISK and CIRCLE are often confused in common speech.)

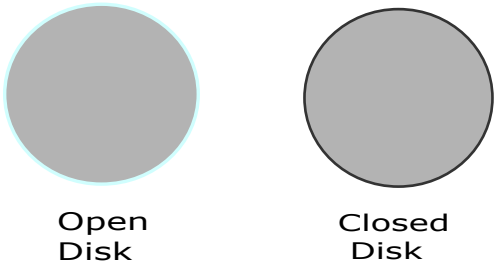
The letter M will be used for a (solid) region of 3 dimensional space, for example the unit ball, $M = \{x \in \mathbb{R}^3 \,|\, |x| \leq 1\}$ whose boundary is the unit sphere $\partial M = \{x \in \mathbb{R}^3 \,|\, |x| = 1\}$. (The terms BALL and SPHERE are often confused in common speech, particularly in cases like a beach ball or a basketball since they are filled with air.)

The letter S will be used for a (2 dimensional) surface in three dimensional space, for example the upper half of the unit sphere. The boundary of this S would be a circle in the x, y plane.

If we do not wish to specify dimension, we will use the letter K. The use of K indicates that the formula will work in any dimension, and this usually means *any* dimension, not just 1, 2 or 3 dimensional space. Naturally ∂K means the boundary of K.

The ball and sphere have analogs in every dimension. It is customary to refer to the ball in $\mathbb{R}^n$ as the n-ball and its boundary as the $(n-1)$-sphere. For example, the unit disk is the 2-ball and its boundary, the unit circle, is the 1-sphere. Note that the m-sphere lives in $\mathbb{R}^{m+1}$. It is called the m-sphere because it requires m variables to describe it, like latitude and longitude on the 2-sphere.

Also useful to know are the terms open and closed. This is a tricky topological concept, so we will treat it only intuitively in the first chapter; it is treated much more fully in Chapter 2. K is *closed* if it includes its boundary.



Open and closed disks

Thus the unit disk and unit ball are closed. If we remove the boundary ∂K from K the resulting set K° is called *open*. Thus for the unit ball in $\mathbb{R}^3$ we have

$$M \quad = \quad \{x \in \mathbb{R}^3 \,|\, |x| \leq 1\} \quad \text{closed 3-ball}$$

$$\begin{aligned} M^\circ &= \{x \in \mathbb{R}^3 \,|\, |x| < 1\} \quad \text{open 3-ball} \\ \partial M &= \{x \in \mathbb{R}^3 \,|\, |x| = 1\} \quad \text{2-sphere} \end{aligned}$$

We want to give a real world example here but remember it must be inexact since real world objects are granular (atomic) in constitution, so they can only approximate perfect mathematical objects. Some people prefer to eat the closed peach (with fuzzy skin), some people prefer the open peach (fuzzy skin removed, peach$^\circ$) and the boundary of the peach, ∂peach, is the fuzzy skin. Perhaps this will help you remember. Deeper knowledge of these matters can be found in [5] and in the wonderful book [2].

For functions we will use a slightly augmented variant of the physics convention. When we write $f : S \to \mathbb{R}$ we mean a function whose input is a point $p \in S$ for some set S and whose output is a real number. ($\mathbb{R}$ will always be the set of real numbers.) This is theoretically useful but not suitable for calculation. When we wish to calculate, we need to introduce coordinates. If we are dealing with the upper half of the unit sphere (set of points in $\mathbb{R}^3$ whose distance from the origin is exactly one and for which $z \geq 0$) then we might write $f(x, y)$ if we choose to represent points in the x, y coordinate system. Notice, and this is an important point, that the coordinate x takes as input $p \in S$ and outputs a real number, it's x coordinate. Hence the coordinates x and y are functions just like f. If S is the upper half of the unit sphere in $\mathbb{R}^3$ then x and y are not actually good coordinates. It would be be better to use longitude and colatitude for our coordinates and then we would write $f(\phi, \theta)$ [1].

It is important to note that phycics and mathematics have different *conventions* in the use of function letters. A function letter, e.g. f, in physics represents a *quantity*, and thus the letter remains the same no matter what coordinate system is being used. In mathematics a letter represents a functional relationship, and when you change the coordinates you (theoretically) must change the function letter. However, in neither case is total consistency in this matter achieved. We will usually use the physics convention. Note that $f(.5, .3)$ is ambiguous in physics until you have specified the coordinate system, whereas in mathematics the coordinate system was specified with the function so $f(.5, .3)$ is unambiguous.

Finally, we will almost always use the letters f, g, h for functions on A, M, S, K. Mostly these will occur in coordinate form, for example $f(x, y, z)$ for a function on M.

Problems for Section 1.2

1. Draw a picture showing a disk and indicating its boundary (which is a circle).

2. Draw $A = \{x \in \mathbb{R}^2 |\ 1 \leq |z| \leq 2\}$ (a two dimensional donut or more properly torus). Indicate the boundary (two circles) and draw an arrowhead

[1] BEWARE. ϕ is longitude in physics but colatitude in mathematics. θ is colatitude in physics but longitude in math. We will use the physics convention.

on each so that when you walk around the circle in the direction indicated your left arm points into the torus and your right arm into empty space.

3. Draw a picture of the unit ball and indicate it's boundary (the unit sphere).

4. Try to draw a picture of $M = \{(x, y, z) \in \mathbb{R}^3 |\ 1 \leq |(x, y, z)| \leq 2\}$ which is a peach with the pit magically removed. Not so easy to draw is it? What is the boundary.

5. Draw a picture of M, a bagel or donut. Indicate the boundary $S = \partial M$ which is called the 2-torus. M itself, the entire bagel or donut, is called the *solid torus*. The plural is *tori*. Tori are a big deal in abstract mathematics and there is one for every dimension. Note it is not so easy to describe M or S using (x, y, z) coordinates.

1.3 Some Formulas to Recall

You are all familiar with the dx, dy, dz which occur in the derivative notation $\frac{dy}{dx}$ and the integral notation

$$\int_A f(x, y)\, dxdy$$

$$\int_A f(x, y, z)\, dxdydz$$

and you recall the Green, divergence and Stokes theorems, which I list here for convenience:

Green's theorem

$$\int_{\partial A} f(x, y)\, dx + g(x, y)\, dy \quad = \quad \int_A \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\, dxdy$$

The divergence theorem or Gauss's theorem

$$\int_{\partial M} f(x, y, z)\, dydz + g(x, y, z)\, dzdx + h(x, y, z)\, dxdy \quad = \quad \int_M \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z}\, dxdydz$$

The Classical Stokes' theorem

$$\int_{\partial S} f(x, y, z)\, dx + g(x, y, z)\, dy + h(x, y, z)\, dz$$

$$= \int_M \left(\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}\right) dydz + \left(\frac{\partial f}{\partial z} - \frac{\partial h}{\partial x}\right) dzdx + \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dxdy$$

You might be more familiar with the last two in the vector forms

$$\int_{\partial M} \mathbf{v} \cdot \mathbf{dS} = \int_M \text{div } \mathbf{v}\, dV$$

and

$$\int_{\partial S} \mathbf{v} \cdot \mathbf{d}\ell = \int_S \text{curl } \mathbf{v} \cdot \mathbf{dS}$$

where $\mathbf{v} = f\,\hat{\mathbf{i}} + g\,\hat{\mathbf{j}} + h\,\hat{\mathbf{k}}$ and $\mathbf{dS} = \hat{\mathbf{n}}\,dS$ where $\hat{\mathbf{n}}$ is the *exterior* unit normal.

There are some conventions on integrals that we will mention now. In former times when integrating over a three dimensional object we would write $\int\int\int_M \text{div } \mathbf{v}\,dV$ This is now completely antiquated, and we will not do it.

On the other hand, there is a convention that when integrating around curves or surfaces that have no boundary we put a small circle on the integral, so that we write

$$\oint_{\partial M} \mathbf{v} \cdot \mathbf{dS} \text{ for } \int_{\partial M} \mathbf{v} \cdot \mathbf{dS}$$

Since this is favored by the physics community we will mostly use it. Notice that if a geometric object is the boundary of something, then it itself has no boundary, and so we will use the circled integral almost exclusively with boundaries.

For our purposes we will define a differential form to be an object like

$$f(x,y)\,dx \qquad f(x,y,z)\,dydz \qquad f(x,y,z)\,dxdydz$$

which we find as integrands in the written out forms of the Green, divergence and Stokes theorem above. If ω is a sum of such objects, it turns out that the three theorems collapse to one mighty theorem, called *the generalized Stokes theorem*, which is valid for all dimensions:

$$\oint_{\partial S} \omega = \int_S d\omega$$

To use this theorem and for other purposes it is only necessary to

1. Learn the algebra that the dx, dy, dz satisfy which is almost the same as ordinary algebra with one important exception.

2. Learn the rules for the operator d which are almost trivial.

Once these are learned differential forms can be manipulated easily and with confidence. It is also useful to learn how various things that happen in vector analysis can be mimicked by differential forms, and we will do this, naively at first and then in much more detail.

If you are concerned about what differential forms ARE, the answer is a little tricky and we are going to put it off until Chapter 2 where we will discuss the surprisingly dull answer to this question. Incidentally, the difficulty in explaining what they really are is one reason they have not become more common in elementary textbooks despite their extreme usefulness.

Just to give a tiny hint of the geometrical interpretation of differential forms: a two form measures the density of lines of force of a field, as discussed by James Faraday a century and a half ago. For more on this subject see [4] or [9]. We will discuss it further in chapter 2 when we have more machinery.

Problems for Section 1.3

The problems are just practise with Green's, divergence (= Gauss's) and Stoke's theorems. These problems will go easier if you use the coordinate forms of the the theorems rather than the vector forms.

1. Verify Green's theorem for $\mathbf{v} = x^2y\,\hat{\mathbf{i}} + xy^2\,\hat{\mathbf{j}}$ and A is the square with corners (0,0), (1,0), (1,1), (0,1). Remember to go around the square in that order, which involves putting the upper and lower limits on the integrals in the correct order.

2. Verify the divergence (= Gauss's) theorem for $\mathbf{v} = x^2yz\,\mathbf{i} + xy^2z\,\mathbf{j} + xyz^2\,\mathbf{k}$ and M is the cube with corners (0,0,0), (1,0,0), (0,1,0), (0,0,1), (0,1,1), (1,0,1), (1,1,0), (1,1,1).

3. Verify Stokes' theorem for $\mathbf{v}$ as in 2. and S the box *with no top* and corners as in 2. ∂S will then be five squares, (sides, bottom but no top).

4. Let A be an area in the plane. Use Green's theorem to find what

$$\frac{1}{2}\int_{\partial A} x\,dy - y\,dx$$

is equal to. There are machines that utilize this to find *what* by running the machine over ∂A?

1.4 Coordinate systems

Our first job is to talk a little about coordinates. You already know most of this so we can do it quickly. As already specified, we will use the notation A for a finite region in $\mathbb{R}^2$ (which is the standard notation for ordinary two dimensional real space). We will use M for a three dimensional finite region of $\mathbb{R}^3$ and S for a curved surface in $\mathbb{R}^3$. Standard coordinates for A would be x, y, but we might want to use polar coordinates r, θ or even more general coordinates u, v. The important thing for A is that there be *two* coordinates and that they be "independent", so that v is not a function of u.

In the case of M, a three dimensional region of $\mathbb{R}^3$, we will need three coordinates x, y, z or r, θ, ϕ or more generally u, v, w to describe it.

Since S is a surface in $\mathbb{R}^3$, it will be described by two coordinates. In elementary cases this will often be x, y but in more complicated situations it is often wise to taylor[2] the coordinates to the geometric object, for example if one is working with the Unit Sphere in $\mathbb{R}^3$ then the appropriate coordinates would be θ, ϕ (longitude and colatitude). Proper choice of coordinates can make a nasty problem much more pleasant.

It is important to be able to move from one coordinate system for a geometric situation to a different coordinate system and we will discuss this later.

Problems for Section 1.4

[2] bad pun

1. How many coordinates does it take to describe an egg including the inside? How many to describe the shell (ignore thickness)?

2. How many coordinates does it take to describe the position of the points in $\mathbb{R}^4$? How many coordinates to describe position on the three sphere

$$(x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2 = 1$$

3. Imagine you have been drafted and are attending artillery school. You are studying the flight of cannonballs. How many coordinates does it take to describe what is happening with a cannonball in flight? You must have coordinates for the position, the velocity, and better put in the time too.

4. Same problem as 3, but assume you are living in a 2 dimensional world. Now how many coordinates? OK, now you are living in a four dimensional world. Now how many coordinates?

5. Same problem as 3, but assume you are not shooting a cannonball but a cat. For some reason it is important that you know how Fluffy is oriented as he flies through the air. Now how many coordinates, and be careful; the first guess here is often not right.

1.5 The Algebra of Differential Forms

We now begin our discussion of the algebra of differential forms. The type of multiplication we will use is called *exterior multiplication*. The term goes back to Grassmann and he invented it to contrast with interior multiplication, which is what we now refer to as the *inner product*. The multiplication symbol used is $\wedge$ as in $dx \wedge dy$ but this is very often omitted. We will use it for a while, and then omit it when we get bored with it, but return to using it whenever we think it clarifies matters.

Let us start with a function in $f(x, y, z)$ on $\mathbb{R}^3$. You already know how to form df:

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

Recalling that, like f, the coordinate x is also a function on $\mathbb{R}^3$ the previous formula writes the differential of f in terms of the differentials of the three *special* functions x, y, z. So we note that there is not much difference between df and dx; they are the same kind of object. All objects of this type (differentials of functions) are collected together in the set

$$\Lambda^1(\mathbb{R}^3)$$

and are called 1-forms. It is important to understand that not all one forms have the form df for some function f. Anything that looks like

$$g_1(x, y, z)\, dx + g_2(x, y, z)\, dy + g_1(x, y, z)\, dz$$

is a 1-form and so is in $\Lambda^1(\mathbb{R}^3)$.

If we multiply a couple of these together we get objects like

$$f\,dx \wedge g\,dy = fg\,dx \wedge dy \in \Lambda^2(\mathbb{R}^3)$$

Note that functions f commute with dx: $f\,dx = dx\,f$; see below. Linear combinations of such objects are called 2-forms. And of course there are 3-forms

$$f\,dx \wedge dy \wedge dz \in \Lambda^3(\mathbb{R}^3)$$

To complete the system we will place the functions in the basement of the building: $f \in \Lambda^0(\mathbb{R}^3)$. It is customary to omit the wedge when multiplying by a function; we write

$$f\,dx \wedge dy \quad \text{for} \quad f \wedge dx \wedge dy$$

There is no significance to this; it is *just* convention.

The algebra of these objects is just like ordinary algebra except for the changes caused by the rule

$$dg \wedge df = -\,df \wedge dg \qquad \text{(anti-commutativity)}$$

for the 1-forms df and dg. An algebra satisfying this rule is called an exterior or Grassmann algebra. This algebra was invented by Hermann Guenther Grassmann about 1840 in an attempt to find an algebra mirroring elementary geometry. It is sufficient to postulate this rule for the coordinate differentials only,

$$dy \wedge dx = -\,dx \wedge dy \quad \text{etc.}$$

since the general rule will follow by linearity.

Thus the exterior algebra is *not* commutative. Our rule is often called anti-commutative and is the simplest generalization of commutative, but it has many consequences which seem strange to the beginner. For example, if we substitute f for g in the rule $dg \wedge df = -\,df \wedge dg$ we get

$$df \wedge df = -\,df \wedge df$$

so moving the right side to the left side by the usual algebraic processes which are all valid here we get

$$\begin{aligned} df \wedge df + df \wedge df &= 0 \\ 2\,df \wedge df &= 0 \\ df \wedge df &= 0 \end{aligned}$$

Thus *the product of a one form with itself is* 0, which is very important. Let's look at another example

$$\begin{aligned} (f\,dx + g\,dy) \wedge (f\,dx + g\,dy) &= ff\,dx \wedge dx + fg\,dx \wedge dy + gf\,dy \wedge dx + gg\,dy \wedge dy \\ &= 0 + fg(dx \wedge dy + dy \wedge dx) + 0 \\ &= fg\,0 \;=\; 0 \end{aligned}$$

as promised.

We also see from this that there are no four forms in 3-space, since if we multiply four of the objects dx, dy, dz together there will be a repetition which will kill the form:

$$dx \wedge dy \wedge dz \wedge dx = -\,dx \wedge dy \wedge dx \wedge dz = dx \wedge dx \wedge dy \wedge dz = 0 \wedge dy \wedge dz = 0$$

In general, for a space K of dimension n there will be forms $\omega \in \Lambda^j(K)$ for $j = 0, \ldots, n$. *It is not true in general that for r-forms ω with $r > 1$ that $\omega \wedge \omega = 0$ although this is fortuitously true for dimensions $n \leq 3$.* There is an example in dimension 4 where $\omega \wedge \omega \neq 0$ in the problems.

Now let us start to deal a little more abstractly, so we will use a region K of dimension n, and consider $\omega \in \Lambda^j(K)$ and $\eta \in \Lambda^k(K)$. Then a little thought shows that

$$\eta \wedge \omega = (-1)^{jk} \omega \wedge \eta$$

For example, with $\omega = dx \in \Lambda^1(\mathbb{R}^3)$ and $\eta = dy \wedge dz \in \Lambda^2(\mathbb{R}^3)$ we have

$$\begin{aligned} (dy \wedge dz) \wedge dx &= dy \wedge dz \wedge dx \\ &= -\,dy \wedge dx \wedge dz \\ &= dx \wedge dy \wedge dz \end{aligned}$$

and if you look at how this special case works you will see why the general case works.

Note that nothing prevents us from mixing things up as in

$$2\,dx + 3\,dx \wedge dy$$

but such things do not occur in practise. Forms where each term has the same number of differentials (forms of the same *degree*) are called *homogeneous*, and we almost always use homogeneous expressions.

1.6 The Operator d

Our next project is the differential operator d, which we introduce according to the following four rules, where K is a space with coordinates $u^1, \ldots, u^n$:

d is the unique operator that satisfies the following laws

1. d is a linear operator: $d(\alpha\omega + \beta\eta) = \alpha\, d\omega + \beta\, d\eta$

2. On the functions $\Lambda^0(K)$, df is given by the advanced calculus formula

$$df = \frac{\partial f}{\partial u^1}\, du^1 + \ldots + \frac{\partial f}{\partial u^n}\, du^n$$

3. if $\omega \in \Lambda^j(K)$ and $\eta \in \Lambda^k(K)$ then (Leibniz's Rule)

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^j \omega \wedge d\eta$$

4. $dd = 0$ (and thus $ddf = 0$ and $ddu^i = 0$)

We have used coordinates $u^1, \ldots, u^n$ rather than $x^1, \ldots, x^n$ to emphasize that these coordinates need not be rectangular or even orthogonal.

Let's look at some examples in 3-space of how these rules make everything work. First we examine $d(f\,dx)$. Since $f \in \lambda^0(\mathbb{R}^3)$ we have, using rule 3,

$$d(f\,dx) = df \wedge dx + (-1)^0 f \wedge ddx = df \wedge dx + 0$$

we also used rule 4, $ddx = 0$, in the last equation. This derives the practical rule in a special case, and the general case (see problems) will be the same.

$$d(f\,dx^{i_1} \wedge \ldots \wedge dx^{i_k}) = df \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

This is the practical rule for d and the one you will use for almost everything you do, so learn it well.

Now let $\omega = f\,dx$ and $\eta = g\,dy$. Then we have

$$\begin{aligned} d(f\,dx \wedge g\,dy) &= d(f\,dx) \wedge g\,dy + (-1)^1 f\,dx \wedge d(g\,dy) \\ &= (df \wedge dx) \wedge g\,dy - f\,dx \wedge (dg \wedge dy) \\ &= df\,g \wedge dx \wedge dy + f\,dg \wedge dx \wedge dy \\ &= (df\,g + f\,dg) \wedge dx \wedge dy \\ &= d(fg)\,dx \wedge dy \end{aligned}$$

just as we expected from the practical rule since $f\,dx \wedge g\,dy = fg\,dx \wedge dy$. This also illustrates how rule 2 is a generalization of Leibniz's formula for the derivative of a product.

Problems for Section 1.6

Several of the problems below are connected to the potential of an electric field of a point source in two, three or n dimensions.

In $\mathbb{R}^2$ find the following

1 a. $d(-y\,dx + x\,dy)$

b. $d(y\,dx + x\,dy)$

c. Find a function $f(x, y)$ so that $df = y\,dx + x\,dy$

d. $d(-\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy)$

e. This one is harder. Find a function θ in the x, y plane so that $d\theta = -\frac{y}{x^2+y^2}\,dx + \frac{x}{x^2+y^2}\,dy$. Hint: Let θ be the polar coordinate variable and express it in terms of x and y. There is an apparant problem when $x = 0$. Ignore it. There is a much bigger problem since θ is *not* a single valued function on the plane. This means that our solution works locally for certain regions in $\mathbb{R}^2$ but not globally in $\mathbb{R}^2$.

f. Let $r = (x^2 + y^2)^{\frac{1}{2}}$. Note $r^2 = (x^2 + y^2)$ so $\frac{\partial r}{\partial x}$ is most easily found by implicit differentiation. Find dr in terms of dx and dy. Compute $d(dr)$ which of course is 0. Now find $d(-\ln r) = d\ln(\frac{1}{r})$.

The following problems refer to $\mathbb{R}^3$.

2 a. $d((x^2 + y^2)\,dx + (x^2 + z^2)\,dy + (x^2 + y^2)\,dz$

2 b. $d(2xy^2z\,dx + 2x^2yz\,dy + x^2y^2\,dz)$

2 c. Find $f(x, y, z)$ so that $df = 2xy^2z\,dx + 2x^2yz\,dy + x^2y^2\,dz$ by guessing the answer and checking if it's correct.

2 d. $d(x\,dydz + y\,dzdx) + z\,dxdy)$. Here when I write $dydz$ I of course mean $dy \wedge dz$. This abbreviation is used very frequently when calculating with differential forms[3]. It is of course critical to keep the differentials in the given order with or without the wedge. Also, you are expected to simplify the answer when possible; in this case the final answer has one term.

2 e. $d\big((x + y^2)\,dydz + (y + z^2)\,dzdx + (z + x^2)\,dxdy\big)$

Let $r = (x^2 + y^2 + z^2)^{\frac{1}{2}}$

3 a. Find $\frac{\partial r}{\partial x}$. You may want to square r and use implicit differentiation.

3 b. Use symmetry to find $\frac{\partial r}{\partial y}$ and $\frac{\partial r}{\partial z}$.

3 c. What is $\frac{\partial}{\partial x}(\frac{1}{r})$? What is $d(\frac{1}{r})$? This is related to the potential of an electric particle, e.g. an electron.

The philosopher Kant realized that this material could be generalized to $\mathbb{R}^n$. We use $r = (x^{12} + x^{22} + \cdots x^{n2})^{\frac{1}{2}}$ and the potential is given by $d(\frac{1}{r^{n-2}})$ for $n \geq 3$. Thus electrons in five spacial dimensions have potential proportional to $\frac{1}{r^3}$.

4 a. Using 3. as a model, compute $d(\frac{1}{r^{n-2}})$.

1.7 Orientation

At this point we must deal with one of the less pleasant aspects of elementary geometry which is orientation. An orientation, to speak loosely, is a sense of twist in a space. For example, in $\mathbb{R}^2$ our standard sense of twist is counterclockwise; we measure angles from the x axis in the direction of the y axis. If we reverse either axis, we get the opposite twist. If the y axis went *down* instead of *up* then we would measure angles in a clockwise direction and $\mathbb{R}^2$ would have the *opposite* orientation. If we reverse both axes then the sense of twist returns to counterclockwise.

[3] Alfred Tarki referred to this as the method of systematic forgetting.

In $\mathbb{R}^3$ the standard orientation is given by the following rule: if you place the fingers of your **RIGHT** hand so that the fingers curl from the x to the y coordinate axes then your thumb points in the direction of the z axis. This is called the *right hand rule*. It has become standard to use this in Calculus books only since the 1940's, and the opposite convention is still found in Italian books, so it is wise to check. To appreciate the subtlety of this concept, think of trying to communicate it to inhabitants of a planet in another galaxy. Since nature is almost symmetric in terms of left and right, the only way we know to clarify this is certain non-symmetric aspects of beta decay. This is referred to in physics as parity. Hopefully parity properties remain the same from galaxy to galaxy.

This idea of orientation, though subtle, is strongly coupled with differential forms and is the reason for the anticommutativity. It also is a prime source of mistakes, and great care must be taken to keep things in correct order. In $\mathbb{R}^2$ (and it's subspaces) the correct order, which expresses the orientation properly, is $dx \wedge dy$. Thus $dy \wedge dx$ is in incorrect order as indicated by the minus sign in

$$dy \wedge dx = -\, dx \wedge dy$$

Recall that when using Green's theorem

$$\oint_{\partial A} f(x,y)\, dx + g(x,y)\, dy = \int_A \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\, dxdy$$

it is *critical* that the direction around the boundary of the left integral be *counterclockwise.* If it is taken clockwise then the two sides of the equation will have opposite signs. This is again due to the orientation which is built into $\mathbb{R}^2$ but which we seldom notice explicitly. There are similar worries in the use of the divergence theorem and Stokes theorem.

In applications, the principal place where orientation occurs in $\mathbb{R}^n$ is in n-forms and $(n-1)$-forms. We will first tell you the the general formula and then give you practical methods to make orientation (relatively) easy to deal with. Let the variables used in $\mathbb{R}^n$ be $u^1, u^2, \ldots, u^n$. (We have switched from $x^1, x^2, \ldots, x^n$ to $u^1, u^2, \ldots, u^n$ to emphasize that the variables can be general; not necessarily rectangular or orthogonal.) The use of superscripts to number the variables is to conform to tensor analysis standards and we don't need to go into the reasons for it here; just do it! And remember u^3 is the third variable, not the cube of u. If we choose an *order* for the variables, which we did by numbering them, this chooses one of the two orientations. Then

$$\begin{aligned} du^1 \wedge du^2 \wedge \ldots \wedge du^n &\text{ is in correct order} \\ du^2 \wedge du^1 \wedge \ldots \wedge du^n &\text{ is in incorrect order} \end{aligned}$$

because

$$du^2 \wedge du^1 \wedge \ldots \wedge du^n = -\, du^1 \wedge du^2 \wedge \ldots \wedge du^n$$

As you can see with a little practice, interchanging any two of the du^i reverses the sign and changes correct to incorrect order or incorrect to correct order. If you are familiar with permutations, odd permutations of $du^1 \wedge du^2 \wedge \ldots \wedge du^n$ give incorrect order and even permutations give correct order.

That part is easy. The tricky part is the $(n-1)$-forms. Here the correct order is (with du^i missing from the list)

$$(-1)^{i-1} du^1 \wedge \ldots \wedge du^{i-1} \wedge du^{i+1} \wedge \ldots \wedge du^n$$

The reason for this choice is so that

$$du^i \wedge ((-1)^{i-1} du^1 \wedge \ldots \wedge du^{i-1} \wedge du^{i+1} \wedge \ldots \wedge du^n) = du^1 \wedge \ldots \wedge du^n$$

which is correct because the du^i must hop over the $i-1$ elements $du^1, \ldots, du^{i-1}$ in order to get back into correct order and each hop contributes a minus sign.

So much for theory. In $\mathbb{R}^3$ correct order is

$$dx \wedge dy \wedge dz \qquad \text{correct order}$$

and for 2-forms we have

$$dy \wedge dz, \quad -dx \wedge dz, \quad dx \wedge dy \qquad \text{correct order}$$

For practical use it is much better to write these in cyclic order.[4]

$$dy \wedge dz, \quad dz \wedge dx, \quad dx \wedge dy \qquad \text{correct order}$$

and the correct order can be easily remembered by writing

$$dxdydzdxdydz$$

and noting that the order of a wedge of two differentials is *correct* if it occurs in this list, for example $dz \wedge dx$ is correct but $dx \wedge dz$ is incorrect since $dxdz$ does *not* occur in the list. Other incorrects are $dy \wedge dx$ and $dz \wedge dy$. The use of differential forms in $\mathbb{R}^3$ relies critically on writing things with correct order.

1.8 Differential Forms and Vectors

Standard 3-dimensional vector analysis was cobbled together by Josiah Willard Gibbs in the 1890s using pieces from a variety of sources. While it works well for many practical purposes it has large deficiencies from a theoretical standpoint. Those parts of it which concern the dot (scalar) product are fine, but those parts which concern the cross (vector product $\mathbf{v} \times \mathbf{w}$) are mathematically clumsy. To see this, consult any textbook for a proof of the vector triple product law

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$$

[4] Cyclic order is a 3-space concept and does not generalize to n-space at all well.

It is often said that the cross product cannot be generalized to higher dimensions but this is not true; what is true is that the analog of the cross product in n dimensions involves not two but $n-1$ vectors. Thus the elementary geometric applications of the cross product can often be reproduced, but the physical applications not so much, which is one of the reasons that for relativity (4 dimensions and space-time rather than just space) we fall back on Tensor Analysis.

In 3 dimensions there are many formal analogies between differential forms and vector analysis. That is, differential forms will act like vectors in many ways. It is a little difficult to find good mathematical reasons for these analogies and we will not concern ourselves here with what these reasons might be, although we will return to the matter later. The practical consideration is that things will work well if we keep in mind the following rules. For 1-forms we have the correspondences

$$dx \leftrightarrow \hat{\mathbf{i}} \qquad dy \leftrightarrow \hat{\mathbf{j}} \qquad dz \leftrightarrow \hat{\mathbf{k}}$$

For 2-forms we have the correspondences

$$dy \wedge dz \leftrightarrow \hat{\mathbf{i}} \qquad dz \wedge dx \leftrightarrow \hat{\mathbf{j}} \qquad dx \wedge dy \leftrightarrow \hat{\mathbf{k}}$$

Note that in the second case we have been careful to place the 2-forms in proper order. This is critical. If you screw this up you will get the wrong sign. Watch the $dz \wedge dx$ term particularly carefully.

With these correspondences kept in mind, we can easily derive many formulas of vector analysis in simple ways. Many things which appear different in vector analysis can be treated in a unified way with differential forms.

1.9 grad, curl and div

In this section we show how the three vector operators grad, curl and div occur naturally in the context of differential forms. Since we are in $\mathbb{R}^3$ we have 0-forms (functions), 1-forms, 2-forms and 3-forms. The d operator vanishes on 3-forms as previously discussed. Hence we look at d on 0-forms, 1-forms and 2-forms.

First 0-forms. Let f be a function (0-form) of the coordinates x, y, z. Then df is just the old Calculus df:

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

and we see here the components of the gradient

$$\text{grad} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$$

Next we do 1-forms where we omit terms that are 0:

$$\begin{aligned} d(f\,dx + g\,dy + h\,dz) &= df \wedge dx + dg \wedge dy + dh \wedge dz \\ &= \left(\frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz\right) \wedge dx \end{aligned}$$

$$
\begin{aligned}
&+ \quad \left(\frac{\partial g}{\partial x}\,dx + \frac{\partial g}{\partial y}\,dy + \frac{\partial g}{\partial z}\,dz\right) \wedge dy \\
&+ \quad \left(\frac{\partial h}{\partial x}\,dx + \frac{\partial h}{\partial y}\,dy + \frac{\partial h}{\partial z}\,dz\right) \wedge dz \\
&= \quad \frac{\partial f}{\partial y}\,dy \wedge dx + \frac{\partial f}{\partial z}\,dz \wedge dx \\
&+ \quad \frac{\partial g}{\partial x}\,dx \wedge dy + \frac{\partial g}{\partial z}\,dz \wedge dy \\
&+ \quad \frac{\partial h}{\partial x}\,dx \wedge dz + \frac{\partial h}{\partial y}\,dy \wedge dz \\
&= \quad \left(\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}\right) dy \wedge dz \\
&+ \quad \left(\frac{\partial f}{\partial z} - \frac{\partial h}{\partial x}\right) dz \wedge dx \\
&+ \quad \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right) dx \wedge dy
\end{aligned}
$$

Now recall that if $\mathbf{v} = f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$ then

$$
\begin{aligned}
\text{curl } \mathbf{v} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f & g & h \end{vmatrix} \\
&= \left(\frac{\partial h}{\partial y} - \frac{\partial g}{\partial z}\right)\mathbf{i} + \left(\frac{\partial f}{\partial z} - \frac{\partial h}{\partial x}\right)\mathbf{j} + \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\right)\mathbf{k}
\end{aligned}
$$

Thus we see we have counterfeited the curl on vectors provided we keep the 2-forms in correct order (so the signs come out right) and we use the correspondence

$$dy \wedge dz \leftrightarrow \mathbf{i} \qquad dz \wedge dx \leftrightarrow \mathbf{j} \qquad dx \wedge dy \leftrightarrow \mathbf{k}$$

Finally, we want to see the differential forms turn up the divergence. Let once again $\mathbf{v} = f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$ and using the above correspondence let us apply d to the form $f\,dy \wedge dz + g\,dz \wedge dx + h\,dx \wedge dy$. Then we get

$$
\begin{aligned}
d(f\,dy \wedge dz + g\,dz \wedge dx + h\,dx \wedge dy) &= df \wedge dy \wedge dz + dg \wedge dz \wedge dx + dh \wedge dx \wedge dy \\
&= \frac{\partial f}{\partial x}\,dx \wedge dy \wedge dz + \frac{\partial g}{\partial y}\,dy \wedge dz \wedge dx \\
&\quad + \frac{\partial h}{\partial z}\,dz \wedge dx \wedge dy \\
&= \left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} + \frac{\partial h}{\partial z}\right) dx \wedge dy \wedge dz
\end{aligned}
$$

and we see div $\mathbf{v}$ is the coefficient of $dx \wedge dy \wedge dz$. Of course $dx \wedge dy \wedge dz$ corresponds to the volume element dV in vector analysis. We will have much more to say about the relations between these operators and differential forms in the section on change of coordinates.

1.10 The Poincaré Lemma and it's Converse

Of the many theorems about differential forms, the three most important are the converse of the Poincaré lemma, Stokes' theorem and the Frobenius theorem. Here we treat the Poincaré lemma; Stokes's theorem will be treated in the next few sections.

The Poincaré lemma was used by Poincaré in his work on the Calculus of Variations. It is very simple; it says that if ω is a differential form then $dd\omega = 0$. Let us see why this works for functions. We will do it in $\mathbb{R}^3$ space but it works the same in any dimension.

$$\begin{aligned} df &= \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy + \frac{\partial f}{\partial z}\,dz \\ ddf &= \left(\frac{\partial^2 f}{\partial y \partial z} - \frac{\partial^2 f}{\partial z \partial y}\right) dy \wedge dz + \left(\frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial x \partial z}\right) dz \wedge dx \\ &+ \left(\frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x}\right) dx \wedge dy \\ &= 0 \end{aligned}$$

because of the equality of mixed partials: $\frac{\partial^2 f}{\partial y \partial z} = \frac{\partial^2 f}{\partial z \partial y}$. It is worth noticing that this is one of the places where we need the functions to have two continuous derivatives to guarantee this equality.

So much for the easy stuff. The converse of the Poincaré Lemma says that if ω is a differential form and $\omega \in \Lambda^r(K)$ and $d\omega = 0$ then there exists a differential form $\alpha \in \Lambda^{r-1}(K)$ for which $d\alpha = \omega$. Sadly this converse is not always true. For it to be true in general, we need to know that K is not too complicated a region. Specifically we need to know that K is *simply connected*. This will be a new concept for most of you. We will try to clarify it with some examples which are of importance in themselves.

First let us look at the unit disk in $\mathbb{R}^2$ which is all the points in $\mathbb{R}^2$ whose distance from the origin is less than or equal to 1. We will call this $D_1(0)$. Let us draw a curve (say a circle of radius $1/2$) then we may "shrink" the curve in a continuous manner until is down to just a point. (The technical terminology is that the curve is homotopic to a point.) A *simple closed curve* is a continuous curve with no end points and no intersections with itself. It should be intuitively clear that any simple closed curve in $D_1(0)$ can be shrunk to a point.

For contrast, consider the *annulus* consisting of points in $\mathbb{R}^2$ that have distance from the origin less than or equal to 1 and greater than or equal to $1/2$. For this region K some simple closed curves can be shrunk to points and others (those that go round the central "hole") cannot be shrunk to points. As a third example consider the unit disk with just the center removed. Then, exactly like the annulus where a whole disk was removed, a curve round the origin cannot be shrunk to a point, because the point we need, the origin, is not in K.

Def A region K is *simply connected* if and only if every simple closed curve in K can be shrunk to a point in K.

A few more examples, this time in $\mathbb{R}^3$. A ball (solid sphere) is simply connected, as is also a ball with the center removed. However, a ball around the origin with the z axis removed is *not* simply connected and neither is the whole of $\mathbb{R}^3$ with the z axis removed. The unit sphere (points in $\mathbb{R}^3$ at unit distance from origin) *is* simply connected. (Note the contrast to the unit circle in $\mathbb{R}^2$ which is *not* simply connected.

These kinds of considerations are called *topological*, and there is a wonderful branch of mathematics which studies such things called topology. Due to the enormous amount of material needed in an engineering curriculum, it is not customary for topology to be included, but if you wish to learn more we recommend the books [2] and [5], which are specifically written with the engineer and physicist in mind. We also provide a "short" introduction to topology in Chapter 2.

Now that we know some topology, we can state the theorem:

Theorem (Converse of Poincaré Lemma) Let K be a simply connected region and $\omega \in \Lambda^r(K)$ and $d\omega = 0$. Then there is an $\alpha \in \Lambda^{r-1}(K)$ such that $d\alpha = \omega$.

It should be mentioned that the α is highly non-unique; there are many α's that will work.

In the problems we will see examples where the region is not simply connected and, though $d\omega = 0$ there is no α for which $d\alpha = \omega$. Practically speaking, this usually takes the form of α being multi-valued, like $\sqrt{x}$ or polar coordinate θ, and thus not being a proper function. In this case, we can often manage to get some use out of the multi-valued function provided we restrict our attention to a simply connected subregion. The vector potential of the magnetic field of an infinite vertical current carrying wire is the paradigm example and we will look at it in the problems.

Next we will derive two popular theorems of vector analysis (in $\mathbb{R}^3$) from the converse of the Poincaré lemma.

For the first we again let $\mathbf{v} = f\mathbf{i} + g\mathbf{j} + h\mathbf{k}$ and we suppose that curl $\mathbf{v} = 0$. We then form the corresponding differential 1-form $\omega = f\,dx + g\,dy + h\,dz$. From the results of the last section, we see that the condition curl $\mathbf{v} = 0$ translates into the condition $d\omega = 0$. Assuming the region $K \subseteq \mathbb{R}^3$ is simply connected we may conclude that there is a 0-form (function) $\phi(x, y, z)$ for which $d\phi = \omega$, that is

$$d\phi \;=\; \frac{\partial \phi}{\partial x}\,dx + \frac{\partial \phi}{\partial y}\,dy + \frac{\partial \phi}{\partial z}\,dz \;=\; f\,dx + g\,dy + h\,dz$$

Translating back into vector notation we have

$$\begin{aligned} \frac{\partial \phi}{\partial x}\mathbf{i} + \frac{\partial \phi}{\partial y}\mathbf{j} + \frac{\partial \phi}{\partial z}\mathbf{k} &= f\mathbf{i} + g\mathbf{j} + h\mathbf{k} \\ \text{grad } \phi &= \mathbf{v} \end{aligned}$$

We have thus proved the vector analysis theorem

Corollary Over a simply connected region $K \subseteq \mathbb{R}^3$ a curl free vector is a gradient.

There is a very similar theorem which you will prove in the problems. It reads

Corollary Over a simply connected region $K \subseteq \mathbb{R}^3$ a divergence free vector is a curl.

More explicitly, if div $\mathbf{v} = 0$ then there is a vector $\mathbf{w}$ so that curl $\mathbf{w} = \mathbf{v}$.

Although formulas for finding the objects whose existence is asserted exist, they are of no great practical utility and we will not discuss them.

Other formulas which are easily derived at this point and which you will derive in the problems are

$$\begin{aligned} \text{curl grad } \phi &= 0 \\ \text{div curl } \mathbf{v} &= 0 \end{aligned}$$

These are both consequences of $dd = 0$ which you have probably already figured out. Also available are results about path independent line integrals which we will look at in the problems.

1.11 Boundaries

Before we can deal with Stokes theorem we must talk a bit about boundaries and their orientation, and also how to compute integrals of differential forms.

The boundary of a area or surface is simply its edge. We will be satisfied with an intuitive understanding of this and not go into the topological details here; they are covered in detail in Chapter 2. The symbol of the boundary of a region or surface K is ∂K. For example the boundary of the unit disk in $\mathbb{R}^2$ is the unit circle, and the boundary of the upper half of the unit sphere in $\mathbb{R}^3$ is the unit circle in the x, y-plane. The boundary of the unit ball in $\mathbb{R}^3$ is the unit sphere. It is common for there to be no boundary; the boundary of the unit sphere in $\mathbb{R}^3$ is empty, which we write $\partial S_1(0) = \emptyset$ where $\emptyset$ is the empty set. An important fact is that *a boundary itself has no boundary,* which we can write as $\partial\partial\, K = \emptyset$.

For Stokes theorem it is important that the boundary ∂K be oriented correctly relative to K itself. We will do this with examples, but the methods are selected so that they work in any dimension. If K is an n-dimensional region in n-space, like A in $\mathbb{R}^2$ or M in $\mathbb{R}^3$, then it inherits an orientation from the n-space, that determined by the order x, y for A or x, y, z for M and analogously for higher dimensions. To orient the boundary of such regions, which necessarily is described by $n-1$ parameters, one uses the exterior normal vector $\hat{\mathbf{n}}$ to K so that $\hat{\mathbf{n}}$ followed by the parameters has the *same* orientation as K. (We will clarify this with examples.) This gives an order to the parameters which is then the order used below in the integrals. There is no way to get around the fact that this is a bit tricky.

For our first example let A be the unit disk in $\mathbb{R}^2$, with boundary the unit circle and parameter t for the unit circle. Let $\hat{\mathbf{n}}$ be the unit normal to the unit

circle. Let $\mathbf{r}(t)$ trace out the unit circle. Then

$$\hat{\mathbf{T}} = \frac{\frac{d\mathbf{r}}{dt}}{\left|\frac{d\mathbf{r}}{dt}\right|}$$

is the *unit tangent vector*. There are two directions one may go along a curve, and the parameterization must be chosen so that $\hat{\mathbf{n}}, \hat{\mathbf{T}}$ has the same orientation as x, y. This amounts to, first, the angle from $\hat{\mathbf{n}}$ to $\hat{\mathbf{T}}$ is a *positive* right angle, and second that as t increases we trace the boundary *counterclockwise* around A. So this comes down to just going around A counterclockwise.

For M a bounded three dimensional region of $\mathbb{R}^3$, the situation is more difficult to see since the boundary will be two dimensional. Suppose the parameters are u and v. Let the surface ∂M be given by $\mathbf{r}(u, v)$. Then when we increase u and hold v fixed, a curve will be traced out in certain direction and it will have tangent vector $\mathbf{r}_u = \frac{\partial \mathbf{r}}{\partial u}$ pointing in that direction. Similarly for $\mathbf{r}_v$. Then the requirement is that $\hat{\mathbf{n}}, \mathbf{r}_u, \mathbf{r}_v$ have the same orientation as x, y, z. If the orientation is opposite to x, y, z, reverse the order of the parameters u and v.

On a practical note, one usually finds the $\hat{\mathbf{n}}$ for ∂M by forming

$$\hat{\mathbf{n}} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$$

so that one really has only to check whether this $\hat{\mathbf{n}}$ is pointing *in* or *out* of M and if it is pointing *in* then change the order of u and v. Also, one does not actually have to *calculate* $\mathbf{r}_u \times \mathbf{r}_v$, only figure out its direction.

As an example, take the upper half of the unit sphere and let us use for coordinates longitude ϕ and colatitude θ. A tangent vector to a *phi* curve goes to the right (looking at the sphere from outside) and a tangent vector to the θ curve goes down. Crossing these gives a vector pointing *in* so we have the wrong order: the order should be θ, ϕ not ϕ, θ. Any 2-forms that occur should thus be presented as $f(\theta, \phi)\, d\theta \wedge d\phi$. This will be critically important when we form integrals.

Also as a practical matter, for $\partial M \subset \mathbb{R}^3$ one often uses two of x, y, z as parameters. If this is the case, the proper orders are $dy \wedge dz,\ dz \wedge dx,\ dx \wedge dy$ as one can determine by using the above rules.

The last case we are going to discuss is S, a surface in $\mathbb{R}^3$. If the surface has no boundary then the situation is just that of ∂M above (practically speaking) and we have already covered the situation; an example would be the unit sphere. However, if the surface has a boundary, then there is no obvious way to decide which way the normal should point. Mathematically one just has to choose for oneself either a direction for the normal $\hat{\mathbf{n}}$ or, equivalently, an order for the coordinates u, v. Physically there may be an obvious way to make the choice. Make sure that your $\hat{\mathbf{n}}$ is consistent with your choice of u, v so that $\hat{\mathbf{n}}, \mathbf{r}_u, \mathbf{r}_v$ is oriented like x, y, z. Once this is settled it only remains to orient the boundary ∂S. Since ∂S is one dimensional, it is only a question of choosing which way to go around it. Here is the rule.

Walk around the boundary of the surface S with your body pointing the same way as $\hat{\mathbf{n}}$ and your **LEFT** hand pointing toward the surface. The direction you are walking is the correct orientation of the boundary.

Note that if you switch the direction of the normal it will cause the proper direction of going round the serface to also switch.

This is the orientation you must use for the correct application of Stokes theorem. If you mess it up the sign will be wrong.

Of course in physics and engineering the sign often comes out wrong, and one just reverses it at the end. Still, it is reassuring when it comes out correct without human intervention.

1.12 Integrals of Forms

This is extremely simple. First, the integrals of forms must be taken over regions of the same dimension as the degree of the form. Thus

a 1-form must be integrated	over	a curve C
a 2-form must be integrated	over	a surface S or A
a 3-form must be integrated	over	a region M

To calculate the integral we recall that in a multiple integral the order of the differentials doesn't matter but for forms the order matters, as the sign will change if two differentials are interchanged. The calculation of an integral is reduced to four steps.

0. Choose correct orientations for the variables.

1. Rearrange the integral so that the differentials are in correct order.

2. Remove the wedges to get an ordinary multiple integral

3. Calculate the multiple integral by the appropriate iterated integral

This should become clear from an example. We wish to calculate the integral

$$\int_A x\,dy \wedge dx$$

where A is the unit disk in $\mathbb{R}^2$. Step 1 is to realize that the differentials are not in the right order. We fix.

$$\int_A x\,dy \wedge dx = -\int_A x\,dx \wedge dy$$

Now the differentials under the integral are in the correct order. Step 2 is to throw out the wedges

$$-\int_A x\,dx \wedge dy = -\int_A x\,dxdy$$

where the last integral is an ordinary double integral from your advanced calculus course. Step 3 is then to go over to the an iterated integral

$$-\int_A x\,dxdy = -\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} x\,dxdy$$

Step 4, which we did not mention above, is then to plug it into your fancy calculator and get the answer. Another method is to calculate the double integral by switching to polar coordinates

$$-\int_A x\,dxdy = -\int_0^{2\pi}\int_0^1 (r\,\cos\theta) r\,drd\theta$$

from which we see immediately that the result is 0.

This pretty much is the whole story. With a modest amount of care this will almost always come out right. If it does not, you might want to take the time to figure out *why* just in case it is important.

1.13 Variable Changes

We begin this section with the following observations. Suppose we are calculating

$$\int_A f(x,y)\,dx\wedge dy$$

Now suppose we find it convenient to change to polar coordinates or some other coordinates u, v. Then the old x, y coordinates are functions of the new u, v coordinates.

$$\begin{aligned} x &= x(u,v) \\ y &= y(u,v) \end{aligned}$$

Differential Forms allow us to convert the integral immediately. We have

$$\begin{aligned} f(x,y) &= f(x(u,v),y(u,v)) = \tilde{f}(u,v) \\ dx\wedge dy &= \left(\frac{\partial x}{\partial u}\,du + \frac{\partial x}{\partial v}\,dv\right)\wedge\left(\frac{\partial y}{\partial u}\,du + \frac{\partial y}{\partial v}\,dv\right) \\ &= \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right) du\wedge dv \end{aligned}$$

Thus

$$\int_A f(x,y)\,dx\wedge dy = \int_A \tilde{f}(u,v)\left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right) du\wedge dv$$

Naturally there are generalizations to higher dimensions. Before we deal with that, though, we should systematize what we have here. The cleverer students,

or those with really good memories, will recognize that expression in brackets as the determinant of the matrix

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

which is called the *Jacobian matrix*[5] of the variable change. Thus we can write the above equation, with $\det(\cdot)$ signifying the determinant, as

$$\begin{aligned} \int_A f(x,y)\,dx\wedge dy &= \int_A \tilde{f}(u,v)\det(\frac{\partial(x,y)}{\partial(u,v)})du\wedge dv \\ &= \int_A \tilde{f}(u,v)\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} du\wedge dv \end{aligned}$$

Naturally the description of A will be quite different in the x, y and u, v variables. If we now go over to *unoriented integrals*, that is ordinary double integrals, we will have, assuming that x, y and u, v give the same orientation to A,

$$\int_A f(x,y)\,dxdy = \int_A \tilde{f}(u,v)\det(\frac{\partial(x,y)}{\partial(u,v)})dudv$$

Since x, y and u, v give the same orientation to A, the determinant in the integral will have a positive value. If the orientations are reversed, it will all be taken care of in the oriented integrals but in the ordinary integrals we need to make the determinant positive by putting in an absolute value, so we have

$$\int_A f(x,y)\,dxdy = \int_A \tilde{f}(u,v)\ |\det(\frac{\partial(x,y)}{\partial(u,v)})|\ dudv$$

This is the famous change of variable formula. In most Calculus books they do not prove it because the proof without differential forms is somewhat difficult. Using differential forms it is quite easy, as you have seen. The reason for this is that differential forms keep control of the orientation, whereas old techniques have to deal with it in an ad hoc and confusing way.

Everything we have done here in 2 dimensions will work perfectly well in n-dimensions. Enthusiasts may wish to work it out for 3 dimensions. You will then see the connection between differential forms and determinants. In fact,we can assert, with certainty of contradiction from some member of the audience, that determinants are important *because* they are the coefficients in Grassmann Algebra calculations. We will see more of this.

1.14 Surface integrals

It is convenient to work out the connection between surface integrals in vector form and differential forms at this point, so we have it available for Stokes

[5] properly pronounced Yacobian, although we know it's hopeless

theorem. We will work it out for surfaces in $\mathbb{R}^3$ but the techniques applied will work just as well for $(n-1)$-surfaces in $\mathbb{R}^n$. This is one of the places in the theory which, while not difficult, is a little tricky.

While it may not be possible to cover the surface with a single coordinate patch, it is always possible to break up the integral into integrals over each patch and add them. Hence we only need to work out the formula for a single coordinate patch.

The standard thing to have to work out for surface integrals is

$$\int_S P\,dy \wedge dz + Q\,dz \wedge dx + R\,dy \wedge dz$$

If we let $\mathbf{v} = P\,\hat{\mathbf{i}} + Q\,\hat{\mathbf{j}} + R\,\hat{\mathbf{k}}$ then the above is a disguised form of our old friend from advanced calculus

$$\int_S \mathbf{v} \cdot \hat{\mathbf{n}}\,dS = \int_S \mathbf{v} \cdot \mathbf{dS}$$

First we need to get some idea of what $\mathbf{dS}$ should be. We will be using the variables u, v as coordinates on the surface. If a normal is already available be sure to pick the order of the variables so that $\hat{\mathbf{n}}, du, dv$ has the same orientation as x, y, z. The surface will then be described by $\mathbf{r}(u, v)$. If we think of moving from a point p in the u and v directions we will have vectors

$$\begin{aligned} \mathbf{r}(u+\Delta u, v) - \mathbf{r}(u,v) &\approx \frac{\partial \mathbf{r}}{\partial u}\,\Delta u \\ \mathbf{r}(u, v+\Delta v) - \mathbf{r}(u,v) &\approx \frac{\partial \mathbf{r}}{\partial v}\,\Delta v \end{aligned}$$

and so the bit of oriented area can be obtained from the cross product

$$\mathbf{\Delta S} = \left(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right) \Delta u \Delta v$$

Now remembering that

$$\mathbf{r}(u,v) = x(u,v)\,\hat{\mathbf{i}} + y(u,v)\,\hat{\mathbf{j}} + z(u,v)\,\hat{\mathbf{k}}$$

we have

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial u} &= \frac{\partial x}{\partial u}\,\hat{\mathbf{i}} + \frac{\partial y}{\partial u}\,\hat{\mathbf{j}} + \frac{\partial z}{\partial u}\,\hat{\mathbf{k}} \\ \frac{\partial \mathbf{r}}{\partial v} &= \frac{\partial x}{\partial v}\,\hat{\mathbf{i}} + \frac{\partial y}{\partial v}\,\hat{\mathbf{j}} + \frac{\partial z}{\partial v}\,\hat{\mathbf{k}} \\ \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} &= \left(\frac{\partial y}{\partial u}\frac{\partial z}{\partial v} - \frac{\partial z}{\partial u}\frac{\partial y}{\partial v}\right)\hat{\mathbf{i}} + \left(\frac{\partial z}{\partial u}\frac{\partial x}{\partial v} - \frac{\partial x}{\partial u}\frac{\partial z}{\partial v}\right)\hat{\mathbf{j}} \\ &\quad + \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}\right)\hat{\mathbf{k}} \\ &= \frac{\partial(y,z)}{\partial(u,v)}\,\hat{\mathbf{i}} + \frac{\partial(z,x)}{\partial(u,v)}\,\hat{\mathbf{j}} + \frac{\partial(x,y)}{\partial(u,v)}\,\hat{\mathbf{k}} \end{aligned}$$

Going over from Δ to d we have at last

$$\begin{aligned}\mathbf{dS} &= \left(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right) dudv \\ &= \left(\frac{\partial(y,z)}{\partial(u,v)}\,\hat{\mathbf{i}} + \frac{\partial(z,x)}{\partial(u,v)}\,\hat{\mathbf{j}} + \frac{\partial(x,y)}{\partial(u,v)}\,\hat{\mathbf{k}}\right) dudv\end{aligned}$$

Now we also have

$$\hat{\mathbf{n}} = \frac{\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}}{|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}|}$$

Finally we have

$$\begin{aligned}&\int_S P\,dy \wedge dz + P\,dz \wedge dx + P\,dx \wedge dy \\ &= \int_S P\,\frac{\partial(y,z)}{\partial(u,v)}\,du \wedge dv + Q\,\frac{\partial(z,x)}{\partial(u,v)}\,du \wedge dv + R\,\frac{\partial(x,y)}{\partial(u,v)}\,du \wedge dv \\ &= \int_S (P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}) \cdot \left(\frac{\partial(y,z)}{\partial(u,v)}\hat{\mathbf{i}} + \frac{\partial(z,x)}{\partial(u,v)}\hat{\mathbf{j}} + \frac{\partial(x,y)}{\partial(u,v)}\hat{\mathbf{k}}\right) du \wedge dv \\ &= \int_S (P\,\mathbf{i} + Q\,\mathbf{j} + R\,\mathbf{k}) \cdot \left(\frac{\partial(y,z)}{\partial(u,v)}\hat{\mathbf{i}} + \frac{\partial(z,x)}{\partial(u,v)}\hat{\mathbf{j}} + \frac{\partial(x,y)}{\partial(u,v)}\hat{\mathbf{k}}\right) dudv \\ &= \int_S \mathbf{v} \cdot \mathbf{dS}\end{aligned}$$

In the next to the last step we switched from an oriented to an unoriented integral since everything was carefully set up to be correctly oriented.

We can sweat a little more out of this. The scalar element of area is of course the absolute value of $\mathbf{dS}$ so that

$$\begin{aligned}dS &= |\mathbf{dS}| = \left|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right| dudv \\ &= \sqrt{\left(\frac{\partial(y,z)}{\partial(u,v)}\right)^2 + \left(\frac{\partial(z,x)}{\partial(u,v)}\right)^2 + \left(\frac{\partial(x,y)}{\partial(u,v)}\right)^2}\,dudv\end{aligned}$$

This is itself a very important formula. Also since

$$\hat{\mathbf{n}} = \frac{\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}}{|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}|}$$

we have

$$\mathbf{dS} = \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\,dudv = \hat{\mathbf{n}} \left|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right| dudv = \hat{\mathbf{n}}\,dS$$

we can write

$$\int_S \mathbf{v} \cdot \mathbf{dS} = \int_S \mathbf{v} \cdot \hat{\mathbf{n}}\,dS$$

It would be wise to note that the integrals we have been working with in this section are often called *flux integrals*. In contrast, there is another kind of surface integral which has little to do with what we are doing. This kind of integral is not vectorial in nature. They look like

$$\int_S f(x,y,z)\,dS$$

A problem that utilized this kind of integral would be finding the center of mass of a hemispherical shell. We have developed the necessary tools to evaluate these integrals, although for us this is a side issue and we only mention it since students should be aware of the two distinct kinds of surface integral.

To evaluate this kind of integral, choose parameter u, v for the surface, (or a portion of the surface,) and then use the above above formula for dS to get

$$\begin{aligned}&\int_S f(x,y,z)\,dS\\ &= \int f(x(u,v),y(u,v),z(u,v))\sqrt{\left(\frac{\partial(y,z)}{\partial(u,v)}\right)^2+\left(\frac{\partial(z,x)}{\partial(u,v)}\right)^2+\left(\frac{\partial(x,y)}{\partial(u,v)}\right)^2}\,dudv\end{aligned}$$

The evaluation of these kinds of integral is very sensitive to the choice of parameters, and an integral that might be extremely difficult with parameters x, y (so that $z = z(x, y)$) might be quite easy with, say, θ, ϕ as parameters.

1.15 The Generalized Stokes Theorem

The generalized Stokes theorem is one of the triumphs of elementary mathematics, subsuming as it does the fundamental theorem of Calculus, Green's theorem, the divergence theorem (= Gauss's theorem), and the ordinary Stokes's theorem. Moreover, it is dimension independent; it works as well in 26 dimensions as in 2 or 3.

Theorem (Generalized Stokes Theorem)
Let K be a oriented subset of $\mathbb{R}^d$ of dimension n and let ∂K be the properly oriented boundary of K. Let ω be an (n-1)-form. Then

$$\int_K d\omega = \oint_{\partial K} \omega$$

One of the important qualities of this theorem is that it is very easy to remember. Hence it can be used to write down the Green, divergence and Stokes' theorems even if you don't precisely remember where the signs go; it figures that out for you. We will now look at the Green, divergence and Stokes' theorems one at a time and see how they fall out of the generalized Stokes theorem. At the end, for fun, we will come back and look at the fundamental theorem of Calculus, which is tricky because it is too simple.

Green's theorem Let $A \subset \mathbb{R}^2$ be a region with boundary ∂A oriented counterclockwise. Then

$$\oint_{\partial A} P(x,y)\,dx + Q(x,y)\,dy = \int_K \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\,dxdy$$

Proof Let $\omega = P(x,y)\,dx + Q(x,y)\,dy$. Applying the general Stokes' theorem we have

$$\begin{aligned}
\oint_{\partial A} P(x,y)\,dx + Q(x,y)\,dy &= \int_{\partial A} \omega \\
&= \oint_A d\omega \\
&= \int_A d(P(x,y)\,dx + Q(x,y)\,dy) \\
&= \int_A \left(\frac{\partial P}{\partial y}\,dy \wedge dx + \frac{\partial Q}{\partial x}\,dx \wedge dy\right) \\
&= \int_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy \\
&= \int_A \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy
\end{aligned}$$

where we used $dy \wedge dx = -\,dx \wedge dy$ and in the last line we used the rule from converting from oriented integrals to ordinary double integrals.

Notice that it was not necessary for us to remember where the sign goes in Green's theorem; our methodology automatically puts it in the correct place.

We will do Stokes' theorem next since it looks so similar in many ways to Green's theorem.

Stokes' Theorem Let S be a surface in $\mathbb{R}^3$ with definite choice of normal $\hat{\mathbf{n}}$ and correctly oriented boundary ∂S. Let $v = P\,\hat{\mathbf{i}} + Q\,\hat{\mathbf{j}} + R\,\hat{\mathbf{k}}$ Then

$$\begin{aligned}
&\oint_{\partial S} P\,dx + Q\,dy + R\,dz = \\
&= \int_S \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dydz + \int_S \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dzdx + \int_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy
\end{aligned}$$

or in vector notation

$$\oint_{\partial S} \mathbf{v} \cdot \mathbf{dl} = \int_S \text{curl } \mathbf{v} \cdot \hat{\mathbf{n}}\,dS$$

Recall that the boundary of S is oriented so that when you walk around the boundary with your body in the direction of $\hat{\mathbf{n}}$ and your left hand reaching in toward the surface then you are walking in the direction of the orientation. (This can be reformulated in terms of fingers and thumb but we find this less confusing.)

Proof: Let ω be the 1-form

$$\omega = P\,dx + Q\,dy + R\,dz$$

Then applying Stokes' theorem

$$\begin{aligned}
&\oint_{\partial S} P\,dx + Q\,dy + R\,dz = \\
&= \oint_{\partial S} \omega \\
&= \int_S d\omega \\
&= \int_S d(P\,dx + Q\,dy + R\,dz) \\
&= \int_S \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz + \int_S \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz \wedge dx + \int_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy \\
&= \int_S \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dydz + \int_S \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dzdx + \int_S \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dxdy \\
&= \int_S \text{curl } \mathbf{v} \cdot \mathbf{dS}
\end{aligned}$$

where we have used the rules governing the transition from oriented to unoriented (ordinary) integrals. Recall that in unoriented integrals the order of the differentials does not matter, but we have left them in cyclic form as a matter of good practise. The transition to vector form was discussed in the previous section.

Now to the divergence theorem. This is quite easy.

The Divergence theorem (Gauss's theorem) Let M be a region of $\mathbb{R}^3$ and ∂M the surface which is its correctly oriented boundary. Then

$$\begin{aligned}
&\oint_{\partial M} Pdydz + Qdzdx + Pdxdy = \\
&= \int_M \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dxdydz
\end{aligned}$$

or in vector notation

$$\oint_{\partial M} \mathbf{v} \cdot \hat{\mathbf{n}}\, dS = \int_M \text{div } \mathbf{v}\; dxdydz$$

Proof Let ω be the 2-form

$$\omega = P\,dy \wedge dz$$

Then applying the generalized Stokes' theorem

$$\oint_{\partial M} P\,dxdy \quad = \quad \oint_{\partial M} P\,dx \wedge dy$$

$$\begin{aligned} &= \oint_{\partial M} \omega \\ &= \int_M d\omega \\ &= \int_M d(P dy \wedge dz) \\ &= \int_M \frac{\partial P}{dx}\, dx \wedge dy \wedge dz \\ &= \int_M \frac{\partial P}{dx}\, dx dy dz \end{aligned}$$

The other two terms are handled similarly. Notice that we have proved slightly more than the theorem states; we have proved the analog of the theorem *for each term*. This can occasionally be useful.

Problems for Section 1.15 (the Generalized Stokes's theorem)

In the following problems use the generalized Stokes's theorem

$$\int_{\partial K} \omega = \int_K d\omega$$

1. Show that if $M \subseteq \mathbb{R}^3$ and bounded then the volume of M is

$$\text{vol}(M) = \int_{\partial M} x\, dy \wedge dz$$

2. Show that if $M \subseteq \mathbb{R}^3$ and bounded then the volume of M is

$$\text{vol}(M) = \frac{1}{3} \int_{\partial M} x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy$$

   This form is often preferable to the form in 1) because, due to symmetry, objectionable terms may well cancel out.

3. Show that now if $M \subseteq \mathbb{R}^n$ an n-dimensional bounded region then

$$\text{vol}(M) = C \int_{\partial M} \sum_{i=1}^{n} (-1)^{i-1} x^i\, dx^1 \wedge \ldots \wedge dx^{i-1} \wedge dx^{i+1} \wedge \ldots \wedge dx^n$$

   What is the constant C?

4. Suppose that ω is a nice smooth $(n-1)$-form on $\mathbb{R}^n$ which is 0 outside of a ball of radius R. Show that

$$\int_{\mathbb{R}^n} d\omega = 0$$

   Hint: You may want to consider the problem on a ball of radius $R_1 > R$. Remember ω is *smooth*.

5. a. Suppose that C_1 and C_2 are two curves in $\mathbb{R}^2$ that each go from point P_1 to P_2 and suppose (for simplicity; it doesn't actually matter) that C_1 and C_2 do not intersect. Suppose also that ω is a 1-form and $d\omega = 0$. Show that
$$\int_{C_1} \omega = \int_{C_2} \omega$$
Hint: draw picture. We say that the integral is *path independent*.

b. Suppose now C_1 and C_2 are two curves on a surface $S \subseteq \mathbb{R}^3$ with parameters u^1, u^2. Let, as before, C_1 and C_2 run from point P_1 on S to point P_2 on S without intersection. Let ω be a one form on S with $d\omega = 0$. Show
$$\int_{C_1} \omega = \int_{C_2} \omega$$
Is the proof very different from a)?

1.16 Curvilinear Coordinates I: preliminary formulas

In this section we will derive formulas for the gradient, curl, divergence and Laplacian in curvilinear coordinates. This is not nearly as straightforward as one might expect, but you will not see these troubles because of good organization. The central trick in this development is to express each of the differential operators in terms of operators that are manifestly coordinate independent. Each of these operators has a coordinate formulation and when they are put together the desired expression arises.

Unfortunately the derivation of some of the basic formulas is a little more conceptual and difficult than most of the material in these notes. To spare those who are interested only in the results, we have put these derivations into Chapter 2 (Mathematical Theory). There is also a lower level introduction to the $*$ operator in the section after the Curvilinear coordinate sections.

To express grad, divergence and curl invariantly we need two functions, $\Phi : V \to V^*$ and $* : \Lambda_r(V) \to \Lambda^{n-r}(V)$. Both of these functions are independent of the choice of the coordinate system, as can be seen in Chapter 2. Although we are just using these formulas to derive the formulas for curvilinear coordinates, they have much wider application in modern mathematics, and effort put into learning them may well pay off in other circumstances. And you will be glad to know that in our circumstances there is not much difficulty.

We will write the formulas we need in n-space because it is no more difficult than writing them in 3-space, and indeed some things are actually clearer. Also it is important to know which things work generally and which things, like curl, work only in three space.

We also note that once you know the position vector $\mathbf{r}$ in terms of the curvilinear coordinates you essentially know everything. From the $e_i = \frac{\partial \mathbf{r}}{\partial u^i}$ you can calculate $g_{ij} = (e_i, e_j)$ and from them everything else. In 1856 Bernhard

Riemann realized that the surrounding n-space was surperfluous and the g_{ij} alone would get you anywhere you wanted to go, and thus invented Riemannian Geometry. We will discuss some of this in a later chapter.

Let $u^1, u^2, \ldots, u^n$ be curvilinear coordinates. Then we have a position vector $\mathbf{r}$ and its derivatives which are more or less the base vectors. For physics they require normalization later, but it would be a big mistake to normalize them now.

$$\begin{aligned} \mathbf{r} &= \mathbf{r}(u^1, \ldots, u^n) \\ \mathbf{e}_i &= \frac{\partial \mathbf{r}}{\partial u^i} \\ g_{ij} &= (\mathbf{e}_i, \mathbf{e}_j) \;=\; \mathbf{e}_i \cdot \mathbf{e}_j \\ (g^{ij}) &= (g_{ij})^{-1} \end{aligned}$$

We are going to mostly use the parentheses notation $(\mathbf{e}_i, \mathbf{e}_j)$ instead of $\mathbf{e}_i \cdot \mathbf{e}_j$ because it keeps things grouped together better. It is just notation. The g_{ij} are called the *metric coefficients* and allow one to work with distance an angles in the coordinates $u^1, u^2, \ldots, u^n$. We think of vectors as being based at particular points which are determined by $\mathbf{r}$. This is an elementary form of the concept *tangent space* which we will introduce later.

Any vector $\mathbf{v}$ based at $\mathbf{r}(u^1, \ldots, u^n)$ can be expressed in terms of the basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ which are also thought of as being based at $\mathbf{r}$. Thus

$$\mathbf{v} = \sum_{i=1}^{n} v^i \mathbf{e}_i = v^i \mathbf{e}_i \qquad \text{(sum sign omitted)}$$

The sum sign is often omitted when the same indexing letter appears both up and down; since Einstein invented it, it is probably a good idea, but you do need to keep it in mind. This simplifies the notation considerably.

Next we need a function where you input the vector $\mathbf{v}$ and it outputs the i^{th} coordinate v^i. You have known this function for years, but never knew what it was. It is du^i. Thus

$$du^i(\mathbf{v}) = v^i$$

This is indeed what du^i really is, but with the current organization of the math curriculum it is not convenient to explain this at the entry level. However, now you know! Incidentally, it *is* possible to make this consistent with the idea that du^i is a little bit of u^i. It works like this. Nearby points are connected by vectors which are very short. If $\mathbf{v}$ is such a short vector, then $v^i = du^i(\mathbf{v})$ is also small. So if we abbreviate $du^i(\mathbf{v})$ by du^i it will behave as a small number, sort of. It's surprising that this works as well as it does. Leibniz, who invented du^i, would be amused.

Since $du^i : V \to \mathbb{R}$ (and is linear) it is a linear functional and thus $du^i \in V^*$, where V^* is the *vector space of linear functionals on* V. A wedge product of three du^i, for example $du^1 \wedge du^3 \wedge du^5$ would then be in $\Lambda^3(V^*)$.

The $du^1, \ldots, du^n$ are the *dual basis* to $e_1, \ldots, e_n$ and form a basis for V^*.

Now we discuss $\Phi : V \to V^*$ and $\Phi^{-1} : V^* \to V$. We have

$$\begin{aligned} \Phi(\mathbf{v}) &= \Phi(v^i e_i) &= g_{ij} v^i du^j \\ \Phi^{-1}(\omega) &= \Phi^{-1}(\lambda_i du^i) &= g^{ij} \omega_i e_j \end{aligned}$$

Here ω is an arbitrary element of V^* and the λ_i are its coordinates in the basis $du^1, \ldots, du^n$.

It is worth mentioning that for linear algebra purposes the inner product represented by g_{ij} and the function Φ determine one another, as is reasonably obvious from the above formulas. There is a term, *correlation*, for the function Φ, and some people consider the correlation to be more basic conceptually than the inner product.

Recall that the g_{ij} give an inner (dot) product for V and in Chapter two we show that this may be "lifted" to V^*. If

$$\begin{aligned} \lambda &= \lambda_i du^i \qquad \lambda \in V^* \qquad \lambda_i \in \mathbb{R} \\ \mu &= \mu_i du^i \qquad \mu \in V^* \qquad \mu_i \in \mathbb{R} \end{aligned}$$

then

$$(\lambda, \mu) = g^{ij} \lambda_i \mu_j$$

just like

$$(\mathbf{u}, \mathbf{v}) = g_{ij} u^i v^j$$

but using the inverse matrix.

Now we digest the information on the $*$ operator, which is derived in a subsequent section and in Chapter 2.

$$\begin{aligned} *du^i &= \sum_k (-1)^{k-1} g^{ik} \sqrt{g}\, du^1 \wedge \ldots \wedge du^{k-1} \wedge du^{k+1} \wedge \ldots \wedge du^n \\ **\omega &= (-1)^{r(n-r)} \omega \qquad \text{for } \omega \in \Lambda^r(V^*) \\ \Omega_0 &= \sqrt{g}\, du^1 \wedge \ldots \wedge du^n \\ *\Omega_0 &= 1 \\ *1 &= \Omega_0 \end{aligned}$$

If the coordinates are orthogonal then things simplify. Setting

$$\begin{aligned} (\mathbf{e}_i, \mathbf{e}_j) &= g_{ij} &= h_i^2\, \delta_{ij} &= \begin{cases} h_i^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \\ (du^i, du^j) &= g^{ij} &= \frac{1}{h_i^2}\, \delta_{ij} &= \begin{cases} \frac{1}{h_i^2} & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \\ \sqrt{g} &= h_1 h_2 \ldots h_n \end{aligned}$$

We know that a basis for $\Lambda^r(V^*)$ is given by

$$du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_r} \quad \text{where} \quad i_1 < i_2 < \cdots < i_r$$

Let $\{k_1, \ldots, k_{n-r}\} = \{1, 2, \ldots, n\} - \{i_1, \ldots, i_r\}$ so that we can form the permutation

$$\left(\begin{array}{ccc|ccc} 1 & \ldots & r & r+1 & \ldots & n \\ i_1 & \ldots & i_r & k_1 & \ldots & k_{n-r} \end{array} \right)$$

of $\{1, 2, \ldots, n\}$. The *sign* of this permutation is $(-1)^s$ where s is the number of interchanges necessary to rearrange the bottom line into the top line (or vice versa). The vertical line is just for the convenience of the reader; it divides the first part of the permutation from the second. Then

$$*(du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_r}) = (-1)^s \frac{h_{k_1} \cdots h_{k_{n-r}}}{h_{i_1} \cdots h_{i_r}} \, du^{k_1} \wedge du^{k_2} \wedge \ldots \wedge du^{k_{n-r}}$$

as we show in Chapter 2.

Now it is time for two examples. The results will just be listed but it is easy to verify they are correct; mostly they can be done in ones head. *We will now begin to omit the wedges between differentials; the reader may assume that if differentials are being multiplied that wedges have been omitted for ease of reading.*

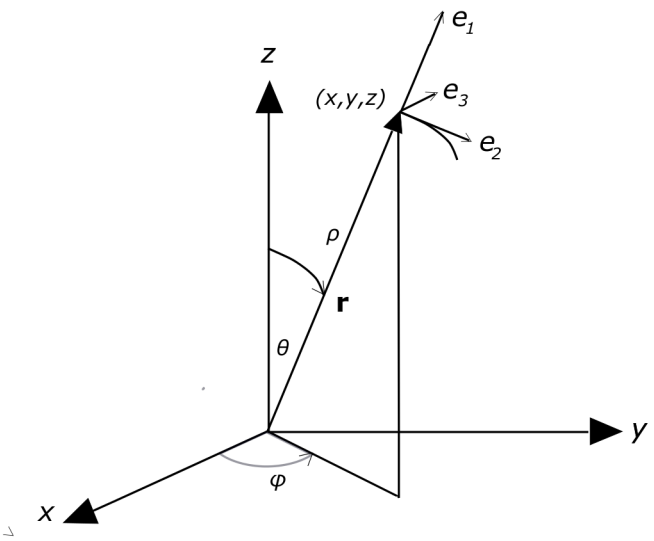
Example 1. $n = 3$ and coordinates are x, y, z.

$$\begin{aligned} \mathbf{r} &= (x, y, z) \\ \mathbf{e}_1 = \frac{\partial \mathbf{r}}{\partial x} &= (1, 0, 0) \\ \mathbf{e}_2 = \frac{\partial \mathbf{r}}{\partial y} &= (0, 1, 0) \\ \mathbf{e}_3 = \frac{\partial \mathbf{r}}{\partial z} &= (0, 0, 1) \\ g_{ij} = e_i \cdot e_j = (e_i, e_j) &= \delta_{ij} = \left\{ \begin{array}{ll} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{array} \right. \end{aligned}$$

Thus we have orthogonal coordinates and also $h_i^2 = g_{ii} = 1$. Thus our formulas give

$$\begin{array}{cccccccccc} *dx & = & dy \wedge dz & \qquad & *dy & = & (-1)dx \wedge dz & \qquad & *dz & = & dx \wedge dy \\ *dy \wedge dz & = & dx & & *dx \wedge dz & = & (-1)dy & & *dx \wedge dy & = & dz \end{array}$$

Naturally in practise we prefer $dz \wedge dx$ (cyclic order) to $(-1)dx \wedge dz$.



Spherical Coordinates

Example 2. $n = 3$ and coordinates are ρ, θ, ϕ.
Here θ (physics standard) is the angle off the z-axis. The natural basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ defined below are shown but in order not to complicate the picture they are shown shorter than they actually are. Also shown is a coordinate line for θ, where θ varies but ρ and ϕ are held contant. Note that $\mathbf{e}_2$ is tangent to the coordinate line at the point being examined. Note that ρ is the distance from the origin to (x, y, z) but $\mathbf{r}$ is the fundamental position vector.

$$\begin{aligned}
\mathbf{r} &= (\rho\sin\theta\cos\phi, \rho\sin\theta\sin\phi, \rho\cos\theta) \\
\mathbf{e}_1 = \frac{\partial\mathbf{r}}{\partial\rho} &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) && h_1^2 = g_{11} = 1 \\
\mathbf{e}_2 = \frac{\partial\mathbf{r}}{\partial\theta} &= (\rho\cos\theta\cos\phi, \rho\cos\theta\sin\phi, -\rho\sin\theta) && h_2^2 = g_{22} = \rho^2 \\
\mathbf{e}_3 = \frac{\partial\mathbf{r}}{\partial\phi} &= (-\rho\sin\theta\sin\phi, \rho\sin\theta\cos\phi, 0) && h_3^2 = g_{33} = \rho^2\sin^2\theta
\end{aligned}$$

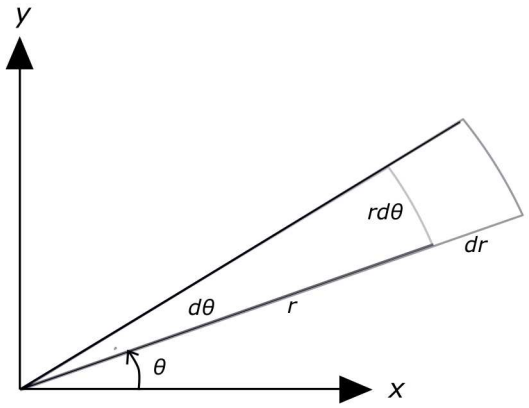
$$\begin{aligned}
g = \det(g_{ij}) &= h_1^2h_2^2h_3^2 \\
\sqrt{g} = h_1h_2h_3 &= \rho^2\sin\theta \\
\Omega_0 = \sqrt{g}d\rho d\theta d\phi &= \rho^2\sin\theta\, d\rho d\theta d\phi \\
*d\rho = \frac{h_2h_3}{h_1}d\theta d\phi &= \rho^2\sin\theta\ d\theta d\phi \\
*d\theta = -\frac{h_1h_3}{h_2}d\rho d\phi &= \frac{h_1h_3}{h_2}d\phi d\rho = \sin\theta\, d\phi d\rho \\
*d\phi = \frac{h_1h_2}{h_3}d\rho d\theta &= \frac{1}{\sin\theta}\, d\rho d\theta
\end{aligned}$$

$$\begin{aligned} *(d\theta d\phi) &= \frac{1}{\rho^2 \sin\theta}\, d\rho \\ *(d\phi d\rho) &= \frac{1}{\sin\theta}\, d\theta \\ *(d\rho d\theta) &= \sin\theta\, d\phi \end{aligned}$$

For the last three equations use $**\omega = (-1)^{r(n-r)}\omega$

Problems for Section 1.16
You will need these results for the problems for Section 1.17.

1. $\mathbb{R}^2$ and polar coordinates r, θ.



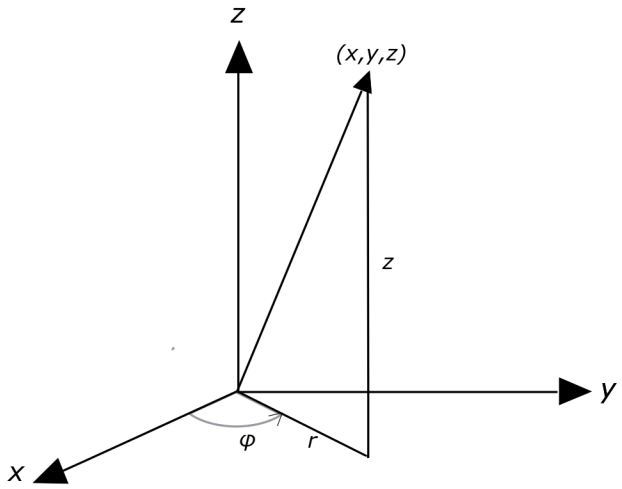
Polar Coordinates

r is given by

$$\mathbf{r}(r,\theta) = \langle r\,\cos\theta, r\,\sin\theta\rangle$$

Find $\mathbf{e}_1 = \frac{\partial \mathbf{r}}{\partial r}$ and $\mathbf{e}_2 = \frac{\partial \mathbf{r}}{\partial \theta}$, Check that $\mathbf{e}_1$, $\mathbf{e}_2$ has the proper orientation for $\mathbb{R}^2$. Now find g_{ij} and the h_i. Find $\Omega_0 = \sqrt{g}\, dr \wedge d\theta$. This is the area element for $\mathbb{R}^2$ in polar coordinates.

2. $\mathbb{R}^3$ and cylindrical coordinates r, ϕ, z.



Cylindrical Coordinates

r is given by

$$\mathbf{r}(r, \phi, \theta) = \langle r\,\cos\theta, r\,\sin\theta, z\rangle.$$

Find $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and check that $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ has the correct orientation for $\mathbb{R}^3$. Find g_{ij}, g^{ij}, h_i and $\sqrt{g}$. Present g_{ij} and g^{ij} in matrix form. Find Ω_0 which gives you the element of volume for $\mathbb{R}^3$ in cylindrical coordinates.

1.17 Curvilinear Coordinates II: the calculations

Now we must consider the strategy of how we will find the formulas for grad, div and curl in curvilinear coordinates. This is quite simple once you know how. We look at the formulas in x, y, z coordinates and figure how to express the x, y, z expressions in terms of the operators Φ and $*$. Since these are coordinate independent, we can use the formulas we find in *any* coordinate system. The rest is just relatively easy calculations.

We will start with curl. Recall that the input and output of curl are vectors and that if we set $\omega = v^1\,dx + v^2\,dy + v^3\,dz$ then

$$d\omega = \left(\frac{\partial v^3}{\partial y} - \frac{\partial v^2}{\partial z}\right) dydz + \left(\frac{\partial v^1}{\partial z} - \frac{\partial v^3}{\partial x}\right) dzdx + \left(\frac{\partial v^2}{\partial x} - \frac{\partial v^1}{\partial y}\right) dxdy$$

which looks a lot like curl. Now if we are starting with a vector $\mathbf{v} = v^1\mathbf{e}_1 + v^2\mathbf{e}_2 + v^3\mathbf{e}_3$ we can get to ω by using Φ. Indeed, using the formulas from the last section,

$$\begin{aligned}\Phi(\mathbf{v}) &= g_{ij}v^i du^j \\ &= \delta_{ij}v^i du^j\end{aligned}$$

$$\begin{aligned} &= v^1\,dx + v^2\,dy + v^3\,dz \\ &= \omega \end{aligned}$$

Then $d\omega$ gives us the above 2-form. However, we cannot use Φ^{-1} to descend again to vectors because Φ^{-1} eats only 1-forms. However $*$ gets us from 2-forms to 1-forms, and we have

$$\begin{aligned} *d\omega &= \left(\frac{\partial v^3}{\partial y} - \frac{\partial v^2}{\partial z}\right) dx + \left(\frac{\partial v^1}{\partial z} - \frac{\partial v^3}{\partial x}\right) dy + \left(\frac{\partial v^2}{\partial x} - \frac{\partial v^1}{\partial y}\right) dz \\ \Phi^{-1}(*d\omega) &= \left(\frac{\partial v^3}{\partial y} - \frac{\partial v^2}{\partial z}\right) \vec{e}_1 + \left(\frac{\partial v^1}{\partial z} - \frac{\partial v^3}{\partial x}\right) \vec{e}_2 + \left(\frac{\partial v^2}{\partial x} - \frac{\partial v^1}{\partial y}\right) \vec{e}_3 \\ \Phi^{-1}(*d\Phi(\mathbf{v})) &= \text{curl } \mathbf{v} \end{aligned}$$

Thus we see, using $\circ$ for function composition,

$$\text{curl} = \Phi^{-1} \circ * \circ d \circ \Phi$$

Since Φ, $*$, and d are all independent of the choice of coordinate system, so is curl$= \Phi^{-1} \circ * \circ d \circ \Phi$. Thus we can use this formula for curl in any coordinate system; we just have to be able to calculate the functions Φ, $*$, and d in the new coordinates and this is trivial. We will now use our method to get formulas for curl in any *orthogonal* coordinate system. We could do it in an arbitrary coordinate system also, but this would be beyond the needs of most engineers and scientists, and would also come out messy.

We will find the formula for curl in the orthogonal coordinate system u^1, u^2, u^3. Although we will eventually have to change over to physical coordinates, we will begin with $\mathbf{v}$ expressed in the natural coordinates system

$$\begin{aligned} \mathbf{e}_i &= \frac{\partial \mathbf{r}}{\partial u_i} \\ \mathbf{v} &= v^1\vec{e}_1 + v^2\vec{e}_2 + v^3\vec{e}_3 \end{aligned}$$

Since the coordinate system is orthogonal, we have $g_{ij} = 0$ for $i \neq j$ and $g_{ii} = h_i^2$. Thus we have

$$\begin{aligned} \mathbf{v} &= v^1\vec{e}_1 + v^2\vec{e}_2 + v^3\vec{e}_3 \\ \Phi(\mathbf{v}) &= g_{ij}v^i du^j \;=\; h_1^2 v^1 du^1 + h_2^2 v^2 du^2 + h_3^2 v^3 du^3 \\ d\Phi(\mathbf{v}) &= \left(\frac{\partial(h_3^2 v^3)}{\partial u^2} - \frac{\partial(h_2^2 v^2)}{\partial u^3}\right) du^2 du^3 + \text{etc.} \\ *d\Phi(\mathbf{v}) &= \frac{h_1}{h_2 h_3}\left(\frac{\partial(h_3^2 v^3)}{\partial u^2} - \frac{\partial(h_2^2 v^2)}{\partial u^3}\right) du^1 + \text{etc.} \\ \text{curl } \mathbf{v} \;=\; \Phi^{-1} * d\Phi(\mathbf{v}) &= \frac{1}{h_1^2}\frac{h_1}{h_2 h_3}\left(\frac{\partial(h_3^2 v^3)}{\partial u^2} - \frac{\partial(h_2^2 v^2)}{\partial u^3}\right) \mathbf{e}_1 + \text{etc.} \\ &= \frac{1}{h_1 h_2 h_3}\left(\frac{\partial(h_3^2 v^3)}{\partial u^2} - \frac{\partial(h_2^2 v^2)}{\partial u^3}\right) \mathbf{e}_1 + \text{etc.} \end{aligned}$$

This is correct but it is in terms of the natural basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ whereas in physics it is usual to express the vector in terms of the normalized vectors

$$\hat{\mathbf{e}}_1 = \frac{1}{h_1}\mathbf{e}_1, \quad \hat{\mathbf{e}}_2 = \frac{1}{h_2}\mathbf{e}_2, \quad \hat{\mathbf{e}}_3 = \frac{1}{h_3}\mathbf{e}_3$$

Then we have

$$\begin{aligned} \mathbf{v} &= v^1\mathbf{e}_1 + v^1\mathbf{e}_2 + v^1\mathbf{e}_3 \\ &= v^1 h_1\hat{\mathbf{e}}_1 + v^2 h_2\hat{\mathbf{e}}_2 + v^3 h_3\hat{\mathbf{e}}_3 \\ &= \tilde{v}^1\hat{\mathbf{e}}_1 + \tilde{v}^2\hat{\mathbf{e}}_2 + \tilde{v}^3\hat{\mathbf{e}}_3 \end{aligned}$$

where

$$\tilde{v}_1 = h_1 v^1 \qquad \tilde{v}_2 = h_2 v^2 \qquad \tilde{v}_3 = h_3 v^3$$

In terms of the $\tilde{v}^i$ and the $\hat{\mathbf{e}}_i$ the formula for curl becomes

$$\begin{aligned} \text{curl } \mathbf{v} &= \frac{1}{h_1 h_2 h_3}\left[\left(\frac{\partial(h_3^2 v^3)}{\partial u^2} - \frac{\partial(h_2^2 v^2)}{\partial u^3}\right) h_1\hat{\mathbf{e}}_1 + \text{etc.}\right. \\ &= \frac{1}{h_1 h_2 h_3}\left[\left(\frac{\partial(h_3 \tilde{v}^3)}{\partial u^2} - \frac{\partial(h_2 \tilde{v}^2)}{\partial u^3}\right) h_1\hat{\mathbf{e}}_1 + \text{etc.}\right. \\ &= \frac{1}{h_1 h_2 h_3}\begin{vmatrix} h_1\hat{\mathbf{e}}_1 & h_2\hat{\mathbf{e}}_2 & h_3\hat{\mathbf{e}}_3 \\ \frac{\partial}{\partial u^1} & \frac{\partial}{\partial u^1} & \frac{\partial}{\partial u^1} \\ h_1\tilde{v}^1 & h_1\tilde{v}^2 & h_1\tilde{v}^3 \end{vmatrix} \end{aligned}$$

This is the formula you find in reference books.

As an example, let us write the formula for Spherical Coordinates using the material for Spherical coordinates at the end of the previous section. Recall that $h_1 = 1,\ h_2 = \rho,\ h_3 = \rho\sin\theta$. Then

$$\text{curl } \mathbf{v} = \frac{1}{\rho^2\sin\theta}\begin{vmatrix} \hat{\mathbf{e}}_1 & \rho\,\hat{\mathbf{e}}_2 & \rho\sin\theta\,\hat{\mathbf{e}}_3 \\ \frac{\partial}{\partial\rho} & \frac{\partial}{\partial\theta} & \frac{\partial}{\partial\phi} \\ \tilde{v}^1 & \rho\tilde{v}^2 & \rho\sin\theta\tilde{v}^3 \end{vmatrix}$$

While the curl will only work in 3 dimensions, we can get formulas for grad and div in n dimensions with no extra effort at all. This is important since we might need either of them in 2 as well as 3 dimensions, and maybe even 4 dimensions for relativity. So we shall do them all at once.

There is no difficulty at all with grad. It inputs a function and outputs a vector so the immediate candidate for grad is $\Phi^{-1} \circ d$. This obviously works in rectangular coordinates since then $\Phi(\mathbf{e}_i) = du^i$. and thus

$$\begin{aligned} \text{grad } f &= \Phi^{-1}(\frac{\partial f}{\partial u^i}\, du^i) \qquad \text{(sum on } i \text{ understood)} \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial u^i}\,\mathbf{e}_i \end{aligned}$$

The last is our old friend the advanced calculus gradient. Notice that we could not use the summation convention here since both the i's in the last term count as low indices. (Recall that *high* index in a denominator counts as *low*.) Watch carefully in the next developments to see why this has happened.

Next recall that for general (not necessarily orthogonal) coordinates we have

$$\Phi(\mathbf{e}_i) = g_{ij}du^j \qquad \Phi^{-1}(du^i) = g^{ij}\mathbf{e}_j$$

and so

$$\begin{aligned} \text{grad } f &= \Phi^{-1}(df) \\ &= \Phi^{-1}(\frac{\partial f}{\partial u^i}du^i) \\ &= \frac{\partial f}{\partial u^i}g^{ij}\mathbf{e}_j \end{aligned}$$

which is nice and simple. For *orthogonal* coordinates we have

$$g^{ij} = \begin{cases} 0 & \text{if } i \neq j \\ \frac{1}{g_{ii}} = \frac{1}{h_i^2} & \text{if } i = j \end{cases}$$

and thus, and thus with natural basis vectors $\mathbf{e}_i$ and physical (normalized) basis vectors $\hat{\mathbf{e}}_i = \frac{1}{h_i}\mathbf{e}_i$ we have

$$\begin{aligned} \text{grad } f &= \sum_{i=1}^{n} \frac{1}{h_i^2}\frac{\partial f}{\partial u^i}\mathbf{e}_i \\ &= \sum_{i=1}^{n} \frac{1}{h_i}\frac{\partial f}{\partial u^i}\hat{\mathbf{e}}_i \end{aligned}$$

The second formula, with physical basis vectors $\hat{\mathbf{e}}_i$, is what one generally sees in reference books.

Our next task is the divergence. This is harder, since we must use the $*$ operator, but we can break it down into easy steps. The first observation we make is important for many purposes. Let

$$\omega = \sum_{i=1}^{n}(-1)^{i-1}v^i du^1 \wedge \cdots \wedge du^{i-1} \wedge du^{i+1} \wedge \cdots \wedge du^n$$

It is customary to write this as

$$\omega = \sum_{i=1}^{n}(-1)^{i-1}v^i du^1 \wedge \cdots \wedge \widehat{du^i} \wedge \cdots \wedge du^n$$

where the hat on the $\widehat{du^i}$ indicates that it is NOT THERE. This is just notation; it tends to make everything look cleaner, but you must keep your wits about

you and watch for it. Now when we find $d\omega$ it comes out very nice:

$$d\omega = \sum_{i,j=1}^{n} (-1)^{i-1} \frac{\partial v^i}{\partial u^j} du^j \wedge du^1 \wedge \cdots \wedge \widehat{du^i} \wedge \cdots \wedge du^n$$

Now note that if $j \neq i$ then the differential contain a repetition and thus give 0. Hence only the terms with $j = i$ need be retained and we have

$$\begin{aligned} d\omega &= \sum_{i,j=1}^{n} (-1)^{i-1} \frac{\partial v^i}{\partial u^j} du^j \wedge du^1 \wedge \cdots \wedge \widehat{du^i} \wedge \cdots \wedge du^n \\ &= \sum_{i=1}^{n} (-1)^{i-1} \frac{\partial v^i}{\partial u^i} du^i \wedge du^1 \wedge \cdots \wedge \widehat{du^i} \wedge \cdots \wedge du^n \\ &= \sum_{i=1}^{n} \frac{\partial v^i}{\partial u^i}\, du^1 \wedge \cdots \wedge du^i \wedge \cdots \wedge du^n \\ &= \left(\sum_{i=1}^{n} \frac{\partial v^i}{\partial u^i} \right) du^1 \wedge \cdots \wedge du^n \end{aligned}$$

where in one of the steps the du^i had to hop over $i-1$ other differentials to find its proper slot. You will notice has this has the look of divergence about it.

To utilize this formula we must get from our vector to an $(n-1)$-form. We can get from the vector to a 1-form using Φ and then to an $(n-1)$-form using $*$. From there, d will take us to an n-form and then another $*$ will get us back to a 0-form, or scalar, so $* \circ d \circ * \circ \Phi$ will get us from a vector to a scalar as we wish for divergence. Thus $* \circ d \circ * \circ \Phi$ is a good candidate for divergence. Let's try it in rectangular coordinates. In rectangular coordinates the natural basis $\mathbf{e}_1, \ldots \mathbf{e}_n$ is just $\mathbf{i}, \mathbf{j}, \mathbf{k}$ when $n = 3$

$$\begin{aligned} \mathbf{v} &= v^i \mathbf{e}_i \\ \Phi(\mathbf{v}) &= v^i du^i \\ *\Phi(\mathbf{v}) &= \sum_{i=1}^{n} v^i (-1)^{i-1} du^1 \wedge \cdots \wedge \widehat{du^i} \wedge \cdots \wedge du^n \\ d * \Phi(\mathbf{v}) &= \left(\sum_{i=1}^{n} \frac{\partial v^i}{\partial u^i} \right) du^1 \wedge \cdots \wedge du^n \end{aligned}$$

as we saw above. Finally

$$* d * \Phi(\mathbf{v}) = \sum_{i=1}^{n} \frac{\partial v^i}{\partial u^i} = \text{div } \mathbf{v}$$

which shows us that indeed the invariant form of div is $* \circ d \circ * \circ \Phi$. If you are worried about the expressions for $*$, they follow immediately from the general expressions below.

Before we launch into the final calculation for div, let's recall the formulas for the $*$ operator. For $\omega = \omega_i du^i$ we have

$$*\omega = \sum_{i,k=1}^{n} (-1)^{k-1} g^{ik} \sqrt{g}\, \omega_i \, du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n$$

where, as before, the hat on $\widehat{du^j}$ indicates the term is NOT THERE. This is derived in the Chapter 2. Also derived there are

$$\begin{aligned} \Omega_0 &= \sqrt{g}\, du^1 \wedge \ldots \wedge du^n \\ *\Omega_0 &= 1 \\ *1 &= \Omega_0 \end{aligned}$$

We also require the formula in general coordinates for Φ.

$$\Phi(\mathbf{v}) = \Phi(v^i \mathbf{e}_i) = g_{ij} v^i du^j$$

Now we are ready for the final calculation for div.

$$\begin{aligned} \text{div } \mathbf{v} &= *d * \Phi(\mathbf{v}) \\ &= *d * (g_{ij} v^i du^j) \\ &= *d \left(\sum_{k=1}^{n} (-1)^{k-1} g^{jk} \sqrt{g}\, (g_{ij} v^i) du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n \right) \end{aligned}$$

Since $g^{jk} g_{ij} = \delta_i^k$, the only non-zero terms are when $k = i$, so the above simplifies to

$$\begin{aligned} \text{div } \mathbf{v} &= *d \left(\sum_{i=1}^{n} (-1)^{i-1} \sqrt{g}\, v^i du^1 \wedge \ldots \wedge \widehat{du^i} \wedge \ldots \wedge du^n \right) \\ &= * \left(\sum_{i,j=1}^{n} (-1)^{i-1} \frac{\partial(\sqrt{g}\, v^i)}{\partial u^j} du^j \wedge du^1 \wedge \ldots \wedge \widehat{du^i} \wedge \ldots \wedge du^n \right) \\ &= * \left(\sum_{i=1}^{n} \frac{\partial(\sqrt{g}\, v^i)}{\partial u^i} du^1 \wedge \ldots \wedge du^i \wedge \ldots \wedge du^n \right) \\ &= * \left(\frac{1}{\sqrt{g}} \sum_{i=1}^{n} \frac{\partial(\sqrt{g}\, v^i)}{\partial u^i} \right) \sqrt{g}\, du^1 \wedge \ldots \wedge du^n \\ &= \left(\frac{1}{\sqrt{g}} \sum_{i=1}^{n} \frac{\partial(\sqrt{g}\, v^i)}{\partial u^i} \right) \end{aligned}$$

and this is the formula for div when $\mathbf{v}$ is expressed in natural coordinates $\mathbf{v} = v^i \mathbf{e}_i$. For orthogonal coordinates we have as usual $\mathbf{e}_i = h_i \hat{\mathbf{e}}_i$ and

$$\mathbf{v} = \sum_{i=1}^{n} v^i \mathbf{e}_i$$

$$\begin{aligned} &= \sum_{i=1}^{n} v^i h_i \hat{\mathbf{e}}_i \\ &= \sum_{i=1}^{n} \tilde{v}^i \hat{\mathbf{e}}_i \end{aligned}$$

where $\tilde{v}^i = h_i v^i$ are the physical coordinates with respect to the ortho*normal* basis $\hat{\mathbf{e}}_i$. Now using

$$\begin{aligned} \sqrt{g} &= h_1 h_2 \cdots h_n \\ \tilde{v}^i &= h_i v^i \qquad \text{no sum on } i \end{aligned}$$

we have

$$\begin{aligned} \text{div } \mathbf{v} &= \frac{1}{h_1 h_2 \cdots h_n} \sum_{i=1}^{n} \frac{\partial}{\partial u^i} \left(h_1 h_2 \cdots h_n v^i \right) \\ &= \frac{1}{h_1 h_2 \cdots h_n} \sum_{i=1}^{n} \frac{\partial}{\partial u^i} \left(h_1 \cdots h_{i-1} h_{i+1} \cdots h_n \tilde{v}^i \right) \end{aligned}$$

Lastly we want to give a formula for the Laplacian. In physics the formula for the Laplacian *in rectangular coordinates* is

$$\triangle f = \sum_{i=1}^{n} \frac{\partial f}{\partial u^i}$$

The normal notation for the Laplacian is $\triangle f$. However, in recent years in mathematics there has been a big tendency to use the symbol $\triangle f$ for the negative of the sum above. There are very good reasons for this[6]. So to make sure there will be no confusion, we will use the symbol $\triangle f$ for the Laplacian as customary in physics. It will be awhile before we need to deal with the mathematical Laplacian.

To find the Laplacian of functions in any coordinate system is now absurdly easy, because we have

$$\triangle f = \text{div grad } f$$

Thus we need only recall the two previously derived formulas for grad and div

$$\begin{aligned} \text{grad } f &= \frac{\partial f}{\partial u^i} g^{ij} \mathbf{e}_j \\ \text{div } \mathbf{v} &= \frac{1}{\sqrt{g}} \sum_{j=1}^{n} \frac{\partial (\sqrt{g}\, v^j)}{\partial u^j} \end{aligned}$$

[6]For example, the mathematical Laplacian has *positive* or 0 eigenvalues; the physics one has negative or 0 eigenvalues

Thus it comes down to substituting

$$v^j = \frac{\partial f}{\partial u^i} g^{ij}$$

into the formula for div $\mathbf{v}$ to get

$$\triangle f \quad = \quad \frac{1}{\sqrt{g}} \sum_{i,j=1}^{n} \frac{\partial}{\partial u^j} \left(\sqrt{g}\, g^{ij} \frac{\partial f}{\partial u^i} \right)$$

One sees immediately that in orthogonal coordinates we have

$$\triangle f = \frac{1}{h_1 \cdots h_n} \sum_{i=1}^{n} \frac{\partial}{\partial u^i} \left(h_1 \cdots h_{i-1} h_{i+1} \cdots h_n \frac{\partial f}{\partial u^i} \right)$$

Problems for Section 1.17

You will need the results you obtained in the problems for section 1.16.

1. $\mathbb{R}^2$ and polar coordinates r, θ. Derive the formula for grad f when expressed in terms of the normalized physical coordinates $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2$. With $\mathbf{v}$ expressed in terms of $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2$, $\mathbf{v} = \hat{\mathbf{v}}^1\, \hat{\mathbf{e}}_1 + \hat{\mathbf{v}}^2\, \hat{\mathbf{e}}_2$ find the formula for div f. Put these together to get the formula for $\triangle f$.

2. $\mathbb{R}^3$ and cylindrical coordinates r, ϕ, z. Using the techniques in this section derive the formulas for grad f, div $\mathbf{v}$ and $\triangle f$ all in terms of the normalized physical coordinates $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$.

1.18 Surfaces and Manifolds

Up to this point we have confined ourselves mostly to $\mathbb{R}^n$ but now it is time to take a more general viewpoint, especially since it will require no additional effort. We have occasionally talked about two dimensional surfaces in $\mathbb{R}^3$ but there is nothing special about that configuration. A convenient place to start our discussion is $m-$dimensional subspaces in $\mathbb{R}^n$. The word *surface* is usually used for a 2-dimensional object; the corresponding word for an $m-$dimensional subspace of $\mathbb{R}^n$ in *embedded manifold*. (If $m = n-1$ then the embedded manifold is often called a *hypersurface*. This is a special case where things are generally nicer than they are in general.) The *embedded* in *embedded manifold* refers to the surrounding $n-$dimensional space. There are manifolds which are not embedded, (for example, the 4-dimensional space of the Universe of General Relativity is usually thought of as not embedded in a higher dimensional space,) and we will briefly discuss this later and also in the second chapter. The advantage of embedding is that a) it seems familiar and b) it is easy to define tangent vectors.

Our discussion here will be more or less intuitive, and we will not discuss various pathologies that can occur to make life difficult, as these are not usual

in ordinary physics. Also, we will imitate some of the constructions that are used in the more general case, of nonembedded manifolds in order to make the transition easier.

First, it is important that the the surface neither cross itself, nor approach itself. The first is easy to avoid; the second is more tricky. Recall that a ball in $\mathbb{R}^n$ is the set of points lying within a certain distance of the central point. More explicitly $B_r(p) = \{x \in \mathbb{R}^n \big| \, |x - p| < r\}$. Also recall that a homeomorphism is a one to one onto map from one set to another which is continuous and has a continuous inverse. Now let p be a point on the manifold. We require that there is an r small enough so that the the intersection of $B_r(p)$ with the manifold be homeomorphic to a disk in $\mathbb{R}^m$. Put another way, every point p on the manifold has a neighborhood that looks like a disk in $\mathbb{R}^m$ (and the neighborhood is found by intersecting the manifold with a $B_r(p)$). That is why the manifold has m dimensions.

Now we need to superimpose coordinates on this situation. That's easy; the disk in $\mathbb{R}^m$ has coordinates $u^1, \ldots, u^n$ and if q on the manifold corresponds to y in the disk which has coordinates $u^1, \ldots, u^m$, which we can then use for p. Home free? By no means. The most important thing comes next. Suppose $\tilde{p}$ is in the neighborhood of p (as above) and on the manifold. Then $\tilde{p}$ also has a neighborhood defined by $B_s(\tilde{p})$ and a homomorphism of its own to a disk in $\mathbb{R}^m$, which will give the neighborhood of $\tilde{p}$ it's own set of coordinates, say $\tilde{u}^1, \ldots, \tilde{u}^m$. Then, each point in the overlap of $B_r(p)$ and $B_s(\tilde{p})$ will have *two* sets of coordinates $u^1, \ldots, u^m$ and $\tilde{u}^1, \ldots, \tilde{u}^m$. Each set will be a function of the other:

$$\begin{aligned} \tilde{u}^i &= \tilde{u}^i(u^1, \ldots, u^m) \\ u^i &= u^i(\tilde{u}^1, \ldots, \tilde{u}^m) \end{aligned}$$

The critical thing here is that these functions have many derivatives. Many means, for physics purposes, at least three continuous derivatives. The manifold is then called a C^3-manifold. In mathematics it is customary to avoid the boring details and require the functions to have infinitely many continuous derivatives so it is C^∞-manifold. This seemingly boring requirement is the key to the whole subject; it is what makes the manifold smooth like a ball rather than unsmooth like a cube. It takes some contemplation to see this but it's true.

One more comment; in classical tensor analysis they are continually talking about coordinate changes as if they made some big difference. They do; the talk of coordinate changes in tensor analysis is just how that subject deals with the material in the previous paragraph[7].

There is a little more to the story, because, after all, we have to put the manifold inside $\mathbb{R}^n$ and this requires a little more to be put into the definition. The requirement is easy to state. Since the manifold is inside $\mathbb{R}^n$ and p will thus have coordinates $x^1, \ldots, x^n$. These will be functions of p's disk coordinates

[7] This is not a well known fact even among the tensor users. Reveal it only to trusted friends

$u^1, \ldots, u^m$; that is

$$x^i = x^i(u^1, \ldots, u^m) \qquad i = 1, \ldots, n$$

The requirement is that the $n \times m$ matrix

$$\begin{pmatrix} \frac{\partial x^1}{\partial u^1} & \cdots & \frac{\partial x^1}{\partial u^m} \\ \frac{\partial x^2}{\partial u^1} & \cdots & \frac{\partial x^2}{\partial u^m} \\ \vdots & \vdots & \vdots \\ \frac{\partial x^n}{\partial u^1} & \cdots & \frac{\partial x^n}{\partial u^m} \end{pmatrix}$$

have rank m. Since $m \leq n$, m is the largest rank it *could* have, so we often say *the matrix has maximal rank*. The reason this is important is that we could set up a coordinate system around any p for $\mathbb{R}^n$ that used $u^1, \ldots, u^m$ for the first m coordinates and then $n - m$ other coordinates to finish the job, but we do not need to pursue this.

We should also mention that while in theory we have set the system up to use disks in $\mathbb{R}^m$ to make coordinates, you can actually do it with any regions you like rather than disks. We only did it with disks to make it precise and visual. Things are rather looser in practise.

The test to see that this condition is satisfied is to check that all the $m \times m$ minors of the matrix are non-zero. It is now time for an example. We will use (as usual) the sphere in 3-space, which illustrates the concepts nicely. The coordinates will be θ, ϕ where ϕ is the longitude and θ is the angle off the z-axis[8].

The $\mathbb{R}^3$ coordinates of the point with coordinates θ, ϕ are

$$\begin{aligned} x &= \sin\theta\cos\phi \\ y &= \sin\theta\sin\phi \\ z &= \cos\theta \\ \mathbf{R} &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \end{aligned}$$

The natural basis of the tangent space is

$$\begin{aligned} \mathbf{e}_1 = \frac{\partial \mathbf{R}}{\partial\theta} &= (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta) \\ \mathbf{e}_2 = \frac{\partial \mathbf{R}}{\partial\phi} &= (-\sin\theta\sin\phi, \sin\theta\cos\phi, 0) \\ \hat{\mathbf{n}} = \frac{e_1 \times e_2}{|e_1 \times e_2|} &= (\sin^2\theta\cos\phi, \sin^2\theta\sin\phi, \sin\theta\cos\theta)/\sin\theta \\ &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \end{aligned}$$

The matrix discussed above is then

$$\begin{pmatrix} \cos\theta\cos\phi & -\sin\theta\sin\phi \\ \cos\theta\sin\phi & \sin\theta\cos\phi \\ -\sin\theta & 0 \end{pmatrix}$$

[8]Mathematicians beware; math convention is opposite!

The minors are

$$\begin{aligned}
\text{Row } 23 &= \begin{pmatrix} \cos\theta\sin\phi & \sin\theta\cos\phi \\ -\sin\theta & 0 \end{pmatrix} = \sin^2\theta\cos\phi \\
\text{Row } 13 &= \begin{pmatrix} \cos\theta\cos\phi & -\sin\theta\sin\phi \\ -\sin\theta & 0 \end{pmatrix} = -\sin^2\theta\sin\phi \\
\text{Row } 12 &= \begin{pmatrix} \cos\theta\cos\phi & -\sin\theta\sin\phi \\ \cos\theta\sin\phi & \sin\theta\cos\phi \end{pmatrix} = \sin\theta\cos\theta
\end{aligned}$$

There is a lot that can be learned from this example. Notice first that the minors are all 0 at the poles where $\theta = 0$ or $\theta = \pi$. This is reflected also in $\mathbf{e}_2$ which is 0 at the poles and thus not much of a basis vector for the tangent space. Note that at the poles ϕ is not defined either. Note that in $\hat{\mathbf{n}} = (\mathbf{e}_1 \times \mathbf{e}_2)/|\mathbf{e}_1 \times \mathbf{e}_2|$ both numerator and denominator are 0, so $\hat{\mathbf{n}}$ is not well defined though it seems to be after the 0's disappear because of the cancellation of $\sin\theta$.

The proper response to this is to select a second coordinate system, perhaps one with east and west poles, to cover the two offending points with proper coordinates. Needless to say, this is hardly ever done in this and many other cases, because with a little sympathetic treatment it is possible to get around the troubles at the poles. However, it is really important to notice such bad points and to make sure whatever you are doing makes sense at such bad points. Often it does, sometimes it doesn't. Beware.

Now on to other things. Notice in the example how easy it was to get the basis of tangent vectors $\mathbf{e}_1, \mathbf{e}_2$, one for each coordinate, at each point of the manifold. The vector space spanned by $\mathbf{e}_1, \mathbf{e}_2$ is called the Tangent plane, and is denoted by $T_p(S)$ (where S is the sphere). In the more general case of an embedded m-dimensional manifold S in n-dimensional space, we will have m tangent vectors at each point, $\mathbf{e}_1, \ldots, \mathbf{e}_m$, and they will be a basis for an m-dimensional space $T_p(S)$ called the *Tangent Space* to the manifold S at p. The condition on the minors guarantees that the $\mathbf{e}_i$ will be linearly independent.

Consider an m-dimensional manifold S with coordinates $u^1, \ldots, u^m$ and basis vectors $\mathbf{e}_i = \partial\mathbf{R}/\partial u^i$. Now let $\mathbf{v} \in T_p(S)$. Then $\mathbf{v} = \sum v^i \mathbf{e}_i$ and we need a function whose input is $\mathbf{v}$ and whose output is v^i. This function is du^i. Although the idea is new to you, this is what du^i actually is. Thus the defining equation is

$$du^i(\mathbf{v}) = v^i$$

(We have seen this before; I'm just reminding you in this new context.) If you are familiar with the the space of linear functionals of a vector space, then $du^1, \ldots, du^m$ form a basis for this space of linear functionals (called the *dual space* and denoted by $T_p^*(S)$.

When you follow this path for awhile, you become aware of what a small role the embedding space $\mathbb{R}^n$ plays in the theory. Riemann also wondered about this, and eventually showed how to get around it. Let us ask what the important contributions of the embedding space $\mathbb{R}^n$ are. There are two most important contributions. First, the embedding space provides an inner product for the

$T_p(S)$ which live inside it. Second, it provides us with a convenient way (visible in the sphere example) of finding tangent vectors and thus the tangent space $T_p(S)$. The first is a critical idea; the second just requires technical trickery. For the first, Riemann proposed attaching an inner product in the form of an $m \times m$ positive definite matrix to each point of S. In doing this Riemann invented the Riemannian Manifold. For the second, the tangent vectors, Tensor Analysis defined a vector as an object that changes in certain ways under coordinate changes. This works but is a bit unsatisfying. The more modern way is to define tangent vectors as directional derivatives on the manifold. We will look at this further in chapter 2 where we will define differentiable manifolds, although most of the work has already been done in this section. For the remainder of this chapter we will stay with embedded manifolds.

We must also mention that the theory we develop subsequently must be modified if the manifold is not *orientable*. Examples of non-orientable manifolds are the Möbius strip and the Klein bottle. The problem is that when a right handed orthonormal basis of the tangent space is slid around the median line of the Möbius strip it returns as a left handed orthonormal basis. The easiest definition of orientable is that an n-dimensional manifold K (embedded or not) is orientable if an only if it has a never vanishing form $\omega \in \Lambda^n(K)$. Such a form is called a topform. From it one can always manufacture a volume form if K has an inner product (e.g. is a Riemannian Manifold). We will not consider non-orientable manifolds in what follows and some of the material will not work on them. See Chapter 2 for more discussion of this matter.

1.19 The Dualizing Operator $*$

One cannot get very far in differential forms without the dualizing operator $* : \Lambda^r(\mathbb{R}) \to \Lambda^{n-r}(\mathbb{R})$. This operator reflects analytically certain geometric properties studied in higher geometry courses called duality properties. Sadly we cannot take the time to talk about this geometry, which would take many pages to deal with and anyway is not a matter of great practical importance for most of physics. In this section we will give a brief introduction to $*$ which will suffice for most of the more elementary applications, as for example Maxwell's equations which we present in a later section.

In most elementary applications of $*$ we deal with the cases

$$\begin{aligned} r = 0 \qquad & * : \Lambda^0(\mathbb{R}) \to \Lambda^n(\mathbb{R}) \\ r = n \qquad & * : \Lambda^n(\mathbb{R}) \to \Lambda^0(\mathbb{R}) \end{aligned}$$

and

$$\begin{aligned} r = 1 \qquad & * : \Lambda^1(\mathbb{R}) \to \Lambda^{n-1}(\mathbb{R}) \\ r = n-1 \qquad & * : \Lambda^{n-1}(\mathbb{R}) \to \Lambda^1(\mathbb{R}) \end{aligned}$$

In each case it is only necessary to derive one of the two formulas; the other is

then obtained from

$$** \omega = (-1)^{r(n-r)}\omega \qquad \text{for } \omega \in \Lambda^r(\mathbb{R})$$

To define $*$ it is necessary first to put an inner product on each Λ^r. This is easily done; We define the inner product with

$$\omega = \omega_i du^i \quad \text{and} \quad \eta = \eta_i du^i$$

by

$$(\omega, \eta) = g^{ij}\omega_i\eta_j$$

We can now put an inner product on each Λ^r as follows

$$(\omega_1 \wedge \ldots \wedge \omega_r,\ \eta_1 \wedge \ldots \wedge \eta_r) \ = \ \begin{vmatrix} (\omega_1, \eta_1) & \ldots & (\omega_1, \eta_r) \\ \vdots & \vdots & \vdots \\ (\omega_r, \eta_1) & \ldots & (\omega_n, \eta_r) \end{vmatrix}$$

and extending by linearity. This defines an inner product on Λ^r for $1 \leq r \leq n$, but we also need it on Λ^0 which is just the scalars. For these we define

$$(r, s) \ = \ rs \qquad \text{for } r, s \text{ scalers}$$

For completeness, we remark that if $\omega \in \Lambda^r$ and $\eta \in \Lambda^s$ and $r \neq s$, $0 \leq r, s \leq n$, then we put

$$(\omega,\ \eta) = 0$$

Our most important use of this formula is the following case where $r = n$:

$$(\omega_1 \wedge \ldots \wedge \omega_n,\ \eta_1 \wedge \ldots \wedge \eta_n) \ = \ \begin{vmatrix} (\omega_1, \eta_1) & \ldots & (\omega_1, \eta_n) \\ \vdots & \vdots & \vdots \\ (\omega_n, \eta_1) & \ldots & (\omega_n, \eta_n) \end{vmatrix}$$

Noting that the above formula gives us

$$(du^r, du^s) = (\delta^i_r du^i, \delta^j_s du^j) = g^{ij}\delta^i_r\delta^j_s = g^{rs}$$

we have with the inner product in Λ^n

$$\begin{aligned} (du^1 \wedge \ldots \wedge du^n,\, du^1 \wedge \ldots \wedge du^n) \quad &= \quad \begin{vmatrix} (du^1, du^1) & \ldots & (du^1, du^n) \\ \vdots & \vdots & \vdots \\ (du^n, du^1) & \ldots & (du^n, du^n) \end{vmatrix} \\ &= \quad \begin{vmatrix} g^{11} & \ldots & g^{1n} \\ \vdots & \vdots & \vdots \\ g^{n1} & \ldots & g^{nn} \end{vmatrix} \ = \ \det\,(g^{i,j}) \ = \ \frac{1}{g} \end{aligned}$$

Thus

$$(\sqrt{g}\, du^1 \wedge \ldots \wedge du^n,\, \sqrt{g} du^1 \wedge \ldots \wedge du^n) = 1$$

We will define

$$\Omega_0 \stackrel{def}{=} \sqrt{g}\, du^1 \wedge \ldots \wedge du^n$$

and refer to it as the *normalized topform*. The normalized topform is unique up to sign; it will change sign if two variables interchange their numbers. Choice of a sign is the same as choice of an orientation. The reason it is essentially unique is that Λ^n is one-dimensional, so there are only two elements of size 1 and they are negatives of one another.

If M is a n-dimensional region than the volume of M is

$$\text{vol}(M) = \int_M \Omega_0$$

There are of course subtleties here; for example the sign could come out wrong. And of course you have to prove this works using whatever your definition of volume is. We will ignore these problems. Also we mention that if f is the density of something, electrical charge for example, or gas, then the amount of the stuff will be

$$\text{Amount of stuff in } M = \int_M f\, \Omega_0$$

Next we want the official definition of $*$ which is at last possible since we have the inner product on Λ^r. For any $\alpha \in \Lambda^{n-r}$ we will always have

$$\omega \wedge \alpha = k\, \Omega_0$$

for some constant k. Then for $\eta \in \Lambda^r$ we define $*\eta$ as that unique element of Λ^{n-r} for which the constant k is (ω, η), which comes down to

$$\boxed{\omega \wedge *\eta \;=\; (\omega, \eta)\, \Omega_0}$$

This is the most important equation involving $*$; if you remember this you can derive everything else. The existence and uniqueness of $*\eta$ are derived in Chapter 2, but it is not a difficult matter; it comes down to the representation of a linear functional in an inner product space.

Some equations readily follow from the basic equation. For example, we have

$$\begin{aligned} \omega \wedge *\eta \;=\; (\omega,\, \eta)\, \Omega_0 \quad &= \quad (\eta,\, \omega)\, \Omega_0 \;=\; \eta \wedge *\omega \\ \omega \wedge *\eta \quad &= \quad \eta \wedge *\omega \end{aligned}$$

which has important uses, for example in the next section on the codifferential. Other important equations like

$$** \omega = (-1)^{r(n-r)}\omega \qquad \text{for } \omega \in \Lambda^r$$

do not follow so readily from the basic equation. We will have more to say about this equation later.

Next, using the basic equation $\omega \wedge *\eta = (\omega, \eta)\,\Omega_0$ we want to derive the formulas for $*$ for general coordinates and 1-forms. Recall that a hat over a term means that it is *missing*. We have $\{du^1, \ldots, du^n\}$ is a basis for Λ^1. A basis for Λ^{n-1} is

$$du^1 \wedge \ldots \wedge \widehat{du^i} \wedge \ldots \wedge du^n \qquad 1 \leq i \leq n$$

Thus, since $du^i \in \Lambda^1$, we have $*du^i \in \Lambda^{n-1}$ and then we can express $*du^i$ as

$$*du^i = \sum_{j=1}^{n} (-1)^{j-1} a^j du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^n$$

where the a^j are functions of $u^1, \ldots, u^n$ which we must determine and the $(-1)^{j-1}$ is inserted for convenience. We now compute both sides of the basic equation and setting them equal will give us the a^j. The basic equation is $du^k \wedge *du^i = (du^k, du^i)\Omega_0$. Computing the left side of the basic equation we have

$$\begin{aligned} du^k \wedge *du^i &= du^k \wedge \left(\sum_{j=1}^{n} (-1)^{j-1} a^j du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^n \right) \\ &= \sum_{j=1}^{n} (-1)^{j-1} a^j du^k \wedge du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^n \end{aligned}$$

All terms on the right side in the sum will be 0 except the one where $j = k$ because if $j \neq k$ there will be repetition in the differentials killing the term. Thus

$$du^k \wedge *du^i = a^k du^1 \wedge \ldots \wedge du^k \wedge \ldots \wedge du^n$$

Notice how the $(-1)^{j-1}$ was used to return du^k to its proper place in the product of differentials. This is a good trick to remember, though it is never important, just convenient. Next we compute the right side of the of the basic equation. We have

$$(du^k, du^i)\Omega_0 = g^{ki}\sqrt{g}\, du^1 \wedge \ldots \wedge du^n$$

Comparing the two expressions we see that

$$a^k = g^{ki}\sqrt{g}$$

and thus

$$\begin{aligned} *du^i &= \sum_{k=1}^{n} (-1)^{k-1} a^k du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n \\ &= \sum_{k=1}^{n} (-1)^{k-1} g^{ki}\sqrt{g} du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n \end{aligned}$$

This equation

$$\boxed{*du^i = \textstyle\sum_{k=1}^n (-1)^{k-1} g^{ki} \sqrt{g} du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n}$$

is of fundamental importance. You have already seen it used in the derivation of the formulas for curvilinear coordinates.

It is possible to determine $*du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n$ in the same way but that method leads into a forest of determinants. Instead, we will use the formula $**\omega = (-1)^{r(n-r)}\omega$ with $r = 1$ and some trickery involving the inverse matrices (g^{ji}) and (g_{ik}).

$$\begin{aligned}
*du^i &= \sum_{j=1}^n (-1)^{j-1} g^{ji} \sqrt{g}\, du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^n \\
\sum_{i=1}^n g_{ik} * du^i &= \sum_{i,j=1}^n (-1)^{j-1} g^{ji} g_{ik} \sqrt{g}\, du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^n \\
&= \sum_{j=1}^n (-1)^{j-1} \left(\sum_{j=1}^n g^{ji} g_{ik} \right) \sqrt{g}\, du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^n \\
&= \sum_{j=1}^n (-1)^{j-1} \delta_k^j \sqrt{g}\, du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^n \\
&= (-1)^{k-1} \sqrt{g}\, du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n
\end{aligned}$$

Hence, starring both sides, we have

$$\begin{aligned}
(-1)^{k-1} \sqrt{g} * du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n &= \sum_{i=1}^n g_{ik} * * du^i \\
&= \sum_{i=1}^n g_{ik} (-1)^{1(n-1)}\, du^i \\
*du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n &= (-1)^{-k+1} (-1)^{n-1} \sum_{i=1}^n g_{ik} du^i
\end{aligned}$$

giving us our final answer

$$\boxed{*du^1 \wedge \ldots \wedge \widehat{du^k} \wedge \ldots \wedge du^n = (-1)^{n-k} \textstyle\sum_{i=1}^n g_{ik} du^i}$$

This is not quite sufficient for our needs. We also need formulas for $*$ when $r = 0$ and when $r = n$. These are easy. Recall that Λ^0 is just the scalars $\mathbb{R}$ and that the inner product is just the ordinary product in $\mathbb{R}$ and that a basis for $\mathbb{R}$ is just the single scalar 1. Then

$$*1 = \Omega_0$$

since

$$1 \wedge *1 = 1 \wedge \Omega_0 = \Omega_0 = (1,1)\Omega_0$$

as required. Then

$$*\Omega_0 = **1 = (-1)^{0(n-0)}1 = 1$$

We have determined the formulas in generalized coordinates for 0-forms, 1-forms, $(n-1)$-forms and n-forms. It is possible to work out similar formulas for r-forms, but these are not as important in applied mathematics and to derive them we would have to introduce some equipment involving minors of determinants. Hence we will put this off to Chapter 2 and concentrate here on the more practically oriented formulas for *orthogonal* coordinates which are relatively easy to derive. Recall that for orthogonal coordinates we have for the position vector $\mathbf{R} = \mathbf{R}(u^1, \ldots, u^n)$

$$\begin{aligned}
\mathbf{e}_j &= \frac{\partial \mathbf{R}}{\partial u^i} \\
h_i &= (\mathbf{e}_i, \mathbf{e}_i) \\
\hat{\mathbf{e}}_i &= \frac{1}{h_i}\mathbf{e}_i \\
(du^i, du^j) &= g^{ii} = \frac{1}{h_i^2} \\
(h_i du^i, h_i du^j) &= 1 \\
E_i &= h_i du^i
\end{aligned}$$

so that $\{h_1 du^1, \ldots, h_n du^n\}$ form an orthonormal set. For convenience let us set $E_i = h_i du^i$. Then $(E_i, E_j) = \delta_{ij}$. Now from the basic equation $\omega \wedge *\eta = (\omega, \eta)\Omega_0$ we see immediately that

$$*(E_{i_1} \wedge \cdots \wedge E_{i_r}) = \operatorname{sgn}(\sigma)(E_{j_1} \wedge \cdots \wedge E_{j_{n-r}})$$

where $\{j_1, \ldots, j_{n-r}\} = \{1, \ldots, n\} - \{i_1, \ldots, i_r\}$ and σ is the permutation

$$\sigma = \left(\begin{array}{ccc|ccc} 1 & \ldots & r & r+1 & \ldots & r \\ i_1 & \ldots & i_r & j_1 & \ldots & j_{n-r} \end{array} \right)$$

Here the convention is $1 \leq i_1 < \ldots < i_r \leq n$ and $1 \leq j_1 < \ldots < j_{n-r} \leq n$ but in fact the formula will work just as well with the i's and j's in any order. Indeed it suffices to check this for ω running through a basis $\{E_{k_1} \wedge \cdots \wedge E_{k_r} \quad 1 \leq i_1, \ldots, \leq k_r\}$ of Λ^r.

$$(E_{k_1} \wedge \cdots \wedge E_{k_r}) \wedge *(E_{i_1} \wedge \cdots \wedge E_{i_r}) \quad = \quad E_{k_1} \wedge \cdots \wedge E_{k_r} \wedge \operatorname{sgn}(\sigma)(E_{j_1} \wedge \cdots \wedge E_{j_{n-r}})$$

Now if $\{k_1, \ldots, k_r\} \neq \{i_1, \ldots, i_r\}$ there will be repetitions on the right side and the result will be 0. Hence we will take $\{k_1, \ldots, k_r\} = \{i_1, \ldots, i_r\}$ and since

both are in increasing order we have $i_\ell = k_\ell$ and

$$\begin{aligned}(E_{i_1} \wedge \cdots \wedge E_{i_r}) \wedge *(E_{i_1} \wedge \cdots \wedge E_{i_r}) &= E_{i_1} \wedge \cdots \wedge E_{i_r} \wedge \operatorname{sgn}(\sigma)(E_{j_1} \wedge \cdots \wedge E_{j_{n-r}}) \\ &= E_1 \wedge \cdots \wedge E_n \\ &= (h_1 du^1) \wedge \cdots \wedge (h_n du^n) \\ &= h_1 \cdots h_n \, du^1 \wedge \cdots \wedge du^n \\ &= \sqrt{g}\, du^1 \wedge \cdots \wedge du^n \\ &= \Omega_0\end{aligned}$$

On the other hand

$$(E_{k_1} \wedge \cdots \wedge E_{k_r},\, E_{i_1} \wedge \cdots \wedge E_{i_r})\Omega_0 \quad = \quad \det\Big((E_{k_\ell}, E_{i_m})\Big)\Omega_0$$

Now if $\{k_1, \ldots, k_r\} \neq \{i_1, \ldots, i_r\}$ then there will be a row in the determinant which is entirely 0. Hence we take $\{k_1, \ldots, k_r\} = \{i_1, \ldots, i_r\}$ and since both are in increasing order we have $i_\ell = k_\ell$, the determinant has 1's on the main diagonal and 0's elsewhere, so the determinant is 1 and the result is

$$(E_{i_1} \wedge \cdots \wedge E_{i_r},\, E_{i_1} \wedge \cdots \wedge E_{i_r})\Omega_0 \quad = \quad \Omega_0$$

Hence if $\{k_1, \ldots, k_r\} \neq \{i_1, \ldots, i_r\}$ both sides of the fundamental equation are 0 and if $\{k_1, \ldots, k_r\} = \{i_1, \ldots, i_r\}$ then both sides are Ω_0, proving that

$$\boxed{*(E_{i_1} \wedge \cdots \wedge E_{i_r}) = \operatorname{sgn}(\sigma)(E_{j_1} \wedge \cdots \wedge E_{j_{n-r}})}$$

Replacing the E_i by $h_i du^i$ we have

$$\boxed{*(h_{i_1} du^{i_1} \wedge \cdots \wedge h_{i_r} du^{i_r}) = \operatorname{sgn}(\sigma)(h_{j_1} du^{j_1} \wedge \cdots \wedge h_{j_{n-r}} du^{j_{n-r}})}$$

and from this

$$*(du^{i_1} \wedge \cdots \wedge du^{i_r}) = \operatorname{sgn}(\sigma)\frac{h_{j_1} \cdots h_{j_{n-r}}}{h_{i_1} \cdots h_{i_r}}(du^{j_1} \wedge \cdots \wedge du^{j_{n-r}})$$

where σ is the permutation

$$\sigma = \left(\begin{array}{ccc|ccc} 1 & \ldots & r & r+1 & \ldots & n \\ i_1 & \ldots & i_r & j_1 & \ldots & j_{n-r} \end{array}\right)$$

This gives us a formula valid for any r

This is also a convenient moment to prove the formula

$$** \omega = (-1)^{r(n-r)}\omega$$

We can do this using the basis $\{E_1, \ldots, E_n\}$. We define the *reverse* $\tilde{\sigma}$ of σ to be, with the above σ,

$$\tilde{\sigma} = \left(\begin{array}{ccc|ccc} 1 & \ldots & n-r & n-r+1 & \ldots & n \\ j_1 & \ldots & j_{n-r} & i_1 & \ldots & i_r \end{array} \right)$$

Now if we move the r i's each past the $n-r$ j's there will be a total of $r(n-r)$ hops to get from $\tilde{\sigma}$ back to σ. Hence if it takes s hops to return σ to the identity, $\mathrm{sgn}(\sigma) = (-1)^s$ and

$$\begin{aligned} \mathrm{sgn}(\tilde{\sigma}) &= (-1)^{s+r(n-r)} = (-1)^{r(n-r)}\mathrm{sgn}(\sigma) \\ \mathrm{sgn}(\sigma)\mathrm{sgn}(\tilde{\sigma}) &= (-1)^{r(n-r)} \end{aligned}$$

Then, with $\omega = E_{i_1} \wedge \cdots \wedge E_{i_r}$ we have

$$\begin{aligned} **\omega &= **E_{i_1} \wedge \cdots \wedge E_{i_r} \\ &= \mathrm{sgn}(\sigma) * E_{j_1} \wedge \cdots \wedge E_{j_{n-r}} \\ &= \mathrm{sgn}(\sigma)\mathrm{sgn}(\tilde{\sigma}) E_{i_1} \wedge \cdots \wedge E_{i_r} \\ &= (-1)^{r(n-r)} E_{i_1} \wedge \cdots \wedge E_{i_r} \\ &= (-1)^{r(n-r)}\omega \end{aligned}$$

Since this is true for the elements of a basis, it will be true for all ω by linearity.

1.20 The Codifferential δ

To deal with the Laplacian for forms, we need the codifferential $\delta : \Lambda^r \to \Lambda^{r-1}$. We will use this in our treatment of Maxwell's equations. For example, the condition of Lorenz is naturally expressed in terms of the codifferential.

In contrast to previous constructions, δ depends for its very definition on the presence of an inner product for the forms on a manifold M. Using the $*$ we can easily define such an inner product. Let ω, $\eta \in \Lambda^r$. Then the inner product is defined by either of the following

$$((\omega, \eta)) = \int_M \omega \wedge *\eta = \int_M (\omega, \eta)\Omega_0$$

Some points to remember. The inner product (ω, η) is an inner product that lives in $T_p(M)$ for each $p \in M$. It and $*$ are strictly local, being algebraic constructions on each tangent space $T_p(M)$ separately. On the other hand, $((\omega, \eta))$ depends on integrating the information given at each p over the manifold. It is thus a global object. Second, although we express it in our formulas for a single coordinate system, in many cases it will be necessary to cut M into separate pieces each of which lives inside the domain of a coordinate system, and to use that coordinate system when integrating over that piece. Sometimes, as for the sphere, we can use a single coordinate system even though that system is bad at specific points, like the North and South pole. This is best considered dumb

luck, and one should always be careful when doing it. But it does work often in applications as long as nothing is discontinuous at the bad points.

Now that we have the inner product on forms, we can consider $((\omega, d\eta))$ where $\omega \in \Lambda^r$ and $\eta \in \Lambda^{r-1}$. We will consider two possible scenarios

a The manifold M has no boundary; $\partial M = \emptyset$ Such a manifold is often referred to as *closed*. The idea "no boundary" includes "no boundary at infinity" so that in this case the manifold M is compact.

b Either ω or η vanishes off a compact subset M_0 of M, so the integral will always be finite.

With one of these situations in place, we can do an integration by parts in the way you have seen done in differential equations, and come up with formal adjoint δ for d. After all this preamble, the actual calculation is quite short. We need to recall that, with $\omega \in \Lambda^{r-1}$ and $\eta \in \Lambda^r$

$$\begin{aligned} \omega \wedge *\eta &= \eta \wedge *\omega \\ d(\omega \wedge *\eta) &= d\omega \wedge *\eta + (-1)^{r-1}\omega \wedge d * \eta \end{aligned}$$

Also note because of the above assumptions on ω and η we have by Stokes theorem

$$\int_K d(\omega \wedge \eta) = \int_{\partial K} \omega \wedge \eta = 0$$

We can now derive the formula for the (formal) adjoint of d.

$$\begin{aligned} ((d\omega, \eta)) &= \int_K d\omega \wedge *\eta \\ &= \int_K d(\omega \wedge \eta) - (-1)^{r-1} \int_K \omega \wedge d * \eta \\ &= 0 + (-1)^r \int_K \omega \wedge d * \eta \\ &= (-1)^r (-1)^{(n-r+1)(r-1)} \int_K \omega \wedge * * d * \eta \end{aligned}$$

Now we must simplify the exponent. Recall that

$$\begin{aligned} k^2 &\equiv k \qquad (\text{mod } 2) \\ -k &\equiv k \\ 2k &\equiv 0 \\ k(k-1) &\equiv 0 \end{aligned}$$

Thus

$$r + (n-r+1)(r-1) \equiv r + (n-r)(r-1) + r - 1 \qquad (\text{mod } 2)$$

$$\begin{aligned} &\equiv (n-r)(r-1)-1 \\ &\equiv n(r-1)-r(r-1)-1 \\ &\equiv n(r-1)-1+2 \\ &\equiv n(r-1)+1 \end{aligned}$$

Hence

$$\begin{aligned} ((d\omega,\eta)) &= (-1)^{n(r-1)+1}\int_K \omega\wedge **d*\eta \\ &= \int_K \omega\wedge *\Big((-1)^{n(r-1)+1}*d*\eta\Big) \end{aligned}$$

Thus if we set

$$\delta : \Lambda^r \to \Lambda^{r-1}$$

defined by

$$\boxed{\delta\eta = (-1)^{n(r-1)+1}*d*\eta \qquad \eta\in\Lambda^r}$$

we have

$$\boxed{((d\omega,\eta)) = ((\omega,\delta\eta)) \qquad \omega\in\Lambda^{r-1},\ \eta\in\Lambda^r}$$

as we wanted. Naturally δ is only a *formal* adjoint to d because we have not considered any boundary conditions, but this is not unusual in mathematical physics.

It is worth noticing that the formula for δ simplifies if we consider the cases of even and odd dimensional spaces separately. We have

$$n \text{ odd:} \qquad \delta\omega = (-1)^r * d * \omega$$

$$n \text{ even:} \qquad \delta\omega = - * d * \omega$$

There are some useful identities connecting $*$, d, and δ which are consequences of $*$ being almost an involution, which we now derive. First we have

$$\begin{aligned} *\delta\omega &= (-1)^{n(r-1)+1} **d*\omega \qquad \omega\in\Lambda^r \\ &= (-1)^{n(r-1)+1}(-1)^{(n-r+1)(r-1)}\, d*\omega \qquad \omega\in\Lambda^r \\ &= (-1)^r\, d*\omega \qquad \omega\in\Lambda^r \end{aligned}$$

since

$$\begin{aligned} n(r-1)+1+(n-r+1)(r-1) &\equiv (n+n-r+1)(r-1)+1 \qquad \bmod 2 \\ &\equiv -(r-1)(r-1)+1 \\ &\equiv (r-1)^2+1 \\ &\equiv (r-1)+1 \\ &\equiv r \end{aligned}$$

Applying our identity to $*\omega$ we have

$$\begin{aligned}
*\,\delta * \omega &= (-1)^{n-r} d * * \omega \\
&= (-1)^{n-r+r(n-r)} d\,\omega \\
&= (-1)^{(n-r)(r+1)} d\,\omega \\
* * \,\delta * \omega &= (-1)^{(n-r)(r+1)} * d\,\omega \\
(-1)^{(n-r-1)(r+1)} \delta * \omega &= (-1)^{(n-r)(r+1)} * d\,\omega \\
\delta * \omega &= (-1)^{(n-r)(r+1)+(n-r-1)(r+1)} * d\,\omega \\
\delta * \omega &= (-1)^{(r+1)(n-r+n-r-1)} * d\,\omega \\
\delta * \omega &= (-1)^{r+1} * d\,\omega
\end{aligned}$$

The r in these formulas always refers to the degree of ω, and this must be carefully remembered when applying the formulas. It is easy to make mistakes applying these formulas. For ease of reference we display the formulas

$$\boxed{d * \omega = (-1)^r * \delta\omega}$$

$$\boxed{\delta * \omega = (-1)^{r+1} * d\,\omega}$$

1.21 The Laplacian

One of the most important operators in mathematical physics is the Laplacian. We have derived formulas for the Laplacian on functions but using the codifferential it is possible to give formulas for the Laplacian on differential forms also. This is important, for example, in Electromagnetic theory where we will need the Laplacian of the one form which expresses the vector potential of the magnetic field.

The Laplacian

$$\triangle f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x^{i2}}$$

has a long history in mathematical physics and almost as long a history in pure mathematics. New vistas for the Laplacian were opened by Beltrami in the 1860's when he defined it on a surface. Eventually it was found to have an unexpected geometric meaning when defined on differential forms. We will have the briefest of looks at this much later in the book. In earlier ages the Laplacian was defined by use of Tensors, but in the 20th Century it was realized it could be defined in terms of the $*$ operator and d, which made for a more elegant exposition, though problems remain in the calculation of formulas. The mathematical definition on forms is now

Def $$\triangle = d\delta + \delta d$$

on any Riemannian or pseudo-Riemannian manifold. Note that we must have an inner product on the surface or manifold in order to define $*$; no inner

product, no Laplacian! The pseudo in pseudo-Riemannian means that the inner product must be non-degenerate but need not be positive definite as is the case in Relativity. We will go into this is some detail in the section of Electromagnetics in four-dimensional space-time.

Note also the extremely important point that by it's definition the Laplacian is independent of the coordinate system. It is even independent of the choice of orientation, since reversing the orientation reverses the sign of each of the two $*$'s found in each term of it's definition and since there are two minus signs each term will come out the same.

Most annoyingly, the mathematical operator $\triangle$ does not quite coincide with the operator $\triangleq$ used in classical mathematical physics and we must be a bit careful about this; the relationship is $\triangleq = -\triangle$. In order to keep it absolutely clear *which* Laplacian we are using I have invented the symbol $\triangleq$ for the familiar **P**hysics Laplacian.

There are a couple of things that can be said in favor of the mathematics notation. First, the eigenvalues λ of $\triangle$ are positive (or 0) while those of $\triangleq$ are negative (or 0). Connected with this is the fact that $((\triangle\omega, \omega)) \geq 0$ while $((\triangleq\omega, \omega)) \leq 0$. The wave equations are

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} + \triangle\omega = 0 \qquad\qquad \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \triangleq\omega = 0$$

and the Schrödinger equations

$$i\hbar\frac{\partial\Psi}{\partial t} = \frac{\hbar^2}{2m}\triangle\Psi + V\Psi \qquad\qquad i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\triangleq\Psi + V\Psi$$

and Poisson's equation for the potential of charge distribution

$$\triangle\,\phi = \frac{\rho}{\epsilon} \qquad\qquad \triangleq\phi = -\frac{\rho}{\epsilon}$$

with the math forms perhaps marginally more elegant. There may be other reasons too for suspecting that Laplace chose the wrong sign. All we can do at the present time is to be careful about which one we are using.

Notice that $d : \Lambda^r \to \Lambda^{r+1}$ and $\delta : \Lambda^r \to \Lambda^{r-1}$ so that the combined $\triangle : \Lambda^r \to \Lambda^r$. Notice also that the definition of the Laplacian defines $\triangle$ on forms of all degrees, not just on functions. This is very important.

Another point of interest is that $\triangle$ is a sort of *square*. Recall that $d^2 = dd = 0$ and $\delta^2 = \delta\delta = 0$. Thus we have

$$\begin{aligned}(d+\delta)^2 &= (d+\delta)(d+\delta) = dd + d\delta + \delta d + \delta\delta \\ &= d\delta + \delta d = \triangle\end{aligned}$$

However, this is not as interesting as it looks because

$$d + \delta : \Lambda^r \to \Lambda^{r+1} \oplus \Lambda^{r-1}$$

and it is hard to see how to get practical use out of that.

Before going on it will be instructive to calculate a couple of Laplacians from scratch. Although we could make use of formulas derived in previous sections, I think it is interesting to see the whole calculation in one place. For functions $f \in \Lambda^0$ we have $\delta f = 0$ (since $*f \in \Lambda^n$ and thus $d * f = 0$) which simplifies the calculation greatly. Thus we have

$$\begin{aligned}
\triangle f &= (d\delta + \delta d)f \;=\; \delta df \\
&= (-1)^{n(1+1)+1} * d * (df) \\
&= - * d * \left(\sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\, dx^i \right) \\
&= - * d \left(\sum_{i=1}^{n} (-1)^{i-1} \frac{\partial f}{\partial x^i}\, dx^1 \wedge \ldots \wedge dx^{i-1} \wedge dx^{i+1} \wedge \ldots \wedge dx^n \right) \\
&= - * \left(\sum_{i,j=1}^{n} (-1)^{i-1} \frac{\partial^2 f}{\partial x^i \partial x^j}\, dx^j \wedge dx^1 \wedge \ldots \wedge dx^{i-1} \wedge dx^{i+1} \wedge \ldots \wedge dx^n \right)
\end{aligned}$$

The only non zero term (due to repetitions) is when $j = i$, hence

$$\begin{aligned}
\triangle f &= - * \left(\sum_{i=1}^{n} (-1)^{i-1} \frac{\partial^2 f}{\partial x^{i2}}\, dx^i \wedge dx^1 \wedge \ldots \wedge dx^{i-1} \wedge dx^{i+1} \wedge \ldots \wedge dx^n \right) \\
&= - * \left(\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x^{i2}}\, dx^1 \wedge \ldots \wedge dx^{i-1} \wedge dx^i \wedge dx^{i+1} \wedge \ldots \wedge dx^n \right) \\
&= - \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x^{i2}}
\end{aligned}$$

which is just what we were expecting. A very similar calculation will get $\triangle f$ for general coordinates, a result which we already have found by other methods, and which you can do in the problems.

For $r = 1$ the computation is a lot more complicated and we will only do it for $n = 3$, which avoids unpleasantness and suffices for our applications. (The big advantage of doing it for $n = 3$ is that we can exploit cyclic order.) We will first compute $\delta d\omega$, then compute $d\delta\omega$, and then put them together to get $\triangle\omega$. Although there are some interesting features, you might want to just skim this calculation.

$$\begin{aligned}
\omega &= \omega_1\, dx + \omega_2\, dy + \omega_3\, dz \\
d\omega &= \left(\frac{\partial \omega_3}{\partial y} - \frac{\partial \omega_2}{\partial z} \right) dy \wedge dz + \left(\frac{\partial \omega_1}{\partial z} - \frac{\partial \omega_3}{\partial x} \right) dz \wedge dx + \left(\frac{\partial \omega_2}{\partial x} - \frac{\partial \omega_1}{\partial y} \right) dx \wedge dy \\
* d\omega &= \left(\frac{\partial \omega_3}{\partial y} - \frac{\partial \omega_2}{\partial z} \right) dx + \left(\frac{\partial \omega_1}{\partial z} - \frac{\partial \omega_3}{\partial x} \right) dy + \left(\frac{\partial \omega_2}{\partial x} - \frac{\partial \omega_1}{\partial y} \right) dz \\
d * d\omega &= \left[\frac{\partial}{\partial y} \left(\frac{\partial \omega_2}{\partial x} - \frac{\partial \omega_1}{\partial y} \right) - \frac{\partial}{\partial z} \left(\frac{\partial \omega_1}{\partial z} - \frac{\partial \omega_3}{\partial x} \right) \right] dy \wedge dz
\end{aligned}$$

$$
\begin{aligned}
&+\left[\frac{\partial}{\partial z}\left(\frac{\partial\omega_3}{\partial y}-\frac{\partial\omega_2}{\partial z}\right)-\frac{\partial}{\partial x}\left(\frac{\partial\omega_2}{\partial x}-\frac{\partial\omega_1}{\partial y}\right)\right]dz\wedge dx\\
&+\left[\frac{\partial}{\partial x}\left(\frac{\partial\omega_1}{\partial z}-\frac{\partial\omega_3}{\partial x}\right)-\frac{\partial}{\partial y}\left(\frac{\partial\omega_3}{\partial y}-\frac{\partial\omega_2}{\partial z}\right)\right]dx\wedge dy\\
\,d\,d\omega &= \left[\frac{\partial^2\omega_2}{\partial y\partial x}-\frac{\partial^2\omega_1}{\partial y^2}-\frac{\partial^2\omega_1}{\partial z^2}+\frac{\partial^2\omega_3}{\partial z\partial x}\right]dx\\
&+\left[\frac{\partial^2\omega_3}{\partial z\partial y}-\frac{\partial^2\omega_2}{\partial z^2}-\frac{\partial^2\omega_2}{\partial x^2}+\frac{\partial^2\omega_1}{\partial x\partial y}\right]dx\\
&+\left[\frac{\partial^2\omega_1}{\partial x\partial y}-\frac{\partial^2\omega_3}{\partial x^2}-\frac{\partial^2\omega_3}{\partial y^2}+\frac{\partial^2\omega_2}{\partial y\partial z}\right]dx
\end{aligned}
$$

OK, $*d*d$ is done. Now we need $d*d*$, and then we can put them together. This one is easier.

$$
\begin{aligned}
\omega &= \omega_1\,dx+\omega_2\,dy+\omega_3\,dz\\
*\,\omega &= \omega_1\,dy\wedge dz+\omega_2\,dz\wedge dx+\omega_3\,dx\wedge dy\\
d*\,\omega &= \left(\frac{\partial\omega_1}{\partial x}+\frac{\partial\omega_2}{\partial y}+\frac{\partial\omega_3}{\partial z}\right)dx\wedge dy\wedge dz\\
\,d\,\omega &= \left(\frac{\partial\omega_1}{\partial x}+\frac{\partial\omega_2}{\partial y}+\frac{\partial\omega_3}{\partial z}\right)\\
d*\,d*\,\omega &= \left(\frac{\partial^2\omega_1}{\partial x^2}+\frac{\partial^2\omega_2}{\partial x\partial y}+\frac{\partial^2\omega_3}{\partial x\partial z}\right)dx+\left(\frac{\partial^2\omega_1}{\partial y\partial x}+\frac{\partial^2\omega_2}{\partial y^2}+\frac{\partial^2\omega_3}{\partial y\partial z}\right)dy\\
&\quad+\left(\frac{\partial^2\omega_1}{\partial z\partial x}+\frac{\partial^2\omega_2}{\partial z\partial y}+\frac{\partial^2\omega_3}{\partial z^2}\right)dz
\end{aligned}
$$

$$
\begin{aligned}
\delta d\omega &= ((-1)^{3(2+1)+1}*d*)d\omega \;=\; *d*d\omega\\
d\delta\omega &= d((-1)^{3(1+1)+1}*d*)\omega \;=\; -d*d*\omega\\
\triangle\omega &= (\delta d+d\delta)\omega\\
&= *d*d\omega-d*d*\omega\\
&= -\left(\frac{\partial^2\omega_1}{\partial x^2}+\frac{\partial^2\omega_1}{\partial y^2}+\frac{\partial^2\omega_1}{\partial z^2}\right)dx-\left(\frac{\partial^2\omega_2}{\partial x^2}+\frac{\partial^2\omega_2}{\partial y^2}+\frac{\partial^2\omega_2}{\partial z^2}\right)dy\\
&\quad-\left(\frac{\partial^2\omega_3}{\partial x^2}+\frac{\partial^2\omega_3}{\partial y^2}+\frac{\partial^2\omega_3}{\partial z^2}\right)dz
\end{aligned}
$$

This is certainly an interesting result. We denote the Laplacian on functions momentarily by $\triangle_0:\Lambda^0\to\Lambda^0$ then the result can be written

$$\triangle\,\omega=(\triangle_0\omega_1)dx+(\triangle_0\omega_2)dy+(\triangle_0\omega_3)dz$$

How general is this result? It certainly cannot work for general coordinates, but it *does* work for rectangular coordinates on $\Lambda^r(\mathbb{R}^n)$. Unfortunately we do not know any better way to prove this than by brute force. For our applications what we have done suffices.

There is a vector form of the above calculation based on a well known formula in three dimensions

$$\text{curl curl}(\mathbf{v}) = \text{grad div}(\mathbf{v}) - \bigtriangledown^2(\mathbf{v})$$

which we rewrite as

$$\triangleq(\mathbf{v}) = \text{grad div}(\mathbf{v}) - \text{curl curl}(\mathbf{v})$$

where in rectangular coordinates

$$\bigtriangledown^2(\mathbf{v}) = (\bigtriangledown^2 v^1)\,\hat{\mathbf{i}} + (\bigtriangledown^2 v^2)\,\hat{\mathbf{j}} + (\bigtriangledown^2 v^3)\,\hat{\mathbf{k}}$$

In general coordinates and three dimensions we can use the above formula as a definition of $\triangleq(\mathbf{v})$ since we have formulas for grad div$(\mathbf{v})$ and curl curl$(\mathbf{v})$. If you examine the differential form calculation above carefully, you will be able to see the $d\delta\omega$ corresponds to grad div$(\mathbf{v})$ and $\delta d\omega$ corresponds to curl curl$(\mathbf{v})$.

Soon we will show that $* \triangle \omega = \triangle * \omega$ in all circumstances. From this we can easily derive that in $\Lambda^2(\mathbb{R}^3)$

$$\triangle(\omega^1\, dydz + \omega^2\, dzdx + \omega^3\, dxdy) = \triangle_0(\omega^1)\, dydz + \triangle_0(\omega^2)\, dzdx + \triangle_0(\omega^3)\, dxdy$$

and in $\Lambda^3(\mathbb{R}^3)$

$$\triangle(f\, dxdydz) = \triangle_0(f)\, dxdydz$$

which tells us all we need to know about $\Lambda^r(\mathbb{R}^3)$.

Now let's develop a bit more theory. First, recall that the global inner product (on all of K) is

$$((\omega, \eta)) = \int_K (\omega, \eta)\Omega_0 = \int_K \omega \wedge *\eta \qquad \omega, \eta \in \Lambda^r(K)$$

provided the integral is finite, which it will be if the support

$$\text{supp}(\omega) = (\text{closure of } \{x \in K \big| \omega(x) \neq 0\})$$

of ω or the support of η is compact. (If K itself is compact this condition is automatically valid. Compact for our purpose means closed and bounded.) We are assuming here that all functions and coefficients of forms have enough continuous derivatives to make the calculations make sense. In general three is enough[9]

Recall from the chapter on the codifferential that

$$((d\omega, \eta)) = ((\omega, \delta\eta)) \qquad \omega \in \Lambda^r,\ \eta \in \Lambda^{r+1}$$

[9]In mathematics these objects are usually taken to have infinitely many derivatives, so as to avoid thinking about such things.

From this we have

$$\begin{aligned}((\triangle\omega,\eta)) &= ((d\delta\,\omega,\eta)+((\delta d\,\omega,\eta)) \\ &= ((\delta\omega,\,\delta\,\eta)+((d\,\omega,d\eta))\end{aligned}$$

This shows that if ω has coefficients with two continuous derivatives that

$$((\triangle\omega,\omega))=0 \text{ if and only if } \delta\omega=0 \text{ and } d\omega=0$$

and this is also clearly equivalent to $\triangle\omega=0$. This leads to the definition

Def ω (with at least one continuous derivative) is *Harmonic* if and only if $\delta\omega=0$ and $d\omega=0$.

and then to the

Theorem If ω has at least two continuous derivatives then ω is Harmonic if and only if $\triangle\omega=0$.

We prefer to use $\delta\omega=0$ and $d\omega=0$ as the definition of Harmonic and $\triangle\omega=0$ as a consequence because of the number of derivatives necessary for these two conditions. The definition as it stands is competing with the old definition $\triangle\omega=0$; bet on the former.

Next, from the previous equation we have

$$\begin{aligned}((\triangle\omega,\,\eta)) &= ((\delta\omega,\,\delta\eta))+((d\omega,\,d\eta)) \\ &= ((\omega,d\delta\eta))+((\omega,\,\delta d\eta)) \\ &= ((\omega,\,\triangle\eta))\end{aligned}$$

where we are still assuming that supp(ω) or supp(η) is compact. The equation

$$((\triangle\omega,\,\eta))=((\omega,\,\triangle\eta))$$

with the condition on the supports means that $\triangle$ is a formally self adjoint operator. It is standard in elementary mathematical physics to leave things here. To get *real* self adjointness requires consideration of the boundary conditions for the operator, and to deal properly with this requires functional analysis and Sobolev spaces, which is a can of worms we wish to *keep* in the can here.

Our final Laplacian duty is the equation $*\triangle\omega=\triangle*\omega$ which we can do with a lovely little calculation. Recall from the chapter on the codifferential that for $\omega\in\Lambda^r(K)$

$$\begin{aligned}d*\omega &= (-1)^r*\delta\omega \\ \delta*\omega &= (-1)^{r+1}*d\omega\end{aligned}$$

Then we have

$$\triangle*\omega = (d\delta+\delta d)*\omega$$

$$\begin{aligned}
&= d(\delta * \omega) + \delta(d * \omega) \\
&= d((-1)^{r+1} * d\omega) + \delta((-1)^r * \delta\omega) \\
&= (-1)^{r+1}(d * d\omega) + (-1)^r(\delta * \delta\omega) \\
&= (-1)^{r+1}(-1)^{r+1}(*\delta d\omega) + (-1)^r(-1)^r(*d\delta\omega) \\
&= *(\delta d\omega + d\delta\omega) \\
&= * \triangle \omega
\end{aligned}$$

Notice that in going from step 4 to step 5 that $d\omega \in \Lambda^{r+1}$ and $\delta\omega \in \Lambda^{r-1}$ which is what makes the signs come out right.

We will show later that this formula is equivalent to Poincaré duality for the de Rham cohomology groups of K, assuming that K is a compact manifold.

1.22 Maxwell's Equations in 3-space

In this section we will show how the standard vector analysis treatment of Maxwell's equations can be recast in the notation of differential forms. Notation has been chosen to make it easy to switch over to a tensor treatment, which we provide in an addendum to the section.

The prerequisites for this section include the $*$ operator, the codifferential which in *three dimensions* is $\delta = *d*$ and the Laplacian $\triangle = \delta d + d\delta$ (math form) and $\triangle = -\triangle$ (physics form).

A standard form of Maxwell's equations using vector formulation is

$$\begin{array}{lcllcllcl}
\text{div } \mathbf{D} & = & \rho & \text{curl } \mathbf{E} & = & -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} & \mathbf{D} & = & \epsilon \mathbf{E} \\
\text{div } \mathbf{B} & = & 0 & \text{curl } \mathbf{H} & = & \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} + \frac{1}{c}\mathbf{j} & \mathbf{B} & = & \mu \mathbf{H}
\end{array}$$

The form of these equations indicates how they should be put into differential forms. The divergence is mirrored by d on 2-forms and the curl by d on 1-forms. This suggests that $\mathbf{D}$ and $\mathbf{B}$ should be 2 forms and $\mathbf{E}$ and $\mathbf{H}$ should be 1-forms. We also note that this suggests the e.g. $\mathbf{D}$ and $\mathbf{E}$ are *not* trivial variants of one another, which was the attitude expressed by Maxwell and Faraday. We set

$$\begin{aligned}
D &= D^1 dydz + D^2 dzdx + D^3 dzdx & E &= E_1 dx + E_2 dy + E_3 dz \\
B &= B^1 dydz + B^2 dzdx + B^3 dzdx & H &= H_1 dx + H_2 dy + H_3 dz
\end{aligned}$$

Since $\mathbf{j}$ is a current density (to be integrated over a surface) it should be a two form

$$j = j^1 dydz + j^2 dzdx + j^3 dzdx$$

and since ρ is a charge density it should be integrated over a region and thus should be a three form, for which we will use the letter P, an upper case Greek ρ.

$$P = \rho\, \Omega_0$$

In rectangular coordinates $P = \rho\, dxdydz$. The matter equations can also be written as forms, for which we require the $*$ operator:

$$\begin{aligned} D &= \epsilon * E \\ B &= \mu * H \end{aligned}$$

and then the differential equations become

$$\begin{aligned} dD = P \qquad & dE = -\frac{1}{c}\frac{\partial B}{\partial t} \\ dB = 0 \qquad & dH = \frac{1}{c}\frac{\partial D}{\partial t} + \frac{1}{c} j \end{aligned}$$

The equations of most significance derived from these are the equation of continuity and the potential equations. The equation of continuity is easy:

$$\begin{aligned} dH &= \frac{1}{c}\frac{\partial D}{\partial t} + \frac{1}{c}\, j \\ 0 = ddH &= \frac{1}{c}\frac{\partial}{\partial t}\, dD + \frac{1}{c}\, dj \\ 0 &= \frac{1}{c}\frac{\partial}{\partial t} P + \frac{1}{c}\, dj \\ 0 &= \frac{\partial P}{\partial t} + dj \end{aligned}$$

This is the form that the equation of continuity takes in our treatment. We note however that if we just consider the coefficient of $dxdydz$ from the equation it becomes

$$0 = \frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j}$$

Now we start on the potential equations.

$$dB = 0 \quad \Longrightarrow \quad B = dA$$

(by the converse of the Poincaré lemma), where A is a 1-form. A is not uniquely determined; we can add dG to A for any $G \in \Lambda^0$ (that is, a function), since

$$d(A + dg) = dA + ddG = dA + 0 = B$$

We can use this G to modify A and this is called *changing the gauge*. We will return to this matter later on. Next we have

$$\begin{aligned} dE &= -\frac{1}{c}\frac{\partial}{\partial t}\, B \\ &= -\frac{1}{c}\frac{\partial}{\partial t}\, dA \;=\; -\frac{1}{c}\, d\left(\frac{\partial A}{\partial t}\right) \\ d\left(E + \frac{1}{c}\,\frac{\partial A}{\partial t}\right) &= 0 \end{aligned}$$

Since $E + \frac{1}{c}\frac{\partial A}{\partial t}$ is a 1-form, there must be a 0-form $-\phi$ for which

$$E + \frac{1}{c}\frac{\partial A}{\partial t} = -d\phi$$

(The minus sign is historical and of no theoretical importance.) ϕ is called the *scalar potential*. Often the letter U is used for it instead of ϕ. A is called the *vector potential*.

If A is changed to $A' = A + dG$ there will be a corresponding change in ϕ which we now determine.

$$\begin{aligned}
E + \frac{1}{c}\frac{\partial A}{\partial t} &= -d\phi \\
E + \frac{1}{c}\frac{\partial A'}{\partial t} &= -d\phi' \\
E + \frac{1}{c}\frac{\partial A}{\partial t} + \frac{1}{c}\frac{\partial}{\partial t}dG &= -d\phi' \\
-d\phi + \frac{1}{c}\frac{\partial}{\partial t}dG &= -d\phi' \\
d(\phi' - \phi + \frac{1}{c}\frac{\partial}{\partial t}G) &= 0 \\
\phi' - \phi + \frac{1}{c}\frac{\partial}{\partial t}G &= C
\end{aligned}$$

where C is some constant. Thus

$$\phi' = \phi - \frac{1}{c}\frac{\partial}{\partial t}G + C$$

Digesting, we have

$$\begin{aligned}
A' &= A + dG \\
\phi' &= \phi - \frac{1}{c}\frac{\partial}{\partial t}G
\end{aligned}$$

where we have set $C = 0$ as customary.

Our next job is to derive the potential equations. However, there is so much slop in A that we cannot reasonably expect nice equations without putting some extra conditions on A. We could simply pull the condition out of the air, but it will be more fun to see it appear in context.

Recall that the physics Laplacian $\triangle$ is the negative of the mathematical Laplacian $\triangle = d\delta + \delta d$ where δ in three dimensions is

$$\delta\omega = (-1)^r * d * \omega \qquad \omega \in \Lambda^r$$

Now we have

$$\begin{aligned}
\triangle A &= (\delta d + d\delta)A = \delta dA + d\delta A \\
&= \delta B + d\delta A
\end{aligned}$$

$$\begin{aligned}
&= (-1)^2 * d * B + d\delta A \\
&= *d(\mu H) + d\delta A \;=\; \mu * dH + d\delta A \\
&= \mu * \left(\frac{1}{c}\frac{\partial D}{\partial t} + \frac{1}{c}\, j\right) + d\delta A \\
&= \mu \left(\frac{1}{c}\frac{\partial (*D)}{\partial t} + \frac{1}{c}\, * j\right) + d\delta A \\
&= \mu \left(\frac{1}{c}\frac{\partial (\epsilon E)}{\partial t} + \frac{1}{c}\, * j\right) + d\delta A \\
&= \frac{\mu}{c}\, * j + \frac{\epsilon\mu}{c}\frac{\partial E}{\partial t} + d\delta A \\
&= \frac{\mu}{c}\, * j + \frac{\epsilon\mu}{c}\frac{\partial}{\partial t}\left(-d\phi - \frac{1}{c}\frac{\partial A}{\partial t}\right) + d\delta A \\
&= \frac{\mu}{c}\, * j - \frac{\epsilon\mu}{c^2}\frac{\partial^2 A}{\partial t^2} + d\left(\delta A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t}\right)
\end{aligned}$$

Rearranging and replacing $\triangle A$ by $-\triangle\!\!\!\triangle A$ we have

$$\frac{\epsilon\mu}{c^2}\frac{\partial^2 A}{\partial t^2} - \triangle\!\!\!\triangle A = \frac{\mu}{c}\, * j + d\left(\delta A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t}\right)$$

This would be the familiar wave equation for A

$$\Box\, A = \frac{\epsilon\mu}{c^2}\frac{\partial^2 A}{\partial t^2} - \triangle\!\!\!\triangle A = \frac{\mu}{c}\, * j$$

with velocity $\frac{c}{\sqrt{\epsilon\mu}}$ if not for the term $d\left(\delta A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t}\right)$. Hence, using the slop in A, we will set this to 0.

$$\delta A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t} = 0 \qquad \text{Condition of Lorenz}$$

Of course, we need to know that adding a suitable dG will force the Condition of Lorenz to be true. We will look at this later. (Usually this step is neglected!)

It is also useful to decode the condition of Lorenz. We have

$$\begin{aligned}
(-1)^1 * d * A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t} &= 0 \\
-* d\left(A_1\, dydz + A_1\, dzdx + A_1\, dxdy\right) - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t} &= 0 \\
-*\left(\frac{\partial A_1}{\partial x} + \frac{\partial A_2}{\partial y} + \frac{\partial A_3}{\partial z}\right) dxdydz - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t} &= 0 \\
-\left(\frac{\partial A_1}{\partial x} + \frac{\partial A_2}{\partial y} + \frac{\partial A_3}{\partial z} + \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t}\right) &= 0
\end{aligned}$$

We must also have the potential equation for ϕ. This is derived in a similar manner, but easier. It is interesting that it throws up the same extra terms which

we eliminate with the Condition of Lorenz. This suggests that the time and space variables might have more of a connection than we expect á priori. This was likely one of the things that induced Lorenz develop the Lorenz-Einstein transformation equations. We imitate the previous calculation, noting that $\delta\phi = 0$ since δ is 0 on Λ^0,

$$\begin{aligned}
\triangle\phi &= (\delta d + d\delta)\phi \;=\; \delta d\phi + d\delta\phi \\
&= \delta\left(-E - \frac{1}{c}\frac{\partial A}{\partial t}\right) + 0 \\
&= -\delta E - \frac{1}{c}\frac{\partial(\delta A)}{\partial t} \\
&= -(-1)^1 * d * E - \frac{1}{c}\frac{\partial}{\partial t}\left(\frac{\epsilon\mu}{c}\frac{\partial\phi}{\partial t}\right) - \frac{1}{c}\frac{\partial}{\partial t}\left(\delta A - \frac{\epsilon\mu}{c}\frac{\partial\phi}{\partial t}\right) \\
&= *d\frac{1}{\epsilon}D - \frac{\epsilon\mu}{c^2}\frac{\partial^2\phi}{\partial t^2}
\end{aligned}$$

where we have invoked the Condition of Lorenz $\delta A - \frac{\epsilon\mu}{c}\frac{\partial\phi}{\partial t} = 0$ once again. Continuing

$$\begin{aligned}
\triangle\phi &= \frac{1}{\epsilon} * P - \frac{\epsilon\mu}{c^2}\frac{\partial^2\phi}{\partial t^2} \\
\frac{\epsilon\mu}{c^2}\frac{\partial^2\phi}{\partial t^2} - \text{⧋}\phi &= \frac{1}{\epsilon} * P \;=\; \frac{1}{\epsilon} * (\rho\Omega_0) \\
\frac{\epsilon\mu}{c^2}\frac{\partial^2\phi}{\partial t^2} - \text{⧋}\phi &= \frac{1}{\epsilon}\rho
\end{aligned}$$

which is the familiar wave equation for ϕ in coordinate independent form (except for orientation considerations and provided that ρ really *is* the physical charge density and has not been modified to fit into some special coordinate system). Note that we used $\text{⧋}\phi = -\triangle\phi$ again. Note also that the equation is identical in form to the Dalembertian equation for A. And finally note that the Condition of Lorenz did not pop up so naturally in this derivation as it did in the calculation for A.

Our next job is a look at the Condition of Lorenz. A close examination of the derivation of the wave equation for A will show that if A is a solution of the equation then the Condition of Lorenz must hold. The question is, can we always *force* the Condition of Lorenz to hold by choosing an appropriate G in

$$A' = A + dG \qquad \phi' = \phi - \frac{1}{c}\frac{\partial G}{\partial t}$$

The answer, as we will show, is yes. We need

$$\delta A' - \frac{\epsilon\mu}{c}\frac{\partial\phi'}{\partial t} = 0$$

How do we find the G?. Substituting into this equation we have

$$\delta(A + dG) - \frac{\epsilon\mu}{c}\frac{\partial}{\partial t}\left(\phi - \frac{1}{c}\frac{\partial G}{\partial t}\right) = 0$$

Since $\delta G = 0$, this can be rewritten as

$$\begin{aligned}(\delta d + d\delta)G + \frac{\epsilon\mu}{c^2}\frac{\partial^2 G}{\partial t^2} + \delta A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t} &= 0 \\ \frac{\epsilon\mu}{c^2}\frac{\partial^2 G}{\partial t^2} + \triangle G &= -\left(\delta A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t}\right) \\ \frac{\epsilon\mu}{c^2}\frac{\partial^2 G}{\partial t^2} - \text{⧋} G &= -\left(\delta A - \frac{\epsilon\mu}{c}\frac{\partial \phi}{\partial t}\right)\end{aligned}$$

(recalling that ⧋ $= -\triangle$) and once again we have a wave equation. We solve for G, correct the A and ϕ to A' and ϕ', and then solve the wave equations for A' and ϕ'. Actually we don't have to make the correction, since solving the wave equations for A and ϕ will work just fine; the A and ϕ we find will automatically satisfy the Condition of Lorenz. The importance of the above is to show that A and ϕ we seek actually exist.

1.23 Flux

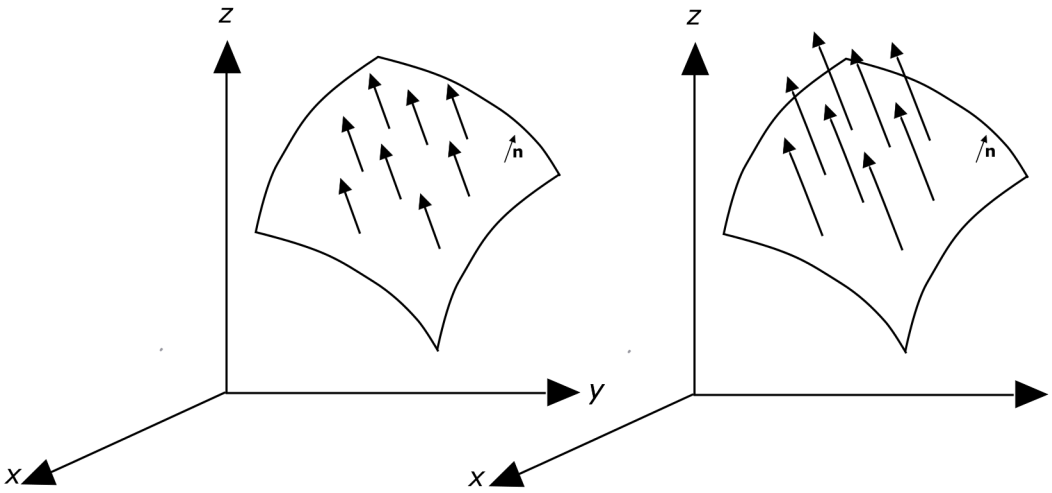
The concept of flux has changed over time. We are going to use the definition common in Electromagnetics. It is wise to check when encountering flux in other areas that the definition is the same.

Def The flux $\Phi_{\mathbf{E}}$ of the electric field $\mathbf{E}$ though a surface S is defined as

$$\Phi_{\mathbf{E}} = \int_S \mathbf{E} \cdot \mathbf{dS} = \int_S \mathbf{E} \cdot \mathbf{n}\, dS$$

Of course the formula is good for any other vector field; not just $\mathbf{E}$.

The following pictures show the vector field $\mathbf{E}$ poking through the surface S from which we compute the flux. The unit normal vector $\mathbf{n}$ is also shown.



Some Flux Twice as much Flux

Some concepts can be handily expressed in terms of flux. For example, the first of Maxwell's equations for Electromagnetism is

$$\text{div } \mathbf{E} = \frac{\rho}{\epsilon}$$

where ρ is the charge density. This can be expressed in terms of flux by

$$\Phi_{\mathbf{E}} = \frac{1}{\epsilon}\,(\text{included charge})$$

where S is any closed surface. By Gauss's theorem we have

$$\int_S \mathbf{E}\cdot\mathbf{dS} = \int_V \text{div } \mathbf{E}\, dv$$

Where V is the volume for which S is the boundary. Thus

$$\begin{aligned}
\Phi_{\mathbf{E}} &= \int_S \mathbf{E}\cdot\mathbf{dS} \\
&= \int_V \text{div } \mathbf{E}\, dv \\
&= \int_V \frac{\rho}{\epsilon}\, dv = \frac{1}{\epsilon}\int_V \rho\, dv \\
&= \frac{1}{\epsilon}\,(\text{included charge})
\end{aligned}$$

The equation

$$\Phi_{\mathbf{E}} = \frac{1}{\epsilon}\,(\text{included charge})$$

is called the integral form of the first Maxwell equation.

In this form, we can find, for example, the electric field of an electron, which has tricky aspects if you use the differential form of the first Maxwell equation because the charge density for electrons seems to be infinite at their position and zero elsewhere. The problems will take you through the details.

Flux is also very useful with Magnetic fields, and there is a great deal of engineering which uses flux of magnetic fields to describe the generation of electricity.

We thought it might be useful if we expressed Flux in terms of our basic operations for differential forms. However, this requires the construction of a *correlation* which is another form of the inner product. We require the representation theorem for linear functionals which is proved in the problems for this section.

Theorem Let V be a finite dimensional vector space with inner product. Let $\ell : V \to \mathbb{R}$ be a linear functional ($\ell \in V^*$). Then there exists a $v \in V$ so that for all $w \in V$

$$\ell(w) = (v, w)$$

It is very important to realize that the theorem is true in *all* finite dimensional spaces regardless of whether the inner product is positive definite or not, but for infinite dimensional inner product spaces the inner product must be positive definite ($(v, v) > 0$ for $v \neq 0$) and V must be complete (see section on topology for complete). We do not need infinite dimensional vector spaces; we included this because everyone should know it.

1.24 Indefinite Inner Product Spaces

This section and the next contain material of far less general usefulness than the preceding sections. They are directed specifically to Relativity, and if you are not interested in this subject then you have little need of these sections.

This first of the two contains the modications necessary when the inner product is not postive definite, so that there are vectors for which $(\mathbf{v}, \mathbf{v}) < 0$. This causes some modifications in the $*$ operator and some changes in sign in the Laplacian. In this section we will handle the situation in general and in the next the specific material for Relativity. We will concentrate again on Maxwell's equations as an example. Naturally developing the theory again with the additional complications makes for a high degree of repetition, and we will be somewhat more terse this time since you have alrady seen the material once.

Let V be an n-dimensional space on which there is an inner product which is non-degnenerate. Thus the inner product satisfies the following

a. $(\alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w}) = \alpha(\mathbf{u}, \mathbf{w}) + \beta(\mathbf{v}, \mathbf{w})$ Linearity

b. $(\mathbf{v}, \mathbf{w}) = (\mathbf{w}, \mathbf{v})$ Symmetry

c. If $(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w}$ then $v = 0$ Non-degenarcy

Notice the c. replaces the usual condition $(\mathbf{v},\mathbf{v}) > 0$ for $\mathbf{v} \neq 0$. In indeffinite inner product spaces there will certainly be vectors for which $(\mathbf{v},\mathbf{v}) = 0$. In Relativity these are called *null vectors* and in math isotropic vectors.

Using the usual methods of linear algebra and being careful to avoid vectors $\mathbf{v}$ for which $(\mathbf{v},\mathbf{v}) = 0$ one can construct without difficulty an orthonormal basis $\mathbf{e}_1,\ldots,\mathbf{e}_n$ for V. Renumbering if necessary they can be arranged in a sequence $\mathbf{e}_1,\ldots,\mathbf{e}_{n-s},\mathbf{e}_{n-s+1},\ldots,\mathbf{e}_n$ for which

$$\begin{aligned}(\mathbf{e}_i,\mathbf{e}_i) &= 1 \quad \text{for } i = 1,\ldots,n-s\\ (\mathbf{e}_i,\mathbf{e}_i) &= -1 \quad \text{for } i = n-s+1,\ldots,n\end{aligned}$$

Here, *orthonormal* means $(\mathbf{e}_i,\mathbf{e}_i) = \pm 1$. Those elements of the basis with $(\mathbf{e}_i,\mathbf{e}_i) = +1$ will be called *positive basis elements* and those elements with $(\mathbf{e}_i,\mathbf{e}_i) = -1$ will be called *negative basis elements*. J. J. Sylverster proved early in the history of linear algebra (1852) that the number s of negative basis elements does not depend on the choice of basis; all orthonormal bases will have the same s. This is called Sylvester's law of inertia. This s will metastisize through all the formulas.

It is worth mentioning that since we have assumed that the inner product is non-degenerate there will indeed by n basis vectors in the orthonormal basis.

For the orthonormal basis defined above we get the usual matrix of metric coefficients $g_{ij} = (\mathbf{e}_i,\mathbf{e}_j)$ and it and its inverse will be

$$E = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0 & 0 & -1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & -1 \end{pmatrix}$$

with $n-s$ ones and s negative ones.

As far as the exterior or Grassmann product (wedge product) of the vectors goes, there will be no changes, since this apparatus works independently of the inner product. We first begin to get effects from the indefiniteness when we introduce the $*$ operator.

We will begin work in $\mathbb{R}^m$ or some n-dimensional submanifold K of it with coordinates $u^1,\ldots,u^n$ and an indefinite inner product on the cotangent space $\Lambda^1(K)$. At each point the cotangent space is spanned by $du^1,,\ldots,du^n$. The topform will be $du^1 \wedge \ldots \wedge du^n$, but the inner product of this with itself may well be negative and thus unsuitable for use in finding volumes. If this should happen, (which it often does in applications) we put in a negative sign to fix it. To see the sign, we reason as follows. There is a process in linear algebra

by which we can manipulate $du^1,,\ldots,du^n$ to give us an orthonormal basis $E^1,\ldots,E^n$ which will have the matrix E above for it's metric. By reverseing the sign of E^n f necessary we can assure that we can assure that $du^1,,\ldots,du^n$ and $E^1,\ldots,E^n$ have the same orientation. Suppose that

$$E^i = \sum \alpha^i_j du^j$$

We know that $\det(\alpha^i_j) > 0$ because the orientations match. Then we have

$$E = (\alpha^i_j)^\intercal\big((du^i, du^j)\big)(\alpha^i_j)$$

and thus

$$\begin{aligned} \det E \;=\; \det\left[(\alpha^i_j)^\intercal\big((du^i, du^j)\big)(\alpha^i_j)\right] &= \det(\alpha^i_j)^\intercal \det\big((du^i, du^j)\big)\det(\alpha^i_j) \\ (-1)^s &= \left[\det(\alpha^i_j)\right]^2 \det(g^{ij}) \end{aligned}$$

Since $[\det(\alpha)]^2$ is positive $(-1)^s$ and $\det(g^{ij})$ have the same sign and we have $(-1)^s\det(g^{ij}) > 0$. Thus to normalize the topform $du^1\wedge\ldots\wedge du^n$ it is natural to take

$$\Omega_0 = \frac{1}{\sqrt{(-1)^s\det(g^{ij})}}\, du^1\wedge\ldots\wedge du^n$$

If we set, as is commonly done, $g = \det(g_{ij}$ then we have $\det(g^{ij} = g^{-1}$ and the formula becomes

$$\Omega_0 = \sqrt{(-1)^s g}\quad du^1\wedge\ldots\wedge du^n$$

There is another useful point. Both Ω_0 and $E^1\wedge\ldots\wedge E^n$ are normalized topforms in the one-dimensional space Λ^n. At worst they could be negatives of one another, but this cannot happen since they have the same orientation. Hence we have

$$E^1\wedge\ldots\wedge E^n = \Omega_0$$

Our next job is to revise the formula for $**\omega$. We take $E^1,\ldots,E^{n-s}$ be the normalized positive basis elements, $((E^i,E^i) = +1)$, and $E^{n-s+1},\ldots,E^n$ to be the normalized negative basis elements, $((E^i,E^i) = -1)$. We now take a typical basis element $E^{i_1}\wedge\ldots\wedge E^{i_r}$ with $i_1 < i_2 < \cdots < i_r$. To this basis element corresponds a permutation

$$\sigma = \left(\begin{array}{cccc|cccc} 1 & 2 & \ldots & r & r+1 & r+2 & \ldots & n \\ i_1 & i_2 & \ldots & i_r & k_1 & k_2 & \ldots & k_{n-r} \end{array}\right)$$

with $k_1 < k_2 < \cdots < k_{n-r}$. We suspect from our previous work that

$$* E^{i_1}\wedge\ldots\wedge E^{i_r} = a\, E^{k_1}\wedge\ldots\wedge E^{k_{n-r}}$$

and we must determine a.

Let s_1 of $E^{i_1}, \ldots, E^{i_r}$ have $(E^i, E^i) = -1$
Let s_2 of $E^{k_1}, \ldots, E^{k_{n-r}}$ have $(E^i, E^i) = -1$

where of course $s_1 + s_2 = s$. Using the definition of $*$ and our guess above about the form of $* E^{i_1} \wedge \ldots \wedge E^{i_r}$ we have

$$\begin{aligned} E^{i_1} \wedge \ldots \wedge E^{i_r} \wedge *(E^{i_1} \wedge \ldots \wedge E^{i_r}) &= \det\big((E^{i_j}, E^{i_\ell})\big)\Omega_0 \\ E^{i_1} \wedge \ldots \wedge E^{i_r} \wedge a\, E^{k_1} \wedge \ldots \wedge E^{i_{n-r}} &= (-1)^{s_1}\Omega_0 \\ a\ \mathrm{sgn}(\sigma) E^1 \wedge \ldots \wedge E^n &= (-1)^{s_1}\Omega_0 \\ a\ \mathrm{sgn}(\sigma)\Omega_0 &= (-1)^{s_1}\Omega_0 \\ a\ \mathrm{sgn}(\sigma) &= (-1)^{s_1} \\ a &= (-1)^{s_1}\ \mathrm{sgn}(\sigma) \end{aligned}$$

and thus

$$\boxed{* E^{i_1} \wedge \ldots \wedge E^{i_r} = (-1)^{s_1}\ \mathrm{sgn}(\sigma)\, E^{k_1} \wedge \ldots \wedge E^{k_{n-r}}}$$

Now we are in a position to find $**\omega$. Let

$$\tilde{\sigma} = \left(\begin{array}{cccc|cccc} 1 & 2 & \ldots & n-r & n-r+1 & n-r+2 & \ldots & n \\ k_1 & k_2 & \ldots & k_{n-r} & i_1 & i_2 & \ldots & i_r \end{array} \right)$$

Then, in a similar fashion to above,

$$* E^{k_1} \wedge \ldots \wedge E^{k_{n-r}} = (-1)^{s_2}\ \mathrm{sgn}(\tilde{\sigma})\, E^{i_1} \wedge \ldots \wedge E^{i_r}$$

Hence

$$\begin{aligned} ** E^{i_1} \wedge \ldots \wedge E^{i_r} &= (-1)^{s_1}\ \mathrm{sgn}\ \mathrm{sgn}(\sigma) * (E^{k_1} \wedge \ldots \wedge E^{k_{n-r}}) \\ &= (-1)^{s_1}\ \mathrm{sgn}(\sigma)\,(-1)^{s_2}\ \mathrm{sgn}(\tilde{\sigma}) E^{i_1} \wedge \ldots \wedge E^{i_r} \\ &= (-1)^{s}(-1)^{r(n-r)} E^{i_1} \wedge \ldots \wedge E^{i_r} \end{aligned}$$

since $s_1 + s_2 = s$ and $\mathrm{sgn}(\sigma)\ \mathrm{sgn}(\tilde{\sigma}) = (-1)^{r(n-r)}$. Since any element $\omega \in \Lambda^r$ can be expressed in terms of these basis elements $E^{i_1} \wedge \ldots \wedge E^{i_r}$ we have

$$** \omega = (-1)^{r(n-r)+s}\, \omega \qquad \text{for } \omega \in \Lambda^r$$

If we now write this using differentials as the basis we have

$$* du^{i_1} \wedge \ldots \wedge du^{i_r} = (-1)^{s_1}\mathrm{sgn}(\sigma)\, du^{k_1} \wedge \ldots \wedge du^{k_{n-r}}$$

where the notation is that used above.

Our next job is the codifferential δ. Because the inner product is no longer positive definite there are some odd effects and operator which corresponds to the Laplacian is now the D'Albertian. We will return to this later. For the moment we will be interested in the formula for the codifferential, which we

derive in a similar manner to the positive definite case. Let $\omega \in \Lambda^{r-1}$ and $\eta \in \Lambda^r$). Recall that

$$d(\omega \wedge *\eta) = d\omega \wedge *\eta + (-1)^{r-1}\omega \wedge d * \eta$$

Then

$$\begin{aligned}((d\omega, \eta)) &= \int_K d\omega \wedge *\eta \\ &= \int_K d(\omega \wedge *\eta) - (-1)^{r-1}\int_K \omega \wedge d * \eta \\ &= \int_{\partial K} \omega \wedge *\eta - (-1)^{r-1}\int_K \omega \wedge d * \eta \\ &= 0 + (-1)^r \int_K \omega \wedge d * \eta\end{aligned}$$

by Stoke's theorem, because we are assuming boundary conditions that kill off the boundary intergral, as is usual in this kind of calculation. Continuing we have

$$\begin{aligned}((d\omega, \eta)) &= (-1)^r(-1)^{(n-r+1)(r-1)+s}\int_K \omega \wedge * * d * \eta \\ &= (-1)^{n(r-1)+s+1}\int_K \omega \wedge * * d * \eta \\ &= \int_K \omega \wedge * \left[(-1)^{n(r-1)+s+1} * d * \eta\right] \\ &= ((\omega, (-1)^{n(r-1)+s+1} * d * \eta))\end{aligned}$$

So if we define

$$\boxed{\delta\eta = (-1)^{n(r-1)+s+1} * d * \eta}$$

we have

$$((d\omega, \eta)) = ((\omega, \delta\eta))$$

Finally we want to introduce the analog of the Laplacian. Since the situation is so different from that of positive definite inner products, it seems reasonable to use a different notation, especially since it (more or less) coincides with that usual in physics. We will use the notation $\Box\omega$ for the new operator, which is defined exactly as was the Laplacian in the positive defintate case, namely

Def $$\Box\omega = (d\delta + \delta d)\omega$$

The notation is meant to suggest a relationship with the classical Dalembertian

$$\boxed{\text{p}}\, f = \frac{1}{c}\frac{\partial^2 f}{\partial t^2} - \frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} - \frac{\partial^2 f}{\partial z^2}$$

The relations

$$\begin{aligned} d*\omega &= (-1)^r * \delta\omega \\ \delta*\omega &= (-1)^{r+1} * d\omega \end{aligned}$$

for $\omega \in \Lambda^r(K)$, remain true in the case of indefinite inner products. The reason for this is slightly interesting so we will present a proof of the first one of these relations. We have, noting that $d*\omega \in \Lambda^{n-r+1}$,

$$\begin{aligned} *\delta\omega &= (-1)^{n(r-1)+1+s} * * d * \omega \\ &= (-1)^{n(r-1)+1+s}(-1)^{(n-r+1)(r-1)+s)} d*\omega \\ &= (-1)^{(n+n-r+1)(r-1)+2s+1} d*\omega \\ &= (-1)^{-(r-1)^2+1} d*\omega \;=\; (-1)^{-(r-1)+1} d*\omega \\ d*\omega &= (-1)^r * \delta\omega \end{aligned}$$

What makes this work is that s occurs once in the exponent of (-1) for δ and once in the formula for $**$ and so it disappears from the final formula. The second formula follows from the first as in the section on the codifferential for the same reason.

The formula

$$\Box(*\,\omega) = *\,(\Box\,\omega)$$

follows from the above two formulas in exactly the same way it did in the chapter on the Laplacian.

1.25 Maxwell's equations in Space-Time

In this section we will use the material in the preceding section to work out Maxwell's equations in Space-Time. We will do this twice. First we will work out the the theory in free space where we assume $\epsilon = \mu = 1$ and then we will work it out in general without the constraints on ϵ and μ.

For Special Relativity, the coordinate system cdt, dx, dy, dz forms an othonormal coordinate system, with that order. Thinking of it as dx^0, dx^1, dx^2, dx^3 for convenience, we have

$$(cdt, cdt) = +1, \quad (dx, dx) = -1, \quad (dy, dy) = -1, \quad (dz, dz) = -1,$$

and

$$(dx^i, dx^j) = 0 \qquad \text{for } i \neq j$$

The matrix for this coordinate system is

$$(g^{ij}) = \big((dx^i, dx^j)\big) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

We can perform the $*$ operation using the following permutations, all of which are even permutations (have $\text{sgn}(\sigma) = +1$).

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 2 & 3 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 3 & 1 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 2 & 3 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 1 & 0 & 2 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 2 & 0 & 3 \end{pmatrix}$$

Using these and the formula

$$* dx^{i_1} \wedge \ldots \wedge dx^{i_r} = (-1)^{s_1} \text{sgn}(\sigma)\, dx^{j_1} \wedge \ldots \wedge dx^{j_r}$$

where

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & r & r+1 & r+2 & \cdots & n \\ i_1 & i_2 & \cdots & i_r & j_1 & j_2 & \cdots & j_{n-r} \end{pmatrix}$$

and s_1 is the number of negative basis elements (that is dx, dy, dz) among $dx^{i_1}, \ldots, dx^{i_r}$, we have the following formulas

$$*1 = cdt \wedge dx \wedge dy \wedge dz$$

$$\begin{aligned} * cdt &= dx \wedge dy \wedge dz & * dx \wedge dy \wedge dz &= cdt \\ * dx &= cdt \wedge dy \wedge dz & * cdt \wedge dy \wedge dz &= dx \\ * dy &= cdt \wedge dz \wedge dx & * cdt \wedge dz \wedge dx &= dy \\ * dz &= cdt \wedge dx \wedge dy & * cdt \wedge dx \wedge dy &= dz \\ * cdt \wedge dx &= -dy \wedge dz & * dy \wedge dz &= cdt \wedge dx \\ * cdt \wedge dy &= -dz \wedge dx & * dz \wedge dx &= cdt \wedge dy \\ * cdt \wedge dz &= -dx \wedge dy & * dy \wedge dz &= cdt \wedge dx \end{aligned}$$

$$* cdt \wedge dx \wedge dy \wedge dz = -1$$

For example, for the 2nd entry we have the permutation

$$\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{pmatrix}$$

with $s_1 = 0$ and $\text{sgn}(\sigma) = +1$, whereas for the fourth entry we have

$$\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 2 & 0 & 3 & 1 \end{pmatrix}$$

with $s_1 = 1$ and $\text{sgn}(\sigma) = -1$, this being the the fourth permutation in the above list with the second and third entries swapped. For the seventh entry we have

$$\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 2 & 3 & 1 \end{pmatrix}$$

so $s_1 = 1$ and $\text{sgn}(\sigma) = 1$, this being the second permutation in the list. The entries in the second column can be derived from those in the first, but beware since the permutation is reversed. The fourth row second column has

$$\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 3 & 1 & 2 \end{pmatrix}$$

and here $s_1 = 2$ and σ is the third entry in the list of permatations so $\text{sgn}(\sigma) = 1$. It is far easier to derive the second column using $**\omega = (-1)^{r(4-r)+3}\omega$.

Now we will start with our four dimensional treatment of Electromagnetics. We recall Maxwell's Equations

$$\begin{array}{lll} \text{div}\,\vec{D} = \rho & \text{curl}\,\vec{E} = -\dfrac{1}{c}\dfrac{\partial \vec{B}}{\partial t} & \vec{D} = \epsilon\vec{E} \\ \text{div}\,\vec{B} = 0 & \text{curl}\,\vec{H} = \dfrac{1}{c}\dfrac{\partial \vec{D}}{\partial t} + \dfrac{1}{c}\vec{j} & \vec{B} = \mu\vec{H} \end{array}$$

For the initial development we will set $\epsilon = \mu = 1$ so that $\vec{D} = \vec{E}$ and $\vec{B} = \vec{H}$ though we will retain the letters for comparison with future work. Also, $\vec{E}$ and $\vec{H}$ will have high indices and $\vec{D}$ and $\vec{B}$ will have low indices. This is not important for us; I'm just maintaining conventions.

The first thing we want to do is to express $\text{div}\,\vec{B} = 0$ and $\text{curl}\,\vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}$ in a single equation involving differential forms. This can be done in several ways, but we want to stay as consistent with classical tensor analysis as possible. So we will use the coefficients from

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & E^1 & E^2 & E^3 \\ -E^1 & 0 & -B_3 & B_2 \\ -E^2 & B_3 & 0 & -B_1 \\ -E^3 & -B_2 & B_1 & 0 \end{pmatrix}$$

which is a representation of a standard Electromagnetic Tensor. We can then form the differential form (summation convention in force)

$$\begin{aligned} F &= \frac{1}{2}F_{\mu\nu}\,dx^\mu dx^\nu \\ &= E^1dx^0dx^1 + E^2dx^0dx^2 + E^3dx^0dx^3 \\ &\quad -B_1dx^2dx^3 - B_2dx^3dx^1 - B_3dx^1dx^2 \end{aligned}$$

Now we compute dF

$$\begin{aligned} dF &= \Big(-\frac{\partial E^3}{\partial x^2} + \frac{\partial E^2}{\partial x^3} - \frac{\partial B_1}{\partial x^0}\Big)dx^0dx^2dx^3 + \Big(-\frac{\partial E^1}{\partial x^3} + \frac{\partial E^3}{\partial x^1} - \frac{\partial B_2}{\partial x^0}\Big)dx^0dx^3dx^1 \\ &\quad + \Big(-\frac{\partial E^2}{\partial x^1} + \frac{\partial E^1}{\partial x^2} - \frac{\partial B_3}{\partial x^0}\Big)dx^0dx^1dx^2 - \Big(\frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3}\Big)dx^1dx^2dx^3 \end{aligned}$$

$$\begin{aligned} &= -\Big((\operatorname{curl}\vec{E})_1 + \frac{1}{c}\frac{\partial B_1}{\partial t}\Big)dx^0dx^2dx^3 - \Big((\operatorname{curl}\vec{E})_2 + \frac{1}{c}\frac{\partial B_2}{\partial t}\Big)dx^0dx^3dx^1 \\ &\quad -\Big((\operatorname{curl}\vec{E})_3 + \frac{1}{c}\frac{\partial B_3}{\partial t}\Big)dx^0dx^1dx^2 - \Big(\operatorname{div}\vec{B}\Big)dx^1dx^2dx^3 \\ &= 0 \end{aligned}$$

in view of the two Maxwell Equations.

Now we know from the converse of the Poincaré lemma that since $dF = 0$ there must be a 1-form A for which $dA = F$. (This is the four dimensional "vector potential".) For historical reasons we write this A as

$$A = \phi\, dx^0 - A^1dx^1 - A^2dx^2 - A^3dx^3$$

For comparison with other treatments we also define the three vector $\vec{A} = (A^1, A^2, A^3)$. We now take the exterior derivative of A and compare it to F.

$$\begin{aligned} dA &= \Big(-\frac{\partial A^1}{\partial x^0} - \frac{\partial \phi}{\partial x^1}\Big)dx^0dx^1 + \Big(-\frac{\partial A^2}{\partial x^0} - \frac{\partial \phi}{\partial x^2}\Big)dx^0dx^2 \\ &\quad +\Big(-\frac{\partial A^3}{\partial x^0} - \frac{\partial \phi}{\partial x^3}\Big)dx^0dx^3 + \Big(-\frac{\partial A^3}{\partial x^2} + \frac{\partial A^2}{\partial x^3}\Big)dx^2dx^3 \\ &\quad +\Big(-\frac{\partial A^1}{\partial x^3} + \frac{\partial A^3}{\partial x^1}\Big)dx^3dx^1 + \Big(-\frac{\partial A^2}{\partial x^1} + \frac{\partial A^1}{\partial x^2}\Big)dx^1dx^2 \\ &= \Big(-\Big(\frac{\partial \vec{A}}{\partial x^0}\Big)_1 - (\operatorname{grad}\phi)_1\Big)dx^0dx^1 + \Big(-\Big(\frac{\partial \vec{A}}{\partial x^0}\Big)_2 - (\operatorname{grad}\phi)_2\Big)dx^0dx^2 \\ &\quad +\Big(-\Big(\frac{\partial \vec{A}}{\partial x^0}\Big)_3 - (\operatorname{grad}\phi)_3\Big)dx^0dx^3 \\ &\quad -\Big((\operatorname{curl}\vec{A})_1\Big)dx^2dx^3 - \Big((\operatorname{curl}\vec{A})_2\Big)dx^3dx^1 - \Big((\operatorname{curl}\vec{A})_3\Big)dx^1dx^2 \end{aligned}$$

Comparing this with

$$\begin{aligned} F &= E^1dx^0dx^1 + E^2dx^0dx^2 + E^3dx^0dx^3 \\ &\quad -B_1dx^2dx^3 - B_2dx^3dx^1 - B_3dx^1dx^2 \end{aligned}$$

we see

$$\begin{aligned} E^i &= -(\operatorname{grad}\phi)_i - \Big(\frac{\partial \vec{A}}{\partial x^0}\Big)_i \\ B_i &= (\operatorname{curl}\vec{A})_i \end{aligned}$$

or more succinctly

$$\begin{aligned} \vec{E} &= -\operatorname{grad}\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} \\ \vec{B} &= \operatorname{curl}\vec{A} \end{aligned}$$

These last equations are not properly part of a four-dimensional theory; they have been included to show that the terms involving A^i actually come from the three dimensional vector potential and the ϕ term in A is the old scalar potential. This is comforting. And to get it to work out right requires considerable fussiness in the definitions, though we made it look easy.

We have now taken care of two of Maxwell's equations and defined the potential A, so it is time to take on the other two equations and connect them to the source terms ρ and $\vec{j} = (j^1, j^2, j^3)$. Then we have to come up with the potential equations but it turns out this is quite easy. First we define the source form

$$J = \rho\, dx^1dx^2dx^3 - \frac{1}{c}j^1\, dx^2dx^3 - \frac{1}{c}j^2\, dx^3dx^1 - \frac{1}{c}j^3\, dx^1dx^2$$

It is not completely obvious what the source form should be; We have chosen the form above but other choices are possible, i.e. multiplying this form by c. We were motivated to choose this form because in a static situation ($dt = 0$) this form, when integrated over a 3-manifold, gives the enclosed charge, which makes it seem natural for a source. The usual form in a tensor treatment (see appendix) corresponds to our choice multiplied by c.

Note that dJ is related to the equation of continuity.

$$\begin{aligned} dJ &= \Big(\frac{\partial\rho}{\partial x^0} + \frac{1}{c}\frac{\partial j^1}{\partial x^1} + \frac{1}{c}\frac{\partial j^2}{\partial x^2} + \frac{1}{c}\frac{\partial j^3}{\partial x^3}\Big)dx^0dx^1dx^2dx^3 \\ &= \frac{1}{c}\Big(\frac{\partial\rho}{\partial t} + \frac{\partial j^1}{\partial x^1} + \frac{\partial j^2}{\partial x^2} + \frac{\partial j^3}{\partial x^3}\Big)dx^0dx^1dx^2dx^3 \\ &= \frac{1}{c}\Big(\frac{\partial\rho}{\partial t} + \operatorname{div}\vec{j}\Big)dx^0dx^1dx^2dx^3 \end{aligned}$$

Next we compute the dual form $*F$. We get

$$\begin{aligned} *F &= *\Big(E^1dx^0dx^1 + E^2dx^0dx^2 + E^3dx^0dx^3 \\ &\quad -B_1dx^2dx^3 - B_2dx^3dx^1 - B_3dx^1dx^2\Big) \\ &= -E^1dx^2dx^3 - E^2dx^3dx^1 - E^3dx^1dx^2 \\ &\quad -B_1dx^0dx^1 - B_2dx^0dx^2 - B_3dx^0dx^3 \end{aligned}$$

However, for comparison with later work it turns out to be useful to use the equations $\vec{D} = \vec{E}$ and $\vec{H} = \vec{B}$ to rewrite $*F$ slightly as

$$\begin{aligned} *F &= -D_1dx^2dx^3 - D_2dx^3dx^1 - D_3dx^1dx^2 \\ &\quad -H^1dx^0dx^1 - H^2dx^0dx^2 - H^3dx^0dx^3 \end{aligned}$$

Next we need the exterior derivative of $*F$:

$$\begin{aligned} d*F = &\Big(-\frac{\partial D_1}{\partial x^0} + \frac{\partial H^3}{\partial x^2} - \frac{\partial H^2}{\partial x^3}\Big)dx^0dx^2dx^3 + \Big(-\frac{\partial D_2}{\partial x^0} + \frac{\partial H^1}{\partial x^3} - \frac{\partial H^3}{\partial x^1}\Big)dx^0dx^3dx^1 \\ &+\Big(-\frac{\partial D_3}{\partial x^0} + \frac{\partial H^2}{\partial x^1} - \frac{\partial H^1}{\partial x^2}\Big)dx^0dx^1dx^2 - \Big(\frac{\partial D_1}{\partial x^1} + \frac{\partial D^2}{\partial x^2} + \frac{\partial D^3}{\partial x^3}\Big)dx^1dx^2dx^3 \end{aligned}$$

$$
\begin{aligned}
&= \Big(-(\frac{\partial \vec{D}}{\partial x^0})_1 + (\operatorname{curl} \vec{H})_1 \Big) dx^0 dx^2 dx^3 + \Big(-(\frac{\partial \vec{D}}{\partial x^0})_2 + (\operatorname{curl} \vec{H})_2 \Big) dx^0 dx^3 dx^1 \\
&\quad + \Big(-(\frac{\partial \vec{D}}{\partial x^0})_3 + (\operatorname{curl} \vec{H})_3 \Big) dx^0 dx^1 dx^2 - \big(\operatorname{div} \vec{D}\big) dx^1 dx^2 dx^3 \\
&= -\rho\, dx^1 dx^2 dx^3 + \frac{1}{c} j^1\, dx^0 dx^2 dx^3 + \frac{1}{c} j^2\, dx^0 dx^3 dx^1 + \frac{1}{c} j^3\, dx^0 dx^1 dx^2 \\
&= -J
\end{aligned}
$$

This is the expression of the second pair of Maxwell's equations using differential forms. We can rewrite this slightly as

$$
\begin{aligned}
\delta F &= *\, d * F = -\rho\, dx^0 + \frac{1}{c} j^1\, dx^1 + \frac{1}{c} j^2\, dx^2 + \frac{1}{c} j^3\, dx^3 \\
&= - * J
\end{aligned}
$$

Now we continue on to the potential equations. The first thing to deal with is the condition of Lorenz[10]. This is very convenient. We form

$$
\begin{aligned}
\delta A &= *\, d * (\phi dx^0 - A^1\, dx^1 - A^2\, dx^2 - A^3\, dx^3) \\
&= *\, d\, (\phi dx^1 dx^2 dx^3 - A^1 dx^0 dx^2 dx^3 - A^2 dx^0 dx^3 dx^1 - A^3 dx^0 dx^3 dx^1) \\
&= * \Big(\frac{\partial \phi}{\partial x^0} + \frac{\partial A^1}{\partial x^1} + \frac{\partial A^2}{\partial x^2} + \frac{\partial A^3}{\partial x^3} \Big) dx^0 dx^1 dx^2 dx^3 \\
&= -\Big(\frac{\partial \phi}{\partial x^0} + \frac{\partial A^1}{\partial x^1} + \frac{\partial A^2}{\partial x^2} + \frac{\partial A^3}{\partial x^3} \Big) \\
&= -\Big(\frac{1}{c} \frac{\partial \phi}{\partial t} + \operatorname{div} \vec{A} \Big)
\end{aligned}
$$

Thus the condition of Lorenz $\frac{1}{c}\frac{\partial \phi}{\partial t} + \operatorname{div} \vec{A} = 0$ is expressed simply by $\delta A = 0$. We discuss later how A may be modified so that it satisfies the condition of Lorenz.

Now we can derive the potential equations for ϕ and $\vec{A}$. We have to compute $\Box A$.

$$
\begin{aligned}
\Box A &= (\delta d + d\delta) A \\
&= \delta dA + 0
\end{aligned}
$$

using the condition of Lorenz. Next we recall $dA = F$ so

$$
\begin{aligned}
\Box A &= \delta dA \\
&= \delta F \\
&= - * J
\end{aligned}
$$

and that is the potential equation. However, to get actual use out of it and to see how it relates to ancient notations, we compute the d'Alembertian on

[10] See the historical note on the *Condition of Lorenz* in an Appendix to this section.

functions:

$$\begin{aligned}
\Box f &= (\delta d + d\delta) f \\
&= \delta df + 0 = *\, d * \Big(\frac{\partial f}{\partial x^0} dx^0 + \frac{\partial f}{\partial x^1} dx^1 + \frac{\partial f}{\partial x^2} dx^2 + \frac{\partial f}{\partial x^3} dx^3 \Big) \\
&= *\, d \Big(\frac{\partial f}{\partial x^0} dx^1 dx^2 dx^3 + \frac{\partial f}{\partial x^1} dx^0 dx^2 dx^3 + \frac{\partial f}{\partial x^2} dx^0 dx^3 dx^1 + \frac{\partial f}{\partial x^3} dx^0 dx^1 dx^2 \Big) \\
&= *\Big(\frac{\partial^2 f}{\partial x^{0\,2}} - \frac{\partial^2 f}{\partial x^{1\,2}} - \frac{\partial^2 f}{\partial x^{2\,2}} - \frac{\partial^2 f}{\partial x^{3\,2}} \Big) dx^0 dx^1 dx^2 dx^3 \\
&= -\Big(\frac{\partial^2 f}{\partial x^{0\,2}} - \frac{\partial^2 f}{\partial x^{1\,2}} - \frac{\partial^2 f}{\partial x^{2\,2}} - \frac{\partial^2 f}{\partial x^{3\,2}} \Big)
\end{aligned}$$

The classical d'Alembertian is

$$\boxed{\text{p}} = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^{1\,2}} - \frac{\partial^2}{\partial x^{2\,2}} - \frac{\partial^2}{\partial x^{3\,2}}$$

so we see that

$$\Box f = -\boxed{\text{p}} f$$

Next we use the formula proved in an appendix to this section that

$$\begin{aligned}
\Box A &= \Box(\phi dx^0 - A^1 dx^1 - A^2 dx^2 - A^3 dx^3) \\
&= (\Box\phi) dx^0 - (\Box A^1) dx^1 - (\Box A^2) dx^2 - (\Box A^3) dx^3
\end{aligned}$$

Now comparing this with

$$\Box A = - * J = -\rho dx^0 + \frac{1}{c} j^1 dx^1 + \frac{1}{c} j^2 dx^2 + \frac{1}{c} j^3 dx^3$$

we see

$$\begin{aligned}
\Box \phi &= -\rho & \qquad \boxed{\text{p}}\, \phi &= \rho \\
\Box A^i &= -\frac{1}{c} j^i & \qquad \boxed{\text{p}}\, A^i &= \frac{1}{c} j^i
\end{aligned}$$

The equations on the right are the classical potential equations.

Next we discuss the matter of how to arrange for the condition of Lorenz to hold. Recall that that $dF = 0$ from which, as a consequence of the converse of the Poincaré lemma, there is 1-form A with $dA = F$. Suppose now we take any differentiable function G and add its differential to A. Then with $A' = A + dG$ we have

$$dA' = d(A + dG) = dA + ddG = dA + 0 = F$$

So the question is, which G should be added to A so that the condition of Lorenz holds. Recall the condition of Lorenz is $\delta A' = 0$, We need

$$\begin{aligned}
\delta A' &= 0 \\
\delta(A + dG) &= 0 \\
\delta dG &= -\delta A \\
\delta dG + d\delta G &= -\delta A \qquad \text{since } \delta G \text{ is } 0 \\
\Box G &= -\delta A
\end{aligned}$$

This is the same type of equation which we found before, and solving it one has G so that $\delta(A + dG) = 0$ as required.

We now want to extend this result to the case where ϵ or μ are not one. This is surprisingly difficult. However, we suspect it can be done because the corresponding equations in three space have the same form for the potential ϕ and the vector potential components A_i. Recall that the form of the equation is

$$\frac{\epsilon\mu}{c^2}\frac{\partial^2 A_i}{\partial t^2} - \frac{\partial^2 A_i}{\partial x^2} - \frac{\partial^2 A_i}{\partial y^2} - \frac{\partial^2 A_i}{\partial z^2} = \frac{\mu}{c} j^i$$

This suggests an electromagnetic wave moving at a speed of

$$k = \frac{c}{\sqrt{\epsilon\mu}}$$

Now we must steer the boat, like Odysseus, between Charybdis and Skilla. Charybdis refers to the fact that for the d'Alembertian to come out with $\frac{1}{k^2} = \frac{\epsilon\mu}{c^2}$ in it, we are going to have to modify the $*$ operator by changing c to k. Skilla refers to the fact that Maxwell's equations have c not k in them. Thus replacing all the c's by k's won't work; we must steer more subtlely. Nevertheless the fact that the equations for ϕ and A_i have the same form in 3-space suggests that it is possible to navigate successfully, which we will now do.

It is worth noting first that the methodology we are about to introduce cannot be the final electromagnetic word, because there are materials for which the relationship between $\vec{D}$ and $\vec{E}$ is not given by a constant ratio, and there are two reasons for this. First, ϵ may vary over space, or over time, but even worse, the two vectors do not point the same way in some materials and circumstances so that the constant ϵ must be replaced by a matrix (ϵ_{ij}) so that $D_i = \epsilon_{ij}E^j$. Naturally the same is true of $\vec{B}$ and $\vec{H}$. So, to be explicit, we assume in what follows that we are working over a region where c, ϵ and μ are all *constant*[11]

There is some possibility of confusion in what follows if we use dx^0. To avoid this, we will revert to using dt, dx, dy, dz. In the following calculations the $*$ operator uses the same equations as before *but* the constant c in those equations is replaced by the constant k. Thus

$$\begin{aligned} *dx^0 dx^2 &= -dx^3 dx^1 \\ *\,c\,dtdy &= -dzdx \end{aligned}$$

changes to

$$*k\,dtdy = -dzdx$$

Except for some trivial algebra the calculation goes as before, and since the ideas are the same we will just present the calculations in the most efficient

[11] We of course know c is constant over the entire age of the entire universe because we have measured it for 150 years and across the solar system.

order. Recalling that $k = c/\sqrt{\epsilon\mu}$, the potential form A is

$$\begin{aligned} A &= \phi\, cdt - A^1\, dx - A^2\, dx - A^3\, dx \\ &= \frac{c}{k}\phi\, kdt - A^1\, dx - A^2\, dx - A^3\, dx \\ &= \sqrt{\epsilon\mu}\,\phi\, kdt - A^1\, dx - A^2\, dx - A^3\, dx \end{aligned}$$

First we go after the codifferential δA,

$$\begin{aligned} *A &= \sqrt{\epsilon\mu}\,\phi\, dxdydz - A^1\, kdtdydz - A^2\, kdtdzdx - A^3\, kdtdxdy \\ d*A &= \left(\frac{\sqrt{\epsilon\mu}}{k}\frac{\partial\phi}{\partial t} + \frac{\partial A^1}{\partial x} + \frac{\partial A^2}{\partial y} + \frac{\partial A^3}{\partial z}\right) kdtdxdydz \\ \delta A = *d*A &= -\Big(\frac{\epsilon\mu}{c}\frac{\partial\phi}{\partial t} + \frac{\partial A^1}{\partial x} + \frac{\partial A^2}{\partial y} + \frac{\partial A^3}{\partial z}\Big) = 0 \end{aligned}$$

by the condition of Lorenz, $\delta A = 0$, which we are assuming, as before. One of the the positive aspects of the four dimensional treatment is that the condition of Lorenz is so simply expressed: $\delta A = 0$ We note in passing that the condition of Lorenz decoded is

$$\frac{\epsilon\mu}{c}\frac{\partial\phi}{\partial t} + \frac{\partial A^1}{\partial x} + \frac{\partial A^2}{\partial y} + \frac{\partial A^3}{\partial z} = 0 \qquad \text{Condition of Lorenz}$$

Next we have, and note we here use c not k so that we can use Maxwell's equations,

$$\begin{aligned} F &= dA \\ &= \left(-\frac{1}{c}\frac{\partial A^1}{\partial t} - \frac{\partial\phi}{\partial x}\right) cdtdx + \left(-\frac{1}{c}\frac{\partial A^2}{\partial t} - \frac{\partial\phi}{\partial y}\right) cdtdy + \left(-\frac{1}{c}\frac{\partial A^3}{\partial t} - \frac{\partial\phi}{\partial z}\right) cdtdy \\ &\quad + \left(-\frac{\partial A^3}{\partial y} + \frac{\partial A^2}{\partial z}\right) dydz + \left(-\frac{\partial A^1}{\partial z} + \frac{\partial A^3}{\partial x}\right) dzdx + \left(-\frac{\partial A_2}{\partial x} + \frac{\partial A^1}{\partial y}\right) dxdy \\ &= E^1\, cdtdx + E^2\, cdtdy + E^3\, cdtdz \\ &\quad -B_1\, dydz - B_2\, dzdx - B_3\, dxdy \\ &= \sqrt{\epsilon\mu}\left(E^1\, kdtdx + E^2\, kdtdy + E^3\, kdtdz\right) \\ &\quad -B_1\, dydz - B_2\, dzdx - B_3\, dxdy \\ *F &= \sqrt{\epsilon\mu}\left(-E^1\, dydz - E^2\, dzdx - E^3\, dxdy\right) \\ &\quad -B_1\, kdtdx - B_2\, kdtdy - B_3\, kdtdz \\ &= \frac{\sqrt{\epsilon\mu}}{\epsilon}\left(-D_1\, dydz - D_2\, dzdx - D_3\, dxdy\right) \\ &\quad +\frac{\mu}{\sqrt{\epsilon\mu}}\left(-H^1\, cdtdx - H^2\, cdtdy - H^3\, cdtdz\right) \\ &= \sqrt{\frac{\mu}{\epsilon}}\,\tilde{F} \end{aligned}$$

where

$$\begin{aligned}\tilde{F} &= \sqrt{\frac{\epsilon}{\mu}} * F \\ &= -H^1\, cdtdx - H^2\, cdtdy - H^3\, cdtdz \\ &\quad -D_1\, dydz - D_2\, dzdx - D_3\, dxdy\end{aligned}$$

Now we find $d\tilde{F}$,

$$\begin{aligned}d\tilde{F} &= \left(-\frac{1}{c}\frac{\partial D_1}{\partial t} + \left(\frac{\partial H_3}{\partial y} - \frac{\partial H_2}{\partial z}\right)\right) cdtdydz + \left(-\frac{1}{c}\frac{\partial D_2}{\partial t} + \left(\frac{\partial H_1}{\partial z} - \frac{\partial H_3}{\partial x}\right)\right) cdtdzdx \\ &+ \left(-\frac{1}{c}\frac{\partial D_3}{\partial t} + \left(\frac{\partial H_2}{\partial x} - \frac{\partial H_1}{\partial y}\right)\right) cdtdxdy - \left(\frac{\partial D_1}{\partial x} + \frac{\partial D_2}{\partial y} + \frac{\partial D_3}{\partial z}\right) dxdydz \\ &= \left(-\frac{1}{c}\frac{\partial D_1}{\partial t} + (\operatorname{curl}\vec{H})_1\right) cdtdydz + \left(-\frac{1}{c}\frac{\partial D_2}{\partial t} + (\operatorname{curl}\vec{H})_2\right) cdtdzdx \\ &+ \left(-\frac{1}{c}\frac{\partial D_3}{\partial t} + (\operatorname{curl}\vec{H})_3\right) cdtdxdy - (\operatorname{div}\vec{D})\, dxdydx \\ &= -\rho\, dxdydz + \frac{1}{c}\left(j_1\, cdtdydz + j_2\, cdtdzdx + j_3\, cdtdxdy\right) \;=\; -J \\ &= -\rho\, dxdydz + \frac{\sqrt{\epsilon\mu}}{c}\left(j_1\, kdtdydz + j_2\, kdtdzdx + j_3\, kdtdxdy\right)\end{aligned}$$

We note in passing that the source form J has showed up here but it is not so useful in this context because ϵ and μ keep it from playing it's former role. Finally, we have

$$\begin{aligned}\Box A &= (\delta d + d\delta)A \\ &= \delta dA + 0 \qquad \text{(conditon of Lorenz: } \delta A = 0) \\ &= \delta F \;=\; *\, d * F \\ &= *\, d\sqrt{\frac{\mu}{\epsilon}}\,\tilde{F} \;=\; \sqrt{\frac{\mu}{\epsilon}} *\, d\tilde{F} \\ &= \sqrt{\frac{\mu}{\epsilon}} * \left(-\rho\, dxdydz + \frac{\sqrt{\epsilon\mu}}{c}\left(j_1\, kdtdydz + j_2\, kdtdzdx + j_3\, kdtdxdy\right)\right) \\ &= \sqrt{\frac{\mu}{\epsilon}}\left(-\rho\, kdt + \frac{\sqrt{\epsilon\mu}}{c}\left(j_1\, dx + j_2\, dy + j_3\, dz\right)\right) \\ &= \sqrt{\frac{\mu}{\epsilon}}\left(-\frac{1}{\sqrt{\epsilon\mu}}\,\rho\, cdt + \frac{\sqrt{\epsilon\mu}}{c}\left(j_1\, dx + j_2\, dy + j_3\, dz\right)\right) \\ &= -\frac{\rho}{\epsilon}\, cdt + \frac{\mu}{c}\left(j_1\, dx + j_2\, dy + j_3\, dz\right)\end{aligned}$$

As before, to make use of this equation we must translate it into component form. The derivation in the appendix of

$$\Box A = (\Box A_0)\, dx^0 + (\Box A_1)\, dx^1 + (\Box A_2)\, dx^2 + (\Box A_3)\, dx^3$$

for $A = A_0\,dx^0 + A_1\,dx^1 + A_2\,dx^2 + A_3\,dx^3$ made no use of c or k. However, we must now find $\Box f$ for a function f in our new circumstances. Since our $*$ formulae use k instead of c, i.e.

$$*\,k\,dtdx^2 = -dx^3dx^1$$

we may expect k to show up in place of c in the following form

$$\Box f = -\frac{1}{k^2}\frac{\partial^2 f}{\partial t^2} + \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

Using this to untangle

$$\Box A = -\frac{\rho}{\epsilon}\,c\,dt + \frac{\mu}{c}\left(j^1\,dx + j^2\,dy + j^3\,dz\right)$$

with $A = \phi cdt - A^1\,dx^1 - A^2\,dx^2 - A^3\,dx^3)$ we have

$$\begin{aligned}\Box\,\phi &= -\frac{\rho}{\epsilon}\\ \Box\,A^i &= -\frac{\mu}{c}\,j^i\end{aligned}$$

Since $k = \frac{c}{\sqrt{\epsilon\mu}}$ and $\boxed{\text{P}}\,f = -\Box\,f$ this translates to

$$\begin{aligned}\frac{\epsilon\mu}{c^2}\frac{\partial^2\phi}{\partial t^2} - \frac{\partial^2\phi}{\partial x^2} - \frac{\partial^2\phi}{\partial y^2} - \frac{\partial^2\phi}{\partial z^2} &= \boxed{\text{P}}\,\phi = \frac{\rho}{\epsilon}\\ \frac{\epsilon\mu}{c^2}\frac{\partial^2 A_i}{\partial t^2} - \frac{\partial^2 A_i}{\partial x^2} - \frac{\partial^2 A_i}{\partial y^2} - \frac{\partial^2 A_i}{\partial z^2} &= \boxed{\text{P}}\,A^i = \frac{\mu}{c}\,j_i\end{aligned}$$

which are the classical equations and indicate a wave traveling at speed $\frac{c}{\epsilon\mu}$.

Appendix I Derivation of d'Alembertian for Functions and Forms

We are here going to derive a couple of formulas for the d'Alembertian which we used in the main part of the Chapter. This calculation is of almost no interest and unless you are really interested you can skip it. The result can be generalized, which we discuss after the derivation.

We will first calculate the d'Alembertian on functions, which is fairly simple. We revert to using x^0, x^1, x^2, x^3 instead of cdt, dx, dy, dz so that the symmetry of the situation is more obvious. Remembering that $\delta f = *\,d\,*\,f$ is 0 since $*f \in \Lambda^4$ and thus $d * f = 0$, we have

$$\begin{aligned}\Box\,f &= (\delta d + d\delta)f\\ &= *\,d\,*\,df + 0\\ &= *\,d\,*\left(\frac{\partial f}{\partial x^0}dx^0 + \frac{\partial f}{\partial x^1}dx^1 + \frac{\partial f}{\partial x^2}dx^2 + \frac{\partial f}{\partial x^3}dx^3\right)\\ &= *\,d\Big(\frac{\partial f}{\partial x^0}dx^1dx^2dx^3 + \frac{\partial f}{\partial x^1}dx^0dx^2dx^3 + \frac{\partial f}{\partial x^2}dx^0dx^3dx^1 + \frac{\partial f}{\partial x^3}dx^0dx^1dx^2\Big)\\ &= *\Big(\frac{\partial^2 f}{\partial x^{0\,2}} - \frac{\partial^2 f}{\partial x^{1\,2}} - \frac{\partial^2 f}{\partial x^{2\,2}} - \frac{\partial^2 f}{\partial x^{3\,2}}\Big)dx^0dx^1dx^2dx^3\\ &= -\frac{\partial^2 f}{\partial x^{0\,2}} + \frac{\partial^2 f}{\partial x^{1\,2}} + \frac{\partial^2 f}{\partial x^{2\,2}} + \frac{\partial^2 f}{\partial x^{3\,2}}\end{aligned}$$

Our next job is to compute the d'Alembertian on the 1-form

$$A = A_0\,dx^0 + A_1\,dx^1 + A_2\,dx^2 + A_3\,dx^3$$

For the final step of this calculation we need, for clarity, to distinguish the d'Alembertian on functions from that on 1-forms. Thus, temporarily, we will refer to the d'Alembertian on functions $f \in \Lambda^0$ by $\Box_0 f$. We need to compute $(d\delta + \delta d)A$, which we do in stages.

$$\begin{aligned}
\delta A &= *\,d\,*\,A \\
&= *\,d\,*\,(A_0\,dx^0 + A_1\,dx^1 + A_2\,dx^2 + A_3\,dx^3) \\
&= *\,d(A_0\,dx^1dx^2dx^3 + A_1\,dx^0dx^2dx^3 + A_2\,dx^0dx^3dx^1 + A_3\,dx^0dx^1dx^2) \\
&= *\Big(\frac{\partial A_0}{\partial x^0} - \frac{\partial A_1}{\partial x^1} - \frac{\partial A_2}{\partial x^2} - \frac{\partial A_3}{\partial x^3}\Big)dx^0dx^1dx^2dx^3 \\
&= -\frac{\partial A_0}{\partial x^0} + \frac{\partial A_1}{\partial x^1} + \frac{\partial A_2}{\partial x^2} + \frac{\partial A_3}{\partial x^3} \\
d\delta A &= \Big(-\frac{\partial^2 A_0}{\partial x^{0\,2}} + \frac{\partial^2 A_1}{\partial x^0\partial x^1} + \frac{\partial^2 A_2}{\partial x^0\partial x^2} + \frac{\partial^2 A_3}{\partial x^0\partial x^3}\Big)dx^0 \\
&+ \Big(-\frac{\partial^2 A_0}{\partial x^1\partial x^0} + \frac{\partial^2 A_1}{\partial x^{1\,2}} + \frac{\partial^2 A_2}{\partial x^1\partial x^2} + \frac{\partial^2 A_3}{\partial x^1\partial x^3}\Big)dx^1 \\
&+ \Big(-\frac{\partial^2 A_0}{\partial x^2\partial x^0} + \frac{\partial^2 A_1}{\partial x^2\partial x^1} + \frac{\partial^2 A_2}{\partial x^{2\,2}} + \frac{\partial^2 A_3}{\partial x^2\partial x^3}\Big)dx^2 \\
&+ \Big(-\frac{\partial^2 A_0}{\partial x^3\partial x^0} + \frac{\partial^2 A_1}{\partial x^3\partial x^1} + \frac{\partial^2 A_2}{\partial x^3\partial x^2} + \frac{\partial^2 A_3}{\partial x^{3\,2}}\Big)dx^3
\end{aligned}$$

Now we do the other term

$$\begin{aligned}
dA &= d(A_0\,dx^0 + A_1\,dx^1 + A_2\,dx^2 + A_3\,dx^3) \\
&= \Big(\frac{\partial A_1}{\partial x^0} - \frac{\partial A_0}{\partial x^1}\Big)dx^0dx^1 + \Big(\frac{\partial A_2}{\partial x^0} - \frac{\partial A_0}{\partial x^2}\Big)dx^0dx^2 + \Big(\frac{\partial A_3}{\partial x^0} - \frac{\partial A_0}{\partial x^3}\Big)dx^0dx^3 \\
&+ \Big(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}\Big)dx^2dx^3 + \Big(\frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}\Big)dx^3dx^1 + \Big(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}\Big)dx^1dx^2 \\
*\,dA &= -\Big(\frac{\partial A_1}{\partial x^0} - \frac{\partial A_0}{\partial x^1}\Big)dx^2dx^3 - \Big(\frac{\partial A_2}{\partial x^0} - \frac{\partial A_0}{\partial x^2}\Big)dx^3dx^1 - \Big(\frac{\partial A_3}{\partial x^0} - \frac{\partial A_0}{\partial x^3}\Big)dx^1dx^2 \\
&+ \Big(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}\Big)dx^0dx^1 + \Big(\frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}\Big)dx^0dx^2 + \Big(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}\Big)dx^0dx^3 \\
d*\,dA &= \Big(-\frac{\partial^2 A_1}{\partial x^1\partial x^0} + \frac{\partial^2 A_0}{\partial x^{1\,2}} - \frac{\partial^2 A_2}{\partial x^2\partial x^0} + \frac{\partial^2 A_0}{\partial x^{2\,2}} - \frac{\partial^2 A_3}{\partial x^3\partial x^0} + \frac{\partial^2 A_0}{\partial x^{3\,2}}\Big)dx^1dx^2dx^3 \\
&+ \Big(-\frac{\partial^2 A_1}{\partial x^{0\,2}} + \frac{\partial^2 A_0}{\partial x^0\partial x^1} - \frac{\partial^2 A_2}{\partial x^2\partial x^1} + \frac{\partial^2 A_1}{\partial x^{2\,2}} + \frac{\partial^2 A_1}{\partial x^{3\,2}} - \frac{\partial^2 A_3}{\partial x^3\partial x^1}\Big)dx^0dx^2dx^3 \\
&+ \Big(-\frac{\partial^2 A_2}{\partial x^{0\,2}} + \frac{\partial^2 A_0}{\partial x^0\partial x^2} - \frac{\partial^2 A_3}{\partial x^3\partial x^2} + \frac{\partial^2 A_2}{\partial x^{3\,2}} + \frac{\partial^2 A_2}{\partial x^{1\,2}} - \frac{\partial^2 A_1}{\partial x^1\partial x^2}\Big)dx^0dx^3dx^1 \\
&+ \Big(-\frac{\partial^2 A_3}{\partial x^{0\,2}} + \frac{\partial^2 A_0}{\partial x^0\partial x^3} - \frac{\partial^2 A_1}{\partial x^1\partial x^3} + \frac{\partial^2 A_3}{\partial x^{1\,2}} + \frac{\partial^2 A_3}{\partial x^{2\,2}} - \frac{\partial^2 A_2}{\partial x^2\partial x^3}\Big)dx^0dx^1dx^2 \\
\delta dA &= *\,d\,*\,dA
\end{aligned}$$

$$
\begin{aligned}
&= \Big(-\frac{\partial^2 A_1}{\partial x^1 \partial x^0} + \frac{\partial^2 A_0}{\partial x^{1\,2}} - \frac{\partial^2 A_2}{\partial x^2 \partial x^0} + \frac{\partial^2 A_0}{\partial x^{2\,2}} - \frac{\partial^2 A_3}{\partial x^3 \partial x^0} + \frac{\partial^2 A_0}{\partial x^{3\,2}} \Big) dx^0 \\
&+ \Big(-\frac{\partial^2 A_1}{\partial x^{0\,2}} + \frac{\partial^2 A_0}{\partial x^0 \partial x^1} - \frac{\partial^2 A_2}{\partial x^2 \partial x^1} + \frac{\partial^2 A_1}{\partial x^{2\,2}} + \frac{\partial^2 A_1}{\partial x^{3\,2}} - \frac{\partial^2 A_3}{\partial x^3 \partial x^1} \Big) dx^1 \\
&+ \Big(-\frac{\partial^2 A_2}{\partial x^{0\,2}} + \frac{\partial^2 A_0}{\partial x^0 \partial x^2} - \frac{\partial^2 A_3}{\partial x^3 \partial x^2} + \frac{\partial^2 A_2}{\partial x^{3\,2}} + \frac{\partial^2 A_2}{\partial x^{1\,2}} - \frac{\partial^2 A_1}{\partial x^1 \partial x^2} \Big) dx^2 \\
&+ \Big(-\frac{\partial^2 A_3}{\partial x^{0\,2}} + \frac{\partial^2 A_0}{\partial x^0 \partial x^3} - \frac{\partial^2 A_1}{\partial x^1 \partial x^3} + \frac{\partial^2 A_3}{\partial x^{1\,2}} + \frac{\partial^2 A_3}{\partial x^{2\,2}} - \frac{\partial^2 A_2}{\partial x^2 \partial x^3} \Big) dx^3
\end{aligned}
$$

$$
\begin{aligned}
\Box A &= d\delta A + \delta dA \\
&= \Big(-\frac{\partial^2 A_0}{\partial x^{0\,2}} + \frac{\partial^2 A_0}{\partial x^{1\,2}} + \frac{\partial^2 A_0}{\partial x^{2\,2}} + \frac{\partial^2 A_0}{\partial x^{3\,2}} \Big) dx^0 + \Big(-\frac{\partial^2 A_1}{\partial x^{0\,2}} + \frac{\partial^2 A_1}{\partial x^{1\,2}} + \frac{\partial^2 A_1}{\partial x^{2\,2}} + \frac{\partial^2 A_1}{\partial x^{3\,2}} \Big) dx^1 \\
&+ \Big(-\frac{\partial^2 A_2}{\partial x^{0\,2}} + \frac{\partial^2 A_2}{\partial x^{1\,2}} + \frac{\partial^2 A_2}{\partial x^{2\,2}} + \frac{\partial^2 A_2}{\partial x^{3\,2}} \Big) dx^2 + \Big(-\frac{\partial^2 A_3}{\partial x^{0\,2}} + \frac{\partial^2 A_3}{\partial x^{1\,2}} + \frac{\partial^2 A_3}{\partial x^{2\,2}} + \frac{\partial^2 A_3}{\partial x^{3\,2}} \Big) dx^3 \\
&= \Box_0 A_0 dx^0 + \Box_0 A_1 dx^1 + \Box_0 A_2 dx^2 + \Box_0 A_3 dx^3
\end{aligned}
$$

This is the formula that we used in the main text. It may appear completely miraculous. However, something like it is true in general; the difference is that there are lower order terms, so the each term then becomes $(\Box A_1 +$ lower order terms $)dx^1$. This is interesting but too difficult for us to investigate.

Appendix II Tensor form of the Electromagnetic Equations

Persons who have been through Electromagnetics using Maxwell's equations with a tensor approach might be interested in comparing the two treatments. Therefore I provide a very brief look at the tensor approach. We begin by specifying the metric tensor:

$$
(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}
$$

We will use the following descriptions of the Electric and Magnetic fields:

$$
\vec{E} = (E^1, E^2, E^3) \qquad \vec{B} = (B_1, B_2, B_3)
$$

There is no significance to the position of the indices on E^i or b_j. Next we specify the electromagnetic tensors $F_{\mu\nu}$ and $F^{\mu\nu}$ which are connected in the usual way by $F_{\mu\nu} = g_{\rho\mu} g_{\sigma\nu} F^{\rho\sigma}$. (Since $(g_{\mu\nu})$ is summetric the order of its indices does not matter.) The electromagnetic tensors are

$$
(F^{\mu\nu}) = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & -B_3 & B_2 \\ E^2 & B_3 & 0 & -B_1 \\ E^3 & -B_2 & B_1 & 0 \end{pmatrix} \qquad (F_{\mu\nu}) = \begin{pmatrix} 0 & E^1 & E^2 & E^3 \\ -E^1 & 0 & -B_3 & B_2 \\ -E^2 & B_3 & 0 & -B_1 \\ -E^3 & -B_2 & B_1 & 0 \end{pmatrix}
$$

The values in $F^{\mu\nu}$ are selected so that they eventually mesh with Maxwell's equations. There is more than one way to do this but those give above are fairly standard.

More notation: $\partial_\mu = \frac{\partial}{\partial x^\mu}$ and $\partial^\nu = g^{\mu\nu}\partial_\mu$. Also the action of ∂_μ on a tensor is often indicated by adding $,\mu$ to its lower indices, for example $\partial_\mu F_{\rho\sigma} = F_{\rho\sigma,\mu}$

In tensorland there is a theorem that if $F_{\mu\nu,\rho} + F_{\rho\mu,\nu} + F_{\nu\rho,\mu} = 0$ then there is a Tensor A so that $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. (This is nothing but the converse of the Poincare lemma applied to $F = F_{\mu\nu}\, dx^\mu \wedge dx^\nu$.) This is where the vector potential $\vec{A} = (A^1, A^2, A^3)$ enters the game. We define a four-vector A by

$$(A_\mu) = (\phi, -A^1, -A^2, -A^3)$$

Signs are chosen so the usual 3 dimensional formulas come out right. Hopefully we will now get that

$$\boxed{F_{\mu\nu,\rho} + F_{\rho\mu,\nu} + F_{\nu\rho,\mu} = 0}$$

is the tensor form of the two Maxwell equations $\mathrm{div}\vec{B} = 0$ and $\mathrm{curl}\,\vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}$. Also the formula $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ should be the tensor form of $\vec{B} = \mathrm{curl}\,\vec{A}$ and $\vec{E} = -\mathrm{grad}\,\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}$. We now run a few random checks to see this is true.

$$\begin{aligned} F_{12,3} + F_{31,2} + F_{23,1} &= -\frac{\partial B_3}{\partial x^3} - \frac{\partial B_2}{\partial x^2} - \frac{\partial B_1}{\partial x^1} = -\mathrm{div}\,\vec{B} \\ F_{02,3} + F_{30,2} + F_{23,0} &= \frac{\partial E^2}{\partial x^3} - \frac{\partial E^3}{\partial x^2} - \frac{\partial B_1}{\partial x^0} \\ &= -(\mathrm{curl}\,\vec{E})_1 - \frac{1}{c}\frac{\partial B_1}{\partial t} \end{aligned}$$

verifying that the the condition $F_{\mu\nu,\rho} + F_{\rho\mu,\nu} + F_{\nu\rho,\mu} = 0$ is indeed identical to the two Maxwell equations. Next we look at the the potential A.

$$\begin{aligned} E^1 = F_{01} &= \partial_0 A_1 - \partial_1 A_0 = \frac{1}{c}\frac{\partial(-A^1)}{\partial t} - \frac{\partial \phi}{\partial x^1} \\ &= -(\mathrm{grad}\,\phi)_1 - (\frac{1}{c}\frac{\partial \vec{A}}{\partial t})_1 \\ B_1 = F_{32} &= \partial_3 A_2 - \partial_2 A_3 = -\frac{\partial A^2}{\partial x^3} + \frac{\partial A^3}{\partial x^2} \\ &= (\mathrm{curl}\,\vec{A})_1 \end{aligned}$$

So we see $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is indeed equivalent to $\vec{B} = \mathrm{curl}\,\vec{A}$ and $\vec{E} = -\mathrm{grad}\,\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}$. If you are interested in these matters it would be useful for you to work out a few more of the possible choices of indices and verify that all is as it should be.

We have expressed two of the four Maxwell equations in tensor form. Next we need to express the other two. For this we need the current tensor

$$(j^\nu) = (c\rho, j^1, j^2, j^3) \text{ where the current is } \vec{j} = (j^1, j^2, j^3)$$

and ρ is the charge density. (The reason for the c in the charge density part of (j^ν) is that current is moving charge with respect to t not x^0 which perturbs

the formulas. If we were starting from scratch we could fix this but it is too late; the formula as given is standard.) The tensor formulation of the second two Maxwell's equations is simply

$$\boxed{\partial_\mu F^{\mu\nu} = \tfrac{1}{c} j^\nu}$$

as we will now show by calculating a small sample of the possible index choices.

$$\begin{aligned}
\partial_\mu F^{\mu 0} &= \frac{\partial F^{10}}{\partial x^1} + \frac{\partial F^{20}}{\partial x^2} + \frac{\partial F^{30}}{\partial x^3} \\
&= \frac{\partial E^1}{\partial x^1} + \frac{\partial E^2}{\partial x^2} + \frac{\partial E^3}{\partial x^3} \\
&= \operatorname{div} \vec{E} = \rho = \frac{1}{c} j^0 \\
\partial_\mu F^{\mu 1} &= \frac{\partial F^{01}}{\partial x^0} + \frac{\partial F^{21}}{\partial x^2} + \frac{\partial F^{31}}{\partial x^3} \\
&= \frac{1}{c}\frac{\partial(-E^1)}{\partial t} + \frac{\partial B_3}{\partial x^2} + \frac{\partial(-B_2)}{\partial x^3} \\
&= -\frac{1}{c}(\frac{\partial \vec{E}}{\partial t})_1 + (\operatorname{curl} \vec{B})_1 = \frac{1}{c} j^1
\end{aligned}$$

showing that the boxed equation above is indeed equivalent to the Maxwell's equations (with $\epsilon = \mu = 1$, $\vec{D} = \vec{E}$, $\vec{H} = \vec{B}$)

$$\operatorname{div} \vec{E} = \rho \qquad\qquad \operatorname{curl} \vec{B} = \frac{1}{c}\frac{\partial \vec{E}}{\partial t} + \frac{1}{c}\vec{j}$$

Notice the corollary

$$\partial_\nu j^\nu = c\, \partial_\nu \partial_\mu F^{\mu\nu} = 0$$

because $F^{\mu\nu}$ is skew symmetric. This then gives

$$\begin{aligned}
\frac{1}{c}\frac{\partial \rho}{\partial t} + \frac{\partial j^1}{\partial x^1} + \frac{\partial j^3}{\partial x^3} + \frac{\partial j^3}{\partial x^3} &= 0 \\
\frac{1}{c}\frac{\partial \rho}{\partial t} + \operatorname{div} \vec{j} &= 0
\end{aligned}$$

which is the equation of continuity.

The last thing we must treat is the potential equations which in the tensor formulation are quite easy. First note

$$(\partial_\mu) = \Big(\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\Big) \qquad\qquad (\partial^\mu) = \Big(\frac{\partial}{\partial x^0}, -\frac{\partial}{\partial x^1}, -\frac{\partial}{\partial x^2}, -\frac{\partial}{\partial x^3}\Big)$$

since $\partial^\mu = g^{\mu\nu}\partial_\nu$. Next the condition of Lorenz is easily described as

$$\boxed{\partial_\mu A^\mu = 0} \qquad\qquad \text{Condition of Lorenz}$$

indeed this equation decodes as

$$\begin{aligned} \frac{1}{c}\frac{\partial\phi}{\partial t} + \frac{\partial A^1}{\partial x^1} + \frac{\partial A^3}{\partial x^3} + \frac{\partial A^3}{\partial x^3} &= 0 \\ \frac{1}{c}\frac{\partial\phi}{\partial t} + \operatorname{div}\vec{A} &= 0 \end{aligned}$$

Now for the potential equations. First recall the equations $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ from which it is easy to get $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ using $g^{\mu\nu}$. We then have

$$\begin{aligned} \frac{1}{c}j^\nu &= \partial_\mu F^{\mu\nu} \\ &= \partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) \\ &= \partial_\mu\partial^\mu A^\nu - \partial^\nu(\partial_\mu A^\mu) \end{aligned}$$

so, by the condition of Lorenz $\partial_\mu A^\mu = 0$, we have the potential equation

$$\boxed{\partial_\mu\partial^\mu A^\nu = \tfrac{1}{c}j^\nu}$$

which decodes as

$$\begin{aligned} \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} - \frac{\partial^2\phi}{\partial x^{1\,2}} - \frac{\partial^2\phi}{\partial x^{2\,2}} - \frac{\partial^2\phi}{\partial x^{3\,2}} &= \rho \\ \frac{1}{c^2}\frac{\partial^2 A^i}{\partial t^2} - \frac{\partial^2 A^i}{\partial x^{1\,2}} - \frac{\partial^2 A^i}{\partial x^{2\,2}} - \frac{\partial^2 A^i}{\partial x^{3\,2}} &= \frac{1}{c}j^i \end{aligned}$$

which would be written more economically as

$$\boxed{\text{p}}\ \phi = \rho, \qquad\qquad \boxed{\text{p}}\ A^i = \frac{1}{c}j^i$$

Appendix III Historical Note on the Condition of Lorenz

The condition of Lorenz was first used by the Danish physicist Ludvig Lorenz (1829-1891). Lorenz had the misfortune to have almost the same name as the much more famous Dutch physicist Hendrik Lorentz (1853-1928), a friend of Einstein and early exponent of relativity, who also used the condition. This is the source of the confusion in the spelling of the name when citing the condition. It was suggested that the condition be referred to as the Lorenz-Lorentz condition, but this suggestion did not make it into general use, possibly due to considerations of euphony.

1.26 Energy Flow and Energy Density

The force on a particle of charge e and velocity $\vec{v}$ is

$$\vec{F} = e(\vec{E} + \vec{v} \times \vec{B})$$

This is called the *Lorentz force law.* Thus the work done in going from $\vec{R}$ to $\vec{R} + d\vec{\ell}$ is

$$\begin{aligned} dW &= \vec{F} \cdot d\vec{\ell} = e(\vec{E} + \vec{v} \times \vec{B}) \cdot d\vec{\ell} \\ &= e(\vec{E} + \vec{v} \times \vec{B}) \cdot \vec{v}\, dt \\ &= e\, \vec{E} \cdot \vec{v}\, dt \end{aligned}$$

Thus

$$\frac{dW}{dt} = e\, \vec{E} \cdot \vec{v}$$

Now let ρ be a continuous charge distribution. Then, as usual, we replace e by $\rho\, dt$ and integrate to find dW/dt. We a also need $\vec{j} = \rho\vec{v}$. The we have

$$\begin{aligned} \frac{dW}{dt} &= \int_M (\rho\, d\tau)\vec{E} \cdot \vec{v} \\ &= \int_M \vec{E} \cdot (\rho\vec{v})\, d\tau \\ &= \int_M \vec{E} \cdot \vec{j}\, d\tau \end{aligned}$$

Rewriting in terms of differential forms we have

$$\frac{dW}{dt} = \int_M E \wedge j$$

Now we can eliminate j by using the Maxwell equation

$$dH = \frac{1}{c}\frac{\partial D}{\partial t} + \frac{1}{c}\, j$$

which is Ampère's law. Thus

$$\begin{aligned} \frac{1}{c} j &= -\frac{1}{c}\frac{\partial D}{\partial t} + dH \\ \frac{dW}{dt} &= c \int_M E \wedge \frac{1}{c} j \\ &= c \int_M E \wedge \big(-\frac{1}{c}\frac{\partial D}{\partial t} + dH \big) \end{aligned}$$

For the next bit of the calculation we need the formula

$$E \wedge \frac{\partial D}{\partial t} = \frac{1}{2}\frac{\partial}{\partial t}(E \wedge D)$$

We use $D = \epsilon * E$ and the general formula for differential forms $\lambda \wedge *\mu = \mu \wedge *\lambda$. We then have

$$E \wedge \frac{\partial D}{\partial t} = E \wedge \frac{\partial}{\partial t}(\epsilon * E)$$

$$\begin{aligned} &= \epsilon\, E \wedge \frac{\partial}{\partial t}(*E) \\ &= \epsilon\, E \wedge *\frac{\partial E}{\partial t} \\ &= \epsilon\, \frac{\partial E}{\partial t} \wedge *E \\ &= \frac{\partial E}{\partial t} \wedge D \end{aligned}$$

Thus

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial t}(E \wedge D) &= \frac{1}{2}\Big(\frac{\partial E}{\partial t} \wedge D + E \wedge \frac{\partial D}{\partial t}\Big) \\ &= \frac{1}{2}\Big(E \wedge \frac{\partial D}{\partial t} + E \wedge \frac{\partial D}{\partial t}\Big) \\ &= E \wedge \frac{\partial D}{\partial t} \end{aligned}$$

In an exactly similar calculation we have

$$\frac{\partial B}{\partial t} \wedge H = \frac{1}{2}\,\frac{\partial}{\partial t}(B \wedge H)$$

Resuming the previous calculation we have

$$\begin{aligned} \frac{dW}{dt} &= c\int_M E \wedge \big(-\frac{1}{c}\,\frac{\partial D}{\partial t} + dH\big) \\ &= c\int_M -E \wedge \frac{1}{c}\,\frac{\partial D}{\partial t} + E \wedge dH \\ &= c\int_M -E \wedge \frac{1}{c}\,\frac{\partial D}{\partial t} + \big(dE \wedge H - d(E \wedge H)\big) \\ &= c\int_M -E \wedge \frac{1}{c}\,\frac{\partial D}{\partial t} + \big(-\frac{1}{c}\,\frac{\partial B}{\partial t} \wedge H - d(E \wedge H)\big) \\ &= c\int_M -\frac{1}{2c}\frac{\partial}{\partial t}\,(E \wedge D) - \frac{1}{2c}\frac{\partial}{\partial t}\,(B \wedge H) - d(E \wedge H) \\ &= \frac{\partial}{\partial t}\int_M -\frac{1}{2}(E \wedge D) - \frac{1}{2}(B \wedge H) - c\int_{\partial M} E \wedge H \end{aligned}$$

by Stokes' theorem. If we define the Energy Density u of the field as

$$u = -\frac{1}{2}\,(E \wedge D + B \wedge H)$$

and the Flux of energy through the surface ∂M as

$$S = c\,E \wedge H \qquad \text{Poyting's Flux Form}$$

we can interpret the previous equation as the change in the Energy as due to the decrease in the field energy less the flux of energy through the boundary.

Chapter 2

Mathematical Theory

2.1 Introduction

This chapter has two purposes. First we must prove the two major theorems of elementary differential forms, Stoke's theorem and the converse of the Poincaré lemma, and second we want to extend the theory with some deeper methods. For example, we would like to have formulas for the $*$ operator on manifolds, where orthonormal coordinates systems are usually not available, so we need some more flexible methods for that.

We also want to discuss the mathematical theories that underlie the development in the first chapter. In addition, we will look at the theory in a more general setting; that of a differentiable manifold rather than simply n-space and its subsets. A differentiable manifold is like a surface in n-space but with possibly more dimensions and without the n-space. If the manifold is actually inside an n-space $\mathbb{R}^n$ the manifold is called an embedded manifold. In point of fact, all manifolds can be embedded in an n-space of sufficiently high dimension [1], but it is no longer customary to do this. Also it may not be natural to do this; few people think that the four dimensional Space-Time of our universe is embedded in a five (or more) dimensional Euclidean space $\mathbb{R}^n$.

In the course of the next few sections we will introduce some concepts from topology, manifold theory, and tangent spaces. We cannot develop these subjects in detail because that would take a whole book in itself. Instead, we will describe the relevant material in a way that will orient the reader and we will point to sources where she may pursue the material to greater depth. Since in many places this material is presented elsewhere without the slightest motivation, reading the following sections will give a fine overview of the whole subject and should make other treatments much easier to follow. When we have introduced enough preliminary material and have returned once more to differential forms proper, our standards of completeness will again increase.

2.2 Permutations

A critical and not particularly well liked part of this is some of the lore of permutations. A permutation is a rearrangement of the set $\{1, 2, \ldots, n\}$; for example $\{3, 5, 6, 1, 2, 4\}$ is a rearrangement of $\{1, 2, 3, 4, 5, 6\}$. By interchanging two elements, adjacent or not, a permutation may be brought back to $\{1, 2, \ldots, n\}$. There are many ways to do this of course, but it turns out that the number of interchanges to do this is always the same mod 2. Thus no matter how the interchanges are done, $(-1)^s$ will always have the same value, and that value is the *sign* of the permutation. It is convenient for us (though somewhat nonstandard) to write a permutation σ as a function with inputs at the top and outputs at the bottom. Thus the permutation above would be written

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 6 & 1 & 2 & 4 \end{pmatrix}$$

[1] This is accomplished by the Whitney or Nash embedding theorems, depending on the circumstances

Here are a possible sequence of interchanges to get back to the identity.

$$\sigma \longrightarrow \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 4 & 1 & 2 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 4 & 1 & 5 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}$$

We used four interchanges; $s = 4$ (count the arrows). Clearly this can be done in many ways but the number of interchanges s will always be an even number, and thus

$$\operatorname{sgn}(\sigma) = (-1)^s = (-1)^4 = +1$$

is well defined.

There is one other useful observation. Suppose that a permutation has the property that it compounded of two *increasing* subsequences of the numbers $\{1, 2, \ldots, n\}$; for example

$$\left(\begin{array}{cccc|ccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 5 & 7 & 8 & 1 & 2 & 4 & 6 & 9 \end{array}\right)$$

or more generally with $i_1 < \ldots < i_r$ and $i_{r+1} < \ldots < i_n$

$$\sigma = \left(\begin{array}{ccc|ccc} 1 & \ldots & r & r+1 & \ldots & n \\ i_1 & \ldots & i_r & i_{r+1} & \ldots & i_n \end{array}\right)$$

Note that the elements after the vertical line, including their order, are determined by those before it. Now the interesting thing here is that the sign of such a permutation can be found as follows. First set

$$T_r = 1 + 2 + \cdots + r = \frac{r(r+1)}{2}$$

(T_r is the r^{th} triangular number.) Then for such a σ

$$\operatorname{sgn}(\sigma) = (-1)^{\sum_{j=1}^{r} i_j \, - T_r}$$

Thus for the example above

$$\operatorname{sgn}(\sigma) = (-1)^{3+5+7+8 \, - T_4} = (-1)^{23-10} = -1$$

Since this kind of permutation is the most common type in differential forms, this easy determination of the sign of σ is quite handy. We now justify the assertion. Suppose we are trying to find the sign of the permutation σ above. We abbreviate for convenience $\sigma = (3578\,|\,12469) \in \mathcal{S}_{9,4}$. Starting with $(1234\,|\,56789)$ the first step is to move 3 to the front, which takes 2=3-1 hops and results in $(3124\,|\,56789)$. Next 5 must be brought forward which requires 3=5-2 hops, giving $(3512\,|\,46789)$ Then 7 forward requiring 4=7-3 hops and giving $(3571\,|\,246789)$. Lastly 8 comes forward requiring 4= 8-4 hops and giving $\sigma(3578\,|\,124679)$. Thus the total number of hops is

$$3-1+5-2+7-3+9-4 = 3+5+7+8-(1+2+3+4) = 3+5+7+8-T_4 = 23-10 = 13$$

as we saw above. The general proof is exactly the same. To get into positon j, i_j must perform $i_j - j$ hops.

We will need, for several purposes, the *reverse* $\tilde{\sigma}$ of a permutation $\sigma \in \mathcal{S}_{n,r}$. If

$$\sigma = \left(\begin{array}{cccc|ccc} 1 & 2 & \dots & r & r+1 & r+2 & \dots & r+(n-r) \\ i_1 & i_2 & \dots & i_r & j_1 & j_2 & \dots & j_{n-r} \end{array} \right)$$

where $\{i_1, \dots, i_r, j_1, \dots j_{n-r}\} = \{1, 2, \dots, n\}$ and $i_1 < i_2 < \dots < i_r$ and $j_1 < j_2 < \dots < j_{n-r}$ then

$$\tilde{\sigma} = \left(\begin{array}{cccc|ccc} 1 & 2 & \dots & n-r & n-r+1 & n-r+2 & \dots & n-r+r \\ j_1 & j_2 & \dots & j_{n-r} & i_1 & i_2 & \dots & i_r \end{array} \right)$$

It is obvious that $\tilde{\sigma} \in \mathcal{S}_{n,n-r}$. Since $n-r$ j's have to hop over r i's to get from σ to $\tilde{\sigma}$, we have $\text{sgn}(\tilde{\sigma}) = (-1)^{r(n-r)}\text{sgn}(\sigma)$, or more symmetrically

$$\text{sgn}(\sigma) \cdot \text{sgn}(\tilde{\sigma}) = (-1)^{r(n-r)}$$

A harder way to get this result is to use the formula for $\text{sgn}(\sigma)$ given above.

2.3 Grassmann Algebra

2.3.1 General Theory of Grassmann Algebra

The reader has been exposed to a certain amount of Grassmann algebra in the first Chapter. Our purpose here is to present the basic ideas in the more abstract setting which we will need in this chapter. The ideas here, though powerful, are actually quite simple but the notation can be complicated and there is no easy way around this. Indeed the annoyning notations for Grassmann Algebra may be part of the reason for this subject not having the central position in mathematics that it deserves. We hope our use of increasing permutations to index elements in the Grassmann algebra may help to overcome this problem.

Given an n-dimensional vector space V we propose to build a building having a basement, the scalers, a first floor, the vector space V, and $n-1$ upper floors which we will construct. The j^{th} floor will be called $\Lambda^j(V)$, so that $\Lambda^0(V) = \mathbb{R}$ (because we are using $\mathbb{R}$ for the scalers in this book) and $\Lambda^1(V) = V$. The higher floors $\Lambda^j(V)$ we must now construct.

There are two basic laws for the Grassmann Algebra of a vector space. First it is an associative algebra and second it is anticommutative on vectors. An associative algebra is a vector space G over a field (here $\mathbb{R}$) with a multiplication satisfying, for all $A, B, C \in G$ and $\alpha, \beta \in \mathbb{R}$,

1a. $A(B+C) = AB + AC$

1b. $(B+C)A = BA + CA$

2. $\alpha(\beta A) = (\alpha\beta)A$

3. $\alpha(AB) = (\alpha A)B = A(\alpha B)$

4. $1A = A$

5. $A(BC) = (AB)C$

In our case the elements of G will be sums of products of elements of the vector space V and we require that

$$uv = -vu \qquad \text{for } u, v \in V$$

Often this kind of multiplication is written with a wedge $\wedge$ so one writes $u \wedge v$ instead of uv, as we did in Chapter 1. Since this section has a huge number of products it will be more convenient to omit the wedge. Just remember that we are using $uv = -vu$ instead of $u \wedge v = -v \wedge u$.

There are some important things that can be derived from the anticommutativity. first note that if we substitute $v = u$ into $uv = -vu$ we get $uu = -uu$ which gives $2uu = 0$. Thus

$$u^2 = uu = 0 \qquad \text{for any } u \in V$$

This is very important, but should not be taken to mean the product of any two elements of the Grassmann Algebra is 0. For example suppose $u, v, w, x \in V$ and they are a linearly independent set. Then if $A = uv + wx$ then

$$\begin{aligned} A^2 = (uv + wx)(uv + wx) &= uvuv + uvwx + wxuv + wxwx \\ &= -uuvv + uvwx - wuxv - wwxx \\ &= 0 + uvwx + uwxv + 0 \\ &= uvwx - uwvx = uvwx + uvwx = 2uvwx \neq 0 \end{aligned}$$

Hence one cannot depend on $A^2 = 0$ unless $A \in \Lambda^1(V)$.

The $A, B, C \in G$ above will refer to sums of products of vectors. Typical examples would be, with $u, v, w, x \in V$,

$$\begin{aligned} uv + 2wx &\in \Lambda^2(V) \\ 3uvw - 5wxy &\in \Lambda^3(V) \\ 4uvxy - 7xywu &\in \Lambda^4(V) \\ 2u + 3w &\in \Lambda^1(V) \\ 2 &\in \Lambda^0(V) \end{aligned}$$

We will now replace the letter G for the algebra by the symbol

$$\Lambda(V) = \bigoplus_{j=0}^{n} \Lambda^j(V)$$

(We will explain later why we only got up to n in the direct sum) Thus all the items above are in $\Lambda(V)$. By use of the various laws an element in $\Lambda(V)$ can be written in many ways. Also remember that $\mathbb{R} = \Lambda^0(V) \subseteq \Lambda(V)$. I want to also

mention that expressions like $4 + u + 2uv - 3uvw$ are perfectly legitimate since $\Lambda(V)$ is an algebra. However, in the course of daily life such expressions come up exceedingly rarely.

We can extract a little more information about commutativity at this point. Let $A \in \Lambda^j(V)$ and $B \in \Lambda^k(V)$. Then A is a sum of products of j vectors and B is a sum of products of k vectors. In moving vectors around from AB to BA, the k vectors of each term of B will have to hop over the j vectors of each term of A. Hence when a vector from a term of B moves to the front of A it has undergone j hops. Since each of the k vectors of a term of B has to hop over j vectors of the terms of A, the total number of hops is jk and since each hop introduces a sign shift, we have

If $A \in \Lambda^j(V)$ and $B \in \Lambda^k(V)$ then

$$BA = (-1)^{jk} AB$$

If either j or k is even we have $BA = AB$. On the other hand if both j and k are odd, then $BA = -AB$. From this we notice that if $A \in \Lambda^j(V)$ and j is odd then putting $A = B$ and $j = k$ in the above equation gives

$$\begin{aligned} AA &= (-1)^{jj} AA \\ A^2 &= -A^2 \\ A^2 &= 0 \end{aligned}$$

The next topic is when the products are 0. Recall that $v^2 = 0$ for $v \in V$. Suppose now that $v_4 = 2v_1 - 3v_2 + 7v_3$; that is the set $\{v_1, v_2, v_3, v_4\}$ is linearly dependant. What is $v_1v_2v_3v_4$? It is

$$\begin{aligned} v_1v_2v_3v_4 &= 2v_1v_2v_3v_1 - 3v_1v_2v_3v_2 + 7v_1v_2v_3v_3 \\ &= 2v_1v_1v_2v_3v_1 + 3v_1v_2v_2v_3 + 7v_1v_2v_3v_3 \ = \ 0 \end{aligned}$$

because each term has a repeated vector, which gives 0. (It is necessary to move the vectors around to get the form vv, but that only changes the sign.) In a similar way we can see that

If the set of vectors $\{v_1, v_2 \ldots, v_m\}$ is linearly dependent then

$$v_1v_2 \ldots v_m = 0$$

On the other hand, if $\{v_1, v_2 \ldots, v_n\}$ is a linearly independent set of vectors (which makes it a basis for V since V has dimension n.) then there is no reason $v_1v_2 \ldots v_n$ should be 0, and we don't want it to be. In a sophisticated treatment of Grassmann Algebra this is accomplished by the concept of freeness, but that is a little heavy for our treatment, and we can accomplish the same thing by

Axiom If $\{v_1, v_2 \ldots, v_n\}$ is a linearly independent set of vectors in an n-dimensional space then

$$v_1v_2 \ldots v_n \neq 0$$

Actually we could instead use $v_1v_2\dots v_n \neq 0$ for a single set of linearly independent vectors (a basis) and then prove that the product of *any* set $\{w_1, v_2\dots, w_n\}$ of linearly independent vectors is non-zero, but this is a very small gain.

From this we see that for any $m \leq n$ if $\{v_1, v_2\dots, v_m\}$ is a linearly independent set of vectors then $v_1v_2\dots v_m \neq 0$, for if it were 0 we could complete $v_1v_2\dots v_m$ to a linearly independent set $v_1v_2\dots v_n$ and we would have

$$v_1v_2\dots v_n = (v_1v_2\dots v_m)v_{m+1}\dots v_n = 0\cdot v_{m+1}\dots v_n = 0$$

contradicting the axiom.

Next we consider a set $\{v_1, v_2\dots, v_m\}$ of vectors where $m > n = \dim(V)$. Recall from linear algebra that in this case the set $\{v_1, v_2\dots, v_m\}$ *must* be linearly dependent, because the dimension of V is the size of a maximal set of linearly independent vectors. Therefore

$$v_1\cdots v_nv_{m+1}\dots v_m = 0$$

because of the linear dependence and we have shown

$$\text{For } k > n = \dim(V) \text{ we have } \Lambda^k(V) = 0$$

Thus the building that is the Grassmann Algebra has the roof over the n-th floor.

This is pretty much the basic story for the Grassmann Algebra, but we must now go about getting additional information for practical use. The first thing we will deal with is the connection with determinants. First we recall the definition of a determinant. The first thing we recall is the Leibniz (or Laplace) definition of a determinant, which is

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{vmatrix} = \sum_\sigma \text{sgn}(\sigma)\ a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}$$

Here the sum is over the $n!$ permutations of $\{1, 2, \dots, n\}$ and $\text{sgn}(\sigma)$ is the sign of the permutation sigma. For example (I move the indices just for variety) let $n = 2$ so there are 2 permuations, the identity ι with sign +1 and the swap σ with sign -1. Then the formula gives]

$$\begin{aligned} \begin{vmatrix} \alpha^1_1 & \alpha^1_2 \\ \alpha^2_1 & \alpha^2_2 \end{vmatrix} &= \text{sgn}(\iota)\ \alpha^1_{\iota(1)}\alpha^2_{\iota(2)} + \text{sgn}(\sigma)\ \alpha^1_{\sigma(1)}\alpha^2_{\sigma(2)} \\ &= (+1)\alpha^1_1\alpha^2_2 + (-1)\alpha^1_2\alpha^2_1 = \alpha^1_1\alpha^2_2 - \alpha^1_2\alpha^2_1 \end{aligned}$$

which we recognize as the standard formula for a 2 by 2 determinant. 3 by 3 and higher all work exactly the same, but have larger numbers ($n!$) of terms.

For the sake of simplicity, let us now take $\dim(V)$=2 and look at what happens when we multiply $A = 2u + 3v$ and $B = 4u + 5v$. We get

$$\begin{aligned} AB &= (2u + 3v)(4u + 5v) = 2\cdot 5\ uv + 3\cdot 4vu = (2\cdot 5 - 3\cdot 4)uv \\ &= \begin{vmatrix} 2 & 4 \\ 3 & 5 \end{vmatrix} uv \end{aligned}$$

Now if you examine this and the formula for the determinant you will find the minus sign is due to exactly the same reason; the permutation σ corresponds to the swapping $vu = -uv$. This is no accident. Determinants are set up to be exactly what is necessary to get the coefficients in the Grassmann Algebra, and it is very likely that determinants gain most of their importance from this use. We go to great lengths to conceal the Grassmann Algebra taking place all around mathematics, and we dance over this large hole in our technique using the flimsy bridge of determinants.

We will now give a second example similar to the last but a little bit more general. Let $\dim(V) \geq 3$ and u, v, w be linearly independent; let A, B, C be

$$A = 2u + v - 3w \qquad B = 4u - v + 2w \qquad C = u + 3v + 5w$$

We'll compute AB and ABC and compare with determinants. The numbers for A, B, C are written as columns here to be consistent with later usage but they could just as well be written as rows. Then (we leave out terms like uu which are all 0)

$$\begin{aligned} AB &= (2u + v - 3w)(4u - v + 2w) = -2uv + 4uw + 4vu + 2uw - 12wu + 3wv \\ &= (-2 - 4)uv + (4 + 12)uw + (-2 - 3)vw = -6uv + 16uw - vw \end{aligned}$$

Notice that I have chosen alphabetical rather than cyclic order for u, v, w. This corresponds, in ways we will elucidate later, with the increasing permutations of the permutations section. We used cyclic order in the first chapter for practical purposes but for an n-dimensional theory it is no longer feasible and we must give it up.

Now we have a look at the matrix that goes with A and B. Remember we are using *columns* for the numbers.

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

notice that u has its coefficients in the first row, v has its coefficients in the second row, and w in the third. With this in mind let us form, using the minors of the matrix

$$\begin{vmatrix} 2 & 4 \\ 1 & -1 \end{vmatrix} uv + \begin{vmatrix} 2 & 4 \\ -3 & 2 \end{vmatrix} uw + \begin{vmatrix} 1 & -1 \\ -3 & 2 \end{vmatrix} vw = -6uv + 16uw - vw$$

Now we compute ABC. We could do this directly from the matrix

$$\begin{vmatrix} 2 & 4 & 1 \\ 1 & -1 & 3 \\ -3 & 2 & 5 \end{vmatrix}$$

but it is more illumating to use the already calculated AB. We then have

$$ABC = (AB)C = \left(\begin{vmatrix} 2 & 4 \\ 1 & -1 \end{vmatrix} uv + \begin{vmatrix} 2 & 4 \\ -3 & 2 \end{vmatrix} uw + \begin{vmatrix} 1 & -1 \\ -3 & 2 \end{vmatrix} vw \right) C$$

$$\begin{aligned}
&= \left(\begin{vmatrix} 2 & 4 \\ 1 & -1 \end{vmatrix} uv + \begin{vmatrix} 2 & 4 \\ -3 & 2 \end{vmatrix} uw + \begin{vmatrix} 1 & -1 \\ -3 & 2 \end{vmatrix} vw \right) (u + 3v + 5w) \\
&= \left(\begin{vmatrix} 2 & 4 \\ 1 & -1 \end{vmatrix} 5 + \begin{vmatrix} 2 & 4 \\ -3 & 2 \end{vmatrix} (-3) + \begin{vmatrix} 1 & -1 \\ -3 & 2 \end{vmatrix} 1 \right) uvw \\
&= ((-6)5 + (16)(-3) + (-1)1)uvw = (-79)uvw
\end{aligned}$$

The most interesting thing here is that we have duplicated the expansion of the 3x3 determinant by the last column, although the order of the terms is from bottom to top. Notice that in the middle term we get uwv when we multiply out and this is the origin of the minus sign in (-3). Notice how ultimately restructuring each term to uvw puts in the mysterious signs which show up in the expansion of the determinant. This is perfectly general.

Also note the above calculations will work in a space of any dimensions as long as u, v, w are padded out to make a linerarly independent set in V and the unused rows are padded with 0's. Naturally if $\dim(V) > 3$ we cannot use a square determinant directly to get the final answer, but we *can* use the determinant of a square minor, which avoids the 0's.

The assumption that $u, v, w, \ldots$ are linearly independent is totally unnecessary for all this to work, but if they are not linearly independent then many of the determinants come out 0. The methodology remains the same but the example would have been less impressive.

Our next job is to find a basis for the Grassmann Algebra $\Lambda(V)$ of V. This is done using the increasing permuatations from the section on permutations. We will then find the dimensions of each $\Lambda^j(V)$ and finally the dimension of the entire Grassmann Algebra $\Lambda(V)$. The basement and first floor are easily done; a basis for $\Lambda^0(V) = \mathbb{R}$ is 1 and the $\dim(\Lambda^0(V)) = 1$. A basis for $\Lambda^1(V) = V$ is $(e_1, e_2, \ldots, e^n)$ and $\dim(\Lambda^1(V)) = n$.

For the higher floors let's look at some examples first.

SKIPPED MATERIAL

Now we must set up bases for the $\Lambda^k(V)$. To do this we recall a few things from the Permutation section. Recall that a permutation is in $\mathcal{S}_{n,k}$ of it has the form

$$\sigma = \left(\begin{array}{ccc|ccc} 1 & \ldots & k & k+1 & \ldots & n \\ i_1 & \ldots & i_k & i_{k+1} & \ldots & i_n \end{array} \right)$$

where $i_1 < i_2 < \ldots < i_k$ and $i_{k+1} < \ldots < i_n$ and $\{i_1, i_2, \ldots, i_k, i_{k+1}, \ldots, i_n\} = \{1, 2, \ldots, n\}$. For example

$$\sigma = \left(\begin{array}{cccc|ccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 5 & 7 & 1 & 4 & 6 \end{array} \right) \in \mathcal{S}_{7,4}$$

Also recall that the sign of an increasing permutation $\sigma \in \mathcal{S}_{n,k}$ is found easily, using $T_k = \sum_{i=1}^{k} i$, by

$$\operatorname{sgn}(\sigma) = (-1)^{\sum_{i=1}^{k} \sigma(i) - T_k}$$

so that for the $\sigma \in \mathcal{S}_{7,4}$ given above we have

$$\text{sgn}(\sigma) \ = (-1)^{2+3+5+7-T_4} = (-1)^{17-10} = (-1)^7 = -1$$

With this equipment we can now instantly construct a basis for $\Lambda^k(V)$ by using, for $\sigma \in \mathcal{S}_{n,k}$

$$e_\sigma = \text{sgn}(\sigma)\ e_{\sigma(1)} e_{\sigma(2)} \cdots e_{\sigma(k)}$$

The coefficient $\text{sgn}(\sigma)$ is an innovation that will simplify life when we get to the $*$-operator in a later section. For an example we use the σ above to get

$$e_\sigma = \text{sgn}(\sigma)\ e_2 e_3 e_5 e_7 = -e_2 e_3 e_5 e_7$$

It should be clear that all possible products of k distinct basis vectors can be rewritten as an e_σ up to sign, and that if $\sigma, \tau \in \mathcal{S}_{n,k}$ then e_σ and e_τ are distinct for $\sigma \neq \tau$. Since no linear combination of e_σ's can equal an e_τ in a non-trivial way, the $e_\sigma,\ \sigma \in \mathcal{S}_{n,k}$ form a basis for $\Lambda^k(V)$ (where $\dim(V)$=n). Since we have a basis, we can find the dimension by counting the basis. To form a basis element we select k elements from a supply of n elements, and thus get $\binom{n}{k}$ possible basis elements. Hence we have

$$\dim(\Lambda^k(V)) = \binom{n}{k}$$

Since the various levels of the Grassmann Algebra are linearly independent of one another, we have

$$\dim(\Lambda(V)) = \sum_{k=0}^{n} \dim(\Lambda^k(V)) = \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

The last equality in the chain comes from using the binomial theorem

$$2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} 1^k = \sum_{k=0}^{n} \binom{n}{k}$$

Now some examples. For $\dim(V)$=3 and k=2 the three permutations of $\mathcal{S}_{3,2}$ are

$$\sigma_1 = \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 1 & 2 & 3 \end{array}\right) \qquad \sigma_2 = \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 1 & 3 & 2 \end{array}\right) \qquad \sigma_3 = \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 3 & 1 \end{array}\right)$$

with the signs

$$\text{sgn}(\sigma_1) \ = +1 \qquad \text{sgn}(\sigma_2) \ = -1 \qquad \text{sgn}(\sigma_3) \ = +1$$

and so

$$e_{\sigma_1} = e_1 e_2 \qquad e_{\sigma_2} = -e_1 e_3 \qquad e_{\sigma_3} = e_2 e_3$$

which we notice is cyclic order.

For dim(V)=4 and k=2 the six permutations of $\mathcal{S}_{4,2}$ are

$$\sigma_1 = \left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{array}\right) \quad \sigma_2 = \left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{array}\right) \quad \sigma_3 = \left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{array}\right)$$
$$\sigma_4 = \left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{array}\right) \quad \sigma_5 = \left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{array}\right) \quad \sigma_6 = \left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{array}\right)$$

with signs (remember the trick; $T_2 = 3$)

$$\begin{array}{lll} \mathrm{sgn}(\sigma_1) = +1 & \mathrm{sgn}(\sigma_2) = -1 & \mathrm{sgn}(\sigma_3) = +1 \\ \mathrm{sgn}(\sigma_4) = +1 & \mathrm{sgn}(\sigma_5) = -1 & \mathrm{sgn}(\sigma_6) = +1 \end{array}$$

and basis elements for $\Lambda^2(V)$

$$\begin{array}{lll} e_{\sigma_1} = e_1 e_2 & e_{\sigma_2} = -e_1 e_3 & e_{\sigma_3} = e_1 e_4 \\ e_{\sigma_4} = e_2 e_3 & e_{\sigma_5} = -e_2 e_4 & e_{\sigma_6} = e_3 e_4 \end{array}$$

Finally, we do a multiplication example

$$(e_2 + 3e_3 - 3e_4)(e_1 - 2e_2 + e_3 + e_4) = -e_1e_2 - 3e_1e_3 + 3e_1e_4 + 7e_2e_3 - 5e_2e_4 + 6e_3e_4$$

which we can do by hand or using the minors of

$$\left(\begin{array}{cc} 0 & 1 \\ 1 & -2 \\ 3 & 1 \\ -3 & 1 \end{array}\right)$$

When written in terms of the basis elements there are some sign shifts.

$$(e_2 + 3e_3 - 3e_4)(e_1 - 2e_2 + e_3 + e_4) = -e_{\sigma_1} + 3e_{\sigma_2} + 3e_{\sigma_3} + 7e_{\sigma_4} + 5e_{\sigma_5} + 6e_{\sigma_6}$$

Note the sign changes in the second and fifth term. These are *related* to the sign changes associated to cofactors in the Laplace expansion of a determinant by complementary minors. Since there is no determinant here there are no cofactor signs and so the relationship is not so direct and a little complicated. With a little thought, though, you can figure it out. Hint: sums of the row indices, and T_2.

There is an interesting relationship between e_σ and $e_{\tilde{\sigma}}$ which we will now briefly explore. If $\sigma \in \mathcal{S}_{n,r}$ and $\tau \in \mathcal{S}_{n,n-r}$ then ${}_\sigma e_\tau$ will usually be $0 \in \Lambda^n$. The only way to avoid the 0 is for the firs $n - r$ items in τ to be exactly that last $n - r$ items in σ. If this is the case then $\tau = \tilde{\sigma}$, and then

$$\begin{aligned} e_\sigma\, e_{\tilde{\sigma}} &= \mathrm{sgn}(\sigma)\, \mathrm{sgn}(\tilde{\sigma})\, e_{\sigma(1)} \cdots e_{\sigma(r)} \cdot e_{\tilde{\sigma}(1)} \cdots e_{\tilde{\sigma}(n-r)} \\ &= \mathrm{sgn}(\sigma)\, \mathrm{sgn}(\tilde{\sigma})\, \mathrm{sgn}(\sigma)\, \mathrm{sgn}(\tilde{\sigma})\, e_1 \cdots e_n \\ &= e_1 \cdots e_n \end{aligned}$$

and thus in all cases

$$e_\sigma\, e_\tau = \delta_{\tilde{\sigma}\tau}\, e_1 \cdots e_n$$

2.3.2 Determinants

What we have done above indicates that there are very close connections between determinants and Grassmann Algebra. Many complex determinant formulas can be proved relatively easily with Grassmann Algebra methods, for example the Cauchy-Binet theorem. We will give some examples.

Let $A : V \to V$ be a linear transformation and $(e_1, \ldots, e_n)$ be a basis for V. We want to compute $(Ae_1) \cdots (Ae_n)$ where we are going to use the matrix of A determined by

$$Ae_i = \alpha_i^j e_j \qquad \text{Summation convention in effect}$$

We compute as follows

$$\begin{aligned}(Ae_1)(Ae_2)\cdots(Ae_n) &= (\alpha_1^{j_1} e_{j_1})(\alpha_2^{j_2} e_{j_2})\cdots(\alpha_n^{j_n} e_{j_n}) \\ &= \sum_\sigma \alpha_1^{\sigma(1)} e_{\sigma(1)} \alpha_2^{\sigma(2)} e_{\sigma(2)} \cdots \alpha_n^{\sigma(n)} e_{\sigma(n)}\end{aligned}$$

because any terms in the first equation with repeated vectors are 0 and if there are no repeated vectors then we must be dealing with a permutation. Continuing

$$\begin{aligned}(Ae_1)(Ae_2)\cdots(Ae_n) &= \sum_\sigma \alpha_1^{\sigma(1)} \alpha_2^{\sigma(2)} \cdots \alpha_n^{\sigma(n)} e_{\sigma(1)} e_{\sigma(2)} \cdots e_{\sigma(n)} \\ &= \Big(\sum_\sigma \operatorname{sgn}(\sigma)\ \alpha_1^{\sigma(1)} \alpha_2^{\sigma(2)} \cdots \alpha_n^{\sigma(n)} \Big) e_1 e_2 \cdots e_n \\ &= \det(\alpha_i^j)\, e_1 e_2 \cdots e_n\end{aligned}$$

This is an important formula. As a first application we note that $\{Ae_1, \ldots, Ae_n\}$ is a linearly independent set if and only if $\det(\alpha_i^j) \neq 0$. Note that if $\det(\alpha_i^j) = 0$ then $\{Ae_1, \ldots, Ae_n\}$ is a linearly dependent set and there exist $\beta^1, \ldots, \beta^n$ not all 0 for which $\beta^i Ae_i = 0$. But then $A(\beta^i e_i) = 0$ and so the nullspace of A is nontrivial.

Our next job is to show that $\det(\alpha_i^j)$ does not depend upon the basis. For this let $\{e_1, \ldots, e_n\}$ and $\{v_1, \ldots, v_n\}$ be basis of V and let

$$Ae_i = \alpha_i^j e_i \qquad Av_i = \beta_i^j v_i \qquad v_i = \gamma_i^j e_i$$

We have

$$\begin{aligned}(Ae_1)(Ae_2)\cdots(Ae_n) &= \det(\alpha_i^j)\, e_1 e_2 \cdots e_n \\ (Av_1)(Av_2)\cdots(Av_n) &= \det(\beta_i^j)\, v_1 v_2 \cdots v_n \\ v_1 v_2 \cdots v_n &= \det(\gamma_i^j)\, e_1 e_2 \cdots e_n\end{aligned}$$

From the last two we have

$$\begin{aligned}(Av_1)(Av_2)\cdots(Av_n) &= \det(\beta_i^j)\, v_1 v_2 \cdots v_n \\ &= \det(\beta_i^j) \det(\gamma_i^j)\, e_1 e_2 \cdots e_n\end{aligned}$$

But we can also calculate

$$\begin{aligned}(Av_1)(Av_2)\cdots(Av_n) &= A(\gamma_1^{j_1}e_{j_1})A(\gamma_2^{j_2}e_{j_2})\cdots A(\gamma_n^{j_n}e_{j_n})\\ &= \gamma_1^{j_1}e_{j_1}\gamma_2^{j_2}e_{j_2}\cdots\gamma_n^{j_n}A(e_{j_1})A(e_{j_1})\cdots A(e_{j_1})\\ &= \sum_{\sigma\in\mathcal{S}_n}\gamma_1^{\sigma(1)}\gamma_2^{\sigma(2)}\cdots\gamma_n^{\sigma(n)}Ae_{\sigma(1)}Ae_{\sigma(2)}\cdots Ae_{\sigma(n)}\end{aligned}$$

because repeated e_{j_i} will result in 0 terms, and the permutations keep all the items in the product distinct. Continuing

$$\begin{aligned}(Av_1)(Av_2)\cdots(Av_n) &= \sum_{\sigma\in\mathcal{S}_n}\operatorname{sgn}(\sigma)\,\gamma_1^{\sigma(1)}\gamma_2^{\sigma(2)}\cdots\gamma_n^{\sigma(n)}Ae_1\,Ae_2\cdots Ae_n\\ &= \det(\gamma_i^j)Ae_1\,Ae_2\cdots Ae_n\\ &= \det(\gamma_i^j)\det(\alpha_i^j)\,e_1\,e_2\cdots e_n\end{aligned}$$

Comparing our two expressions for $Ae_1\,Ae_2\cdots Ae_n$ we see that

$$\begin{aligned}\det(\beta_i^j)\det(\gamma_i^j)\,e_1e_2\cdots e_n &= \det(\gamma_i^j)\det(\alpha_i^j)\,e_1\,e_2\cdots e_n\\ \det(\beta_i^j)\det(\gamma_i^j) &= \det(\gamma_i^j)\det(\alpha_i^j)\\ \det(\beta_i^j) &= \det(\alpha_i^j)\end{aligned}$$

and so the multiplier doesn't depend on which basis is used. Thus we can make the definition

Def Let A be a linear transformation and let $\{v_1,\ldots,v_n\}$ be any basis. Let $Av_i=\beta_i^jv_i$ and define $\det(A)=\det(\beta_i^j)$.

The above calculations show that $\det(A)$ is well defined, and for any basis $\{v_1,\ldots,v_n\}$ we have

$$(Av_1)(Av_2)\ldots(Av_n)=\det(A)\,v_1v_2\cdots v_n$$

Next we will prove an important determinant theorem. Let $Ae_i=\alpha_i^je_j$ (where $(e_1,\ldots,e_n)$ is a basis of V) and assume $\det(\alpha_i^j)\neq 0$. Set $v_i=Ae_i$. Thus by what we just proved $v_1\cdots v_n=\det(\alpha_i^j)\,e_1\cdots e_n=\det(A)\,e_1\cdots e_n$ and $\{v_1,\ldots,v_n\}$ is a basis for V. Now let $Bv_j=\beta_j^kv_k$, which is possible since the v_i form a basis. Then we have

$$\begin{aligned}(Ae_1)(Ae_2)\ldots(Ae_n) &= \det(A)\,e_1e_2\cdots e_n\\ (Bv_1)(Bv_2)\ldots(Bv_n) &= \det(B)\,v_1v_2\cdots v_n \;=\; \det(B)\det(A)\,e_1e_2\cdots e_n\end{aligned}$$

On the other hand we have

$$\begin{aligned}(Bv_1)(Bv_2)\ldots(Bv_n) &= (BAe_1)(BAe_2)\ldots(BAe_n)\\ &= \det(BA)\,e_1e_2\cdots e_n\end{aligned}$$

Comparing the two expressions for $(Bv_1)(Bv_2)\dots(Bv_n)$ we have

$$\begin{aligned}\det(B)\det(A)\, e_1e_2\cdots e_n &= \det(BA)\, e_1e_2\cdots e_n\\ \det(B)\det(A) &= \det(BA)\end{aligned}$$

The fact that we could prove this theorem in a systematic way using Grassmann Algebra again shows the close connection between determinants and Grassmann Algebra and the efficacy of Grassmann Algebra in proving theorems about determinants. Let us point out that the tricky proof above that one can use any basis to get $\det(A)$ is much easier if you already have the theorem $\det(BA) = \det(B)\det(A)$ but this is the theorem that we wanted to *prove* using Grassmann methods.

Our next expedition is the proof of Laplace expansion by complementary minors. This is a slightly cumbersome proof and you may want to skip to the next section if you already know the theorem, or are willing to take it on faith. We present the proof as an example of how well Grassmann methods work.

The Laplace expansion by complementary minors is nothing more than the equation

$$v_1\cdots v_n = (v_1\cdots v_k)\cdot(v_{k+1}\cdots v_n)$$

When this is worked out in coordinates it gives a special case of the Laplace expansion from which the general case is easily derived.

First it might be helpful to state the theorem in its classical form. From the list $\{1,2,\dots,n\}$ select k integers $I = (i_1 < i_2 < \dots < i_k)$ which we will think of as indices of *columns*. We now select all possible sequences $J = (j_1 < j_2 < \dots < j_k)$ and also form the complementary sets $i_{k+1} < i_{k+2} < \dots < i_n$ and for each selection of J its complementary set $j_{k+1} < j_{k+2} < \dots < j_n$. Then (remember I is fixed)

$$\det(\alpha_i^j) = \sum_J (-1)^{i_1+\dots+i_k+j_1+\dots+j_k} \begin{vmatrix} \alpha_{i_1}^{j_1} & \dots & \alpha_{i_k}^{j_1}\\ \vdots & \vdots & \vdots\\ \alpha_{i_1}^{j_k} & \dots & \alpha_{i_k}^{j_k}\end{vmatrix} \begin{vmatrix} \alpha_{i_{k+1}}^{j_{k+1}} & \dots & \alpha_{i_n}^{j_{k+1}}\\ \vdots & \vdots & \vdots\\ \alpha_{i_{k+1}}^{j_n} & \dots & \alpha_{i_n}^{j_n}\end{vmatrix}$$

The formula works just as well if we select a fixed set of rows and sum over the columns. This is due to $\det(A) = \det(A^\top)$. Laplace is generally credited with the discovery of this formula. It is theoretically valuable but computationally inefficient. The exponent on (-1) we will see to be a form of signs of permutations.

To prove this, let $v_i = \alpha_i^j e_i$ where $e_1,\dots,e_n$ is a basis for V. We know

$$v_1\cdots v_n = \det(\alpha_i^j)e_1\cdots e_n$$

We are now going to compute $(v_1\cdots v_k)\cdot(v_{k+1}\cdots v_n)$.

$$\begin{aligned}v_1v_2\cdots v_k &= (\alpha_1^{j_1}e_{j_1})(\alpha_2^{j_2}e_{j_2})\cdots(\alpha_k^{j_k}e_{j_k})\\ &= \alpha_1^{j_1}\alpha_2^{j_2}\cdots\alpha_k^{j_k}\; e_{j_1}e_{j_2}\cdots e_{j_k}\end{aligned}$$

For $e_{j_1}e_{j_2}\cdots e_{j_k} \neq 0$ it must contain no repetitions. Hence the $j_1, j_2, \ldots, j_k$ must be some permutation of $\sigma(1), \sigma(2), \ldots, \sigma(k)$ for some permutation $\sigma \in \mathcal{S}_{n,k}$, and *all* the permutations of $\mathcal{S}_{n,k}$ will occur in the sum. We can describe the permutations of $\sigma(1), \sigma(2), \ldots, \sigma(k)$ by $\sigma(\tau(1)), \sigma(\tau(2)), \ldots, \sigma(\tau(k))$ where $\tau \in \mathcal{S}_k$ and *all* the permuations of $\mathcal{S}_k$ will occur in the sum. Thus

$$v_1 \cdots v_k = \sum_{\sigma\in\mathcal{S}_{n,k}} \sum_{\tau\in\mathcal{S}_k} \alpha_1^{\sigma(\tau(1))} \cdots \alpha_k^{\sigma(\tau(k))} e_{\sigma(1)} \cdots e_{\sigma(k)}$$

Now the inner sum is simply the Leibniz-Laplace expression for a determinant, so this can be rewritten as

$$v_1 \cdots v_k = \sum_{\sigma\in\mathcal{S}_{n,k}} \begin{vmatrix} \alpha_1^{\sigma(1)} & \ldots & \alpha_k^{\sigma(1)} \\ \vdots & \vdots & \vdots \\ \alpha_1^{\sigma(k)} & \ldots & \alpha_k^{\sigma(k)} \end{vmatrix} e_{\sigma(1)} \cdots e_{\sigma(k)}$$

Exactly the same reasoning gives us

$$v_{k+1} \cdots v_n = \sum_{\tau\in\mathcal{S}_{n,n-k}} \begin{vmatrix} \alpha_{k+1}^{\tau(1)} & \ldots & \alpha_n^{\tau(1)} \\ \vdots & \vdots & \vdots \\ \alpha_{k+1}^{\tau(n-k)} & \ldots & \alpha_n^{\tau(n-k)} \end{vmatrix} e_{\tau(1)} \cdots e_{\tau(n-k)}$$

When we multiply these two expressions together to get $(v_1 \cdots v_k) \cdot (v_{k+1} \cdots v_n)$ we will get many terms of the form $e_{\sigma(1)} \cdots e_{\sigma(k)} e_{\tau(1)} \cdots e_{\tau(k)}$ with $\sigma \in \mathcal{S}_{n,k}$ and $\tau \in \mathcal{S}_{n,n-k}$. But we saw above that such terms can only be non-zero when $\tau = \tilde{\sigma}$, in which case they are

$$e_{\sigma(1)} \cdots e_{\sigma(k)} e_{\tilde{\sigma}(1)} \cdots e_{\tilde{\sigma}(n-k)} = e_{\sigma(1)} \cdots e_{\sigma(k)} e_{\sigma(k+1)} \cdots e_{\sigma(n)}$$

and thus we have

$$\begin{aligned}
&v_1 \cdots v_k v_{k+1} \cdots v_n \\
&= \sum_{\sigma\in\mathcal{S}_{n,k}} \begin{vmatrix} \alpha_1^{\sigma(1)} & \ldots & \alpha_k^{\sigma(1)} \\ \vdots & \vdots & \vdots \\ \alpha_1^{\sigma(k)} & \ldots & \alpha_k^{\sigma(k)} \end{vmatrix} \begin{vmatrix} \alpha_{k+1}^{\tilde{\sigma}(1)} & \ldots & \alpha_n^{\tilde{\sigma}(1)} \\ \vdots & \vdots & \vdots \\ \alpha_{k+1}^{\tilde{\sigma}(n-k)} & \ldots & \alpha_n^{\tilde{\sigma}(n-k)} \end{vmatrix} e_{\sigma(1)} \cdots e_{\sigma(k)} e_{\tilde{\sigma}(1)} \cdots e_{\tilde{\sigma}(n-k)} \\
&= \sum_{\sigma\in\mathcal{S}_{n,k}} \begin{vmatrix} \alpha_1^{\sigma(1)} & \ldots & \alpha_k^{\sigma(1)} \\ \vdots & \vdots & \vdots \\ \alpha_1^{\sigma(k)} & \ldots & \alpha_k^{\sigma(k)} \end{vmatrix} \begin{vmatrix} \alpha_{k+1}^{\tilde{\sigma}(1)} & \ldots & \alpha_n^{\tilde{\sigma}(1)} \\ \vdots & \vdots & \vdots \\ \alpha_{k+1}^{\tilde{\sigma}(n-k)} & \ldots & \alpha_n^{\tilde{\sigma}(n-k)} \end{vmatrix} e_{\sigma(1)} \cdots e_{\sigma(n)} \\
&= \sum_{\sigma\in\mathcal{S}_{n,k}} \operatorname{sgn}(\sigma) \begin{vmatrix} \alpha_1^{\sigma(1)} & \ldots & \alpha_k^{\sigma(1)} \\ \vdots & \vdots & \vdots \\ \alpha_1^{\sigma(k)} & \ldots & \alpha_k^{\sigma(k)} \end{vmatrix} \begin{vmatrix} \alpha_{k+1}^{\tilde{\sigma}(1)} & \ldots & \alpha_n^{\tilde{\sigma}(1)} \\ \vdots & \vdots & \vdots \\ \alpha_{k+1}^{\tilde{\sigma}(n-k)} & \ldots & \alpha_n^{\tilde{\sigma}(n-k)} \end{vmatrix} e_1 \cdots e_n
\end{aligned}$$

Comparing this with $v_1 \cdots v_k v_{k+1} \cdots v_n = v_1 \cdots v_n = \det(\alpha_i^j)\, e_1 \cdots e_n$ we have

$$\det(\alpha_i^j) = \sum_{\sigma \in \mathcal{S}_{n,k}} \operatorname{sgn}(\sigma) \begin{vmatrix} \alpha_1^{\sigma(1)} & \dots & \alpha_k^{\sigma(1)} \\ \vdots & \vdots & \vdots \\ \alpha_1^{\sigma(k)} & \dots & \alpha_k^{\sigma(k)} \end{vmatrix} \begin{vmatrix} \alpha_{k+1}^{\tilde{\sigma}(1)} & \dots & \alpha_n^{\tilde{\sigma}(1)} \\ \vdots & \vdots & \vdots \\ \alpha_{k+1}^{\tilde{\sigma}(n-k)} & \dots & \alpha_n^{\tilde{\sigma}(n-k)} \end{vmatrix}$$

which essentially is the result.

To get the full classical form it is only necessary allow the choice of any k columns rather than fixing on $1, 2, \dots, k$ as we did above. To select k columns, select $\tau \in \mathcal{S}_{n,k}$. Move these k columns to the first k which will require $\sum_{i=1}^{k}(\tau(i) - i)$ hops of columns, just as in the proof of the formula for $\operatorname{sgn}(\tau)$. Thus we have the new determinant

$$\begin{vmatrix} \alpha_{\tau(1)}^1 & \dots & \alpha_{\tau(k)}^1 & \alpha_{\tau(k+1)}^1 & \dots & \alpha_{\tau(n)}^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{\tau(n)}^1 & \dots & \alpha_{\tau(k)}^n & \alpha_{\tau(k+1)}^n & \dots & \alpha_{\tau(n)}^n \end{vmatrix} = \operatorname{sgn}(\tau) \begin{vmatrix} \alpha_1^1 & \dots & \alpha_n^1 \\ \vdots & \vdots & \vdots \\ \alpha_1^n & \dots & \alpha_n^n \end{vmatrix}$$

Applying the previously proved special case to the new determinant we have

$$\det(\alpha_i^j) = \sum_{\sigma \in \mathcal{S}_{n,k}} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \begin{vmatrix} \alpha_{\tau(1)}^{\sigma(1)} & \dots & \alpha_{\tau(k)}^{\sigma(1)} \\ \vdots & \vdots & \vdots \\ \alpha_{\tau(1)}^{\sigma(k)} & \dots & \alpha_{\tau(k)}^{\sigma(k)} \end{vmatrix} \begin{vmatrix} \alpha_{\tau(k+1)}^{\tilde{\sigma}(1)} & \dots & \alpha_{\tau(n)}^{\tilde{\sigma}(1)} \\ \vdots & \vdots & \vdots \\ \alpha_{\tau(k+1)}^{\tilde{\sigma}(n-k)} & \dots & \alpha_{\tau(n)}^{\tilde{\sigma}(n-k)} \end{vmatrix}$$

It only remains to clarify the expression for the sign.

$$\begin{aligned} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) &= (-1)^{\sum_{i=1}^{k}\sigma(i) - T_k}(-1)^{\sum_{i=1}^{k}\tau(i) - T_k} \\ &= (-1)^{\sum_{i=1}^{k}\sigma(i) + \sum_{i=1}^{k}\tau(i) - 2T_k} \\ &= (-1)^{\sum_{i=1}^{k}\sigma(i) + \sum_{i=1}^{k}\tau(i)} \end{aligned}$$

which is exactly the classical form of the Laplace expansion.

2.3.3 Action of linear Mappings on $\Lambda(V)$

Let A be a linear operator $A : V \to W$ where W might be V. Then A induces a linear operator

$$\overset{k}{A}\colon \Lambda^k(V) \to \Lambda^k(W)$$

defined on products by

$$\overset{k}{A}\,(v_1 \wedge v_2 \wedge \ldots \wedge v_k) = Av_1 \wedge Av_2 \wedge \ldots \wedge Av_k$$

and extended to all of $\Lambda^k(V)$ by linearity. It is not immediately clear that the new operator is well defined; in fact proper procedure would have been to define it on basis vectors

$$\overset{k}{A}\,(e_{\sigma(1)} \wedge \ldots \wedge e_{\sigma(k)}) = Ae_{\sigma(1)} \wedge \ldots \wedge Ae_{\sigma(k)}$$

for $\sigma \in \mathcal{S}_{n,k}$ and then show that the above formula works for all $v_1, \ldots, v_k$. Since we did things like this in the previous subsection the reader will have no trouble working this out for herself.

Now that we have defined the induced operator we will make use of what Alfred Tarski referred to as "The method of systematic forgetting" and stop writing the tiny k on top of the A so the formula becomes (with forgetting of the k,)

$$A\left(v_1 \wedge v_2 \wedge \ldots \wedge v_k\right) = Av_1 \wedge Av_2 \wedge \ldots \wedge Av_k$$

The reader should continue to regard the k as being there but invisible, so as to keep the notation as clean as possible.

2.4 Introduction to Topology

This section is included mainly for reference. You could give it a quick skim and return when necessary to clarify the notions. On the other hand it wouldn't hurt if you payed some attention to the earlier parts especially, and compactness.

Topology is our first effort to put additional structure onto sets. The subject has an interesting history and consists of two basic strands. The first works with sets of points and took quite a long time to develop into a pleasant theory, with the first efforts of Weierstrass and Bolzano about 1860. It wasn't until the fundamental work of Poincaré around 1900 that mathematicians felt the need for precise ideas and definitions. There were many false starts and blind alleys, and much separating of wheat from chaff, but ultimately it became clear which ideas were fundamental and which definitions adequately expressed them. In time a beautiful theory evolved, characterized by great abstraction, vast generality, and very unintuitive definitions. It turned out that the most convenient way to develop the subject was by means of the concept of open set. This is the strand of topology we are most interested in.

The second strand, the combinatorial and algebraic strand, goes back all the way to Euler in the 1760s, but was put on a firm footing by Henri Poincaré around the turn of the twentieth century, with one further critical idea from Emmy Noether. It is concerned with the large scale structure of spaces. We will have a bit of a look at this also, as it intersects in an important way with differential forms, but it is a vast subject and we can only glance at it from a distance.

Our Introduction for point set Topology contains more than is strictly necessary for what follows. We felt that we should at least explain the material, even if we cannot develop it in detail, that the reader is likely to meet in other books that work with this material. Hence we recommend that the reader skim the later sections (but hit compactness hard) and return to them should questions arise later.

We have room for only a sketchy description of the world of topology, but this material is critical for mathematicians, useful for physicists and not actually harmful to engineers.

Note the first o in topology is pronounced in American English as weakly as possible: t'pólogy. Full pronunciation of the o gives the impression that the speaker is from the backwoods if the rest of his English is standard American. Europeans tend to pronounce the first o much more fully.

2.4.1 Topological Spaces in General; Open Sets

A topological space is a set K together with a set τ of subsets of K which we call the *open sets*. ($A \in \tau \iff$ A is an open set of K.) τ is required to satisfy the following three axioms

1) $\emptyset, K \in \tau$

2) if $A_\iota \in \tau$ for all $\iota \in I$ then $\bigcup_{\iota \in I} A_\iota \in \tau$

3) if $A_1, A_2 \in \tau$ then $A_1 \cap A_2 \in \tau$

The word *family* is a synonym for set, but is often used to emphasize that the family has *sets* for elements. Using this word, 2 and 3 above are often stated as *the union of an arbitrary family of open sets is open and the intersection of a finite number of open sets is open.* Memorize this mantra!

We say that a set τ satisfying 1,2,3 *defines* a topology on K. We could, for example, take τ to be *all* the subsets of K, and the resulting topology on K is called the discrete topology on K and is not as useless as it looks. We could also take $\tau = \{\emptyset, K\}$ and the resulting topology with just two open sets is called the indiscrete topology and is indeed fairly useless. Or we could take $K = \mathbb{R}$ and the open sets to be open intervals $(a, b) = \{x | \, a < x < b\}$, all unions of these, along with the empty set $\emptyset$ and $\mathbb{R}$ itself. This gives the *standard topology* on the real line. Additional examples will be found in the following section on metric spaces.

Topological spaces range from the completely intuitive, e.g. $\mathbb{R}^2$, to the totally bizarre. Fortunately we will remain towards the intuitive end of the spectrum.

Topological spaces can be further graded by requiring them to satisfy additional properties which make them "nicer". For example

K is a T_1-space $\iff$ given $x, y \in K$ there exists an open set U with $x \in U$ and $y \notin U$

K is a T_2 or Hausdorff space $\iff$ given $x, y \in K$ there exist open sets U and V so that $x \in U$ and $y \in V$ and $U \cap V = \emptyset$

(Recall that if $U \cap V = \emptyset$ we say that U and V are *disjoint*.) In most of mathematics Hausdorff spaces are the norm, but the big exception is the vast ediface of Algebraic Geometry, in which the Zariski topology reigns supreme and it is T_1 but *not* Hausdorff.

There are many more definitions of this type (called separation axioms), for example *completely regular* and *normal*, and we need normal for our purposes, because it guarantees that certain continuous functions that we require will exist.

K is a *normal* space $\iff$ given two closed sets S and T of K there exist two open sets U_S and V_T so that $S \subseteq U_S, T \subseteq V_T$ and $U_S \cap V_T = \emptyset$.

Thus *normal* does for closed sets what Hausdorff does for points.

Closed Sets

Closed sets are complements of open sets in K.

Def Let K be a topological space. A subset $T \subseteq K$ is *closed* $\Longleftrightarrow$ its complement $K - T$ is open.

Naturally the complement of a closed set is an open set. Closed sets satisfy the complementary forms of the defining properties of open sets.

1) $\emptyset, K$ are closed

2) if T_ι are closed for all $\iota \in I$ then $\bigcap_{\iota \in I} A_\iota$ is closed

3) if A_1, A_2 are closed then $A_1 \cup A_2$ is closed.

This says that the intersection of an arbitrary family of closed sets is closed and the union of finitely many closed sets is closed.

It would appear from the definition and properties that closed sets and open sets have a certain symmetry; this is very far from the truth. Psychologically and mathematically open sets and closed sets are very different. Nevertheless, if one wanted to, and some mathematicians have wanted to, topological spaces may be defined by using any set of subsets of K which satisfies the above three properties and calling that set of subsets the closed sets. When you then define open sets as the complements of these closed sets you have a topology. This is an alternate way of starting off topology, and the two methods, with open sets or with closed sets, are clearly equivalent. The consensus has drifted toward the open set definition, perhaps just by luck.

Relative Topology

Let K be a topological space and M be any subset of K. Then a topology may be put on M by defining τ_M to consist of the intersection of M with the open sets of K. If M is open then the open sets of M will also be open in K but in general this is not true.

As an example of this idea, let M be compact in K. Thus an open cover by sets of τ can be reduced to a finite subcover (this is the definition of compact; see later subsection). But this implies that an open cover by sets of τ_M can be reduced to a finite subcover. This also works the other way. Thus M is compact no matter if K's topology or the relative topology on M is used.

Metric Spaces

It would be methodologically purer to continue the general theory of topological spaces, but the problem is that we need examples and these are most readily available in metric spaces, which are somewhat familiar objects. A metric space is one in which any two elements have a distance between them. This is characterized by the metric function for which we will use the letter ρ. The letter

d is often used but we wish to reserve d for the exterior derivative, and ρ is an often used alternate. Thus

Def A set K is a metric space if there is a function $\rho : K \times K \to [0, \infty)$ where $\rho(x, y)$ gives the *distance* between x and y. The ρ must satisfy the following properties

1) $\rho(x, x) = 0$ and $\rho(x, y) > 0 \Longleftrightarrow x \neq y$

2) $\rho(x, y) = \rho(y, x)$

3) $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ triangle inequality

The advantage of metric spaces is that we can get hold of the notion of open set in a very intuitive way. Also there is a natural notion of "close" which makes it easy to define continuous function.

Another very helpful thing about metric spaces is as sort of standard neighbourhood, called a ball. A ball is all the points around x that are closer than a specified distance r. We have, for $x \in K$ and K a metric space

$$\begin{aligned} B_r(x) &= \{y : \rho(y, x) < r\} \\ \overline{B}_r(x) &= \{y : \rho(y, x) \leq r\} \\ S_r(x) &= \{y : \rho(y, x) = r\} = \overline{B}_r(x) - B_r(x) \end{aligned}$$

The first is called the open ball of radius r around x. The second is called the *closed* ball of radius r around x, and is the closure of the first in a sense to be made precise later. The third is called the *sphere* of radius r around x.

A metric space may be made into a topological space by defining which sets are open, and verifying the open sets satisfy the defining relations of a topological space. Then

Def Let K be a metric space and $M \subseteq K$. Then M is *open* $\Longleftrightarrow$ for every $x \in M$ there is an $r > 0$ for which $B_r(x) \subseteq M$.

Thus M is open if and only if it contains a ball around each of it's points, where here and elsewhere we interpret ball to mean ball of non-zero radius. We also explicitly add the empty set to the set of open sets, although logically the above definition covers this case.

Now it is easy to see that the open sets satisfy the topology axioms; the first and second are trivial and the third is a consequence of the triangle inequality for metric spaces and we have

Theorem With the above definition of open set, a metric space is a topological space

We will give a few examples of a metric space here, and more later in the section on continuous functions. Notice that any subset of a metric space is itself a metric space using the metric inherited from the big space.

Example 1) $K = \mathbb{R}^n$ and if $x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$ then

$$\rho(x, y) = \Big[\sum_{i=1}^{n} (x^i - y^i)^2 \Big]^{\frac{1}{2}}$$

Example 2) $K = B_1(0) \subseteq \mathbb{R}^n$ with inherited metric.
Example 3) $K = S_1(0) = \{x : \rho(x, 0) = 1\} \subseteq \mathbb{R}^n$ with inherited metric.
Example 4) K is the set of integrable functions on [0,1] and

$$\rho(f, g) = \Big[\int_0^1 |f(x) - g(x)|^2 \Big]^{\frac{1}{2}}$$

Neighbourhoods

Neighborhood is an easy concept, and in fact is more psychological than mathematical. It indicates an attitude toward a point and an open set which contains it. The psychological import of the use of the term *neighborhood* is that we think of the open set as a *small* open set around x.

Def U is a neighbourhood of $x \iff x \in U$ and U is open.

Thus "Let U be a neighborhood of x" is just a synonym for "let U be an open set containing x.

An open set is a neighbourhood of each of its points, and a set is open if contains a neighbourhood of each of its points. (Just take the union of all the neighbourhoods.) Some books define a neighbourhood of X as any set which contains an open set U for which $x \in U$ but we will insist that a neighbourhood be open, which has become the norm. We may sometimes say "S is a closed neighbourhood of x" if $x \in U \subseteq S$ where U is open and $S = \overline{U}$ (see next section for meaning of $\overline{U}$). We will occasionally use the redundant term "open neighbourhood" if we wish to remind the reader that the neighbourhood is open.

Neighbourhoods are a handy way of talking about many things. Here is a good example.

Def Let $M \subseteq K$ be a set. Then the boundary of M, denoted by ∂M, consists of all points $x \in K$ for which any neighbourhood of x contains points of M and points of the complement $K - M$ of M.

For example in $\mathbb{R}^n$, the boundary $\partial B_1(0)$ of the open unit ball $B_1(0)$ consists of the unit sphere $S_1(0) = \{x |\, \rho(x, 0) = 1\}$. The boundary of the closed unit ball $\overline{B}_1(0)$ is also the unit sphere.

In general boundaries are tricky objects but in this book they will always be quite civilized.

Bases

Let K be a topological space with topology τ. If one wants to introduce or discuss a topology, it may be inconvenient to work with the entire family τ, so we can ask if there is a smaller family of open sets which we can use and which will uniquely determine τ and which might be more convenient. This leads to the concept of *basis of a topology.*

Def A family $\rho \subseteq \tau$ is a *basis* for the topology $\tau \iff$ for every open set U and for ever $x \in U$ there is a $W \in \rho$ for which $x \in W \subseteq U$

There is a simpler way we could have defined basis but our way is more to the point psychologically. The simpler way is given in the following theorem, which is quite easy to prove.

Theorem A subset ρ of τ is a basis for $\tau \iff$ Every open set in τ is a union of sets in ρ.

There is a similar idea at the local level. For any $x \in K$ let $\sigma(x)$ be the open sets of σ that have x as a member. Thus $\sigma(x)$ is the family of neighbourhoods of x. Then we can define a neighbourhood basis of $\sigma(x)$ in the obvious way.

Def $\rho(x) \subseteq \sigma(x)$ is a *neighbourhood basis* of $\sigma(x) \iff$ for every $U \in \sigma(x)$ there is a $W \in \rho(x)$ for which $W \subseteq U$.

A good example of a basis is to take any metric space K and for ρ the balls $B_r(x)$ for any $x \in K$ and $r > 0$, A cleverer method would be to take the balls with *rational radius* $r > 0$. Thus each point has a *countable* neighbourhood basis. This is important in some contexts. And for $K = \mathbb{R}^n$ we can do even better, in that we can use only the balls $B_r(x)$ where $r > 0$ is rational and also x has rational coordinates. In this case we have a *countable* basis for $\mathbb{R}^n$. These are standard desirable qualities and have defnitions to match.

Def A topological space is *first countable* $\iff$ each point has a countable neighborhood basis.

A topological space is *second countable* $\iff$ it has a countable basis.

Second countable is one of a number of ways to mathematically interpret the psychological idea "not too big." Another is paracompact, so be introduced later.

Interiors and Closures

Interior is an often useful concept but closure is one of the most critical concepts of Analysis. We define them here and give a few properties. A little thought will allow the reader to prove any of these properties that interest her.

Def Let K be a topological Space and $M \subseteq K$. The interior of M, denoted by M°, is the set of points that have neighbourhoods $U_x \subseteq M$.

Any point in an open set U is in its interior but closed sets often have no interior. For example the x-axis in $\mathbb{R}^2$ is a closed set with no interior. Or consider the set M in $K = \mathbb{R}^2$ consisting of the union of $\overline{B}_1(0)$ and the set [1,2] on the x-axis. Its interior $M^\circ = B_1(0)$.

Now for the closure of a set

Def Let K be a topological space and $M \subseteq K$. Then x is in the closure of M, denoted by $\overline{M}$, $\iff$ every neighbourhood of x intersects M.

For example, in $\mathbb{R}^2$ the closure of $B_1(0)$ is $\overline{B}_1(0)$ as the notation anticipated when we introduced it. More interesting is, with $K = \mathbb{R}$,

$$M = \{x : 0 < x < 1 \text{ and } x \text{ is a rational number}\}.$$

Then $\frac{1}{2} \in M$ and $\frac{\pi}{6} \notin M$. We now have $\overline{M} = [0, 1]$

Naturally the closure $\overline{M}$ of a set is a closed set, which is easy to prove by showing $K - \overline{M}$ is open.

There is a second way to find the interior and closure of a set. It is easy to prove the equivalence of the two ways.

$$\begin{aligned} M^\circ &= \bigcup_{\substack{U \subseteq M \\ U \text{ open}}} U \\ \overline{M} &= \bigcap_{\substack{M \subseteq S \\ S \text{ closed}}} U \end{aligned}$$

Thus the interior is the union of all open sets contained in M and the closure is the intersection of all closed sets containing M.

We then have

$$\begin{aligned} U \text{ is open} &\iff U = U^\circ \\ S \text{ is closed} &\iff S = \overline{S} \end{aligned}$$

Closure and interior are not usually inverses of one another. For example if $S \subseteq \mathbb{R}$ and $S = (0,1) \cup (1,2)$ then $\overline{S} = [0,2]$ and $(\overline{S})^\circ$ is $(0,2)$. On the other hand, $S^\circ = S$ and $\overline{S^\circ} = \overline{S} = [0,2]$

We want to give some properties of the closure operator but to make this more interesting we introduce a new friend. For any set K, the family (family just means set) of subsets of K is a set called the power set of K and denoted by $\mathcal{P}(K)$. Note that if K is a topological space then $\tau \subseteq \mathcal{P}(K)$. Also note that closure,

$$M \to \overline{M}$$

is an operator (function) from $\mathcal{P}(K)$ to $\mathcal{P}(K)$. It has the following easily proved properties

1) $\overline{\emptyset} = \emptyset$

2) $M \subseteq \overline{M}$

3) $\overline{M \cup N} = \overline{M} \cup \overline{N}$

4) $\overline{\overline{M}} = \overline{M}$

What is far more interesting is that an operator from $\mathcal{P}(K)$ to $\mathcal{P}(K)$ having these four properties can be used to define a topology on K. The closed sets are the sets

$$\{\overline{M} : M \subseteq K\}$$

and the open sets of the topology are

$$\{K - \overline{M} : M \subseteq K\}$$

This, our last method of introducing a topology, was pioneered by Kasimerz[2] Kuratowski (1896-1980).

Another approach to closure is via adherent points and points of accumulation.

Def Let K be a topological space and $M \subseteq K$. A point $x \in K$ is an *adherent point* of $M \iff$ every neighbourhood of x contains a point of M.

Clearly every point of M is an adherent point of M.

Def Let K be a topological space and $M \subseteq K$. A point $x \in K$ is an *point of accumulation* of $M \iff$ every neighbourhood of x contains a point of M other than x itself.

Since a neighbourhood of x may be found that excludes the new point given by the definition, neighbourhoods of x contain many points of M. Intuitively the points of M swarm around x, hence the name.

Points of M that are *not* points of accumulation of M are called *isolated points* of M. Thus an isolated point x is one for which there is a neighbourhood U of x for which $U \cap M = \{x\}$. If all the points of M are isolated then M is called a *discrete* set (in K) and the relative topology of M in K is the discrete topology.

Let M^a be the set of adherent points of M and M^{ac} be the set of accumulation points of M. Then the following are easy to prove.

$$\begin{aligned} \overline{M} &= M \cup M^a = M \cup M^{ac} = M \cup \partial M \\ \partial M &= \overline{M} \cap \overline{K - M} \end{aligned}$$

The second equation is little more than a rewrite of the definition of ∂M.

Connectedness

The intuitive idea here is that a topological space K is connected if and only if it is not in two or several pieces. As an example of non-connected set, let $M = B_1((0,0) \cup B_1((2,0)) \subseteq \mathbb{R}^2$

There are two definitions of connectedness, connectedness and arcwise connectedness. Connectedness is easy but not too intuitive. It is critical to remember that in this definition the concept "open" refers to the set M being considered; not some larger set K which may be providing a relative topology for M. A set is defined to be *open in* M if and only if it is the intersection $M \cap U$ with some open set U in K. Thus, if M is not open in K, the open sets in M may not be open in K. This is called the relative topology on M. As an example let $M = \overline{B}_1((0,0)) \subseteq \mathbb{R}^2$. Then the set $M \cap B_1((1,0))$ is open in M even though this intersection is not open in $\mathbb{R}^2$. Now we can define connectedness, with the understanding that "open" means "open in M".

Def A topological space M is *connected* $\iff$ M is **NOT** the union of two disjoint open sets.

[2] pronounced Kazhmeer, like for sweaters, and the w is pronounced like v

In the example above of a non-connected set, each of the open balls is an open set, in M and thus M is the union of two disjoint open sets. A much more interesting example is the subset M of $\mathbb{R}^2$ given by the union of M_1, the graph of $\sin(\frac{1}{x})$ on the interval $(0, \frac{1}{\pi})$ with M_2, the portion of the y-axis from (0,-1) to (0,1). This is a favorite counterexample. M can be seen to be connected because any open set containing a point of M_2 will include points of M_1. In fact, any point of M_2 is a point of accumulation of M_1. We will refer to this set as the Beartoes set, because it resembles the top part of the pawprint of the infinite toed bear.

This brings us to the second form of connectedness. This is based on arcs, which are continuous (see next section) maps of the interval $[-1, 1] \subseteq \mathbb{R}$ into a topological space M.

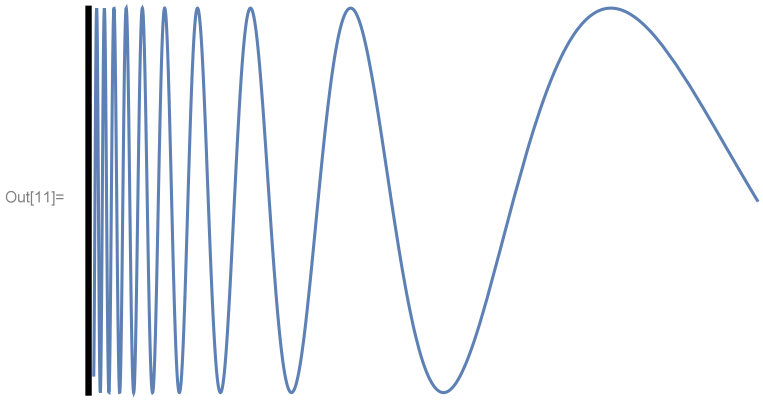
Def An *arc* is a continuous map from $[-1, 1] \subseteq \mathbb{R}$ into a topological space M

Intuitively we replace the function by it's range, which thus looks like a curving line in the space M. It does no harm in this context to think of the arc as smooth. The choice of the interval [-1,1] is arbitrary; any other interval would work as well and perhaps [0,1] is more traditional. But for later purposes I wanted 0 to be interior to the domain of the arc.

Arcwise connected then means we can draw an arc from any point $x_1 \in M$ to any other point $x_2 \in M$.

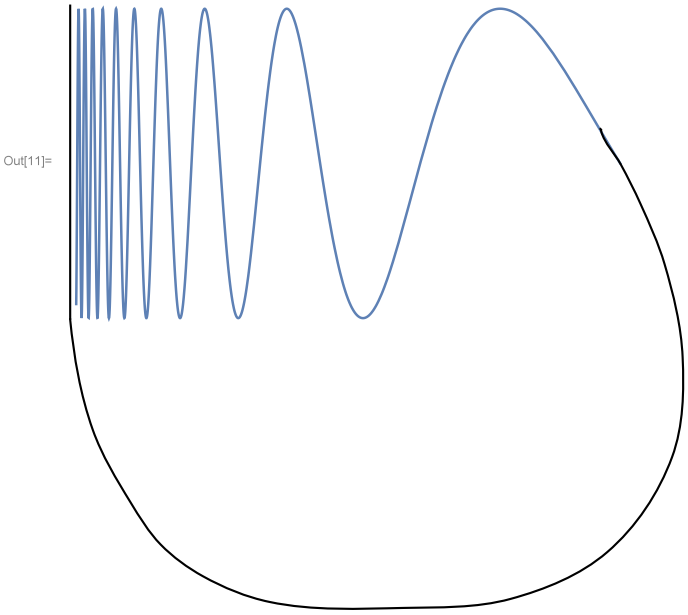
Def A topological space M is *arcwise connected* $\Longleftrightarrow$ for any two distinct points x_1 and x_1 of M there exists an arc for which $f(-1) = x_1$ and $f(1) = x_2$.

Arcwise connected is a stronger condition than connected; if a space is arcwise connected then it is connected, because the two disjoint open sets would "disconnect" the arc and ruin the continuity. The Beartoes set is an example of a connected but not arcwise connected set. No arc can get through all the wiggles to get to the y-axis.



The Beartoes set

This is a handy place to construct another standard counterexample; the paw print of the infinite toed bear[3]. This is constructed in $\mathbb{R}^2$ by drawing a smooth curve in the lower half plane of roughly semicircular shape from the point (0,-1) to the point $(\frac{1}{\pi}, 0)$ and, if you like, connecting the curve so the connections are smooth. Let B (for bear) be the enclosed area. Then B has finite area but its boundary has infinite length. Each point on the boundary has a well defined exterior unit normal vector, though this is not continuous. The value of B as a counterexample is shown below. B is arcwise connected. ∂B, is also arcwise connected, and thus connected, through the bottom roughly circular segment. When you have some conjecture about open sets and their boundaries, always check it against B and ∂B.



B = The paw print of the Infinite Toed Bear

There is a local version of connectedness. A space is locally connected $\Longleftrightarrow$ every point of the space has a local basis of connected neighbourhoods. Similarly for locally arcwise connected. Our friend ∂B is an example of an arcwise connected (and thus connected) space which is neither locally connected nor locally arcwise connected. You can see this by considering the points on the vertical segment from $(0, -1)$ to $(0, 1)$.

2.4.2 Continuous Functions

The most basic idea of Analysis is the *continuous function*. At its most fundamental, continuous means that if x is close to y then $f(x)$ is close to $f(y)$.

[3]Tremarctos multidactylus, Middle Pliocene, Florida, USA.

In topology there is a simple definition of continuous using open sets but, as is not unusual in topology, the definition is technically very convenient but intuitively obtuse. Hence we will take the following path. We will define continuous function in a metric space, where the definition is quite natural, then show that in a metric space this definition is equivalent to one involving only open sets, and then we will use the open set property as the definition of continuous in a general topological space, where it is not obvious what "close" means.

So we begin with metric spaces. Let K and L be two metric spaces and let f be a function from K to L. Then f is continuous at x means that if someone chooses a small $\epsilon > 0$ then by getting y close enough to x we can get $f(y)$ closer than ϵ to $f(x)$. Thus the definition *for metric spaces* is

Def Let K and L be metric spaces and $f : K \to L$. Then f *is continuous at* $x \iff$ for every $\epsilon > 0$ there exists a $\delta > 0$ so that

$$\text{if } \rho(y, x) < \delta \text{ then } \rho(f(y), f(x)) < \epsilon$$

Def f is a continuous map from K to $L \iff f$ is continuous at every $x \in K$.

Recall that for $f : K \to L$ and $M \subseteq K$, the notation $f[M]$ means $f[M] = \{f(x) : x \in M\}$. Then with the notation in the definition, the condition for continuity at x translates to $f[B_\delta(x)] \subseteq B_\epsilon[f(x)]$.

This definition is usually not that difficult to work with although beginning students often find it confusing. Now the wonderful thing is that there is an equivalent toplological definition of continutity which is technically simpler to work with. We will introduce it in the context of metric spaces and then use it as the definition of continuity for general toplogical spaces. First recall the definition of inverse image.

Def Let K and L be any sets and $f : K \to L$. Then for $N \subseteq L$

$$f^{-1}[N] = \{x \in K : f(x) \in N\}$$

Note $f^{-1}[N] \subseteq K$, $f^{-1}[\emptyset] = \emptyset$ and $f^{-1}[L] = K$.

Now we are ready for the following rather surprising theorem.

Theorem Let K and L be metric spaces. Then f is a continuous function from K to $L \iff$ for all $N \subseteq L$, if N is open then $f^{-1}[N]$ is open.

Proof $\Rightarrow$: Let $x \in f^{-1}[N]$. Since N is open and $f(x) \in N$, there exists an $\epsilon > 0$ so that $B_\epsilon(f(x)) \subseteq N$. Select $\delta > 0$ so that if $\rho(x_1, x) < \delta$ then $\rho(f(x_1, f(x)) < \epsilon$. But this just says $f[B_\delta(x)] \subseteq B_\epsilon(f(x))$., so that $B_\delta(x) \subseteq f^{-1}[B_\epsilon(f(x))] \subseteq f^{-1}[N]$. Thus $f^{-1}[N]$ contains a ball around each of its points and thus is open.
$\Leftarrow$: Given $\epsilon > 0$, for any $x \in K$ $B_\epsilon(f(x))$ is an open set in L. Then $f^{-1}[B_\epsilon(f(x))$ is open in K. Since $x \in f^{-1}[B_\epsilon(f(x))$, $f^{-1}[B_\epsilon(f(x))$ contins a ball $B_\delta(x)$ around x. But then $f[B_\delta(x) \subseteq B_\epsilon(f(x))]$. This just says

$$\text{if } \rho(x_1, x) < \delta \text{ then } \rho(f(x_1), f(x)) < \epsilon,$$

which says that f is a continuous function from K to L.

This shows that continuity can be checked by looking at the behaviour of f^{-1} on open sets. This is then perfectly possible in a general topological situation (no metric) and so motivates our final definition of continuous function for topological spaces

Def Let K and L be topological spaces and $f : K \to L$. Then f is a continuous function from K to $L \iff$ for any $N \subseteq L$, if N is open in L then $f^{-1}[N]$ is open in K.

Notice how we arrived here. For metric spaces we had the intuitively clear preliminary definition. This we showed was equivalent (in a metric space) to the present definition, but the present definition is applicable to any pair of topological spaces, making it a candidate for the general definiton of continuity in topological spaces. Experience then shows that it performs as we want it to, and so we adopt it for our general definition. This sort of argument does not always work well, but in this situation it works splendidly, and it is also easy to use.

Homeomorphisms

Def Let K and L be topological spaces and let $f : K \to L$ be a continuous map. f is a *homeomorphism* $\iff$ 1) it is one to one and onto and thus has an inverse map $f^{-1} : L \to K$ and 2) f^{-1} is continuous.

If there is a homeomorphism between K and L then K and L are said to be *homeomorphic*. Homeomorphic spaces are topologically indistinguishable, for example the unit ball and the unit (solid) cube in $\mathbb{R}^3$. It is the topological way of saying the two spaces are *the same*. Of course they may not be the same from *other* points of view, but topologically they are the same. The standard example is a donut (solid) and a coffee cup. Also standard is the description that two spaces are homeomorphic if one can continuously deformed into the other. This description should be understood at the intuitive level; there are technical difficulties with it but it is a good guide.

2.4.3 Compactness

The notion *compactness* is a sort of topological equivalent of the idea *finite* in elementary mathematics. However, this is not at all obvious. For $\mathbb{R}^n$ compactness is equivalent to closed and bounded, but this is not true in general nor is it even true for metric spaces. It took quite a long time to zero in on the best definition for compactness, but in the end the definition that won out is very convenient technically and intuitively supremely opaque, as is not so rare in topology. Once we get beyond the topology section this will stop happening. At any rate, here is the definitions for one of the most important concepts in mathematics.

Def Let K be a topological space. An open cover of K is a family of open subsets A_ι, $\iota \in I$ for which $K \subseteq \bigcup_{\iota \in I} A_\iota$

Def Let K be a topological space. Then K is *compact* $\iff$ For every open cover $\{A_\iota\}$ of K there exists a finite set $\{\iota_1, \ldots, \iota_n\} \subseteq I$ for which $K \subseteq \bigcup_{i=1}^{n} A_{\iota_i}$

The mantra for this is *Every open cover of K can be reduced to a finite subcover*

Our definition above was for the whole space K but we also want a definition for compactness of a subset M of K. This is essentially the same. An open cover

of M is a family of open sets $\{A_\iota,\ \iota \in I\}$ for which $M \subseteq \bigcup A_\iota$. Then M is compact $\Longleftrightarrow$ every open cover of M can be reduced to a finite subcover.

The proof of the following theorem is given because it illustrates the techniques of compactness but is easy.

Theorem A compact subset of a Hausdorff space is closed.

Proof: Let $x \in K - M$. Since K is Hausdorff, for every $y \in M$ there exist open sets U_y and V_y so that $y \in U_y$ and $x \in V_y$ and $U_y \cap V_y = \emptyset$. The family of open sets $\{U_y : y \in M\}$ obviously covers M. Since M is compact, finitely many of these sets, $U_{y_1}, \ldots U_{y_n}$ also cover M. Then $V = V_{y_1} \cap V_{y_2} \cap, \ldots, \cap V_{y_n}$ is an open set (because it is a *finite* intersection) for which $x \in V$ and $V \cap M = \emptyset$. Hence $K - M$ has an open set around each of its points, and thus is open. So K is closed.

Note the critical role here of both Hausdorff and Compactness. The theorem is not true for non-Hausdorff spaces, which considerably lessons the value of compactness in Algebraic Geometry where the topolgy is T_1 but not Hausdorff. The proof is typical (but simpler) of proofs involving compactness. If you enjoyed it you may well enjoy point set topology in general.

Another theorem of great importance (which is very easy to prove) is

Theorem Let N be closed subset of a compact subset M of a topological space K. Then N is also compact.

Next we prove another important theorem about the images of compact sets under continuous maps. This theorem is used very frequently. An illustration will be provided.

Theorem Let $f : K \to L$ be a continuous map and let M be a compact subset of K. Then $f[M]$ is compact.

Proof Let $V_\iota,\ \iota \in I$ be an open cover of $f[M] \subseteq L$. Since f is continuous, $U_\iota = f^{-1}[V_\iota],\ \iota \in I$ form an open cover of $M \subseteq K$. Since M is compact, There exists a finite subset $\iota_1, \ldots, \iota_n$ so that

$$M \subseteq \bigcup_{i=1}^{n} U_{\iota_j}$$

But then

$$f[M] \subseteq \bigcup_{i=1}^{n} V_{\iota_j}$$

since any $y \in f[M]$ is the image of an x in $f^{-1}[M]$ and this x is in some U_{ι_j}.

The mantra for this theorem is the continuous image of a compact set is compact. As an example, we can prove that there is no continuous onto mapping from the unit circle in $\mathbb{R}^2$ to $\mathbb{R}$. This is because the unit circle is compact and $\mathbb{R}$ is not, so such a mapping would contradict the theorem.

Perhaps you want to know why the unit circle is compact. Well the unit interval $[0, 1]$ is compact (see below) and the mapping

$$f(t) = (\cos(2\pi t), \sin(2\pi t))$$

maps $[0,1]$ onto the unit circle. See how handy this theorem is.

Another theorem worth mentioning is that with enough extra assumption we can dispense with the continuity of f^{-1} for homeomorphisms; it is automatic. We will actually prove this theorem because it is a good illustration of how the various concepts fit together.

Theorem Let $f : K \to L$ be a one to one onto continuous map. Let K be compact and L be Hausdorff. Then f^{-1} is continuous and f is a homeomorphism.

Proof: We need to show that $(f^{-1})^{-1} = f$ takes open sets to open sets; this will guarantee f^{-1} is continuous. Since f is one to one, it suffices to show that f takes closed sets to closed sets. But if M is closed in K, then it is compact. Hence $f[M]$ is compact. Since L is Hausdorff, $f[M]$ is closed, and we are done.

The mantra for this theorem is a one to one onto continuous map from a compact space to a Hausdorff space is a homeomorphism.

To actually use compactness it would be nice to know which familiar sets are compact. It is tricky to get the nomenclature correct but in olden times the following went by the name of the Heine-Borel theorem. Recall that a set is *bounded* if and only if it is constained in a sufficiently large ball.

Theorem Let $M \subseteq \mathbb{R}^n$. Then M is compact $\iff$ M is closed and bounded.

This gives us a good supply of compact sets. We specifically mention that this theorem is *not* valid for general metric spaces.

Historically compactness was associated with infinite sequences of points, and an infinite series of points in a compact space would bunch up somewhere. This is true for compact sets in $\mathbb{R}^n$ but is not even true in a general metric space, and the tendency has been away from this approach to the subject. There are interesting theorems about which subsets of a metric space are compact, but it would take us to far afield to consider this. However, one special case of this is worth mentioning. Recall that x is a point of accumulation of M if and only if every neighbourhood of x contains a point of M other than x itself. The following is a form of the Bolzano-Weierstrass theorem.

Thm Let $M \subseteq \mathbb{R}^n$ which is bounded (i.e. its closure is compact). Then any infinite sequence of distinct points of M has a point of accumulation.

The reader has probably seen the last two theorems proved in advanced Calculus courses or perhaps in introductory real analysis.

Local Compactness

Local Compactness is an easy idea;

Def Let K be a topological space. Then K is *Locally Compact* $\iff$ every point $x \in K$ has a neighborhood U whose closure $\overline{U}$ is compact.

An important example is $\mathbb{R}^n$, which is *not* compact.

In particular, the Differentiable Manifolds considered in this book are all locally compact, as are most of the bundles (like the tangent bundle) associated with them. On the other hand, most spaces of functions, which we do not treat,

are not locally compact, and many (most?) metric spaces are also not locally compact.

There is an easy theorem that a locally compact space may be made compact by adding a single point. This is called the one point compactification. Two examples are adding a point at ∞ to the real line, getting the circle, and adding a point at ∞ to the complex plane, getting the Riemann Sphere. In general, though, the construction is not that helpful.

Paracompactness

Paracompactness is an advanced notion and we would not bother with it here if it were not part of the definition of Differentiable Manifold, which starts out with *paracompact Hausdorff space.* This condition has two important consequences. One is that it prevents a Diffentable Manifold from being "too large", but we will not even attempt to clarify this statement. The second is that it permits partitions of unity, which is very important and which we will define below.

Like local compactness, paracompactness is a weakening of the notion of compactness, but in a very different direction. We will not attempt to be more explicit than this. First recall

Def Let K be a topological space. Then $\sigma \subseteq \tau$ is an open cover of $K \iff$ all the V in σ are open and

$$K \subseteq \bigcup_{V \in \sigma} V$$

Next we need the notion of a locally finite cover, which is critical for partitions of unity.

Def Let K be a topological space. An open cover σ of K is *locally finite* $\iff$ every point $x \in K$ has a neighbourhood U_x so that only finitely many of the sets $U_x \cap V$ with $V \in \sigma$ are nonempty.

We say that U_x meets only finitely many sets V of the cover.

Next we need the idea of a refinement of a cover.

Def An open cover σ_1 of K is a *refinement* of an open cover σ of $K \iff$ for every $V \in \sigma_1$ there exists a $U_V \in \sigma$ for which $V \subseteq U_V$

We are now ready to define paracompact

Def Let K be a topological space. Then K is *paracompact* $\iff$ every open cover of K has a locally finite refinement.

If you feel you are missing the intuitive content here, do not feel back; it is a common problem. We mention compact spaces are paracompact, that a closed subset of a paracompact space is paracompact, but a paracompact subspace of a paracompact space need not be closed. Metric spaces are paracompact but this is a somewhat difficult theorem to prove. A topological space is metrizable if and only if it is paracompact and locally metrizable. From this it follows in one step that any differentiable manifold is a metric space. A paracompact space is normal, and this is important for the construction of functions with specific properties on paracompact spaces.

Recall the *support* of a real or complex valued function f on a topological space is the closure of the set on which f is non-zero. It is denoted by $\text{supp}(f)$. The following extremely important theorem allows integration on Differentiable Manifolds to be reduced to integration over small regions.

Def Let σ be an open cover of a topological space K. A *partition of unity* subordinate to the open cover σ is given by

1) a locally finite refinement σ_1 of σ

2) a family of functions f_β, one for each open set $U_\beta \in \sigma_1$ satisfying the following conditions

3) $0 \leq f_\beta \leq 1$ and $\text{supp}(f_\beta) \subseteq U_\beta$ and $\sum_{U_\beta \in \sigma_1} f_\beta(x) = 1$ for all $x \in K$.

And now we have the theorem

Theorem If K is a paracompact topological space and σ is an open cover of K then there exists a partition of unity subordinate to σ

The proof of this theorem is too complicated to present here but the highlights are these. Since a paracompact space is normal, it is possible to construct the functions with the required supports. Since the cover is locally finite, only finitely many of the U_β will have functions which are non-zero on x and thus the sum $\sum_{U_\beta \in \sigma_1} f_\beta(x)$ will be essentially a *finite* sum, and this is critical. The sum may not come out 1 initially, but it can be adjusted by dividing each f_β by the sum and then the new sum of the functions will come out 1. There are some additional complications which we will pass over.

2.4.4 Completeness.

Completeness is not really a topological notion but it's close enough that we include it in the Topology section for convenience. Completeness is a metric space notion, though there are topological forms of it called *Uniform Spaces*. We have no need of the more general notion and will confine ourselves to metric spaces.

What lies behind completeness is the desire to prove a sequence in a metric space converges without having to come up with the element which it converges *to*, which is often difficult. To prove x_n converges to x using the definition of convergence requires us to know x. Since x is often difficult or impossible to identify mathematically, we look for another way to prove convergence.

We will start with convergence in $\mathbb{R}$ and $\mathbb{Q}$. We will consider the sequence $3, 3.1, 3.14, 3.141, 3.1415, 3.14159, \ldots$. Hopefully you know the limit of this sequence; it is π. The situation in $\mathbb{R}$ and $\mathbb{Q}$ is, however, different, in that π is in $\mathbb{R}$ but not in $\mathbb{Q}$. Even though $\pi \notin \mathbb{Q}$ we have the feeling that the sequence is converging. The theory behind that feeling was developed by the great French mathematician A. L. Cauchy[4].

[4]So great that he was the only state functionary who was *not* required to take the loyalty oath. He remained a partisan of the defunct Bourban dynasty until his death.

Here is Caucy's idea. If x_n converges to x then by defnition, for any $\epsilon > 0$ we can find and N so that if $n \geq N$ then $\rho(x_n, x) < \epsilon$. Using $\epsilon/2$ instead of ϵ we see that if $n, m \geq N$ then

$$\rho(x_n, x) < \epsilon/2 \quad \text{and} \quad \rho(x_n, x) < \epsilon/2$$

and thus

$$\begin{aligned} \rho(x_n, x_m) &\leq \rho(x_n, x) + \rho(x, x_m) \quad \text{triangle inequality} \\ &\leq \epsilon/2 + \epsilon/2 \ = \ \epsilon \end{aligned}$$

In words, if a sequence converges then the terms of the sequence get close together. Cauchy's contribution was to assert the opposite: if the terms of the sequence get close together, then the sequence converges. Such a sequence is called a Cauchy (or fundamental) sequence.

Def A sequence in a metric space is Cauchy (or fundamental) $\Longleftrightarrow$ for any $\epsilon > 0$ there exists an N so that

$$\text{if } n, m \geq N \text{ then } \rho(x_n, x_m) < \epsilon$$

We saw with the original example that in $\mathbb{Q}$ a Cauchy sequence need not converge. Spaces in which Cauchy sequences converge are especially nice, and we have a name for them:

Def A metric space is *complete* $\Longleftrightarrow$ every Cauchy sequence in the space has a limit in the space.

Most of the spaces used in Analysis are complete, for example $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{C}$, and $L^2[-\pi, \pi] = \{f | \int_{-\pi}^{\pi} |f|^2 \, dt < \infty\}$ are all complete. But $\mathbb{Q}$ is not. What do we do if we are working with a space that is not complete. We complete it by a standard procedure to be described now. Logically to do procedure one needs to know that $\mathbb{R}$ is complete, but we already know that finite closed intervals of $\mathbb{R}$ are compact, and so by the Bolzano Weierstrass theorem bounded sequences have limits in such intervals. Since Cauchy sequences are bounded, this gives us the completeness of $\mathbb{R}$.

Theorem Let K be a metric space with metric ρ. Then there exists a minimal complete metric space $\tilde{K}$ with metric $\tilde{\rho}$ so that

1. $K \subseteq \tilde{K}$

2. If L is complete and $K \subseteq L$ then $\tilde{K} \subseteq L$

3. If $x, y \in K$ then $\tilde{\rho}(x, y) = \rho(x, y)$

Proof sketch The idea is to construct a new space $\tilde{K}$ from K using equivalence classes of Cauchy sequences. Let $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$ be Cauchy sequences in K. Then the two Cauchy sequences are *equivalent* if and only if $x_n - y_n \to 0$. We now dump all the Cauchy sequences equivalent to $x_1, x_2, \ldots$ into an *equivalence class* which we denote by $[x_1, x_2, \ldots]$, so the elements of $[x_1, x_2, \ldots]$ are Cauchy

sequences equivalent to $x_1, x_2, \ldots$. The set of all equivalence classes of Cauchy sequences is our desired space $\tilde{K}$.

If $[x_1, x_2, \ldots]$ and $[y_1, y_2, \ldots]$ are two equivalence classes in $\tilde{K}$ then we set $\tilde{\rho}([x_1, x_2, \ldots], [y_1, y_2, \ldots]) = \lim_{n\to\infty} \rho(x_n, y_n)$. This is the setup. We then have a lot of boring work showing that $\tilde{\rho}$ does not depend on the choice of Cauchy sequences in an equivalence class, and that $\tilde{K}$, $\tilde{\rho}$ is indeed a metric space, and then finally it is a complete metric space, which is done using a diagonal procedure. If the Cauchy sequence in $\tilde{K}$ is

$$[x_{11}, x_{12}, \ldots], [x_{21}, x_{22}, \ldots], [x_{31}, x_{32}, [x_{33}, \ldots]$$

Then the limit will be the Cauchy sequence

$$[x_{11}, x_{22}, x_{33}, x_{44}, \ldots]$$

There is more work to show that this is indeed a Cauchy sequence, and that we have $[x_{n1}, x_{n2}, \ldots] \to [x_{11}, x_{22}, x_{33}, x_{44}, \ldots]$ etc. So this is quite a little project to do completely, but fortunately no part of it is difficult.

This method is lifted from Cantor's construction of the real numbers from the rationals. A real number is thus an equivalence class of Cauchy sequences of rational numbers.

Although most of the spaces we love in analysis are already complete, often we want to describe new spaces and the above theorem allows us to make these new spaces complete if we wish, so it is quite important for applications.

2.5 Differentiable Manifolds and their Mappings

2.5.1 Definition of Differentiable Manifolds

In order to have a firm basis on which to proceed we need to spend a bit of time on the definition of a differentiable manifold. This material is not very deep but requires a bit of thought. First we want to give a good definition of *differentiable manifold*. Then we want to discuss mappings $f : X \to Y$ between manifolds. This material is useful in the sections that follow, as well as for a great many other things, and should be studied carefully.

We will begin with a topological manifold and then add in the requirements for a differentiable manifold. This is all fairly easy. First we have the concept of *locally Euclidean*

Def A topological space M is *locally Euclidean* if and only if there is a fixed positive integer n so that each $x \in K$ has a neighbourhood U_x which is homeomorphic to an open ball in $\mathbb{R}^n$ with homeomorphism Φ. (If (U_1, Φ_1) and (U_2, Φ_2) are two such open sets which overlap, then we insist that, where defined, $\Phi_2 \circ \Phi_1^{-1}$ is continuous.)

The last sentence is redundant, $\Phi_2 \circ \Phi_1^{-1}$ is automatically continuous because Φ_1 and Φ_2 are homeomorphisms. However, I have put it in so this definition

can be contrasted with the later one for differentiable manifolds. Also I mention that "where defined" means, for $\Phi_2 \circ \Phi_1^{-1}$, the set $\Phi_1[U_1 \cap U_2]$

The n does not depend on the choice of $x \in K$. The neighbourhood is called a local coordinate neighbourhood and we will routinely use the letter Φ for the homeomorphism to the open set in $\mathbb{R}^n$. The colloquial name for local coordinate neighbourhood is *coordinate patch.* For a neighbourhood U_x around $x \in M$ we often set $\Phi(x) = 0$ and we will often forget to mention this.

The requirement that the range of Φ be a ball should be loosely interpreted for practical purposes. Actually, the range of Φ could be any open set homeomorphic to a ball, like the interior of a cube. What we must be careful to avoid, though, is a Φ whose range is the interior of a donut; that is definitely forbidden.

Now we add a few technical considerations and come up with the definition of a *topological manifold.*

Def A *topological manifold* is a locally Euclidean paracompact[5] Hausdorff space.

An example of a 2 manifold would be the boundary of the (solid) unit cube in $\mathbb{R}^3$. But we notice that it has edges and corners, and we don't like them because we want to do Calculus.

We also need the concept of *atlas.* An atlas consists of enough open sets with homeomorphisms to open sets in $\mathbb{R}^n$ so that the open sets cover M. Atlases can be minimal, in the sense that leaving out a set makes the family no longer an atlas. Or it can be maximal, which means than if a pair (U, Φ) can be added to the atlss preserving the requirement that $\Phi \circ \Phi_\iota^{-1}$ is continuous, then it must be added. There are some advantages to using maximal atlases, although minimal atlases seem more intuitively friendly.

Now that we have topological manifolds we can specialize to nicer manifolds. It all depends on the functions $\Phi_2 \circ \Phi_1^{-1}$, which are called *transition maps.* Remember this; we use them very often. A C^0-manifold requires the transition maps to be continuous, so it is just a topological manifold. We can have C^1-manifolds which require $\Phi_2 \circ \Phi_1^{-1}$ to have one continuous derivative. For physics one requires at least C^4-manifolds so one can have a continuous Riemann curvature tensor. However, this is all a lot of dull bookkeeping, so we will instead use in this book C^∞-manifolds, which avoids the bookkeeping. Notice that this is very different from C^ω-manifolds, which require that $\Phi_2 \circ \Phi_1^{-1}$ be real analytic (have power series expansions). C^ω-manifolds are quite different from the kinds of manifolds we use in this book and we will not mention them again. There is not much difference between C^4-manifolds and C^∞-manifolds, but there is a huge difference between them and the C^ω-manifolds.

It is worth mentioning that there is a complex analog of what we are doing here. In this case we have homomorphisms of coordinate patches not to open balls of $\mathbb{R}^n$ but to open polydisks of $\mathbb{C}^n$. The transition maps are then required to be *complex analytic,* and the objects are called *complex manifolds.* These objects will not occur in this book, but they are very interesting. Note that a one-dimension complex manifold ($n = 1$) has two real dimensions, and so

[5] see topological section

to persons of the strictly real persuasion it seems to be a surface. $\mathbb{C}$ itself is a one dimensional complex manifold (with one coordinate patch) which is the reason that complex oriented folks, especially algebraic geometers, refer to $\mathbb{C}$ as the *complex line.* The Riemann Sphere is another one dimensional complex manifold.

When we require the transition maps $\Phi_2 \circ \Phi_1^{-1}$ to always be C^m we may not be able to use the open sets from the original atlas of M as a topological manifold. When we go out and find the sets and functions (U, Φ) suitable for a C^m manifold it is called putting a C^m structure on the topological manifold. (One wonders if one can always do this, but that is way beyond this book.) If we are talking about C^∞ structures we often say things like we are putting a *smooth* structure on the manifold. In this book we will always be talking about smooth (that is C^∞) manifolds. The modifications necessary for lesser degrees of differentiability, though interesting for small m, are of interest only to specialists.

Because a smooth manifold is locally homeomorphic to an open ball in $\mathbb{R}^n$, we can put coordinates on the manifold, so that a point $p \in M$ has a set of numbers $(u^1, \ldots, u^n)$ correlated with it. Formally if $p \in U$ then $\Phi(x) = (u^1, \ldots, u^n)$. If (U_1, Φ_1) and (U_2, Φ_2) are two coordinate patches and $x \in U_1 \cap U_2$, and $\Phi_1(x) = (u^1, \ldots, u^n)$ and $\Phi_2(x) = (v^1, \ldots, v^n)$, then the C^∞ requirement on $\Phi_2 \circ \Phi_1^{-1}$ just says that

$$v^i(u^1, \ldots, u^n), \quad i = 1, \ldots, n \text{ are } C^\infty \text{ functions}$$

At this point we must mention some advance Calculus facts which are of great importance. Because the Φ_i are homeomorphisms, $\Phi_2 \circ \Phi_1^{-1}$ is an invertiable function, which means that we have

$$\begin{aligned} u^i &= u^i(v^1, \ldots, v^n) \\ u^i &= u^i(v^1(u^1, \ldots, u^n), \ldots, v^1(u^n, \ldots, u^n)) \end{aligned}$$

The Jacobian determinant[6] of the transformation $(v^1, \ldots, v^n) \to (u^n, \ldots, u^n)$ is

$$\det \frac{\partial(u^1, \ldots, u^n)}{\partial(v^1, \ldots, v^n)} = \begin{vmatrix} \frac{\partial u^1}{\partial v^1} & \cdots & \frac{\partial u^1}{\partial v^n} \\ \cdots & \cdots & \cdots \\ \frac{\partial u^n}{\partial v^1} & \cdots & \frac{\partial u^n}{\partial v^n} \end{vmatrix}$$

There is a similar Jacobean determinant for $(u^1, \ldots, u^n) \to (v^n, \ldots, v^n)$ and the chain rule then tells us that

$$\det \frac{\partial(u^1, \ldots, u^n)}{\partial(v^1, \ldots, v^n)} \det \frac{\partial(v^1, \ldots, v^n)}{\partial(u^1, \ldots, u^n)} = \det(\delta^i_j) = 1$$

This in turn tells us that

$$\det \frac{\partial(v^1, \ldots, v^n)}{\partial(u^1, \ldots, u^n)} \neq 0$$

[6]The determinant of the *Jacobian matrix*

which is important.

Recall from above the concept of *Atlas.* An Atlas is a set of coordinate patches that covers the manifold. An Atlas is *oriented* if and only if an Atlas exists for which for all coordinate changes from, say, $u^1, \ldots, u^n$ to $v^1, \ldots, v^n$, the Jacobian determinant is not only non-zero as above but *postive*:

$$\det \frac{\partial(v^1, \ldots, v^n)}{\partial(u^1, \ldots, u^n)} > 0$$

This is important. Although in this book we will work only with orientable manifolds, for the sake of contrast the next subsection will briefly discuss non-orientable manifolds. It is not known if the Universe is orientable or non-orientable; in the latter case space travel along certain paths will bring an originally right handed person back as a left handed person, and this is really the essence of orientablity. Since our analysis in the first two chapters is essentially *local*, it remains valid as long as we do not attempt to use it over such large regions that orientablility becomes a consideration. All manifolds are locally orientable.

Now it is easy to see what we mean by a C^∞ real valued function on the manifold M. If F as a function $F : M \to \mathbb{R}$ then we say that F is C^∞ or smooth if and only if $f = F \circ \Phi^{-1}$ is smooth. Note that $f : \Phi[U] \to \mathbb{R}$ where $\Phi[U] \subseteq \mathbb{R}^n$, so $f = f(u^1, \ldots, u^n)$ is just a function of the advanced calculus type and we know exactly what C^∞ means for such functions. (Much of mathematics consists of telling the reader, see, this is just an old friend in disguise.) Now notice that because the transition functions $\Phi_2 \circ \Phi_1^{-1}$ are C^∞, it does not matter which coordinate patch you use to check that a function $F : M \to \mathbb{R}$ is C^∞.

It is worth noting that for each coordinate patch (U, Φ) the coordinate functions $u^i(x)$ are C^∞ real valued functions on U.

The reader can now no doubt see how to define a C^∞ function F from an n-manifold N to an m-manifold M. If $x \in N$ and $F(x) \in M$ we take coordinate patches around x in N and around $F(x) \in M$ so that $\Phi_N(x) = (u^1, \ldots, u^n)$ and $\Phi_M(F(x)) = (v^1, \ldots, v^m)$. Then the v^j become functions of the u^i and we require these functions to be C^∞.

Now we must look at something conceptually less simple. We are going to show that a C^∞ is a metric space. This is the first time we make use of the paracompactness in the definition of C^∞ manifold, but it won't be the last. However, it will always take the form it does here; paracompactness implies the existence of a partition of unity, and that is what we actually need.

Paracompactness implies that if we take an open cover $U_\iota, \iota in I$ of M by coordinate patches then there is a locally finite refinement V_κ $\kappa \in J$ of this cover, and a partition of unity subordinate to this second cover. What this means is that the V_κ cover M, each V_κ is a subset of some U_ι, each point $x \in M$ has a neighbourhood U_x which meets only finitely many V_κ, and there are functions f_κ with range$(f_\kappa) \subseteq [0, 1]$, support$f_\kappa \subseteq V_\kappa$ and for any $x \in M$

$$\sum_\kappa f_\kappa(x) = 1$$

Notice that though J is probably infinite, the sum itself is finite due to the local finiteness of the family V_κ. The set $\{f_\kappa\}$ is called the *partition of unity* subbordinate to the choice of open cover made from coordinate patches. We will use this partition of unity many times in the sequel, and if confused the reader should return here to review the construction. The functions f_κ are constructed to be continuous but not necessarily differentiable. We could use convolution to make them differentiable but there seems no need to do this and it would require some modest effort.

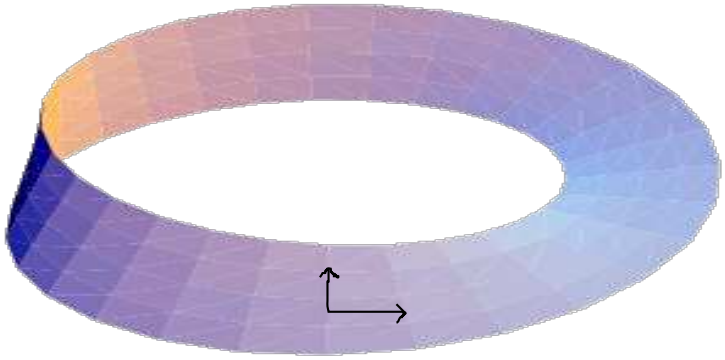
Next we need to make the V_κ into metric spaces. Since each V_κ is a subset of some U_ι, it will suffice to make U_ι into a metric space. But $\Phi_\iota : U_\iota \to W_\iota \subseteq \mathbb{R}^n$ is a homeomorphism. Since $\mathbb{R}^n$ is a metric space, so is W_ι. We can define a metric on U_ι by

$$\rho_\iota(x, y) \stackrel{\text{def}}{=} \rho(\Phi(x), \Phi(y))$$

where the metric on the right is just the usual one in $\mathbb{R}^n$. We have now shown that M is a *locally metrizable* space.

2.5.2 Orientation

In this subsection we briefly discuss non-orientable manifolds. The easiest example is the Moebius strip, which is constructed by cutting out a transparent (theoretically; one can use a piece of ordinary paper and some imagination) strip, giving it a twist, and then taping the ends together. You are probably familiar with this object.

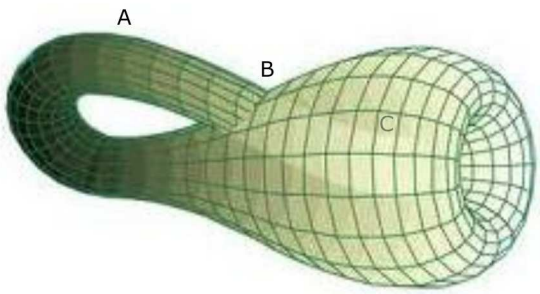


Moebius Strip

Now if we cover the Moebius strip by coordinate patches and move the axes show in the illustration from patch to patch making sure the Jacobian determinant remains positive, when the the axes have gone all the way around the strip the $x-$axis will be pointing in the original direction but the $y-$axis will be pointing in the opposite direction. Of course it will be on the "other side" of the strip but because we made the strip from transparent material we see the two axes as they really are. Note that the Moebius strip has only one

edge. You can verify this by following it around. It also has only one side, since if you go along a road in the middle of the strip you will eventually come back to where you started, but on the "other" side.

The next non-orientable surface is made by gluing the top and bottom of the strip together and then doing the twist for the Moebius strip. Can't do it? The reason you can't do it is that the resulting object, called the Klein Bottle, cannot be embedded in Euclidean 3-space. It require 4 space dimensions if you wish to make a Klein Bottle. The following illustration captures the bottle surprisingly well considering it is a two dimensional representation of an object in four dimensions.



Klein Bottle

I now explain how the picture must be modified (in your head) to be the actual Klein bottle. Note how at B the tail of the bottle appears to go through the side of the fat part. This would occur if you try to make the bottle in 3 dimensions, but this intersection does not occur in four dimensions if you follow these instructions: Set up a four dimensional coordinate system (u, x, y, z). For most of the bottle, that is everything that is not on the tail between A and C, the coordinates are $(0, x, y, z)$ where (x, y, z) refer to a three dimensional coordinate system. Note that C is supposed to be "inside" the fat part, which I have tried to indicate by making it dimmer. For the part of the tail from A to B to C, the u coordinate increases from 0 to 1 between A and B and then decreases back to 0 from B to C. Because all the points of the apparent intersection which are on the fat part of the bottle have $u = 0$ and all the points on the tail between A and C have $u > 0$, the apparent intersection does not exist in the four dimensional construction. With this kept in mind, the illustration is a pretty good picture of the Klein Bottle.

Notice that a fly entering the open front of the flat part of the bottle flies down the the tail and eventually emerges inside the fat part. Thus the Klein Bottle does not separate the space into an inside and an outside, as the surface of the closely related Torus (donut) does. And if you slide an axis at the front

of the fat part over the top and then back along the tail when it returns to it's original position the situation is just like the Moebius strip; the $x-$axis will be pointing in the original direction but the $y-$axis will be pointing in the opposite direction. This shows the non-orientability of the Klein Bottle.

While our treatment has been very brief, there is a lot of fun to be had with these ideas and there are picture books and web pages that show lots of examples. However these ideas are too peripheral to our work for us to go into it further.

2.5.3 Mappings

We now want to consider mappings ϕ from an m-dimensional manifold M to an n-dimensional manifold N; $\phi : M \to N$. It is clear enough what is meant by "mapping"; what we need to do is explain *smooth mapping*. This is natural and easy. Let $p \in M$ and $\phi(p) \in N$. We take coordinate patches (U, Φ) around p in M and (V, Ψ) around $\phi(p)$ in N. Then the map

$$\tilde{\phi} = \Psi \circ \phi \circ \Phi^{-1} : \Phi[U] \to \mathbb{R}^n$$

maps a subset of $\mathbb{R}^m$ to a subset of $\mathbb{R}^n$, and we know exactly when such maps are smooth (meaning C^∞). We thus define

Def With the notations of the preceding paragraph, $\phi : M \to N$ is smooth at p $\iff \tilde{\phi} = \Psi \circ \phi \circ \Phi^{-1}$ is smooth at $\phi(p)$.

Def With the notations of the preceding paragraph, $\phi : M \to N$ is smooth $\iff \phi$ is smooth at every $p \in M$.

It is then easy to prove the usual sort of thing; for example if $\phi : M \to N$ is smooth and $\psi : N \to P$ is smooth then $\psi \circ \phi : M \to P$ is smooth.

2.5.4 Another Point of View

We will now take a moment to present another point of view on manifolds. The prevous presentation is the standard elementary presentation, but here we are going to present another way of thinking about the situation. Since this is tangential to our main concerns, we will present a single example, but that should be sufficient to show how it all works.

We consider a unit sphere. We are going to place two coordinate systems on the unit sphere. These are not very good coordinate systems but that does not matter for our purposes. For the first coordinate system (U_1, Φ_1), we let the sphere centered on the North Pole. This will be mapped into a disk $D_1 = B_r(0)$ in $\mathbb{R}^2$. The second coordinate system (U_2, Φ_2) is similar but centered on the South Pole. The overlap $U_{12} = U_1 \cap U_2$ of the two disks is a stripe of territory on the sphere symmetrically north and south of the equator.

Now let $V_{12} = \Phi_1[U_{12}] \subseteq D_1$. V_{12} will be an annulus (ring) around the outer edge of D_1. Similarly there will be a $V_{21} = \Phi_2[U_{12}] \subseteq D_2$. Then there will be a map $\Phi_{12} = \Phi_2 \circ \Phi_1^{-1} : V_{12} \to V_{21}$ which is just our old transition map taking the annulus in D_1 to the annulus in D_2.

To emphasize the obvious, for each $p \in M$ the transition map Φ_{12} takes $\Phi_1(p)$ to $\Phi_2(p)$. So far everything is just as it was before.

Now comes the new idea. Suppose we are given the two disks, the two annuli, and the transition map, but nothing else. We now manufacture the manifold by taking for the points of the manifold the points of D_1 and D_2 *except* that if a point of $(x, y) \in D_1$ is in the domain of Φ_{12} and a point of D_2 is equal to $\Phi_{12}(x, y)$ then these are to be considered the *same* point of the manifold. In official language we *identify* p and $\Phi_{12}(x, y)$ and in semioffical language we glue p and $\Phi_{12}(x, y)$ together. This amounts to gluing the annulus in D_1 to the annulus in D_2, and that, of course, makes a sphere.

So the basic idea is if we have the open balls in $\mathbb{R}^n$ and the transition maps between open subsets of the balls, we can construct the manifold by gluing the points that are identified by the transition maps and thus *create* the manifold. From the point of view of diffeomorphisms then, everything there is to know about the manifold is contained in the set of balls and their transition maps, and we can construct the manifold by the gluing construction outlined above. This is the modern way of dealing with manifolds and it fits nicely with more advanced constructions, like Čech cohomology.

2.6 Tangent Space and Tangent Bundle

2.6.1 Definition of the Tangent Space at a Point

Now we have to create the *Tangent Space* to a manifold M at a point $p \in M$. If the manifold is embedded in a space $\mathbb{R}^n$ for some $n >$dim(M) there is a natural definition. If $C : [-1, 1] \to M$ is a curve then the coordinates of the curve can be given by $(x^1(t), \ldots, x^n(t))$ and the tangent vector to the curve at $t = 0$ is given by $\frac{d}{dt}(x^1(t), \ldots, x^n(t))\big|_{t=0}$.

Another approach is to take parameters $(u^1, \ldots, u^m)$ where dim(M)=m and then write the coordinates of the embedding in terms of these parameters,

$$x^1 = x^1(u^1, \ldots, u^m),\ x^2 = x^2(u^1, \ldots, u^m), \ldots,\ x^n = x^n(u^1, \ldots, u^m)$$

and then a tangent vector, in fact, a basis $\{e_1, \ldots, e_m\}$ of tangent vectors, at, say $u^1 = \cdots = u^m = 0$, is given by

$$e_i = \frac{\partial}{\partial u^i}\big(x^1(u^1, \ldots, u^m), \ldots, x^n(u^1, \ldots, u^m)\big)\big|_{u^i=0}$$

In our present circumstances where the manifold M is *not* embedded these methods are not available to us directly. So we must be creative and come up with a way to create something that resembles our previous experience in the embedded case but does not require an embedding.

There are two ways to do this. The first way uses curves and the second way uses first order partial differential operators. Both ways are important so we will will look at both of them, but our official way will be the second way.

We do not claim that the second way is superior in any way; we just like it better for reasons we find very hard to articulate. The choice of which method to use in given circumstances depends on the circumstances. For example, to us it seems more natural to use the first way (curves) when dealing with Lie Groups and Lie Algebras.

So we begin with the first way. We define a curve $C : [-1, 1]$ *to*M where we arrange to send 0 to the point p of interest. The local coordinate system (U, Φ) sends a neighbourhood of p to a ball around 0 in $\mathbb{R}^m$ where we arrange to for ϕ to send p to $(0, \ldots, 0)$. Then $c = \Phi \circ C$ is a curve in $\mathbb{R}^m$ with coordinates $(u^1(t), \ldots, u^m(t))$. Then we have

$$c'(0) = \frac{d}{dt}(u^1(t), \ldots, u^m(t))\Big|_{t=0}$$

which is a well defined vector in $\mathbb{R}^m$. Now we form an equivalence class of curves through p (which means $C(0) = p$) which have the *same* $c'(0)$. We call this equivalence class a *tangent vector* and we say the the local coordinates of the tangent vector are $c'(0)$.

With this definition we can produce a vector space $T_p(M)$ of tangent vectors of dimension n which we "visualize" as being "attached" at p in imitation of the embedded situation. To deal with the algebraic structure of the vector space, we have various approaches. For example, if we have a vector $v \in T_p(M)$ it will be given by an equivalence class of curves and let one of these curves by $C(t)$. Then $c(t)$ will be a curve in $\mathbb{R}^m$, as will $2c(t)$. We form a new curve $\tilde{C}(t) = \Phi^{-1}(2c(t))$ with local representation $\tilde{c}(t) = 2c(t)$ and then the equivalence class of $\tilde{C}$ will have for local representation a vector in $\mathbb{R}^m$ twice as long as v. Thus we have defined $2v$, the equivalence class of $\tilde{C}$.

With a certain amount of effort we can define addition of vectors, and then prove that $T_p(M)$ is indeed an m-dimensional vector space. However, we will not go through the details here, since this is really not the path we intend to follow. Many differential geometry books present this construction in excruciating detail for the enthusiast. We hope that we have presented enough so that the construction in the Lie Group/Lie Algebra situation will be clear.

Now it is time to go on to the second method of construction of $T_p(M)$. We will do this is two stages. First we will present a properly invariant definition and show how to work with this. However, for many purposes this development can be short circuited into a shorter and more practical methodology. We will present both methods, the theoretical first.

Theoretical Development

The basic idea is to use what we call (temporarily) a derivation, defined as follows

Def Let M be an m-dimensional manifold, $p \in M$, and U a neighbourhood of p. Then $C^\infty(U)$ is the set of smooth (i.e. C^∞) functions on U.

It is worth noting that $C^\infty(U)$ is a *ring* of functions.

Def A *derivation* at p is a function $L : C^\infty(U) \to \mathbb{R}$ satisfying

1. $L(\alpha F + \beta G) = \alpha L(F) + \beta L(G) \qquad \alpha, \beta \in \mathbb{R} \qquad F, G \in C^\infty(U)$

2. $L(FG) = F(p)L(G) + L(F)G(p)$

Note how requirement 2 anchors the derivation at p.

The derivations will eventually become the vectors of $T_p(M)$.

Now we introduce a coordinate system (u, Φ) around p and for convenience we arrange $\Phi(p) = (0, \ldots, 0)$. Here is an example of derivation L

$$L_i(F) = \frac{\partial}{\partial u^i}(F \circ \Phi)\big|_{u^j=0} \qquad i, j = 1, \ldots, m$$

To simplify the notation let $f = F \circ \Phi^{-1}$ and $g = G \circ \Phi^{-1}$. Thus f and G are functions of $u^1, \ldots, u^m$. Then a little thought will show

$$fg = (F \circ \Phi^{-1})(G \circ \Phi^{-1}) = (FG) \circ \Phi^{-1}$$

and thus

$$\begin{aligned} L_i(FG) &= \frac{\partial}{\partial u^i}(fg)\big|_{u^j=0} = \Big[f(0)\frac{\partial g}{\partial u^i} + \frac{\partial f}{\partial u^i}g(0)\Big]_{u^i=0} \\ &= F(p)L_i(G) + L_i(F)G(p) \end{aligned}$$

showing that L_i is a derivation. Clearly then

$$L = \sum_{i=1}^{m} a^i L_i$$

is also a derivation.

The only tricky thing about the second method is to show that *any* derivation L at p is of the form $L = \sum_{i=1}^m a^i L_i$, which can be done with some easy advanced Calculus techniques. We can do a plausible argument without going so deep. Let $F^i = u^i \circ \Phi$. The the correspoinding f^i is $f^i = F \circ \Phi^{-1} = u^i \circ \Phi \circ \Phi^{-1} = u^i$. and

$$L_j(F^i) = \frac{\partial u^i}{\partial u^j}\big|_{u^k=0} = \delta^i_j$$

On the other hand, for any L, let $a^i = L(F^i)$. Then (summation convention in effect from here on)

$$(L - a^j L_j)(F^i) = a^i - a^j \delta^i_j = a^i - a^i = 0$$

Thus $L - a^j L_j$ is 0 on the "coordinate functions" and this strongly suggests that $L - a^j L_j$ is 0 on all functions, which can be shown using a finite Taylor expansion.

This shows that the L form a finite dimensional vector space of dimension m with basis L_i, $i = 1, \ldots, m$. Thus we define

Def The tangent space $T_p(M)$ to the m-dimensional manifold M at p is the M-dimensional vector space of derivations L at p.

For each coordinate patch there are functions analogous to F^i and elements

$$\begin{pmatrix} a^1 \\ \vdots \\ a^m \end{pmatrix}$$

which we say are the *coordinates of L in that coordinate patch*. We can write this as $(a^1, \ldots, a^m)^\top$ for typographical convenience. We need to find how the coordinates change when we make a change to a new coordinate patch. This is the problem of change of coordinate systems and the resulting changes in the coordinates so familiar from Tensor Analysis. This is another of those areas where you may feel you already know enough; if so you can skip to the next subsection. But be sure you understand how a function from one manifold M to another manifold N acts on the tangent vectors: $df = f_* : T_p(M) \to T_{f(p)}(N)$.

So let (V, Ψ) be a second coordinate system with $p \in V$ and (for convenience) $\Psi(p) = (0, \ldots, 0)$. Then with notation as above we define

$$\begin{aligned} \tilde{f} &= F \circ \Psi^{-1} \\ \tilde{F}^j &= v^j \circ \Psi \\ \tilde{L}_j(F) &= \frac{\partial \tilde{f}}{\partial v^j}\Big|_{v^k=0} \\ \tilde{a}^j &= L(\tilde{F}^j) \end{aligned}$$

Now we have

$$f = F \circ \Phi^{-1} = (F \circ \Psi^{-1}) \circ (\Psi \circ \Phi^{-1}) = \tilde{f} \circ (\Psi \circ \Phi^{-1})$$

Now $\Psi \circ \Phi^{-1}$ is given in coordinates by

$$(v^1(u^1, \ldots u^m), \ldots, v^m(u^1, \ldots u^m))$$

so the previous equations in coordinates just says

$$f(u^1, \ldots u^m) = \tilde{f}(v^1(u^1, \ldots u^m), \ldots, v^m(u^1, \ldots u^m))$$

so that we have

$$\frac{\partial f}{\partial u^i} = \frac{\partial \tilde{f}}{\partial v^j} \frac{\partial v^j}{\partial u^i}$$

by the chain rule for functions from $\mathbb{R}^m$ to $\mathbb{R}$. Next we have for any vector L,

$$\begin{aligned} L &= a^i L_i \\ L(F) &= a^i L_i(f) \;=\; a^i \frac{\partial f}{\partial u^i} \\ &= a^i \frac{\partial \tilde{f}}{\partial v^j} \frac{\partial v^j}{\partial u^i} \end{aligned}$$

But we know

$$L(F) = \tilde{a}^j \frac{\partial \tilde{f}}{\partial v^j}$$

Comparing we see

$$\tilde{a}^j = a^i \frac{\partial v^j}{\partial u^i}$$

which is the desired change of coordinates formula for a vector and is exactly the old tensor formula for a (0,1)-tensor. In matrix form this is

$$\begin{pmatrix} \tilde{a}^1 \\ \vdots \\ \tilde{a}^m \end{pmatrix} = \begin{pmatrix} \frac{\partial v^1}{\partial u^1} & \cdots & \frac{\partial v^1}{\partial u^m} \\ \vdots & \vdots & \vdots \\ \frac{\partial v^m}{\partial u^1} & \cdots & \frac{\partial v^m}{\partial u^m} \end{pmatrix} \begin{pmatrix} a^1 \\ \vdots \\ a^m \end{pmatrix}$$

Now some philosophy. In the tensor theory of the ancients, something was a vector precisely when it had this formula for the change of coordinates. It didn't matter where it came from or where it was going; it only mattered that when you change the coordinates *this* was the way it changed. So for us, the takeaway is this: there are many ways to define $T_p(M)$, and we have seen two, but the only thing that matters is that whatever definition you use, *this* is the change of coordinate formula. All else are matters of convenience and personal taste, and no reasonable case can be made for mine is better than yours, although people will never stop trying.

Our next task is to discover what a map H from Manifold M to manifold N will do to the tangent spaces. (H is not the usual letter for such a map but I need the usual letters here for other purposes.) The idea is that we have a tangent space $T_p(M)$ and H will somehow take the vectors in $T_p(M)$ to vectors in $T_{H(p)}(N)$. This map is called H_* or dH. For the present we will use H_*

This is one of those places where the curve method of constructing the tangent space really works well. For $v \in T_p(M)$ take a curve $C : [-1, 1] \to M$ in the equivalence class which is v. Arrange $C(0) = p$ Then take $\tilde{C} = H \circ C$ which gives a curve in N with $\tilde{C}(0) = H(p)$ and the equivalence class of $\tilde{C}$ at $t = 0$ will be the vector $H_*(v)$ in $T_{H(p)}(N)$. Then set it up in coordinates and use the chain rule and you get the coordinate formulas, which you might be able to guess given the previous section on coordinate change.

However, unfortunately, we decided to use the first order partial differential operator method to construct the tangent spaces so we will have to find the coordinate formula using that method. The definition of H_* is quite easy with this method. Let $F \in C^\infty(V)$ for some open set $V \subseteq N$ containing $H(p)$. Let U be some open set around p where it would be nice if $H[U] \subseteq V$. Let L be a derivation at p. Remember L is a vector in $T_p(M)$; that's how we defined tangent vectors. Now we need a derivation at $H(p)$ to be $H_*(L)$.

Def The derivation $H_*(L)$ is defined, for $F \in C^\infty(V)$, by

$$H_*(L)(F) = L(F \circ H)$$

It is trivial to verify that $H_*(L)(F)$ is a derivation at $H(p)$, so the map does what it is supposed to do: $H_* : T_p(M) \to T_{H(p)}(N)$.

Now we need to find a coordinate formula for H_* that will take the coordinates of L to the coordinates of $H(L)$. We will take (U, Φ) for a coordinate system on M and (V, Ψ) for a coordinate system on N. The coordinats on M will be $(u^1, \ldots, u^m)$ and the coordinates on N will be $(v^1, \ldots, u^n)$. The coordinate form of H will be $h = \Psi \circ H \circ \Phi^{-1}$ given by $(v^1(u^1, \ldots, u^m), \ldots, v^n(u^1, \ldots, u^m))$. $f = F \circ \Psi^{-1}$ will be the coordinate form of F and $\tilde{f} = F \circ H \circ \Phi^{-1}$ will be the coordinate form of $F \circ H$. Then we have

$$\tilde{f} = F \circ H \circ \Phi^{-1} = F \circ \Psi^{-1} \circ \Psi \circ H \circ \Phi^{-1} = f \circ h$$

All this is very much as before. L in coordinate form will be $a^i \frac{\partial}{\partial u^i}$ and $H_*(L)$ in coordinate from will be $b^j \frac{\partial}{\partial v^j}$. Then

$$H_*(L)(F) = L(F \circ H) = a^i \frac{\partial \tilde{f}}{\partial u^i} = a^i \frac{\partial f \circ h}{\partial u^i} = a^i \frac{\partial f}{\partial v^j} \frac{\partial v^j}{\partial u^i}$$

comparing this with

$$H_*(L)(F) = b^j \frac{\partial f}{\partial v^j}$$

we see that

$$b^j = \frac{\partial v^j}{\partial u^i} a^i$$

or in matrix form

$$\begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix} = \begin{pmatrix} \frac{\partial v^1}{\partial u^1} & \cdots & \frac{\partial v^1}{\partial u^m} \\ \vdots & \vdots & \vdots \\ \frac{\partial v^n}{\partial u^1} & \cdots & \frac{\partial v^n}{\partial u^m} \end{pmatrix} \begin{pmatrix} a^1 \\ \vdots \\ a^m \end{pmatrix}$$

From the standpoint of efficiency, it might be better to do this H_* material before the change of coordinate material because then the change of coordinates can be done using the H_* material where $H = I$, the identity mapping. Then the last formula becomes the change of coordinates formula. It would be worth thinking about why this is true. We felt the order we used is a bit more natural if considerably less efficient. Remember we said in the introduction that clarity would trump efficiency.

It is worth mentioning that in general there is no way to pull a vector back from $T_{H(p)}(N)$ to $T_p(M)$. However, if H is a diffeomorphism then it is possible, since then H_* is an isomorphism. This is easily proved from the formula

$$(K \circ H)_* = K_* \circ H_*$$

for $H : M \to N$ and $K : N \to P$. We leave it to the reader to work the details out for herself.

Shorter Methodology

The previous material can be presented in an abbreviated version which is closer to the methods of classical tensor analysis and easier to use for practical purposes. Without the previous material the methodology is suspect, but now that we have the details it is easy to show how to abbreviate it. The abbreviation consists of removing the absolute part of the development, (the important part for some people) and working entirely in coordinates, a cut-out-the-middleman philosophy. Since both of us use these abbreviated methods daily, it seems reasonable to present them and also clarify the relationship of the previous development to classical tensor analysis.

A tangent vector is presented as a derivation on the set $C^\infty(U)$ of smooth functions on some coordinate patch (U, Φ) containing the point $p \in M$ of interest. In our new formalism we present the functions only in their coordinate form $f(u^1, \ldots, u^m)$. The vectors/derivations are presented only in their coordinate forms $v = a^i \frac{\partial}{\partial u^i}$. The mappings from an m-manifold M to an n-manifold N are denoted by a new letter $\phi : M \to N$ and presented as

$$(v^1, \ldots, v^n) = \phi(u^1, \ldots, u^m)$$

which in practise takes the form

$$w^j = w^j(u^1, \ldots, w^m) \qquad j = 1, \ldots, m$$

where (W, Υ) is a coordinate patch around $\phi(p)$ (note abuse of notation).

Now if we wish to change the coordinates from $(u^1, \ldots, u^m)$ to $(v^1, \ldots, v^m)$ we can immediately see how the coordinates change. With

$$v = a^i \frac{\partial}{\partial u^i} = \tilde{a} \frac{\partial}{\partial v^i}$$

we have

$$\begin{aligned} a^i \frac{\partial}{\partial u^i} &= a^i \frac{\partial}{\partial v^j} \frac{\partial v^j}{\partial u^i} = \Big(a^i \frac{\partial v^j}{\partial u^i} \Big) \frac{\partial}{\partial v^j} \\ &= \tilde{a}^j \frac{\partial}{\partial v^j} \end{aligned}$$

Comparing, we see that we have rederived

$$\tilde{a}^j = a^i \frac{\partial v^j}{\partial u^i}$$

The point is that by writing the vectors as first order differential operators and by knowing the form the vectors take as differential operators, the change of coordinate formulas fall out like a stone.

Now lets get ϕ_*. We know $\phi_*(a^i \frac{\partial}{\partial u^i}) = b^j \frac{\partial}{\partial w^i}$. How do we get b^j in terms of a^i? Well $w^j = w^j(u^1, \ldots, u^m)$ so

$$\begin{aligned} b^j \frac{\partial}{\partial w^i} &= \phi_*(a^i \frac{\partial}{\partial u^i}) = a^i \phi_* \Big(\frac{\partial}{\partial u^i} \Big) \\ &= a^i \frac{\partial w^j}{\partial u^i} \frac{\partial}{\partial w^j} \end{aligned}$$

where of course you have to know that $\phi_*\left(\frac{\partial}{\partial u^i}\right)$ acts like a chain rule. From this we see that in coordinates ϕ_* is given by

$$b^j = a^i \frac{\partial w^j}{\partial u^i}$$

Note how similar change of coordinates and ϕ_* turn out. This is because the formalism thinks that the coordinates on N are just another set of coordinates on M after application of ϕ. The formalism is wrong about this, because the two manifolds may have different dimensions, but we can sort of see why it makes this mistake.

Anyway, the point is that the way we have set things up it is remarkably simple to derive formulas of these types once you have gotten used to it. The above is not pure tensor material because of the use of $\partial/\partial u^i,\ i = 1, \ldots .n$ as the basis of the tangent space, but it should be clear from the above how handy an addition this is.

2.6.2 Definition of Vector Bundle

We want to give a quick introducton to the idea of a vector bundle so as to put the Tangent Bundle in context. A vector bundle is a sort of generalization of a cartesian product with special features all its own. Vector bundles are ubiquitous in modern mathematics and physics and the idea *section of a vector bundle* is the mathematical equivalent of vector or tensor field. Fortunately the concepts we need for vector bundles are fairly easy. The letters we use are the standard letters in vector bundle theory.

Let M be a manifold of dimension m and E be a vector space of dimension n. Let $\mathcal{E}$ be a manifold of dimension $m + n$ and $\pi : \mathcal{E} \to M$ be an onto diffeomorphism. We want to define $\mathcal{E}$ is a vector bundle over M with Fibre E. There are two things to look at. First, there is a covering of M by coordinate neighbourhoods U_α so that over each U_α $\mathcal{E}$ is locally isomorphic to a product, and second, we must look at what happens when move from one coordinate patch U_α to another U_β.

For the local isomorphism, we require for each of the coordinate patches U we are using to cover M, we have a diffeomorphism

$$\phi : \pi^{-1}[U] \to U \times E$$

which is linear on each fibre $\pi^{-1}[p]$. This implies $\pi^{-1}[p]$ is isomorphic to $\{p\} \times E$ which is isomorphic to E. Also, if $\pi_1 : U \times E \to U$ is projection onto the first coordinate of $U \times E$, that is $\pi_1(p, e) = p$, then we can factor $\pi = \pi_1 \circ \phi$. We use ξ for elements of $\mathcal{E}$ so if $\xi \in \pi^{-1}[U] \subseteq \mathcal{E}$ and $\pi(\xi) = p$ then $\phi(\xi) = (p, e)$ for some $e \in E$.

Now suppose that $p \in U \times V$ for U and V two coordinate patches in the cover of M. We have a ϕ_U for U and a ϕ_V for V. For ξ lying above p (this means $\pi(\xi) = p$) we have

$$\phi_U(\xi) \quad = \quad (p, e_U)$$

$$\phi_V(\xi) \quad = \quad (p, e_V)$$

and we require, for each pair U, V, the existence of a *linear isomorphism* $\phi_{UV} : E \to E$ for which $\phi_{UV}(e_U) = e_V$. (Note: conventions vary on whether this is called ϕ_{UV} or ϕ_{VU}; we use the former.)

Now, you ask, what happens if $p \in U \cap V \cap W$? Might their be inconsistencies? We can prevent this bad thing by requiring the consistency condition

$$\phi_{WU} \circ \phi_{VW} \circ \phi_{UV} = \text{Identity}_E \qquad \text{The cocycle condition}$$

(Note: This particular equation comes up in the theory of Čech cohomology and is there called the *cocycle condition*. Tragically, we cannot follow this path. "Maybe tragically for YOU," says the coauthor.) From this a couple of things easily follow. If we set $U = W$ in the cocycle condition we get, using the obvious ϕ_{UU} =Identity,

$$\begin{aligned} \phi_{UU} \circ \phi_{VU} \circ \phi_{UV} &= \text{Id}_E \\ \phi_{VU} \circ \phi_{UV} &= \text{Id}_E \\ \phi_{VU} &= \phi_{UV}^{-1} \end{aligned}$$

and now from the cocycle conditon we have

$$\begin{aligned} \phi_{UW} \circ \phi_{WU} \circ \phi_{VW} \circ \phi_{UV} &= \phi_{UW} \circ \text{Id}_E \\ \phi_{VW} \circ \phi_{UV} &= \phi_{UW} \end{aligned}$$

which last equation is the pyschologically more satisfying form of a consistency condition. From it, by reversing the derivation, one can obtain the cocycle condition, which we leave to the reader to verify.

To sum up, a vector bundle is an epimorphic (onto map) diffeomorphism $\pi : \mathcal{E}\ toM$ which is has locally the structure of a product and whose maps ϕ_{UV} satisfy the cocycle condition.

A critical concept for vector bundles is that of *section*.

Def A map $\sigma : M \to \mathcal{E}$ is a *section* $\iff \pi \circ \sigma = Id_M$

Thus a section chooses an element $\sigma(p)$ in the fibre above p in a smooth manner.

Section is the mathematical term for vector field or tensor field. For example, consider an electric field $E(t, x, y, z)$ in the space time manifold we live in. Where does $E(t, x, y, z)$ live? It cannot live in the space time manifold because this is *curved*. So it lives in a 3-dimenstional vector bundle whose underlying manifold is the space time manifold, and the electric field is a *section* of this vector bundle.

Now we want to set things up so we can put coordinates on a vector bundle. We will construct a *local basis of sections* of a vector bundle. We cannot too strongly emphasize that the basis is local. Only rather dull vector bundles have global bases of sections. As before, let (U, Φ) be a local coordinate patch of the type to have a $\phi : \pi^{-1}[U] \to U \times E$ and let $(e_1, \ldots, e_n)$ be a fixed basis for E. We define a section $\sigma : M \to \mathcal{E}$ by

$$\sigma_i(p) = \phi^{-1}(p, e_i)$$

Then $(\sigma_1, \ldots, \sigma_n)$ is called a *local basis of sections of the vector bundle.* Since ϕ is linear on the fibres, we have

if $\phi(\xi) = (p, \alpha^i e_i)$

then $\xi = \alpha^i \sigma_i$

so we have $\mathcal{E}(U) = \pi^{-1}[U])$ doing a fine imitation of a vector space. Recall that (U, Φ) was a coordinate patch on M where $\Phi(p) = (u^1, \ldots, u^m)$. We can now introduce coordinates which depend on Φ and the choice of basis $(e_1, \ldots, e_n)$ of E. We then have

$$\text{The coordinates of } \xi \text{ are } (u^1, \ldots, u^m, \alpha^1, \ldots, \alpha^n)$$

This kind of coordinate system is often called *bundle coordinates.*

We emphasize again that the local basis of sections is truly local; for anything but the simplest vector bundles there is *no* global basis of sections. If a global basis of sections exists, then one easily proves that $\mathcal{E}$ is isomorphic to $M \times E$ and thus is a simple product. Using a couple of neighbourhoods U_α, the reader might try to find a global basis of sections for the Möbius band, which is *not* the product of the circle and [-1,1].

2.6.3 The Tangent Bundle

In a general vector bundle there is no connection between the choice of coordinate system (U, Φ) for a patch and the basis $(e_1, \ldots, e_n)$ for E. In some bundles, however, there is a connection and the tangent bundle is one of those for which this is natural.

The construction of the tangent bundle is obvious. The fibre over each $p \in M$ is the tangent space $T_p(M)$ at that point.

Def The Tangent Bundle $T(M) = \bigcup_{p \in M} T_p(M)$

The epimorphism (onto map) $\pi : T(M) \to M$ is defined by

if $\xi \in T_p(M)$ then $\pi(\xi) = p$.

Since each of the tangent spaces $T_p(M)$ in an m-dimensional vector space, it is natural to take $E = \mathbb{R}^m$. If (U, Φ) is a coordinate patch of sufficiently "small size", we can define ϕ in what seems rather backwards after the last section. If $\Phi(p) = (u^1, \ldots, u^m)$ then there is a natural basis for the tangent space consisting of $(\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^m})$ as we saw in the section on the tangent space. For $\xi \in T(M)$ then $\xi \in T_{\pi(\xi)}$ and we define ϕ by

$$\begin{aligned} \xi &= \alpha^i \frac{\partial}{\partial u^i} \\ \phi(\xi) &= (p, (\alpha^1, \ldots, \alpha^m)^\top) \in M \times E \end{aligned}$$

Next we must consider ϕ_{UV}. Let (U, Φ_U) and (V, Φ_V) be two overlapping coordinate systems and $p \in U \cap V$ Let ϕ_U, ϕ_V the correspoinding local trivializers. Let $\xi \in T(M)$ and $\pi(\xi) = p$. Then

$$\phi_U(\xi) = (p, (\alpha^1, \ldots, \alpha^m)^\top))$$

$$\phi_V(\xi) \quad = \quad (p, (\beta^1, \ldots, \beta^m)^\top))$$

where

$$\xi = \alpha^i \frac{\partial}{\partial u^i} = \beta^j \frac{\partial}{\partial v^j}$$

But we have

$$\alpha^i \frac{\partial}{\partial u^i} = \alpha^i \frac{\partial v^j}{\partial u^i} \frac{\partial}{\partial v^j}$$

so, comparing, we see

$$\beta^j = \frac{\partial v^j}{\partial u^i} \alpha^i$$

or in matrix form

$$\begin{pmatrix} \beta^1 \\ \vdots \\ \beta^m \end{pmatrix} = \begin{pmatrix} \frac{\partial v^1}{\partial u^1} & \cdots & \frac{\partial v^1}{\partial u^m} \\ \vdots & \vdots & \vdots \\ \frac{\partial v^m}{\partial u^1} & \cdots & \frac{\partial v^m}{\partial u^m} \end{pmatrix} \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^m \end{pmatrix}$$

So we see that ϕ_{UV} is simply the Jacobian matrix

$$\phi_{UV} = \frac{\partial(v^1, \ldots, v^m)}{\partial(u^1, \ldots, u^m)}$$

Once again we have derived the tensor change law for vectors = type(1,0) tensors. The reader might find it amusing to contemplate what the cocycle condition amounts to here.

Now we see immediately that a vector field is a section of the Tangent Bundle. Locally the vector field will be described as

$$\alpha^i(u^1, \ldots, u^m) \frac{\partial}{\partial u^i}$$

and this description is often called *natural coordinates*.

2.7 Differential Forms; What Are They Really?

2.7.1 Linear Functionals

This subsection deals with bases for a vector space V and its dual space V^* and what happens when the coordinates are changed. If you are comfortable with this material you can skip the section.

To put differential forms in context we need to talking briefly about linear functionals. The definition is

Def Let V be a vector space and let $\ell : V \to \mathbb{R}$. Then ℓ is a *linear functional* if and only if

1) $\ell(\alpha v) = \alpha \ell(v) \qquad$ for $\alpha \in \mathbb{R}$

2) $\ell(u+v) = \ell(u) + \ell(v)$

The two requirements for a linear functional can be combined into the single requirement $\ell(\alpha u + \beta v) = \alpha\ell(u) + \beta\ell(v)$.

It should be fairly obvious that the linear functionals on V form a vector space. To find the dimension, we resort to a basis, and restrict ourselves to finite dimensional vectors spaces. Then if $\{\vec{e}_1, \cdots, \vec{e}_n\}$ is a basis for V then the vector space V^* of linear functionals will also have dimension n. We remark that this is not true for infinite dimensional vector spaces where the cardinality of of V^*, which is then defined as the space of *continuous* linear functionals, is often larger than the cardinality of V. This suggests that there is no proof that $\dim(V)$=$\dim(V^*)$ without using a basis.

For the finite dimensional space V, each vector v has a unique expression $\vec{v} = v^1\vec{e}_1 + \ldots + v^n\vec{e}_n$ with $v^i \in \mathbb{R}$. We can then define a *dual basis* $\{\underline{e}^1, \ldots, \underline{e}^n\}$ by

$$\underline{e}^j(v) = v^j$$

Notice that $\underline{e}^j(\vec{e}_k) = \delta^j_k$ with δ^j_k being the Kronecker delta. Then for any $\ell \in V^*$ we can get a representation of ℓ in terms of the dual basis by using $\ell_j = \ell(\vec{e}_j)$ and then $\ell = \ell_j\underline{e}^j$. We then have (summation convention in effect!)

$$\ell(\vec{v}) = \ell_j\underline{e}^j(v^k\vec{e}_k) = \ell_j v^k\underline{e}^j()\vec{e}_k) = \ell_j v^k\delta^j_k = \ell_j v^j$$

and this may be represented matricially by

$$\ell(\vec{v}) \longleftrightarrow (\ell_1, \ldots, \ell_n)\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$$

Notice by using a matrix of vectors $(\vec{e}_1, \ldots, \vec{e}_n)$ we may write

$$\vec{v} = (\vec{e}_1, \ldots, \vec{e}_n)\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$$

which, as we will see, can be handy for changing coordinates.

The last thing we need consider in this subsection is what happens when we change the basis. Using the notation introduced above this can be done very efficiently.

Let $\{\vec{f}_1, \ldots, \vec{f}_n\}$ be a new basis for V and express the f_i in terms of the e_j by (summation convention in effect)

$$\vec{f}_i = \gamma^j_i\vec{e}_j$$

or, more explicitly,

$$\begin{array}{rcl} \vec{f}_1 & = & \gamma_1^1\vec{e}_1 + \ldots + \gamma_1^n\vec{e}_n \\ \vdots & \vdots & \vdots \\ \vec{f}_n & = & \gamma_n^1\vec{e}_1 + \ldots + \gamma_n^n\vec{e}_n \end{array}$$

This can be written matricially as

$$(\vec{f}_1, \ldots, \vec{f}_n) = (\vec{e}_1, \ldots, \vec{e}_n) \begin{pmatrix} \gamma_1^1 & \ldots & \gamma_n^1 \\ \ldots & \ldots & \ldots \\ \gamma_1^n & \ldots & \gamma_n^n \end{pmatrix}$$

It is *most critical* to notice that the array of α's in the matrix equation is the transpose of the array in the equations above. This has to do with writing the matrix of vectors as a row rather than a column.

Now the change of basis formula becomes easy. With $\vec{v}$ written in the new basis as $\vec{v} = \beta^j \vec{f}_j$ we have

$$(\vec{f}_1, \ldots, \vec{f}_n) \begin{pmatrix} \beta^1 \\ \vdots \\ \beta^n \end{pmatrix} = \vec{v} = (\vec{e}_1, \ldots, \vec{e}_n) \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^n \end{pmatrix}$$

$$(\vec{e}_1, \ldots, \vec{e}_n) \begin{pmatrix} \gamma_1^1 & \ldots & \gamma_n^1 \\ \ldots & \ldots & \ldots \\ \gamma_1^n & \ldots & \gamma_n^n \end{pmatrix} \begin{pmatrix} \beta^1 \\ \vdots \\ \beta^n \end{pmatrix} = (\vec{e}_1, \ldots, \vec{e}_n) \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^n \end{pmatrix}$$

which shows, (since $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is a basis), that

$$(\gamma_j^i) \begin{pmatrix} \beta^1 \\ \vdots \\ \beta^n \end{pmatrix} = \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^n \end{pmatrix}$$

and thus

$$\begin{pmatrix} \beta^1 \\ \vdots \\ \beta^n \end{pmatrix} = (\gamma_j^i)^{-1} \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^n \end{pmatrix}$$

which is the standard change of basis formula.

Now we must derive a similar formula for linear forms, which is easy. Let $(\vec{e}_1), \ldots, \vec{e}_n)$ and $(\vec{f}_1), \ldots, \vec{f}_n)$ be two bases of V and (γ_j^i) be the change of basis matrix so that $\vec{f}_i = \gamma_i^j \vec{e}_j$. Let $(\underline{e}^1), \ldots, \underline{e}^n)$ and $(\underline{f}^1), \ldots, \underline{f}^n)$ be the dual bases. Let λ be a linear functional and let

$$\lambda = \alpha_i \underline{e}^i = \beta_j \underline{f}^j$$

Notice that

$$\lambda(\vec{e}_k) = \alpha_i \underline{e}^i(\vec{e}_k) = \alpha_i \delta^i_k = \alpha_k$$

and similarly $\lambda(\vec{f}_j) = \beta_j$. Next we have

$$\beta_j = \lambda(\vec{f}_j) = \lambda(\gamma^k_j \vec{e}_k) = \gamma^k_j \alpha_k$$

which we can write as

$$(\beta_1, \ldots, \beta_n) = (\alpha_1, \ldots, \alpha_n) \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_n \\ \cdots & \cdots & \cdots \\ \gamma^n_1 & \cdots & \gamma^n_n \end{pmatrix}$$

Now we multiply on the right and eventually peel off on the left:

$$(\beta_1, \ldots, \beta_n) \begin{pmatrix} \underline{f}^1 \\ \vdots \\ \underline{f}^n \end{pmatrix} = (\alpha_1, \ldots, \alpha_n) \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_n \\ \cdots & \cdots & \cdots \\ \gamma^n_1 & \cdots & \gamma^n_n \end{pmatrix} \begin{pmatrix} \underline{f}^1 \\ \vdots \\ \underline{f}^n \end{pmatrix}$$

But

$$(\beta_1, \ldots, \beta_n) \begin{pmatrix} \underline{f}^1 \\ \vdots \\ \underline{f}^n \end{pmatrix} = \lambda = (\alpha_1, \ldots, \alpha_n) \begin{pmatrix} \underline{e}^1 \\ \vdots \\ \underline{e}^n \end{pmatrix}$$

giving us

$$(\alpha_1, \ldots, \alpha_n) \begin{pmatrix} \underline{e}^1 \\ \vdots \\ \underline{e}^n \end{pmatrix} = (\alpha_1, \ldots, \alpha_n) \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_n \\ \cdots & \cdots & \cdots \\ \gamma^n_1 & \cdots & \gamma^n_n \end{pmatrix} \begin{pmatrix} \underline{f}^1 \\ \vdots \\ \underline{f}^n \end{pmatrix}$$

Since the $(\alpha_1, \ldots, \alpha_n)$ could have any values, we may peel them off from the equation to get

$$\begin{pmatrix} \underline{e}^1 \\ \vdots \\ \underline{e}^n \end{pmatrix} = \begin{pmatrix} \gamma^1_1 & \cdots & \gamma^1_n \\ \cdots & \cdots & \cdots \\ \gamma^n_1 & \cdots & \gamma^n_n \end{pmatrix} \begin{pmatrix} \underline{f}^1 \\ \vdots \\ \underline{f}^n \end{pmatrix}$$

or to write it in non-matricial form

$$\underline{e}^j = \gamma^j_k \underline{f}^k$$

2.7.2 Linear Functionals on the Tangent Space

To finish the section we must relate the above material to the tangent space $T_p(M)$. We reiterate the conventions that we used in secton 2.4. (U, Φ) is a coordinate patch using the variables $(u^1, \ldots, u^n)$. (V, Ψ) is a second such patch using the variables $(v^1, \ldots, v^n)$. $F : U \to \mathbb{R}$ is a smooth function. $F \circ \Phi^{-1}$

is a function from $\Phi[U] :\to \mathbb{R}$ which we will denote, by abuse of notation, by $F(u^1, \ldots, u^n)$. Let $p \in M$ be any point and we arrange that $\Phi(p) = 0$ and $|psi(p) = 0$. The tangent space $T_p(M)$ is made up of first order differential operators which have a basis $L^1, \ldots, L^n$ definted by

$$L_i(F) = \frac{\partial}{\partial u^i} F \circ \Phi^{-1}\Big|_{u^j=0}$$

We abbreviate this by

$$L_i(F) = \frac{\partial}{\partial u^i} F(u^1, \ldots, u^n)$$

which cleans up the notation a bit. We tacitly require that after the differentiation the variables u^j are all set to 0 which localizes us at p.

The whole thing can be repeated for the patch (V, Ψ) and we get $\tilde{L}_i(F) = \frac{\partial}{\partial v^i} F(v^1, \ldots, v^n)$

The map $\Psi \circ \Phi^{-1}$ is given by $(v^1(u^1, \ldots, u^n), \ldots, v^n(u^1, \ldots, u^n))$ Then we have, with the usual advanced Calculus notation

$$\tilde{L}_i(F) = \frac{\partial}{\partial v^i} F(v^1, \ldots, v^n) = \frac{\partial u^j}{\partial v^i} \frac{\partial}{\partial u^j} F(u^1, \ldots, u^n) = \frac{\partial u^j}{\partial v^i} L_i(F)$$

We know that the $\tilde{L}_i$ and L_j, $i, j = 1, \ldots, n$ form bases of $T_p(M)$. For convenience we will now start to use, as is the common habit, $\frac{\partial}{\partial v^i}$ for $\tilde{L}_i$ and $\frac{\partial}{\partial u^j}$ for L_j. Then we can write the previous equation as

$$\frac{\partial}{\partial v^i} = \frac{\partial u^j}{\partial v^i} \frac{\partial}{\partial u^j}$$

Since $(\frac{\partial}{\partial v^1}, \ldots, \frac{\partial}{\partial v^n})$ and $(\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^n})$ are each *bases* of $T_p(M)$, we see $(\frac{\partial u^j}{\partial v^i})$ revealed as the *change of basis matrix.*

Now we need a notation for the dual basis of the basis $(\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^n})$.

Def The dual basis to the basis $(\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^n})$ of $T_p(M)$ will be denoted by

$$(du^1, \ldots, du^n)$$

so that, as is usual for dual bases,

$$du^i(\frac{\partial}{\partial u^j}) = \delta^i_j \qquad \text{(Kronecker delta)}$$

So now you know what du^i is; it is a linear functional on the tangent space. If a tangent vector is $\vec{v} = a^i \frac{\partial}{\partial u^i}$ then

$$du^j(\vec{v}) = du^j(a^i \frac{\partial}{\partial u^i}) = a^i \delta^j_i = a^j$$

Now we apply the formulas from the first subsection to our situation, specifically the formula $\underline{e}^j = \gamma^j_k \underline{f}^k$ which becomes in our situation

$$du^j = \frac{\partial u^j}{\partial v^k} \, dv^k$$

which is magnificently consistent with our usual formulas.

2.7.3 Differential Forms

Everything we have done in this section has been done at a single point, but all the formulas are applicable at any point. In a coordinate patch (U, Φ) we may write an element of the dual of the tangent space in terms of the local basis $\alpha_i(u^1, \ldots, u^n)\, du^i$. Now we can make the p variable so we get an object of the form $\alpha_i(p) du^i$ where of course the du^i is an object in the dual $T_p(M)^*$ of $T_p(M)$. We of course insist that $\alpha(p)$ be a smooth function on M. We have now constructed a first order differential form and we see that it is a section of $T_p(M)^*$.

We need some terminology to describe the situation efficiently. We already have the tangent space at p is $T_p(M)$. The space of linear functionals on $T_p(M)$ will be called $T_p^*(M)$ or $T_p(M)^*$. We also need to think about the the situation with $\alpha(p) du^i$ which lived in a new object.

Def The cotangent bundle of M is

$$\Lambda^1(M) = \bigcup_{p \in M} T_p(M)^*$$

where the $\alpha(p)$ are smooth ($\mathrm{C}^\infty(M)$).

There is an analogous definition for $\Lambda^1(U)$, $U \subseteq M$ where we require the $\alpha(p)$ to be smooth only on U, where U is an open subset of M.

Now for each $p \in M$, $T_p(M)^*$ is a vector space. Using the techniques we learned in the section on Grassmann Algebra, we may construct from the vector space $T_p(M)^*$ of dimension m the Grassmann Algebra layers $\Lambda^k(T_p(M)^*)$ of dimension $\binom{m}{k}$ and the entire Grassmann Algebra $\Lambda(T(M)^*)$ of dimension 2^m.

From these objects it is easy to construct the corresponding bundles.

Def The *The cotangent k-bundle of the cotangent bundle* is

$$\Lambda^k(M) = \bigcup_{p \in M} \Lambda^k(T_p(M)^*)$$

Def A k-form is a section of the cotangent k-bundle.

A 1-form may be written as

$$\sum_{i=1}^{m} \alpha_i(u^1, \ldots, u^m)\, du^i$$

and a k-form may be written as

$$\sum_{i_1, i_2, \ldots, i_k = 1}^{m} \alpha_{i_1, i_2, \ldots, i_k}(u^1, \ldots, u^m)\, du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_k}$$

in which case the coeeficients are usually taken to be skew symmetric in their indices, or

$$\sum_{i_1 < i_2 < \ldots < i_k} \alpha_{i_1 < i_2 < \ldots < i_k}(u^1, \ldots, u^m)\, du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_k}$$

or

$$\sum_{\sigma\in\mathcal{S}_{n,k}} \alpha_\sigma \, du^{\sigma(1)} \wedge du^{\sigma(2)} \wedge \ldots \wedge du^{\sigma(k)}$$

The first expression is consistent with tensor notation but is hard to work with, the second is also hard to work with although with patience and dedication one can fight ones way through the calculations, and the third is the natural way way to express a k-form. However, for technical reasons we will use a slight modification of the third form which will be

$$\sum_{\sigma\in\mathcal{S}_{n,k}} \alpha_\sigma \operatorname{sgn}(\sigma) \, du^{\sigma(1)} \wedge du^{\sigma(2)} \wedge \ldots \wedge du^{\sigma(k)}$$

As an example of the last, we give a 2-form in $\mathbb{R}^3$ as

$$\alpha_\iota \, du^1 \wedge du^2 - \alpha_\sigma \, du^1 \wedge du^3 + \alpha_\tau \, du^2 \wedge du^3$$

where $\iota = (12|3)$, $\sigma = (13|2)$, $\tau = (23|1)$ are the three permuations in $\mathcal{S}_{3,2}$. Note this is coming out in cyclic order due to the presence of $\operatorname{sgn}(\sigma)$.

2.7.4 Behavior of Differential Forms under Mappings

If we have a mapping f between two manifolds M and N, would would expect to find some connection between the differential forms on the two manifolds. There is a natural connection but it goes backwards from what we might expect. Here is what happens:

$$\begin{array}{rcll} f : M & \to & N & \\ f^* : \Lambda(T(N)) & \to & \Lambda(T(M)) & \text{Note reversal of order!} \end{array}$$

The description of f^* is very simple; it is essentially substitution! Let $(U), \Phi)$ be a coordinate patch on M with coordinates $u^1, \ldots u^m$ on M and and for some $x \in U$ let (V, Ψ) be a coordinate patch around $f(x) \in N$. Then locally we have $v^i = v^i(u^1, \ldots u^m)$. Locally a differential form on N looks like

$$\omega = \sum_{\sigma\in\mathcal{S}_{n,k}} a_\sigma(v^1, \ldots, v^n) \operatorname{sgn}(\sigma) \, dv^{\sigma(1)} \wedge \ldots \wedge dv^{\sigma(k)}$$

which we write more conveniently as

$$\omega = \sum_{\sigma\in\mathcal{S}_{n,k}} a_\sigma(v^i) \, dv^\sigma$$

Now we define $f^*\omega \in \Lambda^k(T(M))$ locally by

$$f^*\omega = \sum_{\sigma\in\mathcal{S}_{n,k}} a_\sigma(v^i(u^j)) \operatorname{sgn}(\sigma) \, \frac{\partial v^{\sigma(1)}}{\partial u^{i_1}} \, du^{i_1} \wedge \ldots \wedge \frac{\partial v^{\sigma(k)}}{\partial u^{i_k}} \, du^{i_k}$$

where $i_1, \ldots, i_k$ all run from one to m. Naturally there will be a lot of calculation necessary to grind this down into a proper looking form

$$f^*\omega = \sum_{\tau \in \mathcal{S}_{m,k}} b_\sigma(u^i)\, du^\tau$$

This is how f^* works in terms of coordinates, and this is how you use for a practical calculation. However, this dependence on coordinates is now considered a trifle crude, and we should try to find some coordinate independent definition.

2.7.5 Duality between $\Lambda^k V^*$ and $\Lambda^k V$

In order, in the next section, to understand the interaction of geometric objects like rectangles and parallelepipeds with differential forms it is necessary to look at the duality between $\Lambda^k T(M)^*$ and $\Lambda^k T(M)$. This was essentially worked out by Grassmann in 1842, but Grassmann did not have the concept of the dual space to work with, and we have opted for a slightly more modern approach using linear functionals. However, this is a triviality of presentation and essentially we are using the critical formulas developed by Grassmann.

In this section we will present just the algebraic theory. The application to geometry, manifolds and physics and will be discussed in the following section.

We will first present the formula. Essentially Grassmann's formula is a *definition* of the duality between $\Lambda^k V$ and $\Lambda^k V^*$ but we will discuss why it is a good choice.

We recall the alternate way of writing a linear functional as $< \ell, v >$ rather than $l(v)$, This alternate form is more symmetric, bringing out the dual relation between V and V^* more clearly, but it also has great typographical advantages for the material that follows.

Recall that we already have the duality for $k = 1$, which is simply $< \ell, v >= l(v)$. We need to generalize this for $k > 1$. Let $\ell^1, \ldots, \ell^k \in V^*$ and $v_1, \ldots, v_k \in V$. Then Grassmann's formula is

$$< \ell^1 \wedge \ldots \ell^k,\, v_1 \wedge \ldots \wedge v_k >= \det \begin{pmatrix} < \ell^1, v_1 > & < \ell^1, v_2 > & \ldots & < \ell^1, v_k > \\ \vdots & \vdots & \vdots & \vdots \\ < \ell^k, v_1 > & < \ell^k, v_2 > & \ldots & < \ell^k, v_k > \end{pmatrix}$$

Since $\Lambda^k V^*$ consists of elements which are sums of terms of the form $\ell^1 \wedge \ldots \ell^k$ and $\Lambda^k V$ consists of elements which are sums of the form $v_1 \wedge \ldots \wedge v_k$ we extend the duality to the entire spaces by linearity, which is possible because the determinant is linear in each row and each column.

To get more insight into how the duality formula works we will have a look at how it works on bases. Recall that with a basis $\{e_1, \ldots, e_n\}$ of V we have the dual basis $\{e^1, \ldots, e^n\}$ of V^* where the e^i are defined by

$$< e^j, e_i >= \delta_i^j$$

This has the effect that if $v = v^i e_i$ then

$$< e^j, v >=< e^j, v^i e_i >= v^i < e^j, e_i >= v^i \delta_i^j = v^j$$

Now exactly the same thing can be done for V^*, that is if $\ell = \ell_j e^j$ then

$$< \ell, e_i >=< \ell_j e^j, e_i >= \ell_j < e^j, e_i >= \ell_j \delta^j_i = \ell_i$$

This duality of form extends through the whole theory so that if we prove one of the pair the other can be assumed proved also.

Next recall our notation for bases in $\Lambda^k(V)$. A basis for $\Lambda^k(V)$ is given by

$$e_\sigma = \mathrm{sgn}(\sigma)\ e_{\sigma(1)} \cdots e_{\sigma(k)} \in \Lambda^k V \qquad \text{for } \sigma \in \mathcal{S}_{n,k}$$

(Be careful not to forget the $\mathrm{sgn}(\sigma)$ coefficient of e_σ.) Similarly we have

$$e^\tau = \mathrm{sgn}(\tau)\ e^{\tau(1)} \cdots e^{\tau(k)} \in \Lambda^k V^* \qquad \text{for } \sigma \in \mathcal{S}_{n,k}$$

We are interested in the interaction between the two kinds of basis elements.

$$< e^\tau, e_\sigma >= \mathrm{sgn}(\tau)\mathrm{sgn}(\sigma) \begin{vmatrix} < e^{\tau(1)}, e_{\sigma(1)} > & < e^{\tau(1)}, e_{\sigma(2)} > & \dots & < e^{\tau(1)}, e_{\sigma(k)} > \\ < e^{\tau(2)}, e_{\sigma(1)} > & < e^{\tau(2)}, e_{\sigma(2)} > & \dots & < e^{\tau(2)}, e_{\sigma(k)} > \\ \vdots & \vdots & \vdots & \vdots \\ < e^{\tau(k)}, e_{\sigma(1)} > & < e^{\tau(k)}, e_{\sigma(2)} > & \dots & < e^{\tau(k)}, e_{\sigma(k)} > \end{vmatrix}$$

Suppose first that $\sigma = \tau$. Then the determinant has 1's on the diagonal and 0's elsewhere, and also $\mathrm{sgn}(\sigma)^2 = 1$, so $< e^\sigma, e_\sigma >= 1$ On the other hand if $\sigma \neq \tau$ then there is a $\sigma(j)$ which is not in $\tau(1), \dots, \tau(k)$. Then the column for $\sigma(j)$ in the determinant is a column consisting only of 0's, and thus the determinant is 0; $< e^\tau, e_\sigma >= 0$. We can now write this, using an obvious extension of the Kronecker delta, as

$$< e^\tau, e_\sigma >= \delta^\tau_\sigma$$

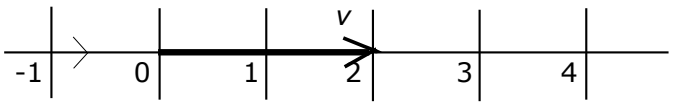
What we have shown is that $\{e^\tau \big|\ \tau \in \mathcal{S}_{k,n}\} \subseteq \Lambda^k(V^*)$ is the *dual basis* of $\{e_\sigma \big|\ \sigma \in \mathcal{S}_{k,n}\} \subseteq \Lambda^k(V)$. This is enough to show that Grassmann's formula indeed gives a duality, which means that if $< L, A >= 0$ for some $L \in \Lambda^k(V^*)$ and *all* $A \in \Lambda^k(V)$ then $L = 0$.

2.7.6 Geometric and Physical Meaning of Differential Forms, Part I

We wish, in this section, to begin the discussion of the geometric and physical meaning of differential forms. However, we do not at this point have enough structure to describe really interesting things like flux. We are limited here to those things that do not require a the space to have a metric. However we can handle more than one might think without a metric. For example, a crude form of density can be handled at this point. For definiteness we will think of density as matter density, although charge density would work as well. Things actually requiring a metric will be discussed in a later section.

The simplest case is the case of one dimension. We will discuss this in some detail because the ideas are quite clear in this case. We consider the x-axis as

a one dimensional vector space and we choose a basis element $\frac{\partial}{\partial x} = e$. Recall that dx is then defined by $\langle dx, \frac{\partial}{\partial x}\rangle = \langle dx, e\rangle = 1$ or in less symmetric notation $dx(\frac{\partial}{\partial x}) = dx(e) = 1$. Let $v = 2e$. Then clearly $dx(v) = 2$ and we want to represent this geometrically.

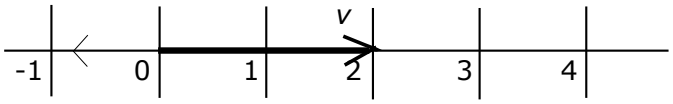


Geometric representation of $dx(v)$

Here dx is represented by the sequence of vertical lines which are erected at the points ke, $k \in \mathbb{Z}$. To find $dx(v)$ one counts the number of intervals across which v stretches, and that gives the value of $dx(v)$ which in this case is 2.

This is the basic idea, but in the geometric representation game there are always details. Note the faint arrowhead between -1 and 0. (This sense arrowhead can be placed anywhere on the x axis.) It gives the sense of dx. Then if $w = -3e$, then w will point in the opposite direction to the sense arrow, and this means that $dv(w) = -3$).

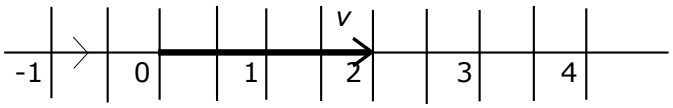
Now we ask, how do we represent $-dx$? This is easy; just reverse the faint arrow.



Geometric representation of $(-dx)(v)$

Here v continues to cover two intervals but the senses are opposite so $(-dx)(v) = -2$. With $-dx$ and w the senses coincide and thus $(-dx)(w) = (-dx)(-3e) = 3$.

Now let us add density of mass into the picture. The base vector e gives a unit length on the x-axis and thus we can use this to make a mass density. Suppose the density is 2. This is represented analytically by $2dx$ and geometrically by doubling the number of vertical lines.



Geometric representation of $(2dx)(v)$

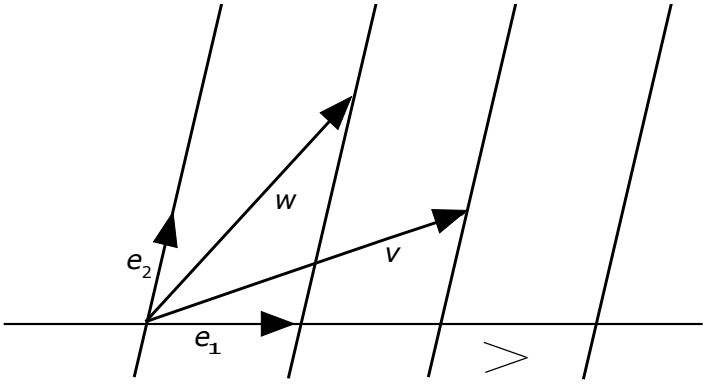
You now find the mass by again counting the number of intervals that v goes across, which in this case is 4, which clearly is correct for the mass.

If the mass has variable density, then the frequency of the verical lines will change as we move along the x-axis.

This completes what we need for the one dimensional case. The reader will easily be able to construct any further elaboration of this case from what has been presented. The takeaway here is that differentials are represented by series

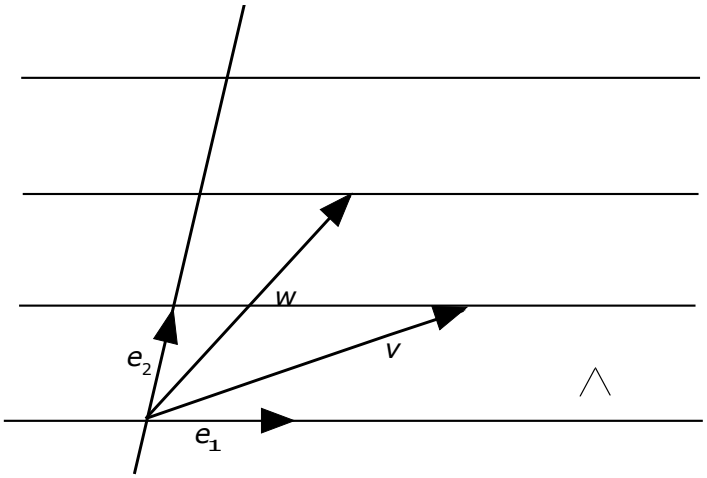
of vertical lines and the value of the differential on a vector is found by counting intervals crossed.

Now we want to look at a differential for dx in a 2-dimensional space. Not too surprisingly we once again have a series of "vertical" lines but they are now parallel to e_2. The angle between the x and y axes is not determinable because there is no inner product with which to measure angles. (I have picked an arbitrary angle for purposes of illustration. It is unwise to use a right angle as it can give false impressions to the unwary.)



Geometric representation of dx

From the illustration is is clear that $dx(v) = 2$ and $dx(w) = 1$ since v crosses two line and w crosses one. The next illustration shows the lines for dy (parallel to e_1).

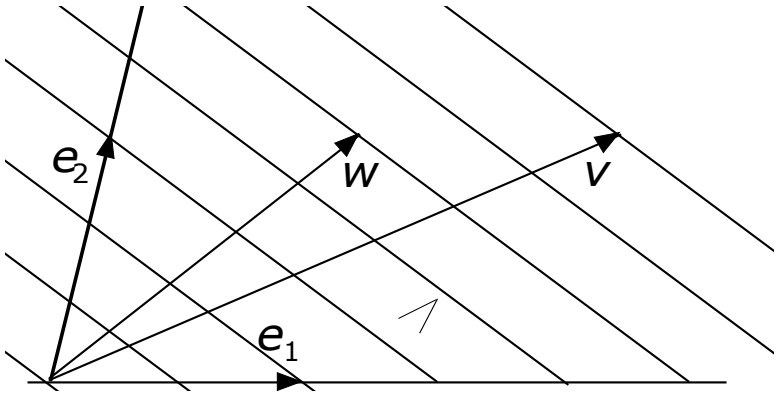


Geometric representation of dy

and we see that $dy(v) = 1$ and $dy(w) = 2$. Thus we know now that $v = 2e_1 + 1e_2$ while $w = 1e_1 + 2e_2$.

The faint arrowheads without bodies in these diagrams indicate the direction of increasing values for the lines. These are as necessary here as they were for the 1-dimensional situation, and it is easy to forget them.

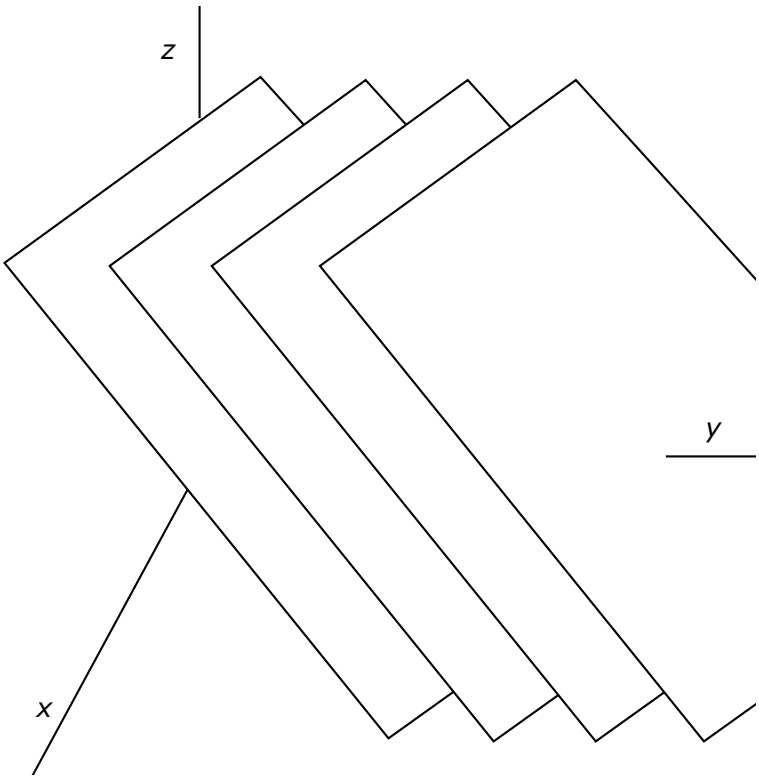
Now you may be wondering how and arbitrary (constant) differential 1-form in a 2-dimensional space may be represented. Let us take as example $\omega = 2dx + 3dy$. We take as reference points the points p on the axes with $\omega(p) = 1$. These would be $\frac{1}{2}e_1$ and $\frac{1}{3}e_2$. Next draw a parallel lines through the origin and then draw a series of equispaced parallel lines using these two as base. This is the series of lines that represents $\omega = 2dx + 3dy$.



Geometric representation of $\omega = 2dx + 3dy$

Now here $v = 2e_1 + e_2$ and $\omega(v) = 2 \cdot 2 + 3 \cdot 1 = 7$ and one sees that v goes through 7 stripes. Similarly $w = e_1 + e_2$ and $\omega(w) = 2 \cdot 1 + 3 \cdot 1 = 5$ and indeed w goes through 5 stripes. (Due to optical illusions it can appear to some people that v is not actually $2e_1 + e_2$ but this impression is incorrect; the illustrations were made using guidelines (no longer visible) and are quite accurate.)

Next you want to know how to represent 1-forms in a 3-dimensional space. Perhaps you have already figured it out, but the illustration may be helpful. We will not do numerical examples here since the general idea should be clear enough; the vector v pierces a certain number of the spaces between the planes and that gives $\omega(v)$.



Geometric representation of 1-form ω in 3-space

This completes our discussion of one forms. Clearly 2-forms and 3-forms will be slightly more difficult, but leveraging what we already know we should be able to navigate the swamp.

We will first have a look at 2-forms. Naturally we must have 2 or more dimensions for the 2-form to live in. Let us consider the easy case of a two form in 2-dimensional space. The basis vectors will be e_1 and e_2 as usual and the dual basis will be dx and dy. The Grassmann Algebra $\Lambda^2(V)$ will be one dimensional, so all elements $A \in \Lambda^2(V)$ will be multiples of a single basis element which we will take to be $e_1 \wedge e_2$. We have $\Lambda^2(V^*)$ is the dual space of $\Lambda^2(V)$ and using the duality defined in the previous section we have

$$\langle dx \wedge dy, e_1 \wedge e_2 \rangle = dx \wedge dy(e_1 \wedge e_2) = \begin{vmatrix} \langle dx, e_1 \rangle & \langle dx, e_2 \rangle \\ \langle dy, e_1 \rangle & \langle dy, e_2 \rangle \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

using the usual properties of the dual basis.

We know, since $\Lambda^2(V)$ is one dimensional, that for any vectors v and w, $v \wedge w = ke_1 \wedge e_2$ for some $k \in \mathbb{R}$. By duality, $k = \langle dx \wedge dy, v \wedge w$, and we can compute k if we know the expansion of v and w in terms of e_1 and e_2. Suppose

$$\begin{aligned} v &= v^1 e_1 + v^2 e_2 \\ w &= w^1 e_1 + w^2 e_2 \end{aligned}$$

Then

$$\begin{aligned} k = \langle dx \wedge dy, v \wedge w \rangle &= \begin{vmatrix} dx(v^1 e_1 + v^2 e_2) & dx(w^1 e_1 + w^2 e_2) \\ dy(v^1 e_1 + v^2 e_2) & dy(w^1 e_1 + w^2 e_2) \end{vmatrix} \\ &= \begin{vmatrix} v^1 & w^1 \\ v^2 & w^2 \end{vmatrix} = v^1 w^2 - v^2 w^1 \end{aligned}$$

We do not have the concept of Area available yet, since we do not have a metric at this point, but when we get a metric then $e_1 \wedge e_2$ will be a unit area parallelogram and $k = v^1 w^2 - v^2 w^1$ will be the (signed) area of the oriented parallelogram bounded by v and w.

We pause in the development for an explanation of some ancient notation. Since $v^1 = dx(v)$ etc. we could write k as

$$k = dx(v)dy(w) - dx(w)dy(v)$$

To be able to drop the w and v the ancients substituted δ for d when d was applied to w, so the equation looked like

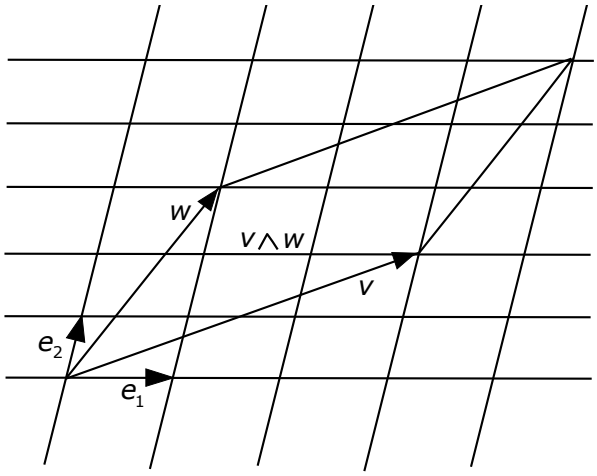
$$k = dx(v)\delta y(w) - \delta x(w)dy(v)$$

In ancient times dx and δx were not thought of as linear functionals but as the actual (infinitely small) quantities $dx(v)$ and $\delta x(v)$ and similarly with y. Thus the equation, with k interpreted as area, came out

$$dS = dx\delta y - \delta x dy$$

and this is a formula found in many ancient books, especially differential geometry books, and in some not so ancient.

Now for a geometric representation, $dx \wedge dy$ will be a represented by a grid of parallelograms with sides parallel to e_1 and e_2 and k will be the number of parallelograms covered by the parallelogram with sides $v = 3e^1 + 2e_2$ and $w = 1e_1 + 3e_2$.



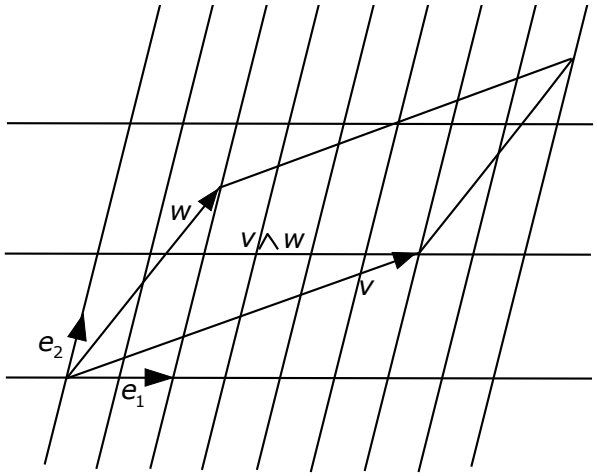
Geometric representation of 2-form $\omega = dx \wedge dy$ in 2-space

We can see that exactly

$$\begin{vmatrix} 3 & 1 \\ 2 & 3 \end{vmatrix} = 7$$

of the fundamental parallelegrams are covered. Oops, you're not so sure it's 7? Isn't it nice that we have analytic techniques to count them.

We must also remember that the representation of a differential form as a wedge product is far from unique. As a trivial example,

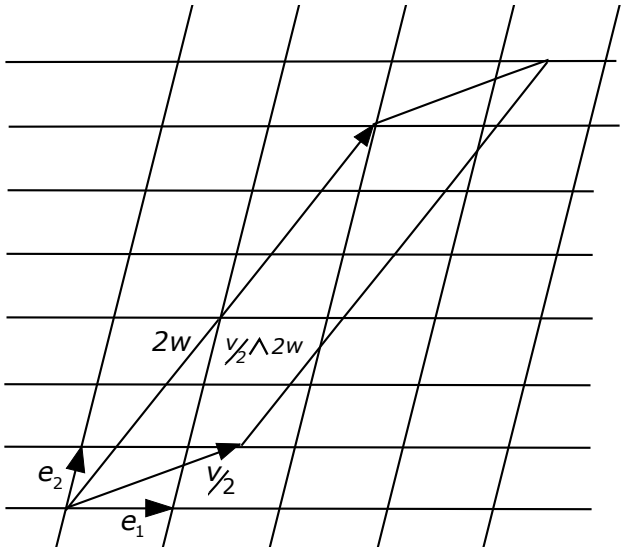


Geometric representation of 2-form $\omega = 2dx \wedge \frac{1}{2}dy$ in 2-space

$\omega = dx \wedge dy = 2dx \wedge \frac{1}{2}dy$, and this of course will have a corresponding geometric representation. And we can do a corresponding calculation.

$$\begin{aligned} \langle 2dx \wedge \frac{1}{2}dy, (3e_1 \wedge 2e_2) \wedge (e_1 \wedge 3e_2) \rangle &= \begin{vmatrix} 2dx(3e_1 + 2e_2) & 2dx(e_1 + 3e_2) \\ \frac{1}{2}dy(3e_1 + 2e_2) & \frac{1}{2}dy(e_1 + 3e_2) \end{vmatrix} \\ &= \begin{vmatrix} 6 & 2 \\ 1 & \frac{3}{2} \end{vmatrix} = 7 \end{aligned}$$

Of course the same thing can be done with the vectors. We have $v \wedge w$ can be written as $\frac{1}{2}v \wedge 2w$ and this can be represented as before.



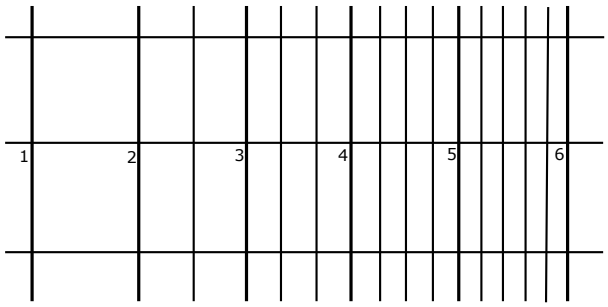
Geometric representation of 2-form $\omega = dx \wedge dy(\frac{1}{2}v \wedge 2w)$

and the computation is

$$\begin{aligned}\langle dx \wedge dy, (\frac{3}{2}e_1 \wedge e_2) \wedge (2e_1 \wedge 6e_2)\rangle &= \begin{vmatrix} dx(\frac{3}{2}e_1 \wedge e_2) & dx(2e_1 \wedge 6e_2) \\ dy(\frac{3}{2}e_1 \wedge e_2) & dy(2e_1 \wedge 6e_2) \end{vmatrix} \\ &= \begin{vmatrix} \frac{3}{2} & 2 \\ 1 & 6 \end{vmatrix} = 7\end{aligned}$$

Clearly we could go on multiplying examples but we think this is enough for you to get the idea. Forms with constant coefficients are represented geometrically by grids of the type appropriate for the dimensions, and if you require such a representation you can probably work it out for yourself.

The representation of forms with variable coefficients does not work as well but we will give one example and explain the difficulties.

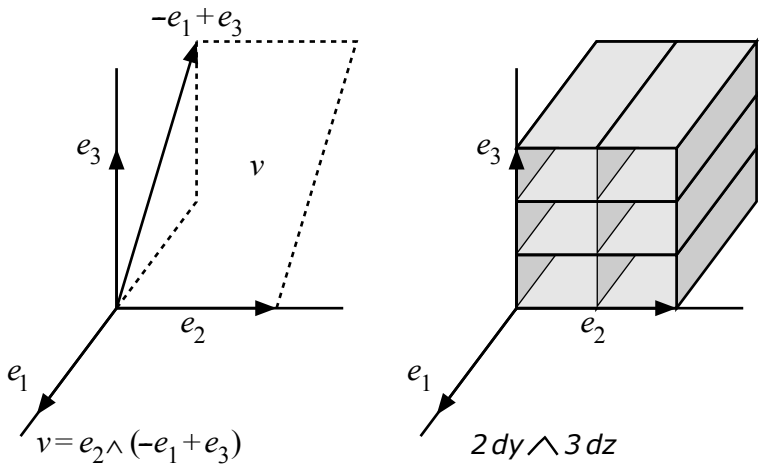


Geometric representation of xdx in $\mathbb{R}^2$

This representation can only be approximate, because the density of the lines is increasing continuously, whereas we have drawn it so that between 1 and 2

the density is 1, between 2 and 3 the density is 2, etc. Thus if we evaluated the form on a bivector $v \wedge w$ the answer would be only approximately correct. To get the exact answer we would need to do an integral.

We have already given an example of a one form in three dimensional space where it was represented as a series of planes. Thus it will come as no surprise that a two form is represented as two series of planes.



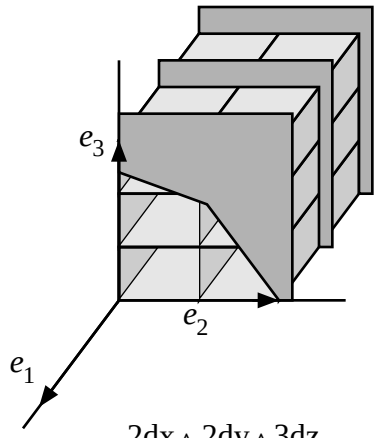
Geometric representation of $e_2 \wedge (-e_1 + e_3)$ and $2dy \wedge 3dz$

The illustration shows that $e_2 \wedge (-e_1 + e_3)$ cuts through 6 "tubes" which shows the $2dy \wedge 3dz\big(e_2 \wedge (-e_1 + e_3)\big) = 6$. And indeed calculation shows that

$$\begin{aligned} 2dy \wedge 3dz\big(e_2 \wedge (-e_1 + e_3)\big) &= 2dy \wedge 3dz(e_2 \wedge (-e_1)) + 2dy \wedge 3dz(e_2 \wedge e_3) \\ &= \begin{vmatrix} 2dy(e_2) & 2dy(-e_1) \\ 3dz(e_2) & 3dz(-e_1) \end{vmatrix} + \begin{vmatrix} 2dy(e_2) & 2dy(e_3) \\ 3dz(e_2) & 3dz(e_3) \end{vmatrix} \\ &= \begin{vmatrix} 2 & 0 \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} 2 & 0 \\ 0 & 3 \end{vmatrix} = 0 + 6 = 6 \end{aligned}$$

The reader may feel a sense of déjà vu from this last example. That might be because at some point in her education she might have encountered Faraday's *tubes of force.* We see here the mathematical form of Faraday's concept. Of course, since the coefficients for Faraday's forms are almost always variable we encounter the same difficulty we say before with $xdx \wedge dy$, but Farady never intended tubes of force representation to be exact.

We finally want to show a representation of a 3-form in 3-space. We will limit ourself to about the simplest possible case.



Geometric representation of $2dx \wedge 2dy \wedge 3dz$ in $\mathbb{R}^3$

Thus a 3-form is represented by a system of cells in 3-space, and a product of three vectors will be a parallelopided which will contain a number of these cells, which gives the value of the 3-form on the product of vectors.

possible example here

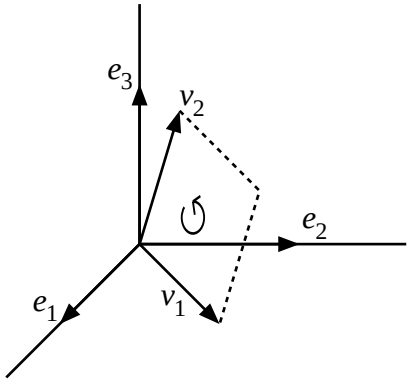
The reader may be wondering when we are going to introduce the concept of *flux.* The answer is, not here, and we now explain why. If you have any experience with flux, for example the flux of a magnetic field through an area, you will recall that the angle at which the lines of magnetic force, or the magnetic vector, meet the area is critical. But at this point in our development; we cannot measure angles; to measure angles we need an inner product. Hence we cannot treat flux until we introduce inner products, and we will do this as soon as we prove a couple of important theorems which do *not depend* on inner products. So fear not; flux will get its time on stage when when we do the Geometric and Physical meaning of Differential Forms, part II.

ZZZZZZZZZZZ Subsequent material needs to be redistributed.

In this section we will discuss how the foregoing material relates to applications. The matter is simple in essence but complex in some aspects. The geometry of the situation will be represented by elements of $\Lambda^k T(M)$ and the physics, like desities and fluxes, will be represented by elements of $\Lambda^k T(M)^*$. The use of two Grassmann Algebras is unusual, but we feel it brings out the situation more clearly. Although much of this section is about simpler matters than differentiable manifolds, we will use the notation appropriate for differentiable manifolds so as not to have to change notation in the middle. There is also the advantage of accustoming the reader to the notation in a simple situation.

Let us begin with the geometry. The basis elements of the geometry will be $\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^n}$ which we earlier called $e_1, \ldots, e_n$. If $n = 3$ then $\frac{\partial}{\partial u^1}, \frac{\partial}{\partial u^2}, \frac{\partial}{\partial u^3}$ are just notation for $\vec{i}, \vec{j}, \vec{k}$. This notation is adapted for use on the tangent space of a manifold but works anywhere.

To illustrate the idea, let $n = 3$. We want to illustrate the rectangle bounded by $v_1 = \frac{\partial}{\partial u^1} + \frac{\partial}{\partial u^2}$ and $v_2 = \frac{1}{3}\frac{\partial}{\partial u^1} + \frac{1}{3}\frac{\partial}{\partial u^2} + \frac{\partial}{\partial u^3}$. This rectangle will be represented by $v_1 \wedge v_2 \in \Lambda^2 T(M)$



The Rectangle $v_1 \wedge v_2$

2.7.7 The Exterior Derivative

The tangent bundle has no natural derivative and to take derivatives derivatives there requires the introduction of additional structures (connections). However, for the cotangent bundle and its Grassmann Algebra $\Lambda(T(M))^*$ we are fortunate in having a natural derivative operation d. Although several people contributed to its invention, the critical person was Élie Cartan. Unlike most great mathematical inventions there were many people who disliked, or at least were unenthusiastic about, the new derivative, and it took a long while to become popular. Possibly part of the reason for this is that it does not integrate well into the notation of differential forms, no matter which variant is used. This can be annoying, and the situation is even worse for it's cousin the codifferential δ which we will treat later in the book.

Although there is a direct definition of d, it uses Lie Derivatives which we want to avoid in this book which is already complicated enough. An indirect definition of d can be used, based on Axioms, and this is the most popular method. There are three axioms, which then allow us to compute d in all circumstances.

1. $df = \frac{\partial f}{\partial u^i}\, du^i$ for $f \in \Lambda^0(M)$. That is f is a function on M.

2. $dd\omega = 0$ for $\omega \in \Lambda(M)^*$.

3. Let $\omega \in \Lambda^j(M)^*$ and $\eta \in \Lambda^k(M)^*$. Then

$$d(\omega \wedge \eta) = (d\omega) \wedge \eta + (-1)^j \omega \wedge (d\eta)$$

The third axiom says that d is an antiderivation on the exterior algebra $\Lambda(M)^*$ of degree 1. This is important for some applications so you should remember it but we will not use antiderivations. It would suffice for the second axiom to

assume it only for $f \in \Lambda^0(M)^*$ but since this is easy to prove I thought it easier to just build $d^2 = 0$ in from the beginning.

RECAKK $f^*(\omega)$ HERE.

2.8 Integration of Differential Forms on Manifolds

2.8.1 Orientability

We will only be discussing the interation of differential forms on *orientable* manifolds. The standard example of a non-orientable manifold is the Möbius strip. The non-orientability is intuitively obvious because if you slide a normal vector around the strip when you return it is pointing the other way. However, this won't do for us because a normal vector needs the surface to be embedded in $\mathbb{R}^{m+1}$. A better way to look at it is (using a transparent Möbius strip) to slide an x, y frame around the strip. When it returns the arrows are pointing wrong; the frame has become left handed. If the Universe were non-orientable then a right handed person travelling along the correct loop would come back left handed.

We would like a way to easily describe this mathematically and there is one. Consider a differentiable manifold M and a collection of coordinate patches that cover it. On the overlap of two patches (U_1, Φ_1) where $\Phi_1 : U_1 \to \mathbb{R}^m$ with coordinates $(u^1, \ldots, u^m)$ and (U_2, Φ_2) where $\Phi_2 : U_2 \to \mathbb{R}^m$ with coordinates $(v^1, \ldots, v^m)$ we insist the transition functions $v^1(u^1, \ldots, u^m), \ldots, v^m(u^1, \ldots, u^m)$ have Jacobians with positive determianants.

$$\det \frac{\partial(v^1, \ldots, v^m)}{\partial(u^1, \ldots, u^m)} = \det \begin{pmatrix} \frac{\partial v^1}{\partial u^1} & \cdots & \frac{\partial v^1}{\partial u^m} \\ \vdots & \vdots & \vdots \\ \frac{\partial v^m}{\partial u^1} & \cdots & \frac{\partial v^m}{\partial u^m} \end{pmatrix} > 0$$

In any particular case if the Jacobian comes out negative one need only swap v^1 and v^2 to make the Jacobian positive. If one can do this over the whole manifold the manifold is orientable. If one cannot do it, then the manifold is non-orientable and not to be condidered in this section on integrability. Of course, if a manifold is *not* orientable we might be able to take a rather large portion of it which *is* orientable, and then we can do integration on that portion.

Now suppose that we have selected a collection of charts that cover M and all of which have positive determinant of the Jacobian of the transition functions with one another (on the overlaps of course; otherwise the condition means nothing.) A collection of charts that covers M is called an *Atlas* and when the charts fulfill the above condition it is called an *oriented Atlas*. An oriented Atlas determines an *orientation* of the Manifold M. There are two possible orientations for the manifold.

A new chart is *positively oriented* if, when added to the original Atlas, the new Atlas remains an oriented Atlas. A new chart is *negatively oriented* if this is not the case. (The new chart is out of synch with the rest of the charts of the oriented Atlas.)

Another way to deal with orientation is to find a never 0 topform. A topform Ω is an element of $\Lambda^m(T(M)^*)$. If such a topform Ω since it can be used to test whether each coordinate patch has properly oriented coordinates and if they all match up with Ω they will match up with each other. On the other hand if we have an oriented Atlas, we have a never zero topform $du^1 \wedge \ldots \wedge du^m$ in each coordinate patch and these can then be put together into a single never 0 topform using a partition of unity.

Positively oriented is a relative concept; there is no way to determine an absolute positive orientation in mathematics.

Note that in physics it would be possible to determine an absolute system of positive orientation based on the lack of parity conservation of beta decay. Thus it is possible to discuss *left* and *right* with distant civilizations, and thus determine an orientation for the Universe, assuming that the failure of parity remains the same everywhere. Like the constancy of the speed of light, our assumption that the parity non-conservation is the same everywhere is based on pretty local information.

2.8.2 Definition of Multiple Integral

In this section we review the definition of multiple integral on a subset M of $\mathbb{R}^m$. We assume for simplicity that M is a "nice" set, without boundary complication or toplogical complications. For example a simply connected open set union with a part or all of its boundary.

We divide M up into little rectangles or parallelepipeds with edges parallel to the coordinate axes. We select in each parallelepiped a point and evaluate the function f at that point. We calculate the area of each parallelepiped by multiplying the lengths Δu^j of its sides, getting the length from the real number coordinates of the edges. We index the areas by i. We then form the sum

$$\sum_i f(u_i^1, u_i^2, \ldots, u_i^m)\, \Delta u_i^1 \Delta u_i^2 \cdots \Delta u_i^m$$

This sum is an approximation to the integral

$$\int_M f(u^1, u^2, \ldots, u^m)\, \underline{d}u^1\, \underline{d}u^2 : \cdots \underline{d}u^m$$

The $d's$ are underlined to make it clear at a glance that this is an advanced Calculus integral and *not* the integral of a differential form. To make this even clearer, when we integrate froms we will put in the wedges, although later we will again forget to do this. The underlining will remain for this kind of advanced Calculus integral.

The approximation will be better if the Δu^i are made smaller. Just for now, let

$$\text{mesh } = \max_i \{\Delta u_i^1, \Delta u_i^2, \cdots, \Delta u_i^m\}$$

Then the definiton of the integral is

$$\int_M f(u^1, u^2, \ldots, u^m)\, \underline{d}u^1\, \underline{d}u^2 : \cdots \underline{d}u^m \; =$$
$$= \lim_{\text{mesh}\to 0} \sum_i f(u_i^1, u_i^2, \ldots, u_i^m)\, \Delta u_i^1 \Delta u_i^2 \cdots \Delta u_i^m$$

In a perfect world we would now prove that this limit exists, but since you saw this proved in your advanced Calculus class or your Analysis class we will omit this detail.

2.8.3 Recipe for Integrating a Differential Form

This is a relatively simple matter and the only difficulty is keeping the orientation straight.

The first matter is to reduce the problem to integration over an M which is contained in a coordinate patch. Let M be a manifold (which by definition is paracompact) and let $\underline{U}_1$ be a cover by coordinate patches and let $\underline{U} = \{U_\iota,\ \iota \in I\}$ be a locally finte subcover (which still consists of coordinate patches) and let ρ_i be a partition of unity subordinate to $\underline{U}$. Then $\sum_{\iota\in I} \rho_\iota = 1$ and $\text{supp}(\rho_\iota) \subseteq U_\iota$. Then $f = \sum_{\iota\in I} \rho_\iota \cdot f$ and $\text{supp}(\rho_\iota \cdot f) \subseteq U_\iota$. Then for an m-form ω

$$\int_M \omega = \sum_\iota \int_{U_\iota} \rho_\iota \cdot \omega$$

where on the right the support of the integrand is contained in a coordinate patch. We are saying nothing about whether the sum converges but it definitely will if the manifold M or the support of ω is compact.

We now consider the case where M and $\text{supp}(\omega) \subseteq M$ are contained in a coordinate patch where the coordinates are $u^1, u^2, \ldots u^m$. Then ω can be written

$$\omega = f du^1 \wedge du^2 \wedge \cdots \wedge du^m$$

and we must define $\int_M \omega$. Recall that we are defining integration of forms *only* on *oriented manifolds*. Had we been smart we would have arranged all the coordinate patches to come from an oriented Atlas. But since we forgot to do this, let deal with the problem now. If the coordinate system $u^1, u^2, \ldots u^m$ is positively oriented, all is well. If however it is negatively oriented then $u^2, u^1, \ldots u^m$ is positively oriented. We can now define the integral. Using a positively oriented coordinate system

Def

$$\int_M \omega = \int_M f\, du^1 \wedge \cdots \wedge du^m = \int_M f\, \underline{d}u^1\, \cdots \underline{d}u^m$$

where of course the integral on the right is the advanced Calculus multiple integral.

Suppose now that the coordinate system was *not* positively oriented. Then

$$\begin{aligned}\int_M \omega &= \int_M f\, du^1 \wedge du^2 \wedge \cdots \wedge du^m \qquad \text{negatively oriented}\\ &= -\int_M f\, du^2 \wedge du^1 \wedge \cdots \wedge du^m \qquad \text{positively oriented}\\ &= -\int_M f\, \underline{d}u^2\, \underline{d}u^1\, \cdots \underline{d}u^m\\ &= -\int_M f\, \underline{d}u^1\, \underline{d}u^2\, \cdots \underline{d}u^m\end{aligned}$$

where the last step uses Fubini's theorem that says that the order of differentials does not matter. (The hypotheses of Fubini's theorem are satisfied because we are in a coordinate patch.) Thus the import of the above is that if the coordinate system is negatively oriented then a minus sign must be inserted when converting to ordinary integrals.

Naturally, once you have converted to the advanced Calculus multiple integral, you can usually evaluate it by converting to iterated integrals but his is not our convern.

What is our concern is how things happen when we switch coordinate systems. To deal with this we must recall another advance Calculus theorem which lesser books merely state and only the bravest books prove. (And even the bravest ones often do only two variables.) The theorem is

Thm If M is described by two coordinates systems (U, Φ) and (V, Ψ) then

$$\int_m f(v^1, \ldots, v^m)\, \underline{d}u^1\, \ldots \underline{d}u^m =$$

$$= \int_m f(v^1(u^1, \ldots, u^m), \ldots, v^m(u^1, \ldots, u^m)) \left| \det \left(\frac{\partial(v^1, \ldots, v^m)}{\partial(u^1, \ldots, u^m)} \right) \right| \underline{d}u^1\, \ldots \underline{d}u^m$$

We may assume that the two coordinate systems are similarly oriented in which case $\partial(v^1, \ldots, v^m)/\partial(u^1, \ldots, u^m) > 0$ and the absolute values signs in the last equation have no effect. (Otherwise we can correct with minus signs as we showed above.) Now for the forms, recall that

$$dv^i = \frac{\partial v^i}{\partial u^j}\, du^j$$

so

$$\begin{aligned}\omega &= f(v^1, \ldots, v^m)\, dv^1 \wedge \ldots \wedge dv^m\\ &= f(v^1(u^1, \ldots, u^m), \ldots, v^m)) \det \left(\frac{\partial(v^1, \ldots, v^m)}{\partial(u^1, \ldots, u^m)} \right) du^1 \wedge \ldots \wedge du^m\end{aligned}$$

and thus using the recipe for evaluating the integrals of the ω in each of the two coordinate systems gives the same result; the integral of the form does not depend on which coordinate system is used provided they are similarly oriented, which they will be if taken from the positive Atlas.

2.9 Proof of Stokes Theorem

We will present in this chapter a proof of Stokes theorem. This is a low level proof using the familiar methods of Advanced Calculus, as seen in a third semester Calculus course. As this is an important theorem we will take it slow. Even though the proof is quite elementary it is valid over the whole range of application of the theorem.

It is rather remarkable that such a ubiquitous, important, and powerful theorem consists of two parts, neither of which is very deep. First there is fundamental theorem of Calculus. The "depth" of the result is all here. Then there is an enormous amount of bookkeeping which is tedious and rather complex but which has no depth at all. It is a remarkable instance of how the range of applicability of a result (in this case the fundamental theorem of Calculus) can be enormously increased by careful bookkeeping.

In this section the coefficients of the differentail forms will be assumed to have two continuous derivatives (C^2 functions) and the functions describing the boundaries will have one continuous derivative except at corners and folds, which will be assumed isolated; for example the corners are not allowed to have points of accumulation. This is usually referred to as *piecewise smooth*.

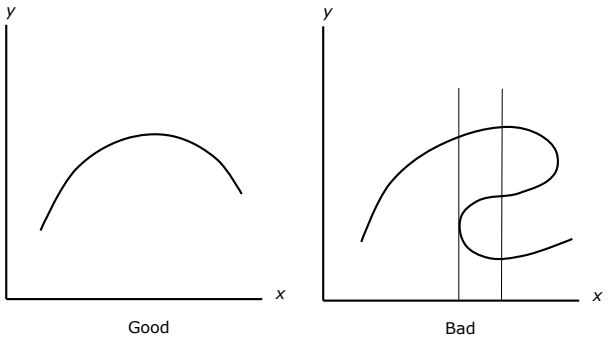
We will first treat Stoke's theorem in two dimensions because some ideas are particularly clear in this case. Let us recall some facts about line intergrals. Let C be an oriented curve and we wish to find, for example $\int_C F(x,y)\,dy$. To do this we find a parametrization of C, for example $x = x(t)$ and $y = y(t)$ for $t_1 \leq t \leq t_2$ and for which the direction of increasing t matches the orientation of C. Then

$$\int_C F(x,y)\,dy = \int_{t_1}^{t_2} F(x(t),y(t))\,\frac{dx}{dt}\,dt$$

However, it would be a little inconvenient (though possible, and perhaps a good idea,) to use this form of parametrization to prove Stokes theorem, and we follow a more classical route. For this we use the parameter x rather than t and the parametrization is then

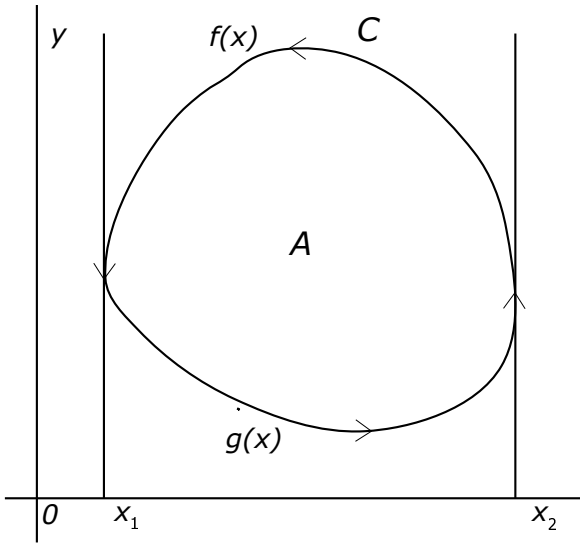
$$\begin{aligned} x &= x \\ y &= f(x) \end{aligned}$$

This changes the game slightly, because for a given x a vertical line may hit the curve C only once, so the class of curves C becomes restricted.



Good and bad curves for the x parametrization

This is an artificial restriction and we will eventually remove it. The following illustration shows the easiest case of Stokes theorem in two dimensions. We need a term for regions of the sort illustrated below, for which both the x parameter and the y parameter will work, so that top and bottom, or right and left, curves can be described functionally. We will temporarily call these regions *nice.*



The simplest 2-dimensional case

Note the arrows on $C = \partial A$ which indicate the orientation, in this case counterclockwise or spin up[7].

Now we calculate a line integral.

$$\begin{aligned}\int_C H(x,y)dx &= \int_{x_1}^{x_2} H(x,g(x))\,dx + \int_{x_2}^{x_1} H(x,f(x))\,dx \\ &= \int_{x_1}^{x_2} H(x,g(x)) - H(x,f(x))\,dx\end{aligned}$$

[7] *Spin up* is the quantum mechanics term for counterclockwise and *spin down* for clockwise.

The minus sign is due to reversing the direction of integration from $x_2 \to x_1$ to $x_1 \to x_2$. Here we see one of the origins of the concept of orientation. Now let us apply this to the following integral

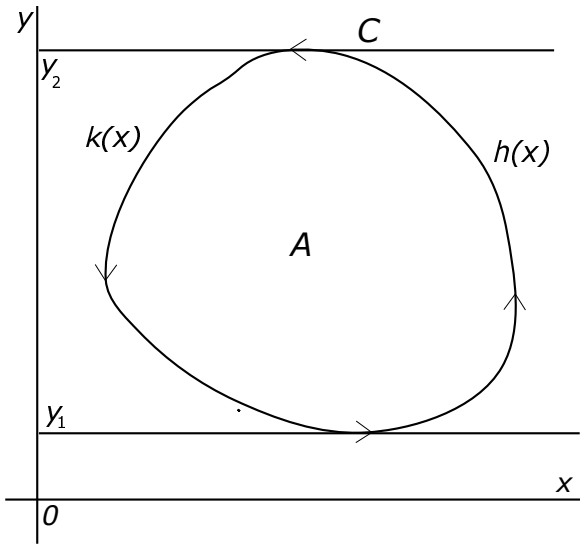
$$\begin{aligned}\int_A \frac{\partial F}{\partial y}\,dx \wedge dy &\stackrel{\text{def}}{=} \int_A \frac{\partial F}{\partial y}\,dxdy \;=\; \int_{x_1}^{x_2}\int_{y=g(x)}^{y=f(x)} \frac{\partial F}{\partial y}\,dydx \\ &= \int_{x_1}^{x_2} F(x,f(x)) - F(x,g(x))dx \qquad \text{(Fund. Th. of Calculus)} \\ &= -\int_C F(x,y)\,dx \;=\; -\int_{\partial A} F(x,y)\,dx\end{aligned}$$

If we set $\omega = F(x,y)\,dx$ then $d\omega = \frac{\partial F}{\partial y} dy \wedge dx = -\frac{\partial F}{\partial y} dx \wedge dy$ and the above can be rewritten as

$$\int_{\partial A} \omega = \int_{\partial A} F(x,y)\,dx = -\int_A \frac{\partial F}{\partial y}\,dx \wedge dy = \int_A d\omega$$

and we have proved a special case of Stokes' Theorem.

Now we do it all again with $\int_C G(x,y)\,dy$. This would not be worth doing if it were not for the subtle effects of orientation which makes it worth a look.



Another case with different orientation

Exactly as before we we have

$$\begin{aligned}\int_A \frac{\partial G}{\partial x}\,dx \wedge dy &\stackrel{\text{def}}{=} \int_A \frac{\partial G}{\partial x}\,dxdy \;=\; \int_{y_1}^{y_2}\int_{k(x)}^{h(x)} \frac{\partial G}{\partial x}\,dxdy \\ &= \int_{y_1}^{y_2} G(h(y),y) - G(k(y),y)\,dy \\ &= \int G(x,y)\,dy\end{aligned}$$

What you need to see here is there is no minus sign this time. Here is why: in the first case $\int_{\partial A} F(x,y)\,dx$, when x increases we go positively at the bottom of the curve and negatively at the top of the curve. In the second case $\int_{\partial A} G(x,y)\,dy$ when y increase we go negatively on the bottom of the curve and positively on the top of the curve. This accounts for the minus sign in the formula for $\int_{\partial A} F(x,y)\,dx$ and the lack of one in $\int_{\partial A} G(x,y)\,dy$, disturbing the symmetry. The asymmetry is an artifact of the choice of orientation; it is due to having made a specific choice of the order of the variables: dx, dy which determines the orientation.

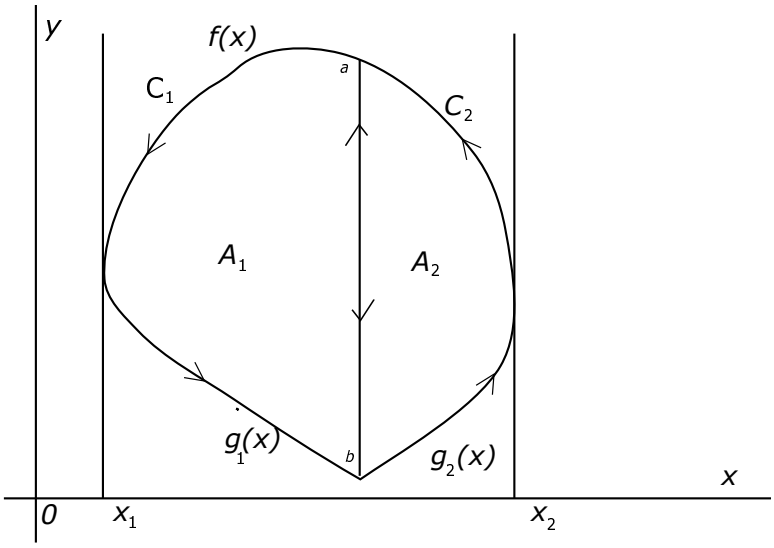
The takeaway here is the fact that the algebraic rules of Grassmann algebra correspond perfectly to the geometric situation. That it is actually possible to get something like this to work is little short of a miracle. That it works in many dimensions if even less short.

If we put the two calculations together and set $\omega = F(x,y)\,dx + G)x,y)\,dy$, which is the general form of a 2-form in $\mathbb{R}^2$, we have

$$\begin{aligned}\int_{\partial A}\omega &= \int_{\partial A} F(x,y)\,dx + G(x,y)\,dy \;=\; \int_A -\frac{\partial F}{\partial y} + \frac{\partial G}{\partial x}\,dxdy \\ &= \int_A -\frac{\partial F}{\partial y} + \frac{\partial G}{\partial x}\,dx\wedge dy \;=\; \int_A d\omega\end{aligned}$$

Thuse we have proved Stokes' Theorem for a nice region in $\mathbb{R}^2$ with smooth boundary.

The smoothness is not important, but we discuss it a bit because it illustrates an important trick. Suppose the bottom of our region has a corner it it.



A slight complication in the 2-dimensional case

From the theoretical point of view this is the same as the previous case, but from the practical point of view one has to find another parametrization. We

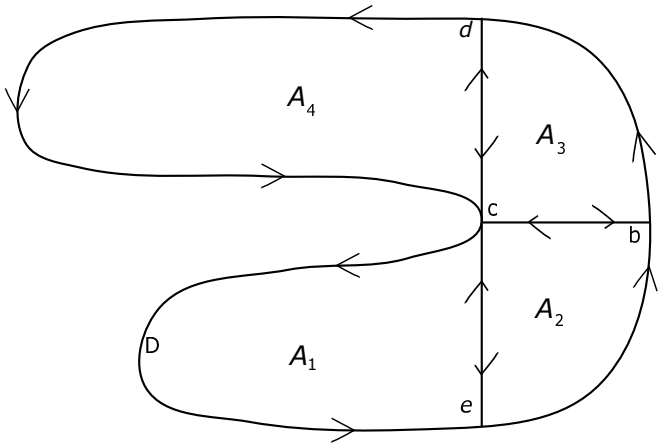
can however reduce this to the previous case of smooth boundary by using the vertical line through the corner to divide the region into two subregions in each of which Stokes' Theorem holds. The midline is part of the boundary of both A_1 and A_2. The unbalanced arrows indicate which region we are regarding it as the boundary of. The large side of the arrow indicates the region. Thus the down pointing arrow indicates a part of the boundary of A_2 and the up pointing arrow indicates a part of the boundary of A_1. Stokes' Theorem holds on each of the regions so we have

$$\begin{aligned}
\int_{\partial A} \omega &= \int_{C_2} \omega + \int_{C_1} \omega \\
&= \int_{C_2} \omega + \int_{ab} \omega - \int_{ab} \omega + \int_{C_1} \omega \\
&= \int_{C_2} \omega + \int_{ab} \omega + \int_{ba} \omega + \int_{C_1} \omega \\
&= \int_{\partial A_2} \omega + \int_{\partial A_1} \omega \\
&= \int_{A_2} d\omega + \int_{A_1} d\omega \;=\; \int_{A} d\omega
\end{aligned}$$

and we have Stokes' Theorem again. The takeaway here is that we can greatly extend the applicability of Stokes' theorem by introducting extra boundaries that are cancelled out when we sum things up. We will use this technique for many purposes.

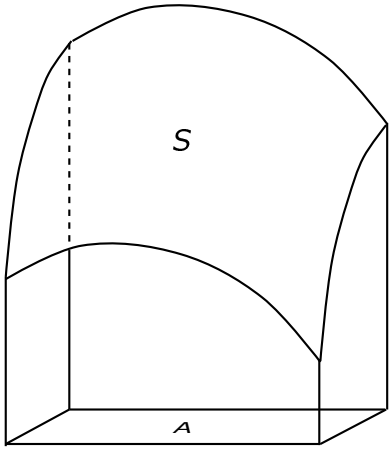
Clearly we accomodate any finite numbers of corners (or cusps) and such boundaries are called *piecewise smooth*. However, there are limits to how far we can stretch Stokes' theorem and it would be very unwise to apply it to the paw print of the infinite toed bear, whose boundary is, among other nasty features, infinitely long.

To accomodate non-nice regions, we merely need cut up the region into nice regions. It may take a lot of subregions but we can always do it. Here is a sample.



Cutting up into nice regions

Next we want to move to three dimensional objects like the 3-Ball $B_1(0)$. We first need to understand how to do surface integrals. The method is just like line integrals. Find a



A surface for a surface integral

parametrization of S by parameters u and v. Express the surface integral in terms of these parameters

$$\begin{aligned}\int_S \omega &= \int_S F(x,y,z)\,dx\wedge dy + G(x,y,z)\,dy\wedge dz + H(x,y,z)\,dz\wedge dx \\ &= \int_A J(u,v)\,du\wedge dv\end{aligned}$$

where the integrands in x, y, z and the differentials have all been reexpressed in terms of the parameters u, v and A is some suitable subset of the parameter domain.

For our purposes we want to use, in this case, x and y as the parameters and we set $\omega = F(x, y, z)\, dx \wedge dy$. The surface S is given by $z = f(x, Y)$ so the integral over S becomes

$$\int_S \omega = \int_A F(x, y, f(x, y)) dx \wedge dy$$

Now we want to relate this to Stokes' theorem. Then if M is the three dimensional manifold bounded by S and the 5 planes in the figure, we want to show that for this M we have $\int_M d\omega = \int_{\partial M} \omega$, where

$$\omega = F(x, y, z)\, dx \wedge dy \quad \text{and} \quad d\omega = \frac{\partial F}{\partial z}\, dz \wedge dx \wedge dy = \frac{\partial F}{\partial z}\, dx \wedge dy \wedge dz$$

Then

$$\begin{aligned}
\int_M d\omega &= \int_M \frac{\partial F}{\partial z}\, dz \wedge dx \wedge dy \\
&= \int_M \frac{\partial F}{\partial z}\, dz dx dy \qquad \text{(ordinary multiple integral)} \\
&= \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z=f(x,y)} \frac{\partial F}{\partial z}\, dz dx dy \\
&= \int_{x_1}^{x_2} \int_{y_1}^{y_2} F(x, y, z)\Big|_{z_1}^{z=f(x,y)}\, dy dz \\
&= \int_A F(x, y, f(x, y))\, dy \wedge dz \\
&= \int_S F(x, y, z)\, dy \wedge dz \\
&= \int_S \omega
\end{aligned}$$

Now let us consider $\int_{A_i}$ for the five plane regions that form the rest of the boundary of M. Take for example the rightmost plane A_2. For this plane x is constant, and hence in evaluating $\int_{A_2} \omega$ we will have $dx = 0$. Hence $\int_{A_2} \omega = 0$. The same will be true of the other four boundary planes, so we will have

$$\begin{aligned}
\int_{\partial M} \omega &= \int_S \omega + \sum_{i=1}^{5} \int_{A_i} \omega \\
&= \int_S \omega + \sum_{i=1}^{5} 0 \;=\; \int_S \omega \\
&= \int_M d\omega
\end{aligned}$$

Thus proving Stokes' Theorem in this particular case. However, as we showed in the two dimensional case, the general case can be proved by carving up any

M into subregions like the case we just handled. When we add the resulting Stokes Theorems for the subregions, the interior planes will cancel one another out, giving the result in general.

We hope it is clear from our examples for two and three dimensions that a similar result can be proved for any n-dimensional region in $\mathbb{R}^n$. The basic idea in always the same, and note it is just the fundamental theorem of Calculus and a lot of bookkeeping.

But we are not done. We need to prove the result for case which generalizes a two dimensional region and its boundary in 3-space. Fortunately this is mostly just a little trickery.

2.10 Proof of the Converse of the Poincare Lemma

We will present two proofs of the Converse of the Poincare Lemma. In this section we will present a proof that uses only material from Advanced Calculus. The proof is thus "elementary" but complicated in detail. The second proof will be presented in Chapter 3, which has the flavor of algebraic topology and is thus an appropriate place for the second proof which uses chain homotopy arguments. In neither place is any special knowledge assumed, only perseverance. The proof presented in this section was developed by one of the authors, though no doubt others have also discovered it.

We will begin this section with the easiest case, partially because it actually isn't covered by the general case and partially because it illustrates the general idea of the proof in a circumstance where it is clearly visible. Then we will take on the general case which is challenging for merely notational reasons. The increasing permatations which have served us so well in the past do not mesh well with the exterior derivative d. We will need to come up with a formula for d that looks different from our previous work, but which it is a good idea to become familiar with. This will help us in proving the theorem.

The general theorem is

Def If U is a simply connected open subset of a manifold and $\omega \in \Lambda^k(U)$ and $d\omega = 0$ then there is and $\alpha \in \Lambda^{k-1}(U)$ for which $d\alpha = \omega$.

The concept *simply connected* was discussed in Chapter 1. We remind the reader that it means any simple closed curve may be continuously deformed to a point while remaining in U. An important non-example is $\mathbb{R}^2 - \{(0,0)\}$ where any curve surrounding the origin is *not* continuouly deformable into a point; the curve can't get round the missing point (0,0). This space will provide us with a very important example where the theorem fails.

The special case we will first look at is $k = 1$. Thus if ω is a 1-form and $d\omega = 0$ there is a function α with $d\alpha = \omega$. You are familiar with this from chapter 1 and you may remember the mantra *a curl free vector is a gradient* which you might have learned in advance Calculus. We will prove this in the case of a parallelepiped $a_i < u^i < b_i$ of dimension m. Once proved for this case it is easy (though annoying) to prove it for the general case.

We select any *base point* $a_i < u_0^i < b_i$ in the parallelepiped which stays fixed for the argument. Moving the base point will result α accumulating an additive constant which causes no problem for $d\alpha = \omega$.

Let

$$\omega = \sum_{i=1}^{m} \omega_i du^i$$

be the one form. Then $d\omega = 0$ decodes as

$$\begin{aligned} d\omega &= \sum_{i=1}^{m} d\omega_j \wedge du^j \\ &= \sum_{i,j=1}^{m} \frac{\partial \omega_j}{\partial u^i}\, du^i \wedge du^j \\ &= \sum_{i<j} \left(\frac{\partial \omega_j}{\partial u^i} - \frac{\partial \omega_i}{\partial u^j} \right) du^i \wedge du^j \end{aligned}$$

This formula is nice but doesn't generalize well for $k > 1$. From $d\omega = 0$ we see

$$\frac{\partial \omega_j}{\partial u^i} = \frac{\partial \omega_i}{\partial u^j}$$

which we use in the key step of the proof.

No we simply write down the formula for α. We remind the reader that the underlined differential mean they are just part of the integration apparatus for ordinary integrals and we are not dealing here with the integration of forms.

$$\alpha = \sum_{k=1}^{m} \int_{u_0^k}^{u^k} \omega_k(u_0^1, \ldots, u_0^{k-1}, u^k, \ldots, u^m) \underline{d}u^k$$

We write out a couple of terms so the following calculations will be easier to understand.

$$\begin{aligned} \alpha &= \int_{u_0^1}^{u^1} \omega_1(u^1, \ldots, u^m) \underline{d}u^1 + \int_{u_0^2}^{u^2} \omega_2(u_0^1, u^2, \ldots, u^m) \underline{d}u^2 \\ &+ \int_{u_0^3}^{u^3} \omega_3(u_0^1, u_0^2, u^3, \ldots, u^m) \underline{d}u^3 + \sum_{k=4}^{m} \int_{u_0^k}^{u^k} \omega_k(u_0^1, \ldots, u_0^{k-1}, u^k, \ldots, u^m) \underline{d}u^k \end{aligned}$$

Note the critical fact that the k^{th} summand does not depend on any of $u^1, u^2, \ldots, u^{k-1}$. Since all functions are C^∞ and the region of integration is a parallelepiped we may diffentiate under the integral sign. We have

$$\begin{aligned} \frac{\partial \alpha}{\partial u^1} &= \omega_1(u^1, \ldots, u^m) \\ \frac{\partial \alpha}{\partial u^2} &= \int_{u_0^1}^{u^1} \frac{\partial \omega_1}{\partial u^2}(u^1, \ldots, u^m) \underline{d}u^1 + \omega_2(u_0^1, u^2 \ldots, u^m) \end{aligned}$$

$$\begin{aligned}
&= \int_{u_0^1}^{u^1} \frac{\partial \omega_2}{\partial u^1}(u^1,\ldots,u^m)\underline{d}u^1 + \omega_2(u_0^1,u^2\ldots,u^m)\\
&= \omega_2(u^1,u^2\ldots,u^m) - \omega_2(u_0^1,u^2\ldots,u^m) + \omega_2(u_0^1,u^2\ldots,u^m)\\
&= \omega_2(u^1,u^2\ldots,u^m)
\end{aligned}$$

By now you are probably getting the idea but we'll do one more.

$$\begin{aligned}
\frac{\partial \alpha}{\partial u^3} &= \int_{u_0^1}^{u^1} \frac{\partial \omega_1}{\partial u^3}(u^1,\ldots,u^m)\underline{d}u^1 + \int_{u_0^2}^{u^2} \frac{\partial \omega_2}{\partial u^3}(u_0^1,u^2,\ldots,u^m)\underline{d}u^2\\
&\quad + \omega_3(u_0^1,u_0^2,u^3\ldots,u^m)\\
&= \int_{u_0^1}^{u^1} \frac{\partial \omega_3}{\partial u^1}(u^1,\ldots,u^m)\underline{d}u^1 + \int_{u_0^2}^{u^2} \frac{\partial \omega_3}{\partial u^2}(u_0^1,u^2,\ldots,u^m)\underline{d}u^2\\
&\quad + \omega_3(u_0^1,u_0^2,u^3,\ldots,u^m)\\
&= \omega_3(u^1,u^2\ldots,u^m) - \omega_3(u_0^1,u^2,\ldots,u^m)\\
&\quad + \omega_3(u_0^1,u_2,\ldots,u^m) - \omega_3(u_0^1,u_0^2,\ldots,u^m)\\
&\quad + \omega_3(u_0^1,u_0^2,u^3,\ldots,u^m)\\
&= \omega_3(u^1,u^2\ldots,u^m)
\end{aligned}$$

Thus the proof depends on coming up with a telescoping sum where everybody except the first term dies. Now that we have seen how it works, we can write out the proof properly.

$$\begin{aligned}
\frac{\partial \alpha}{\partial u^i} &= \sum_{k=1}^{i-1} \int_{u_k^0}^{u^k} \frac{\partial \omega_k}{\partial u^i}(u_0^1,\ldots,u_0^{k-1},u^k,\ldots,u^m)\underline{d}u^k + \omega_i(u_0^1,\ldots,u_0^{i-1},u^i,\ldots,u^m)\\
&= \sum_{k=1}^{i-1} \int_{u_k^0}^{u^k} \frac{\partial \omega_i}{\partial u^k}(u_0^1,\ldots,u_0^{k-1},u^k,\ldots,u^m)\underline{d}u^k + \omega_i(u_0^1,\ldots,u_0^{i-1},u^i,\ldots,u^m)\\
&= \sum_{k=1}^{i-1} \left(\omega_i(u_0^1,\ldots,u_0^{k-1},u^k,\ldots,u^m) - \omega_i(u_0^1,\ldots,u_0^k,u^{k+1},\ldots,u^m\right)\\
&\quad + \omega_i(u_0^1,\ldots,u_0^{i-1},u^i,\ldots,u^m)\\
&= \omega_i(u^1,\ldots,u^m)
\end{aligned}$$

It's simple when you see how it works. Finally

$$\begin{aligned}
d\alpha &= \sum_{i=1}^{m} \frac{\partial \alpha}{\partial u^i}\, du^i\\
&= \sum_{i=1}^{m} \omega_i du^i\\
&= \omega
\end{aligned}$$

as required.

For the general case, we must learn to write $d\omega$ in a new way. We have used the notation

$$\omega = \sum_{\sigma \in \mathcal{S}_{n,k}} \omega_\sigma du^\sigma$$

but we have found no way to integrate this notation with the exterior derivative and we must back up to a cruder notation. In what follows all the subscripted i's will run from 1 to m.

$$\omega = \sum_{i_1<i_2<\ldots i_k} \omega_{i_1 i_2 \ldots i_k}\, du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_k}$$

We then have quite simply and fairly uselessly

$$\begin{aligned} d\omega &= \sum_{i_1<i_2<\ldots i_k} d\omega_{i_1 i_2 \ldots i_k}\, du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_k} \\ &= \sum_{i_1<i_2<\ldots i_k} \sum_{i_0} \frac{\partial \omega_{i_1 i_2 \ldots i_k}}{du^{i_0}}\, du^{i_0} \wedge du^{i_1} \wedge du^{i_2} \wedge \ldots \wedge du^{i_k} \end{aligned}$$

So far so good. Notice that this is a $k+1$ form as expected. We must decode this into something resembling the formula for ω itself and this is where it gets tricky as we must change the labeling in an unintuitive way. The first critical thing to note is that if i_0 is among $i_1, i_2, \ldots, i_k$ then the term is 0, because of repetition in the differentials. Now we look at a particular term in the double sum

$$\frac{\partial \omega_{12478}}{\partial u^5}\, du^5 \wedge du^1 \wedge du^2 \wedge du^4 \wedge du^7 \wedge du^8$$

We prefer this be written as

$$(-1)^3 \frac{\partial \omega_{12478}}{\partial u^5}\, du^1 \wedge du^2 \wedge du^4 \wedge du^5 \wedge du^7 \wedge du^8$$

so the differentials are in proper order. Next we cleverly insert an index that isn't there into the subscript on ω. We indicate it isn't there by putting a caret on top of the relevant i; in this case there is a caret on the 5 in the subscript of ω which indicates the 5 is not there.

$$(-1)^3 \frac{\partial \omega_{124\hat{5}78}}{\partial u^5}\, du^1 \wedge du^2 \wedge du^4 \wedge du^5 \wedge du^7 \wedge du^8$$

Now what would this term look like if it was one of an indexed system of terms with indices $i_0, i_1, \ldots, i_k$? It would be

$$(-1)^3 \frac{\partial \omega_{i_0 i_1 i_2 \widehat{i_3} i_{k-1} i_k}}{\partial u^{i_3}}\, du^{i_0} \wedge du^{i_1} \wedge du^{i_2} \wedge du^{i_3} \wedge \ldots \wedge du^{i_{k-1}} \wedge du^{i_k}$$

where in the example $i_0 = 1$, $i_1 = 2$, $i_2 = 4$, $i_3 = 5$, $i_{k-1} = 7$, $i_k = 8$ and notice $\omega_{i_0 i_1 i_2 \widehat{i_3} i_{k-1} i_k} = \omega_{i_0 i_1 i_2 i_{k-1} i_k} = \omega_{12478}$. Now this is a typical term in $d\omega$, so by summing up all such terms we will get $d\omega$. Thus

$$d\omega = \sum_{i_0<i_1<\ldots<i_k} \sum_{j=0}^{k} (-1)^j \frac{\partial \omega_{i_0 i_1 i_2 \ldots \widehat{i_j} \ldots i_{k-1} i_k}}{\partial u^{i_j}}\, du^{i_0} \wedge du^{i_1} \wedge \ldots \wedge du^{i_{k-1}} \wedge du^{i_k}$$

It is important to know this formula for a couple of reasons. First, it can be useful when, as we will do below, it is necessary to integrate terms from $d\omega$ into further calculations. Second, this formula is often used in other books. In some older books it was even the *definition* of $d\omega$. The downside of course is that it is cumbersome. However, it does have one redeeming feature; if a particular product of differentials $du^{i_0} \wedge du^{i_1} \wedge \ldots \wedge du^{i_{k-1}} \wedge du^{i_k}$ is given (notice that $i_0, i_1, \ldots, i_k$ is fixed but there are still many terms in the sum depending on which j for i_j is chosen,) we can read off its coefficient from the formula, which will be

$$\sum_{j=0}^{k} (-1)^j \frac{\partial \omega_{i_0 i_1 i_2 \ldots \widehat{i_j} \ldots i_{k-1} i_k}}{\partial u^{i_j}}$$

For example, if $\omega = \sum_{i_1} \omega_{i_1} \, du^{i_1}$ is a 1-form then the coefficient of $du^{i_0} \wedge du^{i_1}$ in $d\omega$ is

$$\frac{\partial \omega_{i_1}}{\partial u^{i_0}} - \frac{\partial \omega_{i_0}}{\partial u^{i_1}}$$

We are now ready for the proof of the general theorem; on a simply connected domain M if $\omega \in \Lambda^k$ and $d\omega = 0$ then there exists and $\alpha \in \Lambda^{k-1}$ so that $d\alpha = \omega$. The proof for general k is similar to the proof for $k = 1$ but more complex in its details. As before we present the proof for a parallelepiped $a_i < u^i < b_i$ and utilize a base point $(u_0^1, u_0^2, \ldots, u_0^m)$ which is in the parallelepiped. Then the differential form $\omega \in \Lambda^k$ will be

$$\omega = \sum_{i_1 < i_2 < \ldots < i_k} \omega_{i_1 i_2 \ldots i_k} \, du_1^i \wedge du_2^i \wedge \ldots \wedge du_k^i$$

As we saw above we may write $d\omega$ as

$$d\omega = \sum_{i_0 < i_1 < \ldots < i_k} \sum_{j=0}^{k} (-1)^j \frac{\partial \omega_{i_0 i_1 i_2 \ldots \widehat{i_j} \ldots i_{k-1} i_k}}{\partial u^{i_j}} \, du^{i_0} \wedge du^{i_1} \wedge \ldots \wedge du^{i_{k-1}} \wedge du^{i_k}$$

Recall that the caret on i_j in the subscript $i_0 i_1 i_2 \ldots \widehat{i_j} \ldots i_{k-1} i_k$ indicates that i_j is not actually there and the actual subscript is $i_0 i_1 i_2 \ldots i_{j-1} i_{j+1} \ldots i_{k-1} i_k$.

We are assuming $d\omega = 0$ which amounts to

$$\sum_{j=0}^{k} (-1)^j \frac{\partial \omega_{i_0 i_1 i_2 \ldots \widehat{i_j} \ldots i_{k-1} i_k}}{\partial u^{i_j}} = 0$$

or, what is the same but will be useful later

$$\sum_{j=0}^{k-1} (-1)^j \frac{\partial \omega_{i_0 i_1 i_2 \ldots \widehat{i_j} \ldots i_{k-1} i_k}}{\partial u^{i_j}} + (-1)^k \frac{\partial \omega_{i_0 \ldots i_{k-1}}}{\partial u^{i_k}} = 0$$

We will now write down the solution α of $d\alpha = \omega$.

$$\alpha = \sum_{i_2 < \ldots < i_k} \alpha_{i_2 \ldots i_k} \, du^{i_2} \wedge \ldots \wedge du^{i_{k-1}} \wedge du^{i_k}$$

where

$$\begin{aligned}\alpha_{1i_3\ldots i_k} &= 0\\ \alpha_{ji_3\ldots i_k} &= \sum_{\ell=1}^{j-1}\int_{u_0^\ell}^{u^\ell}\omega_{\ell ji_3\ldots i_k}(u_0^1,\ldots,u_0^{\ell-1},u^\ell,\ldots,u^m)\,\underline{d}u^\ell \qquad 1<j<i_3<\ldots<i_k\end{aligned}$$

Using our new formula for the exterior derivative we have

$$d\alpha=\sum_{i_1<i_2<\ldots<i_k}\sum_{j=1}^{k}(-1)^{j-1}\frac{\partial\alpha_{i_1\ldots\widehat{i_j}\ldots i_k}}{\partial u^{i_j}}\,du^{i_1}\wedge\ldots\wedge du^{i_k}$$

Thus $d\alpha=\omega$ is equivalent to

$$\sum_{j=1}^{k}(-1)^{j-1}\frac{\partial\alpha_{i_1\ldots\widehat{i_j}\ldots i_k}}{\partial u^{i_j}}=w_{i_1\ldots i_k}$$

This, however, is a mere computation which we now perform:

$$\begin{aligned}&\sum_{j=1}^{k}(-1)^{j-1}\frac{\partial\alpha_{i_1\ldots\widehat{i_j}\ldots i_k}}{\partial u^{i_j}}\\ &\quad= \frac{\partial\alpha_{i_2\ldots i_k}}{\partial u^{i_1}}-\frac{\partial\alpha_{i_1i_3\ldots i_k}}{\partial u^{i_2}}+\sum_{j=3}^{k}(-1)^j\frac{\partial\alpha_{i_1\ldots\widehat{i_j}\ldots i_k}}{\partial u^{i_j}}\\ &\quad= \sum_{\ell=1}^{i_2-1}\frac{\partial}{\partial u^{i_1}}\int_{u_0^\ell}^{u^\ell}\omega_{\ell i_2\ldots i_k}(u_0^1,\ldots,u_0^{\ell-1},u^\ell,\ldots,u^m)\underline{d}u^\ell\\ &\quad- \sum_{\ell=1}^{i_1-1}\frac{\partial}{\partial u^{i_2}}\int_{u_0^\ell}^{u^\ell}\omega_{\ell i_1i_3\ldots i_k}(u_0^1,\ldots,u_0^{\ell-1},u^\ell,\ldots,u^m)\underline{d}u^\ell\\ &\quad+ \sum_{j=3}^{k}(-1)^{j-1}\sum_{\ell=1}^{i_1-1}\frac{\partial}{\partial u^{i_j}}\int_{u_0^\ell}^{u^\ell}\omega_{ki_1\ldots\widehat{i_j}\ldots i_k}(u_0^1,\ldots,u_0^{\ell-1},u^\ell,\ldots,u^m)\,\underline{d}u^\ell\\ &\quad= \sum_{\ell1}^{i_1-1}\int_{u_0^\ell}^{u^\ell}\sum_{j=1}^{k}(-1)^{j-1}\frac{\partial\omega_{ki_1\ldots\widehat{i_j}\ldots i_k}}{\partial u^{i_j}}(u_0^1,\ldots u_0^{\ell-1}u^\ell\ldots u^m)\,\underline{d}u^k\\ &\quad+ \sum_{\ell=i_1}^{i_2-1}\frac{\partial}{\partial u^{i_1}}\int_{u_0^\ell}^{u^\ell}\omega_{\ell i_2\ldots i_k}(u_0^1,\ldots u_0^{\ell-1}u^\ell\ldots u^m)\,\underline{d}u^k\end{aligned}$$

Using the equation

$$0=d\omega=\sum_{i_0<i_1<\ldots<i_k}\sum_{j=0}^{k}(-1)^j\frac{\partial\omega_{i_0i_1i_2\ldots\widehat{i_j}\ldots i_{k-1}i_k}}{\partial u^{i_j}}\,du^{i_0}\wedge du^{i_1}\wedge\ldots\wedge du^{i_{k-1}}\wedge du^{i_k}$$

we may replace the sum in the first integral by

$$\frac{\partial\omega_{i_1\ldots i_k}}{\partial u^k}(u_0^1,\ldots u_0^{\ell-1}u^\ell\ldots u^m)$$

and the second summand reduces, for $\ell = i_1$, to

$$\omega_{i_1 i_2 \dots i_k}(u_0^1, \dots, u_0^{i_1 - 1}, u^{i_1}, \dots, u^m)$$

with the remaining terms for $i_1 < \ell \leq i_2 - 1$ being 0 since for these terms the integral does not depend upon i_1.

Thus the right hand side of the previous quation becomes

$$\begin{aligned}
&\sum_{\ell=1}^{i_1-1} \int_{u_0^k}^{u^k} \frac{\partial \omega_{i_1 \dots i_\ell}}{\partial u^\ell}(u_0^1, \dots, u_0^{\ell-1}, u^\ell, \dots, u^m)\, \underline{d}u^\ell + \omega_{i_1, \dots i_k}(u_0^1, \dots, u_0^{i_1-1}, u^{i_1}, \dots, u^m) \\
&= \sum_{\ell=1}^{i_1-1} [\omega_{i_1 \dots i_k}(u_0^1, \dots, u_0^{\ell-1}, u^\ell, \dots, u^m) - \omega_{i_1 \dots i_k}(u_0^1, \dots, u_0^\ell, u^{\ell+1}, \dots, u^m)] \\
&+ \omega_{i_1, \dots i_k}(u_0^1, \dots, u_0^{i_1-1}, u^{i_1}, \dots, u^m) \\
&= \omega_{i_1, \dots i_k}(u^1, \dots, u^m)
\end{aligned}$$

as required.

Uniqueness

The α given by the converse of the Poincaré lemma is never unique. We may immediately find a plethora of α's. Suppose that $\eta \in \Lambda^{k-1}$ and $d(\alpha + \eta) = \omega$. Then since $d\alpha = \omega$ we must have $d\eta = 0$, and any η with $d\eta = 0$ will give another α. One plentiful source of such η is $\eta = d\zeta$ with $\zeta \in \Lambda^{k-2}$. Then $d(\alpha + \eta) = d(\alpha + d\zeta) = d\alpha + dd\zeta = \omega + 0 = \omega$.

If M is simply connected then this exhausts the possible α since $d\eta = 0$ implies that there exists a ζ with $d\zeta = \eta$. If M is *not* simply connected then there will be η which do not come from any $d\zeta$. An example can be constructed from the first example below. These are important matters and will be discussed more completely in Chapter 3.

Examples and Applications

We begin with one of the most important applications which is also a counterexample. We will work in $M = \mathbb{R}^2 - (0,0)$, the plane minus the origin. The 1-form ω is

$$\omega = -\frac{y}{x^2 + y^2}\, dx + \frac{x}{x^2 + y^2}\, dy$$

Then

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

Notice first that ω is not continuous at (0,0) so it is fortunate that $(0,0) \notin M$. Now let us consider the rectangle $.1 < x < 10$ and $-10 < y < 10$. In this

rectangle ω in C^∞ and we may use our formula. We will take the base point (x_0, y_0) to be $(1, 0) \in M$. The converse of the Poincaré lemma says that there exists an α defined on the rectangle for which $d\alpha = \omega$ and we even have a formula for it. Transcribing the formula into the variables x, y we have

$$\begin{aligned}
\alpha &= \int_1^x \omega_1(x,y)\,\underline{d}x + \int_0^y \omega_2(1,y)\,\underline{d}y \\
&= -\int_1^x \frac{y}{x^2+y^2}\,\underline{d}x + \int_0^y \frac{1}{1^2+y^2}\,\underline{d}y \\
&= \arctan\left(\frac{y}{x}\right)\Big|_1^x + \arctan(y)\Big|_0^y \\
&= \arctan\left(\frac{y}{x}\right) - \arctan(y) + \arctan(y) - \arctan(0) \\
&= \arctan\left(\frac{y}{x}\right) \;=\; \operatorname{arccot}\left(\frac{x}{y}\right)
\end{aligned}$$

(The second formula is to be used when $x = 0$, i.e. the y axis.) Now it is easy to see that $d\alpha = \omega$ not only on the rectangle but anywhere in M, which remember excludes (0,0). Now note that M is not simply connected so while we were guaranteed an α on the rectangle we are *not* guaranteed an α on all of M. Of course it looks like we actually do have an α on all of M but this is illusion. We don't. Why not? Well, to see this consider the fact that $\arctan\left(\frac{y}{x}\right)$ is our old friend the polar coordinate angle θ. Why is this bad? Because there is no single valued continuous angle θ on the punctured plane. You have two choices. The first is to put in a discontinuity at, say, the negative x-axis so that points slightly above the negative x-axis have angles close to π and points slightly below the negative x-axis have angles close to $-\pi$. This θ is clearly not continuous on the negative x-axis. The second choice is to admit a multi-valued θ but, as Grassmann pointed out, this cure is far worse than the disease. So, no function θ on M and hence no α. The situation is actually fairly common. Around any point of M we can find a neighbourhood in which there is an α, so $d\alpha = \omega$ always has a *local* solution, but there is no *global* solution for all of M. This could not happen if M were simply connected, but is common when M is not. We will discuss this further in Chapter 3. And a philosophical point. The fact that there is no θ for the punctured plane should not be regarded as a sad thing. It is the source of much that is fine and wonderful and contemplating it was one of the things that led Poincaré to his theories of Homotopy and Homology.

For our next example we do something a little easier. We take

$$\omega = (-2y + yz)\,dx \wedge dy + (-xy)\,dy \wedge dz + 2x\,dx \wedge dz$$

so that

$$\omega_{12} = -2y + yz \qquad \omega_{23} = -xy \qquad \omega_{13} = 2x$$

We take our base point as $(0, 0, 0)$. Then our formula

$$\alpha_{1i_3\ldots i_k} \;=\; 0$$

$$\alpha_{ji_3\ldots i_k} = \sum_{\ell=1}^{j-1} \int_{u_0^\ell}^{u^\ell} \omega_{\ell ji_3\ldots i_k}(u_0^1,\ldots,u_0^{\ell-1},u^\ell,\ldots,u^m)\,\underline{d}u^\ell \qquad 1<j<i_3<\ldots<i_k$$

gives us

$$\begin{aligned}
\alpha_1 &= 0\\
\alpha_2 &= \int_0^x \omega_{12}(x,y,z)\,\underline{d}x = \int_0^x -2y+yz\,\underline{d}x\\
&= -2xy+xyz\\
\alpha_3 &= \int_0^x \omega_{13}(xyz)\,\underline{d}x + \int_0^y \omega_{23}(0yz)\,\underline{d}y\\
&= \int_0^x 2x\,\underline{d}x + \int_0^y 0\,\underline{d}y\\
&= x^2
\end{aligned}$$

giving us

$$\alpha = 0\,dx + (-2xy+xyz)\,dy + x^2\,dz$$

so

$$d\alpha = (-2x+yz)\,dx\wedge dy + (-xy)\,dy\wedge dz + 2x\,dx\wedge dz = \omega$$

2.11 Riemannian and PseudoRiemannian Manifolds

In this section we will introduce Riemannian and pseudo-Riemannian structures which are imposed on a differentiable manifold as additional structure. The additional structure allows us to find angles between vectors, lengths of vectors, and then, with Calculus techniques lengths of curves and geodesics. (Geodesics are curves satisfying certain differential conditions but are naively thought of as being curves of shortest distance, or somewhat less naively as being curves for which the length is stationary when the endpoints are fixed.) Since light travels along geodesics in general relativity, the applicability of the material is obvious. Great circle routes for airplane travel are another practical example of geodesics.

It is possible to put a Riemannian or pseudo-Riemannian structure on any differentiable manifold. Locally one lifts the structure from the range of (U,ϕ) in $\mathbb{R}^n$ back onto the manifold and then globalizes by using a partition of unity subordinate to the coordinate patches, as was done in an earlier section. However, the Riemannian or pseudo-Riemannian structure may not mirror anything of mathematical interest. On the other hand, it may match some physical reality or may be useful in attacking some problem which initially is unrelated to such structures, as in the proof of the Poincaré conjecture.

2.11.1 Definition of pseudo-Riemannian Manifolds

We will begin by defining a pseudo-Riemannian structure. Everything we say about pseudo-Riemannian structures applies to Riemannian structures which are a special case. A pseudo-Riemannian structure by definition lives on the Tangent Bundle, but we remark that the same methodology can be applied to any vector bundle on the manifold. However, this goes a bit beyond our subject of differential forms, so we will sadly let it go.

A pseudo-Riemannian structure on a Differentiable Manifold M is a continuously varying non-degenerate inner product $(\ ,\ )_p$ on each $T_p(M)$. The reader is presumed to know that an inner product takes a pair of vectors into $\mathbb{R}$ and is linear in each variable. Also $(u, v) = (v, u)$ which is referred to as symmetry. Non-degenerate will be discussed later. Here we note that a vector field (= section of the bundle $T(M)$) is continuously varying if $\Phi_* v$ is continuouly varying in $\mathbb{R}^m$ for each coordinate patch (U, Φ). (This is equivalent to the coordinates of the vector in $\mathbb{R}^m$ being continuous.) If v and w are two continuously varying vector fields then we are assuming that (u, v) is a continuously varying real valued function.

The inner product is controlled in coordinates by

$$g_{ij}(u_1, \ldots u_m) = (e_i(u_1, \ldots u_m), e_j(u_1, \ldots u_m))$$

where $e_1(p), \ldots e_m(p)$ form a continuously varying local basis of $T(M)$ over a coordinate patch (U, Φ). Then $(\ ,\ )_p$ being continuouly varying is equivalent to the $g_{ij}(u_1, \ldots u_m)$ being real valued continuous functions.

For many purposes we want the g_{ij} to be *differentiable or smooth* functions but we do not need this yet.

Note the extremely important

$$g_{ji} = g_{ij}$$

since $g_{ji} = (e_i, e_j) = (e_j, e_i) = g_{ji}$.

The most often used local basis e_i is the *natural basis*

$$e_i = \frac{\partial}{\partial u^i}$$

but there is no compulsion to do so. For many purposes it is convenient to use an orthonormal basis for the inner product. This simplifies some things and complicates others. Any smoothly varying local basis will produce a collection of g_{ij}. However, it is important to realize that when moving from coordinate patch (U_1, Φ_1) to coordinate patch (U_2, Φ_2), the g_{ij} will change, and on the overlap there will be two g_{ij}'s available. Since most of what we do is local this will not impact us much.

Now suppose that $v = v^i e_i$ and $w = w^j e_j$. Then

$$(v, w) = (v^i e_i, w^j e_j) = (e_i, e_j) v^i w^j = g_{ij} v^i w^j = \begin{pmatrix} v^1 \\ \vdots \\ v^m \end{pmatrix}^\top (g_{ij}) \begin{pmatrix} w^1 \\ \vdots \\ w^m \end{pmatrix}$$

We are assuming that the inner product is non-degenerate, and this means that for a $u \in T_p(M)$,

$$\text{If } (u,v)_p = 0 \text{ for all } v \in T_p(M) \text{ then } u = 0$$

Non-degenacy is equivalent to $\det(g_{ij}) \neq 0$. Indeed, suppose that $\det(g_{ij}) = 0$. Then regarded as an operator from $\mathbb{R}^m$ to itself, it is not invertible, so there exists a vector $v = (e_1, \dots e_m)(v_1, \dots, v_m)^\top$ for which $g_{ij}v^i = 0$ and $v \neq 0$. But then for all $w \in T_p(M)$ we have $(v,w) = g_{ij}v^iw^j = 0$ and $v \neq 0$. The argument is easily reversed to get the implication in the other direction.

Hence we will always assume that $\det(g_{ij}) \neq 0$, and thus there will be an inverse matrix for (g_{ij}) which we will denote by (g^{ij}). We see immediately that

$$g_{ij}g^{jk} = \delta^j_k$$

which just expresses that the matrices are inverses of one another.

Def A pseudo-Riemannian metric is Riemannian $\Longleftrightarrow$ for any $v \neq 0$, $(v,v) > 0$.

It is a fairly easy theorem in Linear Algebra that for a symmetric bilinear form $g_{ij}v^iw^j$ a basis $e_1, \dots, e_m$ can be found for which

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \dots & \dots & \dots & \dots & \dots & \dots & \vdots \\ 0 & \dots & 1 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & -1 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \dots & \dots & \dots & \ddots & \dots & \dots & \dots & \vdots \\ 0 & \dots & \dots & \dots & \dots & -1 & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 0 \end{pmatrix}$$

with r ones and s negative ones and $m - r - s$ zeros. It's the s that will be important in the formulas. In our case, since $\det(g_{ij}) \neq 0$, there are no zeros on the diagonal and $r + s = m$. With this basis, $\det(g_{ij}) = (-1)^s$.

Now suppose we change the basis so that $e_i = \gamma^j_i f_j$ or

$$(e_1, \dots, e_m) = (e_1, \dots, e_m)(\gamma^j_i)$$

Then, with the (g_{ij}) above, we have

$$\begin{aligned} g_{ij} &= (e_i, e_j) = (\gamma^k_i f_k, \gamma^\ell_i f_\ell) \\ &= \gamma^k_i \gamma^\ell_i (f_k, f_\ell) = \gamma^k_i \gamma^\ell_i \tilde{g}_{k\ell} \end{aligned}$$

This can be written in matrix form as

$$\begin{aligned} (g_{ij}) &= (\gamma^k_i)^\top (\tilde{g}_{k\ell})(\gamma^\ell_i) \\ \det(g_{ij}) &= [\det(\gamma^\ell_i)]^2 \det(\tilde{g}_{k\ell}) \end{aligned}$$

since $\det((\gamma_i^k)^\top) = \det(\gamma_i^k)$. Note that these formula work for *any* change of coordinates, and so would be useful when moving from one coordinate patch to another. Notice that in our case we have

$$(-1)^s = \det(g_{ij}) = [\det(\gamma_i^\ell)]^2 \det(\tilde{g}_{k\ell})$$

Another simple bit of linear algebra shows that s remains the same under change of coordinates (this is Sylvesters law of inertia). We now set

Def $g = \det(g_{ij})$
and we see from the above equation that $(-1)^s g > 0$. The reader may be familiar with formulas containing $\sqrt{g}$, and this in the pseudo-Riemannian case is replaced by $\sqrt{(-1)^s g}$ which is always real, and shows up as $\sqrt{-g}$ in Relativity where $s = 3$.

Example Special Relativity. We have $\mathbb{R}^4$ with coordinates u^0, u^1, u^2, u^3 and a vector is given by

$$v = v^0 \frac{\partial}{\partial u^0} + v^1 \frac{\partial}{\partial u^1} v^2 \frac{\partial}{\partial u^2} + v^3 \frac{\partial}{\partial u^3}$$

Note $v^i = du^i(v)$. We have in special relativity

$$(v, w) = v^0 w^0 - v^1 w^1 - v^2 w^2 - v^3 w^3 = g_{ij} v^i w^j$$

so

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and $s = 3$ and so

$$\det(g_{ij}) = (-1)^3 = -1$$

Notice that (v, w) may be written as

$$(v, w) = g_{ij} du^i(v) du^j(w)$$

which explains why the inner product is often described as $g_{ij} du^i du^j$.

2.11.2 Putting a Metric on $\Lambda^k(T(M)^*)$

In order to define the $*$ operator and for other reasons we need to move the inner product (v, w) from $T(M)$ to $\Lambda^k(T(M)^*)$. We will do this in two stages. First we will move it to $T(M)^*$ and then from $T(M)^*$ to $\Lambda^k(T(M)^*)$.

Metric on $T(M)^*$

For defining the inner product on $T(M)^*$ it is convenient make use of the correlation $\Psi : V \to V^*$ which we will shortly define. A correlation is equivalent

to a (non-degenerate) inner product, in the sense that given one the other can be constructed.

Let $v, w \in T(M)$. The we can define, via the inner product, a linear functional $\ell_v \in T^*(M)$ by

$$\ell_v(w) \stackrel{\text{def}}{=} (v, w)$$

We then define the correlation by

$$\Psi(v) \stackrel{\text{def}}{=} \ell_v$$

and it is mere notation that

$$\begin{aligned} < \ell_v, w > &= (v, w) \\ < \Psi(v), w > &= (v, w) \end{aligned}$$

Suppose now that $\Psi(v) = 0$. Then $(v, w) = 0$ for all w so by nondegeneracy we have $v = 0$. Hence Ψ is one to one (injective) and since $\dim(T(M)) = \dim(T(M)^*)$ we know Ψ is an isomorphism.

We note in passing that given a correlation (an isomorphism $\Psi : V \to V^*$) we can construct an inner product by

$$(u, v) \stackrel{\text{def}}{=} < \Psi(v), w >$$

Nondegeneracy follows from Ψ being an isomorphism. Details are left to the reader.

Our next task is finding the matrix for Ψ. Let $e_1, \ldots, e_m$ be a local basis for $T(M)$ and $e^1, \ldots, e^m$ the dual basis for $T(M)^*$. We want a matrix for Ψ in these bases. As usual $g_{ij} = (e_i, e_j)$. Let $\Psi(v) = \ell$ and $v = v^i e_i$, $\ell = \ell_j e^j$. Then

$$\begin{aligned} \ell_j &= \langle \ell, e_j \rangle = \langle \Psi(v), e_j \rangle = (v, e_j) \\ &= (v^i e_i, e_j) = v^i (e_i, e_j) = g_{ij} v^i \end{aligned}$$

So the matrix of Ψ in the bases $e_1, \ldots, e_m$ for $T(M)$ and $e^1, \ldots, e^m$ for $T(M)^*$ is (g_{ij}). Note that Ψ is an isomorphism if and only if $\det(g_{ij}) \neq 0$ if and only if the inner product is non-degenerate. Note also that Ψ^{-1} exists and its matrix is $(g_{ij})^{-1} = (g^{ij})$. And finally note that $\ell_j = g_{ij} v^i$ is an instance of "index lowering", which is endemic in tensor analysis and now you know what it means. Note also

$$g^{ij} \ell_j = g^{ij} g_{jk} v^k = \delta^i_k v^k = v^i$$

which is an instance of "index raising". In tensor analysis they would use v_j instead of our ℓ_j, since the position of the index tells them what space the element is in. They regarded this as being two descriptions of the same element, which makes at least one of us uncomfortable.

We can now define the inner proudcut on $T(M)^*$ easily using the correlation Ψ. Let $\ell, m \in T(M)^*$.

$$(\ell, m) \stackrel{\text{def}}{=} (\Psi^{-1}(\ell), \Psi^{-1}(m))$$

We check non-degeneracy. Suppose $(\ell, m) = 0$ for all m. Then $(\Psi^{-1}(\ell), \Psi^{-1}(m)) = 0$ for all m and since Ψ^{-1} is an isomorphism we have $(\Psi^{-1}(\ell), w) = 0$ for all w. By nondegeneracy we have $\Psi^{-1}(\ell) = 0$ and since Ψ^{-1} is an isomorphism we have $\ell = 0$ as required. (Note that we used both the onto (surjective) and one to one (injective) properties of Ψ^{-1}.)

Now we would like the matrix of (ℓ, m) in the basis $e^1, \ldots, e^m$ of $T(M)^*$. Let $\ell = \ell_i e^i$ and $m = m_j e^j \in T(M)^*$ and also $\Psi^{-1}(\ell) = v = v^k e_k$ and $\Psi^{-1}(m) = w = w^\ell e_\ell \in T(M)$. Then

$$\begin{aligned}(\ell, m) &= (\Psi^{-1}(\ell), \Psi^{-1}(m)) = (v, w) = (v^k e_k, w^\ell e_\ell) \\ &= (g^{ki}\ell_i e_k, g^{\ell j} m_j e_\ell) = g^{ki} g^{\ell j} \ell_i m_j (e_k, e_\ell) \\ &= g^{ki} g^{\ell j} \ell_i m_j g_{k\ell} = g^{ki} \ell_i m_j \delta^j_k = g^{ik} \ell_i m_k\end{aligned}$$

which gives us the matrix of the inner product in $T(M)^*$. It is nice to know that

$$(e^i, e^j) = (\delta^i_k e^k, \delta^i_k e^\ell) = g^{kl} \delta^i_k \delta^i_k = g^{ij}$$

which we were expecting.

We digest this for the natural basis $\frac{\partial}{\partial u^1}, \ldots, \frac{\partial}{\partial u^m}$

$$\begin{aligned}\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\right) &= g_{ij} \\ (du^i, du^j) &= g^{ij} \\ \Psi\left(\frac{\partial}{\partial u^i}\right) &= \Psi\left(\delta^j_i \frac{\partial}{\partial u^j}\right) = g_{kj}\delta^j_i e^k = g_{ik} e^k \\ \Psi^{-1}(du^i) &= g^{ik} e_k\end{aligned}$$

Note carefully that, in most circumstances, $\Psi\left(\frac{\partial}{\partial u^i}\right) \neq du^i$. This is a common source of error.

Metric on $\Lambda^k(T(M)^*)$

Our next project is to extend the inner product from $\Lambda^1(T(M)^*)$ to $\Lambda^k(T(M)^*)$. The same tricks would let us extend the metric from $\Lambda^1(T(M))$ to $\Lambda^k(T(M))$ but we do not need this here.

There are several ways to do this and we do not claim any special virtue for the path we have chosen. In the end it gives correct results where it intersects with other mathematics and that is about all we can ask for.

Recall

$$e^\sigma = \operatorname{sgn}(\sigma)\ e^\sigma(1) \wedge \ldots \wedge e^\sigma(k) \quad \text{for } \sigma \in \mathcal{S}_{m,k}$$

We now define

$$(e^\sigma, e^\tau) \stackrel{\text{def}}{=} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\ \det \begin{pmatrix} (e^{\sigma(1)}e^{\tau(1)}) & \ldots & (e^{\sigma(1)}e^{\tau(k)}) \\ \cdots & \cdots & \cdots \\ (e^{\sigma(k)}e^{\tau(1)}) & \ldots & (e^{\sigma(k)}e^{\tau(k)}) \end{pmatrix}$$

$$= \quad \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\ \det\begin{pmatrix} g^{\sigma(1)\tau(1)} & \dots & g^{\sigma(1)\tau(k)} \\ \cdots & \cdots & \cdots \\ g^{\sigma(k)\tau(1)} & \dots & g^{\sigma(k)\tau(k)} \end{pmatrix}$$

The determinant is a minor of g^{ij}) and we will have a lot of use for this minor and the accompanying sign, so we will give it a designation $g^{\sigma\tau}$:

$$(e^\sigma, e^\tau) = g^{\sigma\tau} \stackrel{\text{def}}{=} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\ \det(g^{\sigma(i)\tau(j)})$$

Recall the Laplace expansion by complementary minors (with $\tau \in \mathcal{S}_{m,k}$ fixed)

$$\det(g^{ij}) =$$
$$= \sum_{\sigma\in\mathcal{S}_{m,k}} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \begin{vmatrix} g^{\sigma(1)\tau(1)} & \dots & g^{\sigma(1)\tau(k)} \\ \cdots & \cdots & \cdots \\ g^{\sigma(k)\tau(1)} & \dots & g^{\sigma(k)\tau(k)} \end{vmatrix} \begin{vmatrix} g^{\sigma(1)\tau(1)} & \dots & g^{\sigma(1)\tau(k)} \\ \cdots & \cdots & \cdots \\ g^{\sigma(k)\tau(1)} & \dots & g^{\sigma(k)\tau(k)} \end{vmatrix}$$

and recall that $\operatorname{sgn}(\sigma)\operatorname{sgn}(\tilde{\sigma}) = (-1)^{k(m-k)}$. Then the above equation may be written

$$\begin{aligned} \det(g^{ij}) &= \sum_{\sigma\in\mathcal{S}_{m,k}} g^{\sigma\tau}\operatorname{sgn}(\tilde{\sigma})\operatorname{sgn}(\tilde{\tau})\, g^{\tilde{\sigma}\tilde{\tau}} \\ &= \sum_{\sigma\in\mathcal{S}_{m,k}} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\, g^{\sigma\tau} g^{\tilde{\sigma}\tilde{\tau}} \end{aligned}$$

using $\operatorname{sgn}(\sigma)\operatorname{sgn}(\tilde{\sigma}) = (-1)^{k(m-k)}$ twice.

Thus we end up with the inner product defined on the base vectors e^σ of $\Lambda^k(T(M)^*)$ by

$$(e^\sigma, e^\tau) = g^{\sigma\tau} = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\ \det(g^{\sigma(i)\tau(j)})$$

and this can be extended to all of $\Lambda^k(T(M)^*)$ by linearity. Finally, if if $k \neq \ell$ and $e^\sigma \in \Lambda^k(T(M)^*)$ and $\tau \in \Lambda^\ell(T(M)^*)$ then we set $(e^\sigma, e^\tau) = 0$. This gives us an inner product on $\Lambda(T(M)^*)$.

The attentive reader is now thinking about whether this inner product is indeed non-degenerate. It is, but a direct proof would take us a little off track. We can prove this easily with the $*$-operator in the next section. The procedure lacks elegance but makes up for it in the ease with which it is done. We hope you enjoy the methodology.

2.12 The $*$-Operator

One cannot get very far using differential forms without the dualizing operator $*$ which we saw used in the first chapter for deriving formulas for curvilinear coordinates and which gave us nice formulas for electromagnetism. In this section we will give an introduction to the $*$ operator which will be more complete than the treatment in the first Chapter, which relied heavily on an orthonormal

basis, which is the usual method. Here we will develop the theory using an arbitrary basis which can then be used with an arbitrary coordinate system. Thus this section will present the basic ideas and the formulas for the $*$ operator sufficient to justify the work done in the first chapter. We will not repeat material here which was adequately covered there. In general the $*$ operator is a rather complicated beast and so ones desires to do computations are sometime frustrated.

It is important to realize that to have a $*$-operator on a manifold, *the manifold must be oriented*.

Using the Riemannian and pseudo-Riemannian structure we can define the $*$-operator which gives a duality between $\Lambda^k(M)$ and $\Lambda^{m-k}(M)$. This can be used for a number of things, as was illustrated in the electrical sections of Chapter 1, and it is also one road to Poincaré duality for cohomology as we will see in Chapter 3. The formulas for the Beltrami Operator on a Pseudo Riemannian Manifold suffer no changes and thus we will not delve further into them in this chapter.

Our treatment is surprisingly simple if you overlook the unpleasant aspects of computation, which are sometimes tedius. We point out that summing over increasing permutations gives us almost the kind of control one has in classical tensor analysis, although keeping control of the signs of the permutations is a constant annoyance and an rich source of errors.

2.12.1 The Unit Box

The $*$-operator is based upon the *unit box* which we are now going to define. Recall that $\Lambda^m(T(M))$ and $\Lambda^m(T(M))^*$ are one dimensional spaces with an inner product. We will concentrate on $\Lambda^m(T(M))^*$ but the same tricks work for the other case. In each case, there are two elements of unit norm. We can pick out one of the two at any point by using a coordinate system from the orientation Atlas around that point. If the coordinates are $u^1, \ldots, u^m$ (in that order) then we select $\Omega = du^1 \wedge \ldots \wedge du^m$ as a preliminary basis element for $\Lambda^m(T(M)^*)$. However, it will depend on the coordinate patch chosen, and we wish to avoid this. Thus we require, in addition, that the basis element have unit norm. Since we wish to preserve the orientation, the new basis element will be $\Omega_0 = k\Omega$ where $k > 0$. Thus our requirement is that $|(\Omega_0, \Omega_0)| = 1$. We need the absolute value signs since with and indefinite metric we cannot predict the sign of (Ω_0, Ω_0).

Recall that if the inner product, when diagonalized, has s negative ones on the diagonal. Thus $g = \det(g_{ij}) = (-1)^s \cdot (\text{a positive number})$ so

$$(-1)^s g > 0$$

Since $(g^{ij}) = (g_{ij})^{-1}$, $\det(g^{ij}) = \frac{1}{g}$, We can now compute (Ω_0, Ω_0) where $\Omega_0 = k\Omega$ and $\Omega = du^1 \wedge \ldots \wedge du^m$.

$$(\Omega_0, \Omega_0) \quad = \quad (k\Omega, k\Omega) \; = \; k^2(du^1 \wedge \ldots \wedge du^m, du^1 \wedge \ldots \wedge du^m)$$

$$\begin{aligned} &= k^2 \begin{vmatrix} (du^1,du^1) & \dots & (du^1,du^m) \\ \dots & \dots & \dots \\ (du^m,du^1) & \dots & (du^m,du^m) \end{vmatrix} \\ &= k^2 \begin{vmatrix} g^{11} & \dots & g^{1m} \\ \dots & \dots & \dots \\ g^{m1} & \dots & g^{mm} \end{vmatrix} = k^2 \frac{1}{g} = k^2 (-1)^s \frac{1}{(-1)^s g} \\ &= (-1)^s \left[\frac{k}{\sqrt{(-1)^s g}} \right]^2 \end{aligned}$$

Recall $k > 0$ because of orientation. Thus to get $|(\Omega_0, \Omega_0)| = 1$ we must set $k = \sqrt{(-1)^s g}$ and then we have

$$\begin{aligned} \Omega_0 &= k\Omega = \sqrt{(-1)^s g}\, \Omega = \sqrt{(-1)^s g}\, du^1 \wedge du^2 \wedge \ldots \wedge du^m \\ (\Omega_0, \Omega_0) &= (-1)^s \end{aligned}$$

We have assumed that the manifold M over which we are working is oriented, so we may always adjust any coordinate system so that it is positively oriented with respect to the oriented Atlas. Using this coordinate system and the formula above always gives us the same Ω_0, since there are only two elements of $\Lambda^m(T(M)^*)$ with unit size, and the other one has the wrong orientation. (Actually, the other one is $\tilde{\Omega}_0 = \sqrt{(-1)^s g}\, \mathbf{du^2} \wedge \mathbf{du^1} \wedge du^3 \wedge \ldots \wedge du^m$ if you ever need it. Note $\tilde{\Omega}_0 = -\Omega_0$ for many reasons.)

Theorem On an oriented manifold there is a unique topform Ω_0 with the same orientation as the oriented Atlas. The formula, using any properly oriented coordinate system, is

$$\Omega_0 = \sqrt{(-1)^s g}\, du^1 \wedge du^2 \wedge \ldots \wedge du^m$$

We have used $du^1, \ldots, du^n$ as the basis of of $T(M)^*$ in this section, but absolutely nothing would change if we took an arbitrary correctly oriented local basis $e^1, \ldots, e^n$ instead. The identical arguments would then give

$$\begin{aligned} \Omega_0 &= \sqrt{(-1)^s g}\, e^1 \wedge \ldots \wedge e^m \\ e^1 \wedge \ldots \wedge e^m &= \frac{1}{\sqrt{(-1)^s g}}\, \Omega_0 \end{aligned}$$

where of course $g = \det(g_{ij})$, $g_{ij} = (e_i, e_j)$ and $g^{ij} = (e^i, e^j)$.

2.12.2 Definition of the $*$-Operator

We are now in a position to define the $*$-operator. It was invented by Grassmann in 1842, but it's use in the Manifold situation was pioneered by W.V.D. Hodge. There is basically a single formula from which everything algebraic about $*$ can be derived. It is

$$\lambda \wedge *\eta = (\lambda, \eta)\, \Omega_0 \qquad \text{for } \lambda, \eta \in \Lambda^k(T(M)^*)$$

The Ω_0 reminds us that we are working on an *oriented* manifold M. Now there is a question of whether a $*\eta$ exists satisfying this equation. The usual method for proving this is to make use of duality in the inner product space $\Lambda^k(T(M)^*)$ but this is not available to us because we never checked that the inner product on $\Lambda^k(T(M)^*)$ was non-degenerate. Thus we must use a more direct method, which makes up in efficiency what it lacks in elegance. We will assume that $*\eta$ indeed exists, derive a formula for it in coordinates, and then check that the formula indeed produces a $*\eta$ that satisfies $\lambda \wedge *\eta = (\lambda, \eta)\ \Omega_0$. To push through the calculation we will need to do some preliminary work. First recall that for $\sigma \in \mathcal{S}_{n,k}$

$$e^\sigma = \mathrm{sgn}(\sigma)\ e^{\sigma(1)} \wedge \ldots \wedge e^{\sigma(k)}$$

and also note that as τ runs through $\mathcal{S}_{n,k}$, the reverse $\tilde{\tau}$ runs through $\mathcal{S}_{n,n-k}$. Also note that if $\sigma \neq \tau$ then some $\sigma(i)$ will *not* occur in $\tau(1), \ldots, \tau(k)$ and so that $\sigma(i)$ *will* occur in $\tau(k{+}1), \ldots \tau(m)$ and hence *will* occur in $\tilde{\tau}(1), \ldots, \tilde{\tau}(n{-}k)$ and thus $e^\sigma \wedge e^{\tilde{\tau}} = 0$ due to a repeated element in the product. It remains to find $e^\sigma \wedge e^{\tilde{\sigma}}$.

$$\begin{aligned}
e^\sigma \wedge e^{\tilde{\sigma}} &= \mathrm{sgn}(\sigma)\ \mathrm{sgn}(\tilde{\sigma})\ e^{\sigma(1)} \wedge \ldots \wedge e^{\sigma(k)} e^{\tilde{\sigma}(1)} \wedge \ldots \wedge e^{\tilde{\sigma}(n-k)} \\
&= \mathrm{sgn}(\sigma)\ \mathrm{sgn}(\tilde{\sigma})\ e^{\sigma(1)} \wedge \ldots \wedge e^{\sigma(k)} e^{\sigma(k+1)} \wedge \ldots \wedge e^{\sigma(m)} \\
&= (-1)^{k(n-k)} \mathrm{sgn}(\sigma)\ e^1 \wedge \ldots \wedge e^m \\
&= \mathrm{sgn}(\tilde{\sigma})\ e^1 \wedge \ldots \wedge e^m \\
&= (-1)^{k(n-k)} \mathrm{sgn}(\sigma)\ \frac{1}{\sqrt{(-1)^s g}}\ \Omega_0
\end{aligned}$$

The next to the last formula,

$$\boxed{e^\sigma \wedge e^{\tilde{\sigma}} = \mathrm{sgn}(\tilde{\sigma})\ e^1 \wedge \ldots \wedge e^m}$$

is perhaps the handiest for computation.

We now want to use the basic formula $\lambda \wedge *\eta = (\omega,\ \eta)\Omega_0$ to find a formula for $*\eta$. We will concentrate on $*e^\sigma$ because if we know this the rest is just linearity. We set

$$*e^\sigma = \lambda^\sigma_{\tilde{\tau}}\ e^{\tilde{\tau}}$$

Note summation convention is operative with respect to $\tilde{\tau}$. Our job is to determine $\lambda^\sigma_{\tilde{\tau}}$ and we do this using familiar methods of linear algebra, although they look a little different in this context. We take the wedge product of $*e^\sigma$ with e^ρ and evaluate in two different ways. First we have

$$e^\rho \wedge *e^\sigma = (e^\rho,\ e^\sigma)\Omega_0 = g^{\rho\sigma}\Omega_0$$

Next we have

$$\begin{aligned}
e^\rho \wedge *e^\sigma &= e^\rho \wedge \lambda^\sigma_{\tilde{\tau}} e^{\tilde{\tau}} \\
&= \lambda^\sigma_{\tilde{\tau}}\ e^\rho \wedge e^{\tilde{\tau}}
\end{aligned}$$

Now $e^\rho \wedge e^{\tilde{\tau}} = 0$ if $\rho \neq \tau$ so we only get $\lambda^\sigma_{\tilde{\rho}}$ from his equation but it is enough.

$$\begin{aligned} e^\rho \wedge *e^\sigma &= \lambda^\sigma_{\tilde{\rho}}\, e^\rho \wedge e^{\tilde{\rho}} \qquad \text{(no sum on } \rho\text{)} \\ &= \lambda^\sigma_{\tilde{\rho}}\, \mathrm{sgn}(\tilde{\rho})\, \frac{1}{\sqrt{(-1)^s g}}\, \Omega_0 \end{aligned}$$

Comparing the two expressions for $e^\rho \wedge *e^\sigma$ we have

$$g^{\rho\sigma}\, \Omega_0 = \lambda^\sigma_{\tilde{\rho}}\, \mathrm{sgn}(\tilde{\rho})\, \frac{1}{\sqrt{(-1)^s g}}\, \Omega_0$$

from which we immediately get

$$\begin{aligned} \lambda^\sigma_{\tilde{\rho}} &= \mathrm{sgn}(\tilde{\rho})\, g^{\rho\sigma} \sqrt{(-1)^s g} \\ &= (-1)^{k(n-k)} \mathrm{sgn}(\rho)\, g^{\rho\sigma} \sqrt{(-1)^s g} \end{aligned}$$

and thus

$$\begin{aligned} *e^\sigma &= \sum_{\rho \in \mathcal{S}_{n,k}} g^{\rho\sigma} \sqrt{(-1)^s g}\, \mathrm{sgn}(\tilde{\rho})\, e^{\tilde{\rho}} \\ *(\mathrm{sgn}(\sigma)\, e^{\sigma(1)} \wedge \ldots \wedge e^{\sigma(k)}) &= \sum_{\rho \in \mathcal{S}_{n,k}} g^{\rho\sigma} \sqrt{(-1)^s g}\, e^{\rho(k+1)} \wedge \ldots \wedge e^{\rho(m)} \\ *(e^{\sigma(1)} \wedge \ldots \wedge e^{\sigma(k)}) &= \mathrm{sgn}(\sigma) \sum_{\rho \in \mathcal{S}_{n,k}} g^{\rho\sigma} \sqrt{(-1)^s g}\, e^{\rho(k+1)} \wedge \ldots \wedge e^{\rho(m)} \end{aligned}$$

where the first equation

$$\boxed{*e^\sigma = q \textstyle\sum_{\rho \in \mathcal{S}_{n,k}} g^{\rho\sigma} \sqrt{(-1)^s g}\, \mathrm{sgn}(\tilde{\rho})\, e^{\tilde{\rho}}}$$

is perhaps the handiest for computation.

Let us now use the formula in a case where we already know the correct answer. We will compute $*du^i$ where we take $s = 0$ and $k = 1$. For this

$$\begin{aligned} \sigma &= \left(\begin{array}{c|cccccc} 1 & 2 & \ldots & \ldots & \ldots & \ldots & m \\ i & 1 & \ldots & i-1 & i+1 & \ldots & m \end{array} \right) \\ \rho &= \left(\begin{array}{c|cccccc} 1 & 2 & \ldots & \ldots & \ldots & \ldots & m \\ i & 1 & \ldots & j-1 & j+1 & \ldots & m \end{array} \right) \end{aligned}$$

$$\begin{aligned} g^{\rho\sigma} &= \mathrm{sgn}(\rho)\, \mathrm{sgn}(\sigma)\, \det(g^{\rho(1)\sigma(1)}) \\ &= (-1)^{i-1}(-1)^{j-1} g^{ji} \;=\; (-1)^{i+j} g^{ij} \end{aligned}$$

$$\begin{aligned} *(\mathrm{sgn}(\sigma)\, du^\sigma) &= \sum_{\rho \in \mathcal{S}_{m,k}} g^{\rho\sigma} \sqrt{g}\, \mathrm{sgn}(\tilde{\rho})\, du^{\tilde{\rho}} \\ (-1)^{i-1} * du^i &= \sum_{j=1}^{m} (-1)^{i+j} g^{ij} \sqrt{g}\, du^1 \wedge \ldots \wedge du^{j-1} \wedge du^{j+1} \ldots \wedge du^m \\ *du^i &= \sum_{j=1}^{m} (-1)^{j-1} g^{ij} \sqrt{g}\, du^1 \wedge \ldots \wedge \widehat{du^j} \wedge \ldots \wedge du^m \end{aligned}$$

where as usual the hat on $\widehat{du^j}$ means that du^j is *missing*. This is the same result we derived in Chapter 1. In the case of orthongonal coordinates it immediately gives, with $(\frac{\partial}{\partial u^i},\, \frac{\partial}{\partial u^i}) = h_i^2$,

$$*du^i = (-1)^{i-1} \frac{h_1 \cdots \widehat{h_i} \cdots h_m}{h_i}\, du^1 \wedge \ldots \wedge \widehat{du^i} \wedge \ldots \wedge du^m$$

What we have shown is that if $*e^\sigma$ exists, it must be given by

$$*e^\sigma = \sum_{\rho \in \mathcal{S}_{n,k}} g^{\rho\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\, e^{\tilde{\rho}}$$

We now quickly check that this formula actually works.

$$e^\tau \wedge *e^\sigma \quad = \quad \sum_{\rho \in \mathcal{S}_{n,k}} g^{\rho\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\, e^\tau \wedge e^{\tilde{\rho}}$$

The right side will be 0 unless $\rho = \tau$ so

$$\begin{aligned} e^\tau \wedge *e^\sigma &= g^{\tau\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\, e^\rho \wedge e^{\tilde{\rho}} \\ &= g^{\tau\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\, \operatorname{sgn}(\tilde{\rho})\, e^1 \wedge \ldots \wedge e^m \\ &= g^{\tau\sigma} \sqrt{(-1)^s g}\, \frac{1}{\sqrt{(-1)^s g}}\, \Omega_0 \\ &= (e^\rho,\, e^\sigma)\, \Omega_0 \end{aligned}$$

as required. Careful contemplation of this verification will give the attentive reader a good feel for how everything is working.

Next we ask, is $*$ surjective. Indeed it is even better; it is almost an involution. Indeed

$$** \omega = (-1)^{k(n-k)+s}\, \omega \qquad\qquad \text{for } \omega \in \Lambda^k$$

As usual, s is the number of negative entries on the diagonal of a diagonalized (g_{ij}). We verify this formula with a computation. In the computation the critical step is the Laplace expansion by complementary minors, which we remind the reader is

$$\det(g^{ij}) = \delta_{\sigma\tau} \sum_{\rho \in \mathcal{S}_{m,k}} \operatorname{sgn}(\rho)\, \operatorname{sgn}(\tau)\, g^{\rho\sigma} g^{\tilde{\rho}\tilde{\tau}}$$

So off we go

$$\begin{aligned} * e^\sigma &= \sum_\rho g^{\rho\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\, e^{\tilde{\rho}} \\ ** e^\sigma &= \sum_\rho g^{\rho\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\ * e^{\tilde{\rho}} \\ ** e^\sigma &= \sum_{\rho,\tau} g^{\rho\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\, g^{\tilde{\rho}\tilde{\tau}} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tau)\, e^\tau \end{aligned}$$

$$\begin{aligned} &= \sum_\tau (-1)^s g(-1)^{k(n-k)} \left(\sum_\rho \operatorname{sgn}(\rho)\ \operatorname{sgn}(\tau)\ g^{\rho\sigma} g^{\tilde{\rho}\tilde{\tau}} \right) e^{\tilde{\tau}} \\ &= (-1)^{k(n-k)+s} g \sum_\tau \delta_{\sigma\tau} \det(g^{ij}) e^\tau \\ &= (-1)^{k(n-k)+s} g \frac{1}{g}\, e^\sigma \\ &= (-1)^{k(n-k)+s} e^\sigma \end{aligned}$$

as desired. It is interesting that the Laplce expansion, which is the critcal step here, is equivalent to associativity of the wedge product. So the formula we have derived is connected in some subtle way to associativity. We know no more.

We see immediately that

$$*(*(-1)^{k(n-k)+s} e^\sigma) = e^\sigma$$

proving that $*$ is surjective and thus injective and an isomorphism from $\Lambda^k \to \Lambda^{m-k}$ since the two spaces have the same dimension.

We are also in a postion now to settle a question that derailed the abstract development for us, namely is the inner product on $\Lambda^k(T(M)^*)$ non-degenerate? To check this, suppose that $(\lambda, \eta) = 0$ for all λ. By the formula $\lambda \wedge *\eta = (\lambda, \eta)\, \Omega_0$, and the fact that $*$ is surjective, we see that $\lambda \wedge \mu = 0$ for all μ in $\Lambda^{n-k}(T(M)^*)$. But then, if λ is written in terms of coordinates, $\lambda = \lambda_\sigma e^\sigma$, we see, by wedging λ by $e^{\tilde{\tau}}$ that each λ_σ must be 0, so $\lambda = 0$.

2.12.3 Special Case of Orthogonal Coordinates

In practical life the coordinate systems being used often have the coordinate lines intersecting orthogonally, as longitude and latitude lines do on the sphere. In this case great simplifications occur which makes the computations much more pleasant. We discussed this in practical terms in Chapter 1 so here we need only show that the formula given there follows from the general formula. (You recall we briefly mentioned this for $*\omega$ where $\omega \in \Lambda^1$ in the previous section.) It is also not difficult to derive these formulas from the basic formula $\lambda \wedge *\eta = (\lambda,\, \eta)\, \Omega_0$ but this way will give us some practise with the general formula.

The tangent vectors to the coordinate lines will be $e_i = \frac{\partial}{\partial u^i}$ and the orthogonality of the coordinate lines means that for $i \neq j$ we have $(e_i, e_j) = 0$. Hence the matrix elements $g_{ij} = (e_i, e_j)$ will be 0 when $i \neq j$ and the matrix (g_{ij}) will be a diagonal matrix. To match up with the customary notation in physics and engineering we denote the diagonal elements by $\pm\, h_i^2$ so that $g_{ii} = (e_i, e_i) = \pm\, h_i^2$

and

$$(g_{ij}) = \begin{pmatrix} h_1^2 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & h_2^2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \dots & h_r^2 & 0 & \dots & 0 \\ 0 & \dots & \dots & 0 & -h_{r+1}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & -h_{r+s}^2 \end{pmatrix}$$

and $g^{ii} = (e^i, e^i) = \pm 1/h_i^2$ and

$$(g_{ij}) = \begin{pmatrix} \frac{1}{h_1^2} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{h_2^2} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \dots & \frac{1}{h_r^2} & 0 & \dots & 0 \\ 0 & \dots & \dots & 0 & -\frac{1}{h_{r+1}^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \dots & 0 & 0 & \dots & -\frac{1}{h_{r+s}^2} \end{pmatrix}$$

Since all off-diagonal elements are 0, $g^{\sigma\tau} = 0$ for $\tau \neq \sigma$. Let $\ell = s(\sigma) =$ (the number of elements in $h_{\sigma(1)}, h_{\sigma(2)}, \dots, h_{\sigma(k)}$ which are negative). Then

$$\begin{aligned} g^{\sigma\sigma} &= \operatorname{sgn}(\sigma)\, \operatorname{sgn}(\sigma)\, (-1)^{s(\sigma)} \frac{1}{h_{\sigma(1)}^2 h_{\sigma(2)}^2 \cdots h_{\sigma(k)}^2} \\ &= (-1)^{s(\sigma)} \frac{1}{h_{\sigma(1)}^2 h_{\sigma(2)}^2 \cdots h_{\sigma(k)}^2} \end{aligned}$$

Now recall the general formula

$$* e^\sigma = \sum_{\rho \in \mathcal{S}_{n,r}} g^{\rho\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\rho})\, e^{\tilde{\rho}}$$

which in our case becomes

$$* e^\sigma = g^{\sigma\sigma} \sqrt{(-1)^s g}\, \operatorname{sgn}(\tilde{\sigma})\, e^{\tilde{\sigma}}$$

Decoding the symbols we have

$$\begin{aligned} *(\operatorname{sgn}(\sigma)\, e^{\sigma(1)} \cdots e^{\sigma(1)}) &= (-1)^{s(\sigma)} \frac{1}{h_{\sigma(1)}^2 h_{\sigma(2)}^2 \cdots h_{\sigma(k)}^2} \sqrt{h_1^2 \cdots h_m^2}\, e^{\tilde{\sigma}(1)} \cdots e^{\tilde{\sigma}(m-k)} \\ *(e^{\sigma(1)} \cdots e^{\sigma(1)}) &= (-1)^{s(\sigma)} \operatorname{sgn}(\sigma)\, \frac{h_{\tilde{\sigma}(1)} h_{\tilde{\sigma}(2)} \cdots h_{\tilde{\sigma}(m-k)}}{h_{\sigma(1)} h_{\sigma(2)} \cdots h_{\sigma(k)}} e^{\tilde{\sigma}(1)} \cdots e^{\tilde{\sigma}(m-k)} \end{aligned}$$

which can also be written as

$$*(e^{\sigma(1)} \cdots e^{\sigma(1)}) = (-1)^{s(\sigma)} \operatorname{sgn}(\sigma)\, \frac{h_{\sigma(k+1)} \cdots h_{\sigma(m)}}{h_{\sigma(1)} \cdots h_{\sigma(k)}} e^{\sigma(k+1)} \cdots e^{\sigma(m)}$$

2.12.4 Example: Special Relativity

Let us give a few examples of the use of the formula

$$*(e^{\sigma(1)}\cdots e^{\sigma(1)}) = (-1)^{s(\sigma)}\mathrm{sgn}(\sigma)\,\frac{h_{\sigma(k+1)}\cdots h_{\sigma(m)}}{h_{\sigma(1)}\cdots h_{\sigma(k)}}e^{\sigma(k+1)}\cdots e^{\sigma(m)}$$

in special relativity. In this case $m = 4$ and $s = 3$ and all $h_i = 1$. Suppose we want $*dx^0 = *c\,dt$. Then

$$\sigma = \left(\begin{array}{c|ccc} 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{array}\right) \qquad k = 1 \qquad s(\sigma) = 0 \qquad \mathrm{sgn}(\sigma) = 1$$

$$\begin{aligned} *\,dx^0 &= 1\cdot 1\cdot 1\,dx^1\wedge dx^2\wedge dx^3 = dx^1\wedge dx^2\wedge dx^3 \\ *\,dx^1\wedge dx^2\wedge dx^3 &= **\,dx^0 = (-1)^{k(m-k)+s}dx^0 = dx^0 \end{aligned}$$

$$\sigma = \left(\begin{array}{c|ccc} 0 & 1 & 2 & 3 \\ 1 & 0 & 2 & 3 \end{array}\right) \qquad k = 1 \qquad s(\sigma) = 1 \qquad \mathrm{sgn}(\sigma) = -1$$

$$\begin{aligned} *\,dx^1 &= (-1)\cdot(-1)\cdot 1\,dx^0\wedge dx^2\wedge dx^3 = dx^0\wedge dx^2\wedge dx^3 \\ *\,dx^0\wedge dx^2\wedge dx^3 &= **\,dx^1 = (-1)^{k(m-k)+s}dx^1 = dx^1 \end{aligned}$$

$$\sigma = \left(\begin{array}{c|ccc} 0 & 1 & 2 & 3 \\ 2 & 0 & 1 & 3 \end{array}\right) \qquad k = 1 \qquad s(\sigma) = 1 \qquad \mathrm{sgn}(\sigma) = 1$$

$$\begin{aligned} *\,dx^2 &= (-1)\cdot 1\cdot 1\,dx^0\wedge dx^1\wedge dx^3 = -dx^0\wedge dx^1\wedge dx^3 \\ &= dx^0\wedge dx^3\wedge dx^1 \\ *\,dx^0\wedge dx^3\wedge dx^1 &= **\,dx^2 = (-1)^{k(m-k)+s}dx^2 \\ &= (-1)^{3+3}dx^2 = dx^2 \end{aligned}$$

$$\sigma = \left(\begin{array}{cc|cc} 0 & 1 & 2 & 3 \\ 0 & 2 & 1 & 3 \end{array}\right) \qquad k = 2 \qquad s(\sigma) = 1 \qquad \mathrm{sgn}(\sigma) = -1$$

$$\begin{aligned} *\,dx^0\wedge dx^2 &= (-1)\cdot(-1)\,dx^1\wedge dx^3 = -dx^3\wedge dx^1 \\ dx^3\wedge dx^1 &= -**\,dx^0\wedge dx^2 = -(-1)^{k(m-k)+s}dx^0\wedge dx^2 \\ &= dx^0\wedge dx^2 \end{aligned}$$

just as we found in Chapter 1.

2.13 Geometric and Physical Applications of Differential Forms, Part II

squat

Bibliography

[2] Frankel, Theodore THE GEOMETRY OF PHYSICS, 3rd edition, Cambridge University Press, Cambridge,etc, 2012

[3] Gilkey, Peter B. INVARIANCE THEORY, THE HEAT EQUATION AND THE ATIZYA-SINGER INDEX THEOREM, 2nd Edition Boca Raton 1995

[6] Lovelock, David and Rund, Hanno TENSORS, DIFFERENTIAL FORMS AND VARIATIONAL PRINCIPLES, Wiley Interscience

[4] Misner, Thorne and Wheeler, GRAVITATION, W.H. Freeman, San Francisco, 1973

[5] Nakahara, Mikio GEOMETRY, TOPOLOGY, AND PHYSICS, Institute of Physics Publishing, London, 2003

[7] Roe, John, ELLIPTIC OPERATORS, TOPOLOGY AND ASYMPTOTIC METHODS, 2nd Edition, Boca Raton, 1998

[8] Rosenberg, S. THE LAPLACIAN ON A RIEMANNIAN MANIFOLD, Cambridge, 1997

[9] Schulz, William, THEORY AND APPLICATION OF GRASSMANN ALGEBRA, http://www.cefns.nau.edu/ schulz/grassmann.pdf