A Note of Explanation

It has been my experience that many incoming graduate students have not yet been exposed to some of the more advanced mathematics needed to access electrodynamics or quantum mechanics. Mathematical tools such as tensors or differential forms, intended to make the calculations much simpler, become hurdles that the students must overcome, taking away the time and understanding from the physics. The professors, on the other hand, do not feel it necessary to go over these topics. They cannot be blamed, since they are trying to teach physics and do not want to waste time with the details of the mathematical trickery that students should have picked up at some point during their undergraduate years. Nonetheless, students find themselves at a loss, and the professors often don't even know why the students are struggling until it is too late.

Hence this paper. I have attempted to summarize most of the more advanced mathematical trickery seen in electrodynamics and quantum mechanics in simple and friendly terms with examples. I do not know how well I have succeeded, but hopefully students will be able to get *some* relief from this summary.

Please realize that while I have done my best to make this as complete as possible, it is not meant as a textbook. I have left out many details, and in some cases I have purposely glossed over vital subtlties in the definitions and theorems. I do not provide many proofs, and the proofs that are here are sketchy at best. This is because I intended this paper only as a reference and a source. I hope that, in addition to giving students a place to look for a quick reference, it might spark some interest in a few to take a math course or read a book to get more details. I have provided a list of references that I found useful when writing this review at the end.

Finally, I would like to thank Brock Tweedie for a careful reading of the paper, giving very helpful advice on how to improve the language as well as catching a few mistakes.

Good luck, and have fun!

September 19, 2006

In this version (1.5), I have made several corrections and improvements. I would like to write a chapter on analysis and conformal mapping, but unfortunately I haven't got the time right now, so it will have to wait. I want to thank Linda Carpenter and Nadir Jeevanjee for their suggestions for improvement from the original version. I hope to encorporate more of Nadir's suggestions about tensors, but it will have to wait for Version 2...

September 10, 2007

In version 2.0, I have made several vital corrections, tried to incorporate some of the suggestions on tensors alluded to above, and have included and improved the sections on analysis that I always wanted to include. Thanks to Michael Luke for the plot in Chapter 1, that comes from his QFT lecture notes.

List of Symbols and Abbreviations

Here is a list of symbols used in this review, and often by professors on the blackboard.

Symbol	Meaning
$\forall$	For all
$\exists$	There exists
$\exists!$	There exists unique
$a \in A$	a is a member (or *element*) of the set A
$A \subset B$	A is a subset of B
$A = B$	$A \subset B$ and $B \subset A$
$A \cup B$	The set of members of the sets A *or* B
$A \cap B$	The set of members of the sets A *and* B
$\emptyset$	The empty set; the set with no elements
$A \amalg B$	The "disjoint union"; same as $A \cup B$ where $A \cap B = \emptyset$
WLOG	Without Loss of Generality
$p \Rightarrow q$	implies; If p is true, then q is true.
$p \Leftrightarrow q$; iff	p is true if and only if q is true.
$\mathbb{N}$	Set of natural numbers (positive integers)
$\mathbb{Z}$	Set of integers
$\mathbb{Q}$	Set of rational numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers

Contents

Chapter 1

Tensors and Such

1.1 Some Definitions

Before diving into the details of tensors, let us review how coordinates work. In N-dimensions, we can write N unit vectors, defining a **basis**. If they have unit magnitude and point in orthogonal directions, it is called an **orthonormal basis**, but this does not have to be the case. It is common mathematical notation to call these vectors $\vec{e}_i$, where i is an index and goes from 1 to N. It is important to remember that these objects are vectors even though they also have an index. In three dimensions, for example, they may look like:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{1.1.1}$$

Now you can write a general N-vector as the sum of the unit vectors:

$$\vec{r} = x^1\vec{e}_1 + x^2\vec{e}_2 + ... + x^N\vec{e}_N = \sum_{i=1}^{N} x^i\vec{e}_i \tag{1.1.2}$$

where we have written x^i as the coordinates of the vector $\vec{r}$. Note that the index for the coordinate is a superscript, whereas the index for the unit vector is a subscript. This small fact will turn out to be very important! Be careful not to confuse the indices for exponents.

Before going any further, let us introduce one of the most important notions of mathematics: the inner product. You are probably very familiar with the concept of an inner product from regular linear algebra. Geometrically, we can define the inner product in the following way:

$$\langle\vec{r_1}, \vec{r_2}\rangle = \langle\sum_{i=1}^{N} x_1^i\vec{e}_i, \sum_{j=1}^{N} x_2^j\vec{e}_j\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N} x_1^i x_2^j\langle\vec{e}_i, \vec{e}_j\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N} x_1^i x_2^j g_{ij} \tag{1.1.3}$$

where we have defined the new quantity:

$$g_{ij} \equiv \langle\vec{e}_i, \vec{e}_j\rangle \tag{1.1.4}$$

Mathematicians sometimes call this quantity the **first fundamental form**. Physicists often simply call it the **metric**.

Now look at the last equality of Equation 1.1.3. Notice that there is no *explicit* mention of the basis vectors $\{\vec{e}_i\}$. This implies that it is usually enough to express a vector in terms of its components. This is exactly analogous to what you are used to doing when you say, for example, that a vector is (1,2) as opposed to $\mathbf{x}+2\mathbf{y}$. Therefore, from now on, unless it is important to keep track of the basis, we will omit the basis vectors and denote the vector only by its components. So, for example, $\vec{r}$ becomes x^i. Note that "x^i" is not rigorously a vector, but only the vector components. Also note that the choice of basis *is* important when we want to construct the metric g_{ij}. This will be an important point later.

1.2 Tensors

The proper question to ask when trying to do tensor analysis involves the concept of a **transformation**. One must ask the question: *How do the coordinates change ("transform") under a given type of transformation?* Before going any further, we must understand the general answer to this problem.

Let's consider a coordinate system (x^i), and perform a transformation on these coordinates. Then we can write the new coordinates $(x^{i'})$ as a function of the old ones:

$$x^{i'} = x^{i'}(x^1, x^2, ..., x^N) \tag{1.2.5}$$

Now let us consider only infinitesimal transformations, i.e.: *very small* translations and rotations. Then we can Taylor expand Equation 1.2.5 and write the change in the coordinates as:

$$\delta x^{i'} = \sum_{j=1}^{N} \frac{\partial x^{i'}}{\partial x^j} \delta x^j \tag{1.2.6}$$

where we have dropped higher order terms. Notice that the derivative is actually a matrix (it has two indices):

$$A^{i'}_{\cdot j} \equiv \frac{\partial x^{i'}}{\partial x^j} \tag{1.2.7}$$

Now Equation 1.2.6 can be written as a matrix multiplication:

$$\delta x^{i'} = \sum_{j=1}^{N} A^{i'}_{\cdot j} \delta x^j \tag{1.2.8}$$

Before going further, let us consider the new matrix $A^{i'}_{\cdot j}$. First of all, take note that one index is primed, while one is not. This is a standard but very important notation, where the primed index refers to the new coordinate index, while the unprimed index refers to the old coordinate index. Also notice that there are two indices, but one is a subscript, while the other is a superscript. Again, this is not a coincidence, and will be important in what follows. I have also included a period to the left of the lower index: this is to remind you

that the indices should be read as "ij" and *not* as "ji". This can prove to be very important, as in general, the matrices we consider are not symmetric, and it is important to know the order of indices. Finally, even though we have only calculated this to lowest order, it turns out that Equation 1.2.8 has a generalization to large transformations. We will discuss these details in Chapter 2.

Let us consider a simple example. Consider a 2-vector, and let the transformation be a regular rotation by an angle θ. We know how to do such transformations. The new coordinates are:

$$\begin{aligned} x^{1'} &= x^1 \cos\theta + x^2 \sin\theta \sim x^1 + \theta x^2 + \mathcal{O}(\theta^2) \\ x^{2'} &= x^2 \cos\theta - x^1 \sin\theta \sim x^2 - \theta x^1 + \mathcal{O}(\theta^2) \end{aligned}$$

Now it is a simple matter of reading off the $A^{i'}_{.j}$:

$$\begin{aligned} A^{1'}_{.1} = 0 \quad A^{1'}_{.2} = \theta \\ A^{2'}_{.1} = -\theta \quad A^{2'}_{.2} = 0 \end{aligned}$$

Armed with this knowedge, we can now define a tensor:

A **tensor** *is a collection of objects which, combined the right way, transform the same way as the coordinates under infinitesimal (proper) symmetry transformations.*

What are "infinitesimal proper symmetry transformations"? This is an issue that we will tackle in Chapter 2. For now it will be enough to say that in N-dimensions, they are translations and rotations in each direction; without the proper, they also include reflections ($x \to -x$). This definition basically states that a tensor is an object that for each index, transforms according to Equation 1.2.8, with the coordinate replaced by the tensor.

Tensors carry indices. The number of indices is known as the **tensor rank**. Notice, however, that *not everything that caries an index is a tensor!!*

Some people with more mathematical background might be upset with this definition. However, you should be warned that the mathematician's definition of a tensor and the physicist's definition of a tensor are not exactly the same. I will say more on this in Chapter 5. For the more interested readers, I encourage you to think about this; but for the majority of people using this review, it's enough to know and use the above definition.

Now that we have a suitable definition, let us consider the basic examples.

1.2.1 Rank 0: Scalars

The simplest example of a tensor is a rank-0 tensor, or **scalar**. Since each index is supposed to transform like Equation 1.2.8, and there is no index, it must be that a scalar does not transform at all under a coordinate transformation. Such quantities are said to be **invariants**.

Notice that an element of a vector, such as x^1, is *not* a scalar, as it does transform nontrivially under Equation 1.2.8.

We will see that the most common example of a scalar is the length of a vector:

$$s^2 = \sum_{i=1}^{N} x_i^2 \equiv \sum_{i=1}^{N} x_i x^i \tag{1.2.9}$$

Exercise: Prove s is an invariant quantity under translations and rotations. See Section 1.3 for details.

1.2.2 Rank 1: Vectors

The next type of tensor is rank-1, or **vector**. This is a tensor with only one index. The most obvious example of a vector is the coordinates themselves, which by definition satisfy Equation 1.2.8. There are other examples as well. Some common examples of vectors that appear in physics are the momentum vector $\vec{p}$, the electric field $\vec{E}$, and force $\vec{F}$. There are many others.

1.2.3 The General Case

It is a relatively straightforward project to extend the notion of a rank-1 tensor to any rank: just keep going! For example, a rank-2 tensor is an object with two indices (T^{ij}) which transforms the following way:

$$T^{i'j'} = \sum_{k,l=1}^{N} A^{i'}_{.k} A^{j'}_{.l} T^{kl} \tag{1.2.10}$$

An example of a rank-2 tensor is the moment of intertia from classical rigid-body mechanics. A tensor of rank-3 would have three A's in the transformation law, etc. Thus we can define tensors of any rank by insisting that they obey the proper transformation law *in each index*.

1.3 Upstairs, Downstairs: Contravariant vs Covariant

We have put a great deal of emphasis on whether indices belong as superscripts or as subscripts. In Euclidean space, where we are all used to working, the difference is less important (although see below), but in general, the difference is huge. Now we will make this quantitative. For simplicity, I'll only consider rank-1 tensors (1 index), but the generalization is very straightforward.

Recall that we defined a vector as an object which transformed according to Equation 1.2.8. There the index was a superscript. We will call vectors of that form **contravariant vectors**. Now we want to define a vector with an index as a subscript, and we will do it using Equation 1.2.9.

In any reasonable world, the length of a vector should have nothing to do with the coordinate system you are using (the magnitude of a vector does not depend on the location

of the origin, for example). So we will define a vector with a lower index as an object which leaves the length invariant:

$$\sum_{i=1}^{N} x_{i'} x^{i'} = s^2 = \sum_{i'=1}^{N} (\sum_{k=1}^{N} A^{i'}_{.k} x^k)(\sum_{l=1}^{N} x_l B^l_{.i'}) = \sum_{k,l=1}^{N} \left[\sum_{i'=1}^{N} B^l_{.i'} A^{i'}_{.k} \right] x^k x_l \tag{1.3.11}$$

where $B^l_{.i'}$ is the matrix that defines how this new type of vector is supposed to transform. Notice that if the left and right side of this expression are supposed to be equal, we have a constraint on the quantity in brackets:

$$\sum_{i'=1}^{N} B^l_{.i'} A^{i'}_{.k} = \delta^l_k \tag{1.3.12}$$

where δ^l_k is the **Kroneker Delta**: $+1$ if $k = l$, 0 otherwise. In matrix notation, this reads:

$$BA = 1 \Rightarrow B = A^{-1} \tag{1.3.13}$$

So we know what B must be. Reading off the inverse is quite easy, by comparing with the definition of A in Equation 1.2.7:

$$B^j_{.i'} = \frac{\partial x^j}{\partial x^{i'}} \equiv A^j_{.i'} \tag{1.3.14}$$

The last step just came from noting that B is identical to A with the indices switched. So we have defined a new quantity, which we will call a **covariant vector**. It is just like a contravariant vector except that it transforms in exactly the oposite way:

$$x_{i'} = \sum_{j=1}^{N} x_j A^j_{.i'} \tag{1.3.15}$$

There is one more very important point to notice about the contravariant and covariant vectors. Recall from Equation 1.1.3 that the length of a vector can be written as:

$$s^2 = \sum_{i,j=1}^{N} g_{ij} x^i x^j$$

But we also wrote another equation for s^2 in Equation 1.3.11 when we started defining the covariant vector. By comparing these two equations, we can conclude a very useful relationship between covariant and contravariant vectors:

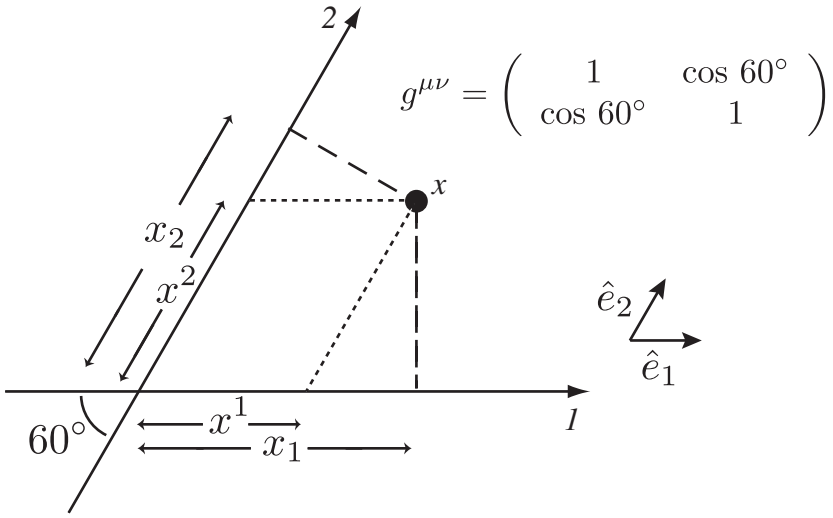

Figure 1.1: A graph showing the difference between covariant and contravariant coordinates.

$$x_i = \sum_{j=1}^{N} g_{ij} x^j \tag{1.3.16}$$

$$x^j = \sum_{k=1}^{N} g^{jk} x_k \tag{1.3.17}$$

$$\sum_{j=1}^{N} g_{ij} g^{jk} = \delta_i^k \tag{1.3.18}$$

So we have found that the metric takes contravariant vectors to covariant vectors and vice versa, and in doing so we have found that the inverse of the metric with lower indices is the metric with upper indices. This is a *very* important identity, and will be used many times.

When working in an orthonormal flat-space coordinate system (so $g_{ij} = \delta_{ij}$), the difference between covariant and contravariant is negligible. We will see in Chapters 3 and 4 some examples of where the difference begins to appear. But even now I can show you a simple nontrivial example. Consider a cartesian-like coordinate system, but now let the coordinate axes cross at an angle of 60 deg (see Figure 1.1). Now the metric is no longer proportional to the Kroneker delta (compare this to Equation (1.1.4), and so there will be a difference between covariant and contravariant coordinates. You can see that explicitly in the figure above, where the same point has $x^1 < x^2$, but $x_1 > x_2$. So we find that covariant and contravariant are the same *only in orthonormal cartesian coordinates!* In the more general case, even in flat space, the difference is important.

As another example, consider polar coordinates (r, θ). Even though polar coordinates describe the flat plane, they are not orthonormal coordinates. Specifically, if $x^i = (r, \theta)$, then it is true that $x_i = (r, r^2\theta)$, so here is another example of how covariant and contravariant makes a difference.

1.4 Mixed Tensors

We have talked about covariant and contravariant tensors. However, by now it should be clear that we need not stop there. We can construct a tensor that is covariant in one index and contravariant in another. Such a tensor would be expected to transform as follows:

$$T^{i'}_{.j'} = \sum_{k,l=1}^{N} A^{i'}_{.k} A^{l}_{.j'} T^{k}_{.l} \tag{1.4.19}$$

Such objects are known as **mixed tensors**. By now you should be able to construct a tensor of any character with as many indices as you wish. Generally, a tensor with n upper indices and l lower indices is called a (n, l) tensor.

1.5 Einstein Notation

The last thing to notice from all the equations so far is that every time we have an index that is repeated as a superscript and a subscript, we are summing over that index. This is not a coincidence. In fact, it is generally true that anytime two indices repeat in a monomial, they are summed over. So let us introduce a new convention:

Anytime an index appears twice (once upstairs, once downstairs) it is to be summed over, unless specifically stated otherwise.

This is known as **Einstein notation**, and will prove very useful in keeping the expressions simple. It will also provide a quick check: if you get an index repeating, but both subscripts or superscripts, chances are you made a mistake somewhere.

1.5.1 Another Shortcut

There is another very useful shortcut when writing down derivatives:

$$f(x)_{,i} = \partial_i f(x) \equiv \frac{\partial f}{\partial x^i} \tag{1.5.20}$$

Notice that the derivative with respect to contravariant-x is covariant; the proof of this follows from the chain rule and Equation (1.2.6) This is very important, and is made much more explicit in this notation. I will also often use the shortand:

$$\partial^2 = \partial_i \partial^i = \nabla^2 \tag{1.5.21}$$

This is just the Laplacian in arbitrary dimensions[1]. It is not hard to show that it is a scalar operator. In other words, for any scalar $\phi(\mathbf{x})$, $\partial^2\phi(\mathbf{x})$ is also a scalar.

[1] In Minkowski space, where most of this is relevant for physics, there is an extra minus sign, so $\partial^2 = \Box = \frac{\partial^2}{\partial t^2} - \nabla^2$,which is just the D'Alembertian (wave equation) operator. I will cover this more in Chapter 3; in this chapter, we will stick to the usual Euclidean space.

1.6 Some Special Tensors

1.6.1 Kroneker Delta

We have already seen the Kroneker Delta in action, but it might be useful to quickly sum up its properties here. This object is a tensor which always carries two indices. It is equal to +1 whenever the two indices are equal, and it is 0 whenever they are not. Notice that in matrix form, the Kroneker Delta is nothing more than the unit matrix in N dimensions:

$$[\delta^i_j] = \begin{pmatrix} 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Notice that the Kroneker delta is naturally a $(1,1)$ tensor.

1.6.2 Levi-Civita Symbol

The Levi-Civita symbol (ϵ) is much more confusing than the Kroneker Delta, and is the source of much confusion among students who have never seen it before. Before we dive into its properties, let me just mention that no matter how complicated the Levi-Civita Symbol is, life would be close to unbearable if it wasn't there! In fact, it wasn't until Levi-Civita published his work on tensor analysis that Albert Einstein was able to complete his work on General Relativity- it's that useful!

Recall what it means for a tensor to have a symmetry (for now, let's only consider a rank-2 tensor, but the arguments generalize). A tensor T_{ij} is said to be **symmetric** if $T_{ij} = T_{ji}$ and **antisymmetric** if $T_{ij} = -T_{ji}$. Notice that an antisymmetric tensor cannot have diagonal elements, because we have the equation $T_{ii} = -T_{ii} \Rightarrow T_{ii} = 0$ (no sum). In general, tensors will not have either of these properties; but in many cases, you might want to construct a tensor that has a symmetry of one form or another. This is where the power of the Levi-Civita symbol comes into play. Let us start with a definition:

The *Levi-Civita symbol in N dimensions* is a tensor with N indices such that it equals +1 if the indices are an **even permutation** and −1 if the indices are an **odd permutation**, and zero otherwise.

We will talk about permutations in more detail in a little while. For now we will simply say $\epsilon_{12\cdots N} = +1$ if all the numbers are in ascending order. The Levi-Civita symbol is a *completely antisymmetric tensor*, i.e.: whenever you switch two adjacent indices you gain a minus sign. This means that whenever you have two indices repeating, it must be zero (see the above talk on antisymmetric tensors).

That was a lot of words; let's do some concrete examples. We will consider the three most important cases (N=2,3,4). Notice that for N=1, the Levi-Civita symbol is rather silly!

N=2: ϵ_{ij}

For the case of two dimensions, the Levi-Civita symbol has two indices, and we can write out very easily what it looks like:

$$\epsilon_{12} = +1$$
$$\epsilon_{21} = -1$$

N=3: ϵ_{ijk}

In three dimensions, it becomes a little harder to write out explicitly what the Levi-Civita symbol looks like, but it is no more conceptually difficult than before. The key equation to remember is:

$$\epsilon_{123} = +1 \tag{1.6.22}$$

What if I write down ϵ_{213}? Notice that I recover ϵ_{123} by switching the first two indices, but owing to the antisymmetric nature, I must include a minus sign, so: $\epsilon_{213} = -1$. There are six possible combinations of indices. Can you decide which ones are $+1$ (even) and which are -1 (odd)? Here's the answer:

$$\begin{aligned} \epsilon_{123} = \epsilon_{231} = \epsilon_{312} &= +1 \\ \epsilon_{321} = \epsilon_{213} = \epsilon_{132} &= -1 \end{aligned}$$

N=4: ϵ_{ijkl}

I am not going to do N=4 explicitly, but hopefully by now you feel comfortable enough to figure it out on your own. Just remember, you always start with an ascending number sequence, so:

$$\epsilon_{1234} = +1$$

and just take it from there! For example: $\epsilon_{2143} = -\epsilon_{2134} = +\epsilon_{1234} = +1$.

The Levi-Civita symbol comes back again and again whenever you have tensors, so it is a very important thing to understand. Remember: if ever you get confused, just write down the symbol in ascending order, set it equal to $+1$, and start flipping (adjacent) indices until you find what you need. That's all there is to it!

Finally, I'd like to point out a small inconsistency here. Above I mentioned that you want to sum over indices when one is covariant and one is contravariant. But often we relax that rule with the Levi-Civita symbol and always write it with lower indices no matter what. There are some exceptions to this rule when dealing with curved space (General Relativity), but I will not discuss that here. From now on, I will always put the Levi-Civita indices downstairs - if an index is repeated, it is still summed over.

1.6.3 Using the Levi-Civita Symbol

Useful Relations

When performing manipulations using the Levi-Civita symbol, you often come across products and contractions of indices. There are two useful formulas that are good to remember:

$$\epsilon_{ijk}\epsilon_{ijl} = 2\delta_{kl} \tag{1.6.23}$$

$$\epsilon_{ijk}\epsilon_{ilm} = \delta^{jk}_{lm} \equiv \delta_{lj}\delta_{mk} - \delta_{lk}\delta_{mj} \tag{1.6.24}$$

Exercise: Prove these formulas.

Cross Products

Later, in Chapter 5, we will discuss how to take a generalized cross-product in n-dimensions. For now, recall that a cross product only makes sense in three dimensions. There is a beautiful formula for the cross product using the Levi-Civita symbol:

$$(\vec{A} \times \vec{B})_i = \epsilon_{ijk} A^j B^k \tag{1.6.25}$$

Using this equation, and Equations (1.6.23) and (1.6.24), you can derive all the vector identities on the cover of Jackson, such as the double cross product. Try a few to get the hang of it. That's a sure-fire way to be sure you understand index manipulations.

As an example, let's prove a differential identity:

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{A}) = \epsilon_{ijk}\partial_j \epsilon_{klm}\partial_l A_m &= \epsilon_{kij}\epsilon_{klm}\partial_j\partial_l A_m \\ = [\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}]\partial_j\partial_l A_m &= \partial_i\partial_m A_m - \partial^2 A_i = \nabla(\nabla\cdot\mathbf{A}) - \nabla^2\mathbf{A} \end{aligned}$$

Exercise: Prove that you can interchange dot and cross products: $\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} = \mathbf{A}\times\mathbf{B}\cdot\mathbf{C}$.

1.7 Permutations

A **permutation** is just a shuffling of objects. Rigorously, consider a set of objects (called S) and consider an automorphism $\sigma : S \to S$ (one-to-one and onto function). Then σ is a permutation. It takes an element of S and sends it uniquely to another element (possibly the same element) of S.

One can see very easily that the set of all permutations of a (finite) set S is a group with multiplication defined as function composition. We will talk more about groups in the next chapter. If the set S has n elements, this group of permutations is denoted S_n, and is often called the **Symmetric group**. It has $n!$ elements. The proof of this is straightforward application of group and number theory, and I will not discuss it here.

The theory of symmetric groups is truly deep and of fundamental importance in mathematics, but it also has much application in other branches of science as well. For that reason, let me give a very brief overview of the theory. For more information, I recomend your favorite abstract algebra textbook.

Specific permutations are usually written as a row of numbers. It is certainly true (and is rigorously provable for the anal retentive!) that any finite set of order n (i.e.: with n elements) can be represented by the set of natural numbers between 1 and n. Then WLOG[2] we can always represent our set S in this way. Then a permutation might look like this:

$$\sigma = (1234)$$

This permutation sends $1 \mapsto 2$,$2 \mapsto 3$,$3 \mapsto 4$,$4 \mapsto 1$. This is in general how it always works. If our set has four elements in it, then this permutation touches all of them. If it has more than four elements, it leaves any others alone (for example, $5 \mapsto 5$).

Working in S_5 for the moment, we can write down the product of two permutations by composition. Recall that composition of functions *always* goes rightmost function first. This is very important here, since the multiplication of permutations is almost never commutative! Let's see an example:

$$(1234)(1532) = (154)$$

Where did this come from? Start with the first number in the rightmost permutation (1). Then ask, what does 1 go to? Answer: 5. Now plug 5 into the secondmost right transposition, where 5 goes to 5. Net result: $1 \mapsto 5$. Now repeat the same proceedure on 5 to get $5 \mapsto 4$, and so on until you have completed the cycle (quick check: does $2 \mapsto 2$,$3 \mapsto 3$?). Multiplying three permutations is just as simple, but requires three steps instead of two.

A permutation of one number is trivial, so we don't worry about it. A **transposition** is a permutation of two numbers. For example: (12) sends $1 \mapsto 2$,$2 \mapsto 1$. It is a truly amazing theorem in group theory that any permutation (no matter what it is) can be written as a product of transpositions. Furthermore, this decomposition is unique up to trivial permutations. Can you write down the decomposition of the above product[3]?

Armed with that knowledge, we make a definition:

Definition: If a permutation can be written as a product of an even number of transpositions, it is called an **even permutation**. If it can be written as an odd number of perumtations, it is called an **odd permutation**. We define the **signature** of a permutation in the following way:

$$\mathrm{sgn}(\sigma) = \begin{cases} +1 & \sigma \quad \text{even} \\ -1 & \sigma \quad \text{odd} \end{cases} \tag{1.7.26}$$

It is pretty obvious that we have the following relationships:

$$\begin{aligned} (even)(even) &= (even) \\ (odd)(odd) &= (even) \\ (odd)(even) = (even)(odd) &= (odd) \end{aligned}$$

[2]Without Loss Of Generality

[3]Answer: (15)(45). Notice that this is *not* the same as (45)(15).

What this tells us is that the set of even permutations is closed, and hense forms a subgroup, called the **Alternating Subgroup** (denoted A_n) of order $\frac{1}{2}n!$. Again, I omit details of the proof that this is indeed a subgroup, although it's pretty easy. Notice that the odd elements do not have this property. The alternating subgroup is a very powerful tool in mathematics.

For an application of permutations, a good place to look is index manipulation (hence why I put it in this chapter!). Sometimes, you might want to sum over all combinations of indices. For example, the general formula for the determinant of an $n \times n$ matrix can be written as:

$$\det A = \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \tag{1.7.27}$$

1.8 Constructing and Manipulating Tensors

To finish this chapter, I will review a few more manipulations that provide ways to construct tensors from other tensors.

1.8.1 Constructing Tensors

Consider two tensors of rank k (K) and l (L) respectively. We want to construct a new tensor (T) from these two tensors. There are three immediate ways to do it:

1. **Inner Product**: We can contract one of the indices from K and one from L to form a tensor T of rank $k + l - 2$. For example, if K is rank 2 and L is rank 3, then we can contract, say, the first index of both of them to form T with rank 3:

   $$T_{\mu\nu\rho} = K^{\alpha}_{.\mu} L_{\alpha\nu\rho}$$

   We could construct other tensors as well if we wished by contracting other indices. We could also use the metric tensor or the Levi-Civita tensor if we wished to form symmetric or antisymmetric combinations of tensors. For example, if A, B are vectors in 4 dimensions:

   $$T_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} A^{\rho} B^{\sigma} \Rightarrow T_{12} = A^3 B^4 - A^4 B^3, \quad \text{etc.}$$

2. **Outer Product**: We can always construct a tensor by combining two tensors without contracting any indices. For example, if A and B are vectors, we can construct: $T_{ij} = A_i B_j$. Because we are using a Cartesian coordinate system, we call tensors of this form **Cartesian Tensors**. When combining two vectors in this way, we say that we have constructed a **Dyadic**.

3. **Symmetrization**: Consider the outer product of two vectors $A_i B_j$. Then we can force the tensor to be symmetric (or antisymmetric) in its indices by construction:

$$A_{(i}B_{j)} \equiv \frac{1}{2}(A_iB_j + A_jB_i) \tag{1.8.28}$$

$$A_{[i}B_{j]} \equiv \frac{1}{2}(A_iB_j - A_jB_i) \tag{1.8.29}$$

Circular brackets about indices mean symmetric combinations, and square brackets about indices mean antisymmetric combinations. You can do this to as many indices as you want in an outer product (for n indices, replace $\frac{1}{2} \to \frac{1}{n!}$).

Dyadics provide a good example of how combining tensors can be more involved than it first seems. Consider the dyadic we have just constructed. Then we can write it down in a fancy way by adding and subtracting terms:

$$T_{ij} = \frac{1}{N}(T^k_k)\delta_{ij} + \frac{1}{2}(T_{ij} - T_{ji}) + [\frac{1}{2}(T_{ij} + T_{ji}) - \frac{1}{N}(T^k_k)\delta_{ij}] \tag{1.8.30}$$

N is the dimension of A and B, and the sum over k in the first and last terms is implied by Einstein notation. Notice that I have not changed the tensor at all, because any term I added I also subtracted. However, I have decomposed the tensor into three different pieces. The first piece is just a scalar (times the identity matrix), and therefore is invariant under rotations. The next term is an antisymmtric tensor, and (in 3D) can be written as a (pseudo) vector- it's a cross product (see Equation (1.6.25)). Finally the third piece is a symmetric, traceless tensor.

Let's discuss the matter concretely by letting $N = 3$ and counting degrees of freedom. If A and B are 3-vectors, then T_{ij} has $3 \times 3 = 9$ components. Of the three pieces, the first is a scalar (1), the second is a 3-vector (3) and the third is a rank-2 symmetric tensor (6) that is traceless (-1); so we have decomposed the tensor according to the following prescription:

$$3 \otimes 3 = 1 \oplus 3 \oplus 5$$

We say that the dyatic form T_{ij} is a **reducible Cartesian tensor**, and it reduces into 3 **irreducible spherical tensors**, each of rank 0, 1 and 2 respectively.

1.8.2 Kroneker Products and Sums

With that introduction, we can talk about Kroneker products and sums. A **Kroneker product** is simply an outer product of two tensors of lesser rank. A dyatic is a perfect example of a Kroneker Product. Writing it out in a matrix, we have (for A, B rank-1 tensors in 3 dimensions):

$$\mathbf{T} = \mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} A_1B_1 & A_1B_2 & A_1B_3 \\ A_2B_1 & A_2B_2 & A_2B_3 \\ A_3B_1 & A_3B_2 & A_3B_3 \end{pmatrix} \tag{1.8.31}$$

A **Kroneker sum** is the concatinated sum of matrices. You can write the Kroneker sum in matrix form using a block-diagonal matrix:

$$\mathbf{T} = \mathbf{C} \oplus \mathbf{D} \oplus \mathbf{E} \;\; = \;\; \begin{pmatrix} C & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & E \end{pmatrix} \tag{1.8.32}$$

where C is an $m \times m$ matrix, D is an $n \times n$ matrix, E is an $l \times l$ matrix,and "0" is just filling in any left over entries with a zero. Notice that the final matrix is an $(m+n+l) \times (m+n+l)$ matrix.

At this point, we can rewrite our equation for the dyadic:

$$\mathbf{A} \otimes \mathbf{B} \sim \mathbf{C} \oplus \mathbf{D} \oplus \mathbf{E} \tag{1.8.33}$$

where $\mathbf{A} \otimes \mathbf{B}$ is a 3×3 matrix, $\mathbf{C}$ is 1×1, $\mathbf{D}$ is 3×3 and $\mathbf{E}$ is 5×5. Notice that the matrix dimensions are not the same on both sides of the equation - these are *not* the same matrix! However, they each have the same number of independent quantities, and both *represent* the same tensor $\mathbf{T}$. For this reason, the study of finding different dimensional matrices for $\mathbf{T}$ goes under the heading of **representation theory**. It is a very active part of modern algebraic research, and is very useful in many areas of science.

Why would one ever want to do this? The answer is that when working with dyadics in practice, very often one is only interested in the part of the tensor that transforms the same way. Notice that when you rotate $\mathbf{T}$, you never mix terms between $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{E}$. This kind of relationship is by no means as obvious when written in its Cartesian form. For that reason, it is often very useful to work with the separate spherical tensors rather than with the entire Cartesian tensor. The best example is in Quantum Mechanics, when you are talking about angular momentum. See Sakurai for more details.

Chapter 2

Transformations and Symmetries

2.1 Symmetries and Groups

Here we will discuss some properties of groups in general that you need to know. Then we will progress to discuss a very special and important kind of group, called a Lie Group. Finally we will look at some examples.

- *Definition:* A **group** G is a set with an operator called multiplication that obeys four axioms: Closure ($g_1, g_2 \in G \Rightarrow g_1 g_2 \in G$), Associative($(g_1 g_2)g_3 = g_1(g_2 g_3)$), an identity (generically called e such that $ge = eg = g$) and inverses (g has inverse g^{-1} such that $gg^{-1} = g^{-1}g = e$). If the group is commutative ($g_1 g_2 = g_2 g_1$), it is said to be an **Abelian group**, otherwise it is Non-Abelian.

- Discrete groups represent operations that either do or do not take place. Continuous groups represent operations that differ infinitesimally from each other.

- Elements in continuous groups can be labeled by a set of continuous parameters. If the set of parameters is finite, the group is said to be a **finite group**. If the ranges of the parameters are closed and bounded, the group is said to be a **compact group**.

- *Definition:* A **subgroup** is a subset of a larger group that is also a group. A subgroup $N \subset G$ is a **normal subgroup** (also called **invariant subgroup**) if $\forall s_i \in N, g^{-1} s_i g \in N \ \forall g \in G$. In other words, elements of N stay in N after being multiplied by elements in G. We write $N \triangleleft G$. There are several definitions for invariant subgroups, but they are all equivalent to this one.

- Consider two groups that commute with each other (G, H). Then you can construct the **direct product group** in the following way: $G \times H \equiv \{(g_j h_k)|(g_j h_k)(g_l h_m) = (g_j g_l h_k h_m)\}$. Be careful not to confuse a direct product group with a Kroneker product of tensors.

- Fact: $G, H \triangleleft G \times H$ are invariant subgroups.

- *Definition:* If a group has no invariant subgroups, it is called a **simple** group.

- *Definition:* Suppose there is a square $n \times n$ matrix M that corresponds to each element of a group G, where $g_1 g_2 = g_3 \Rightarrow M(g_1)M(g_2) = M(g_3)$ and $M(g^{-1}) = [M(g)]^{-1}$. Then M is called a **representation** of the group G, and n is called the **dimension** of the representation. If this identification is one-to-one, the representation is called "faithful". An example of an unfaithful representation is where all elements of a group G go to 1.

- *Definition:* A representation is **reducible** if each matrix can be expressed in block-diagonal form. Using the Kroneker sum notation, we can rephrase this definition: $\exists A, B$ such that $M = A \oplus B$. A and B are also valid representations of the group. Furthermore, they do not talk to each other (think angular momentum in quantum mechanics). If a representation is not reducible, it is said to be **irreducible**.

- Consider an n-tuple ξ_a, not necessarely a vector in the sense of Chapter 1. Then the group element g inan $n-$dimensional representation M acts on the vector in the following way:
$$\xi_a^{'} = M(g)_a^b \xi_b$$
(recall Einstein notation). Since the matrix M must be nonsingular (otherwise there would be no g^{-1} and it wouldn't be a group), $\exists \overline{\xi}^a$ such that $\overline{\xi}^{a'} = \overline{\xi}^b [M(g)^{-1}]_b^a$. $\overline{\xi}$ is called the **adjoint** of ξ.

- *Definition:* The smallest (lowest dimension) nontrivial irreducible faithful representation is called the **fundamental representation**. We will see some examples of this shortly. Here, the "trivial" representation is the one-dimensional representation.

- *Definition:* A **Lie group** is a continuous group that is also a manifold whose multiplication and inverse operations are differentiable. It is named after Sophus Lie, the mathematician who first discovered their properties. This definition is quite a mouthful- don't worry too much about it if you're not a theorist! Things will become a lot simpler once we get to some examples. But before that, here are some important results you should know:

- *Theorem:* All compact Lie groups have unitary representations. In other words: $M(\epsilon) = e^{i\epsilon^j F_j}, j = 1, \ldots, N$, and F_j matrices[1]. The F_j are called **generators**. From this we can see that $M^\dagger = M^{-1}$, as long as $F^\dagger = F$, so we require that the generators be *Hermitian*. Warning: the index on F is not a matrix index - $F_j = (F_j)_{ab}$.

- Many (though not all) Lie groups have the additional property that $\det M(\epsilon) = +1$ $\forall \epsilon$. From the above result, this implies $Tr(F_i) = 0$; this can be proved with the help of the very useful fact
$$detA = e^{\mathrm{Tr}\log A}$$
This is called the **special condition**.

[1]Recall the definition for the exponential of a matrix: $e^A = \sum_{n=0}^{\infty} \frac{(A)^n}{n!}$

- When ϵ is small, we can expand the exponential to get:

$$M(\epsilon) = I + i\epsilon^j F_j + \mathcal{O}(\epsilon^2)$$

This is exactly the same as an infinitesimal transformation, and so we say that F_j *generates* infinitesimal transformations.

- *Theorem:* The local structure of any Lie group is completely specified by a set of commutation relations among the generators:

$$[F_i, F_j] = if_{ijk}F_k \tag{2.1.1}$$

(the sum over k is understood). These commutators specify an algebra spanned by the generators[2], and the f_{ijk} are called the **structure constants** of the Lie group. They are the same for any representation of the group.

- *Theorem:* $f_{ijl}f_{lkm} + f_{kil}f_{ljm} + f_{jkl}f_{lim} = 0$. The proof follows from Jacobi's identity for commutators.

- In general, the structure constants are antisymmetric in their indices.

- Notice that we can rearrange the above version of the Jacobi identity and use the antisymmetry to rewrite it as:

$$(f_i)_{jl}(f_k)_{lm} - (f_k)_{jl}(f_i)_{lm} = -f_{ikl}(f_l)_{jm} \tag{2.1.2}$$

where I have just separated the indices in a convienient way, but nothing more. Then if I define a matrix with the elements $(F_i)_{jk} \equiv -if_{ijk}$, I have reproduced Equation (2.1.1). Therefore, this is a valid representaion! It is so important, it is given a special name - it is called the **regular adjoint representation**.

- *Definition:* The **rank** of a Lie group is the number of simultaneously diagonalizeable generators; that is, the number of generators that commute among themselves. This set of generators forms a subalgebra called the **Cartan Subalgebra**. The rank of the group is the dimension of this subalgebra.

- *Definition:* $C^{(n)} \equiv f_{i_1 j_1 j_2} f_{i_2 j_2 j_3} \cdots f_{i_n j_n j_1} F_{i_1} F_{i_2} \cdots F_{i_n}$ are called the **Casimir operators** notice that all indices are repeated and therefore summed via Einstein convension).

- Casimir operators commute with every generator; hence they must be multiples of the identity. This means that they represent conserved quantities. It is also true that the number of Casimir operators is equal to the rank of the group.

That's enough general stuff. Let's look at some very important examples:

[2] An **algebra** is a vector space with a vector product, where in this case, the basis of this vector space is the set of generators (see Chapter 5). This algebra is called a **Lie Algebra**.

2.1.1 SU(2) - Spin

- Consider a complex spinor

$$\xi = \begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix}$$

and the set of all transformations which leave the norm of the spinor invariant; i.e.:

$$s^2 = |\xi^1|^2 + |\xi^2|^2$$

s invariant under all transformations. We will also require $\det M = 1$. From the above, we know that the generators (call them σ_i) must be linearly independent, traceless, Hermitian 2×2 matrices. Let us also include a normalization condition:

$$Tr(\sigma_i \sigma_j) = 2\delta_{ij}$$

Then the matrices we need are the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.1.3}$$

Then we will chose the group generators to be $F_i = \frac{1}{2}\sigma_i$. The group generated by these matrices is called the Special Unitary group of dimension 2, or **SU(2)**. In the fundamental representation, it is the group of 2×2 unitary matrices of unit determinant.

- Recall from quantum mechanics that the structure constants are simply given by the Levi-Civita tensor:

$$f_{ijk} = \epsilon_{ijk}$$

- Only σ_3 is diagonal, and therefore the rank of SU(2) is 1. Then there is only one Casimir operator, and it is given by:

$$C^{(2)} = F_1^2 + F_2^2 + F_3^2$$

- If we remove the Special condition (unit determinant), then the condition that $Tr\sigma_i = 0$ no longer need hold. Then there is one more matrix that satisfies all other conditions and is still linearly independent from the others- the identity matrix:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Notice, however that this is invariant:

$$e^{i\epsilon^j F_j} e^{i\epsilon^0 F_0} e^{-i\epsilon^j F_j} = e^{i\epsilon^0 F_0} \tag{2.1.4}$$

Therefore, the group generated by $e^{i\epsilon^0 F_0}$ is an invariant subgroup of the whole group. We call it U(1). That it is a normal subgroup should not suprise you, since it is true that $e^{i\epsilon^0 F_0} = (e^{i\epsilon^0/2})\sigma_0$, which is a number times the identity matrix.

- The whole group (without the special) is called U(2). We have found $U(1) \triangleleft U(2)$, $SU(2) \triangleleft U(2)$ and we can write:

$$U(2) = SU(2) \times U(1) \tag{2.1.5}$$

This turns out to be true even in higher dimensional cases:

$$U(N) = SU(N) \times U(1)$$

- Sumarizing: $R = e^{i\frac{\theta}{2}(\vec{n}\cdot\vec{\sigma})}$ is the general form of an element of SU(2) in the fundamental representation. In this representation, $C^{(2)} = 2$, as you might expect if you followed the analogy from spin in quantum mechanics.

- You know from quantum mechanics that there are other representations of spin (with different dimensions) that you can use, depending on what you want to describe - this change in representation is nothing more than calculating Clebsch-Gordon coefficients!

- Example: what if you want to combine two spin-$\frac{1}{2}$ particles? Then you know that you get (3) spin-1 states and (1) spin-0 state. We can rewrite this in terms of representation theory: we'll call the fundamental representation **2**, since it is two-dimensional. The spin-1 is three dimensional and the spin-0 is one dimensional, so we denote them by **3** and **1** respecitively. Then we can write the equation:

$$\mathbf{2} \otimes \mathbf{2} = \mathbf{3} \oplus \mathbf{1} \tag{2.1.6}$$

(Notice once again that the left hand side is a 2×2 matrix while the right hand side is a 4×4 matrix, but both sides only have four independent components). Equation (2.1.6) is just saying that combining two spin-$\frac{1}{2}$ particles gives a spin-1 and a spin-0 state, only using the language of groups. In this language, however, it becomes clear that the four-dimensional representation is reducible into a three-dimensional and a one-dimensional representation. So all matrix elements can be written as a Kroneker sum of these two representations. Believe it or not, you already learned this when you wrote down the angular momentum matrices (recall they are block-diagonal). This also shows that you do not need to consider the entire (infinitely huge) angular momentum matrix, but only the invariant subgroup that is relevant to the problem. In other words, the spin-1 and spin-0 do not talk to each other!

2.1.2 SO(3,1) - Lorentz (Poincaré) Group

Much of this information is in Jackson, Chapter 11, but good ol' JD doesn't always make things very clear, so let me try to summarize.

- The **Poincaré Group** is the set of all transformations that leave the metric of flat space-time invariant (see Chapter 4). The metric in Minkowski (flat) space is given by:

$$ds^2 = c^2dt^2 - d\mathbf{r}^2$$

- Intuitively we know that this metric is invariant under parity (space reversal), time reversal, spacial rotations, space and time translations and Lorentz transformations. Then we know that these operations should be represented in the Poincaré group.

- Ignore translations for now. You can prove to yourself that the matrices that represent the remaining operators must be orthogonal ($O^T = O^{-1}$); the group of $n \times n$ orthogonal matrices is usually denoted by O(n). In this case, $n = 4$; however, there is a minus sign in the metric that changes things somewhat. Since three components have one sign, and one component has another, we usually denote this group as **O(3,1)**, and we call it the **Lorentz group**.

- The transformations we are interested in are the *proper orthochronous* transformations- i.e.: the ones that keep causality and parity preserved. This is accomplished by imposing the "special" condition ($\det O = +1$). However, this is not enough: although this forbids parity and time-reversal, it does not forbid their product (prove this!). So we consider the **Proper Orthochronous Lorentz Group**, or $SO(3,1)_+$, where the "orthochronous" condition (subscript "+" sign) is satisfied by insisting the the (00) component of any transformation is always positive; that along with the special condition is enough to isolate the desired subgroup. As this is quite a mouthful, we usually just refer to SO(3,1) as the Lorentz group, even though it's really only a subgroup.

- We can now go back to Chapter 1 and rephrase the definition of a tensor in flat spacetime as an object that transforms under the same way as the coordinates under an element of the Lorentz group.

- The Lorentz group has six degrees of freedom (six parameters, six generators). For a proof, see Jackson. We can interpret these six parameters as three spacial rotations and three boosts, one about each direction. We will denote the rotation matrix generators as S_i, and the boost generators as K_i, where $i = 1, 2, 3$. These matrices are:

$$S_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad S_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \qquad S_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$K_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad K_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad K_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

- In general, we can write any element of the Lorentz group as $e^{-[\theta \cdot \mathbf{S} + \eta \cdot \mathbf{K}]}$, where θ are the three rotation angles, and η are the three boost parameters, called the **rapidity**. You should check to see that this exponential does in fact give you a rotation or Lorentz transformation when chosing the proper variables. See Jackson for the most general transformation.

- The boost velocity is related to the rapidity by the equation:

$$\beta_i \equiv \tanh \eta_i \tag{2.1.7}$$

- Exercise: what is the matrix corresponding to a boost in the $\hat{\mathbf{x}}$-direction? Answer:

$$\begin{pmatrix} \cosh\eta & \sinh\eta & 0 & 0 \\ \sinh\eta & \cosh\eta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

  where $\gamma = (1-\beta^2)^{-1/2}$. This is what you're used to seeing in Griffiths!

- Notice that you can always boost along a single direction by combining rotations with the boosts. This is certainly legal by the nature of groups!

- The angles are bounded: $0 \leq \theta \leq 2\pi$, but the rapidities are not: $0 \leq \eta < \infty$. The Lorentz group is *not* a compact group! And therefore it does not have faithful unitary representations. This is a big problem in Quantum Mechanics and comes into play when trying to quantize gravity.

- The Lorentz algebra is specified by the commutation relations:

$$\begin{aligned} [S_i, S_j] &= \epsilon_{ijk} S_k \\ [S_i, K_j] &= \epsilon_{ijk} K_k \\ [K_i, K_j] &= -\epsilon_{ijk} S_k \end{aligned}$$

  Some points to notice about these commutation relations:

  1. The rotation generators form a closed subalgebra (and hense generate a closed subgroup) – it is simply the set of all rotations, SO(3). It is *not* a normal subgroup, however, becuase of the second commutator.
  2. A rotation followed by a boost is a boost, as I mentioned above.
  3. Two boosts in diferent directions correspond to a rotation about the third direction.
  4. The minus sign in the last commutator is what makes this group the noncompact SO(3,1), as opposed to the compact SO(4), which has a + sign.

- Notice that you could have constructed the generators for the Lorentz group in a very similar way to how you construct generators of the regular rotation group (angular momentum):

$$J^{\mu\nu} = x \wedge p = i\hbar(x^\mu \partial^\nu - x^\nu \partial^\mu) \tag{2.1.8}$$

where $i, j = 1, 2, 3$, $\mu, \nu = 0, 1, 2, 3$ and you can think of the wedge product as a generalization of the cross product (see Chapter 5). In this notation, $J^{0i} \equiv K_i$ and $J^{ij} \equiv \epsilon_{ijk} S_k$. These six matrices are equivalent to the six above ones. Notice that, as before, the greek indices are *not* matrix indices.

2.2 Transformations

- We wish to understand how transformations affect the equations of motion. The most important kinds of transformations are translations and rotations. These are called **proper transformations**. Other transformations might include parity, time reversal and charge inversion. We also might be interested in internal symmetries such as gauge transformations in E&M, for example.

- When discussing the theory of transformations, it is useful to consider how objects transform under a very small change; such changes are called **infinitesimal transformations**. Let's consider these first. In what follows, I will be supressing indices (to make the notation simpler).

- We want to invoke a transformation on a coordinate: $x \to x + \delta x$. Then we would expect a change in any function of x: $f(x) \to f(x) + \delta f$. What is the form of δf? If f represents a tensor, then it should have the same form as δx by definition.

- Example: Consider an infinitesimal rotation about a point in the i direction. The transformation will look like:

$$x \to x + i\epsilon J_i x + \mathcal{O}(\epsilon^2) = [1 + i\epsilon J_i + \mathcal{O}(\epsilon^2)]x$$

  where J_i is the generator of the rotation (angular momentum).

- Recall that performing two rotations in the same direction is just multiplying the two rotations together. Then performing N infinitesimal rotations of size $\epsilon = a/N$, where a is a finite number, and ϵ is infinitesimal in the limit $N \to \infty$, we have:

$$\lim_{N\to\infty} \left(1 + \frac{iaJ_i}{N}\right)^N = e^{iaJ_i} \Rightarrow x \mapsto e^{iaJ_i} x \tag{2.2.9}$$

  Hense we have constructed a **finite transformation** from an infinite number of infinitesimal ones. Notice that if you expand the exponential and keep only linear terms (for a very small) you reproduce the infinitesimal rotation.

2.3 Lagrangian Field Theory

Here is a brief overview of the important things you should know about Lagrangian field theory.

- Recall in Lagrangian mechanics, you define a Lagrangian $L(q(t), \dot{q}(t))$ and an action $S = \int L dt$. Now we wish to decribe a Lagrangian density $\mathcal{L}(\phi(x), \partial_\mu \phi(x))$ where $L = \int \mathcal{L} d^3x$, and ϕ is a field which is a funtion of spacetime. The action can be written as $S = \int \mathcal{L} d^4x$. The action completely specifies the theory.

- The equations of motion can now be found, just as before by the condition $\delta S = 0$. This yields the new Euler-Lagrange equations:

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \phi_{,\mu}} \right) = 0 \tag{2.3.10}$$

  Notice that there is the sum over the index μ, following Einstein notation.

- Example: $\mathcal{L} = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2$ is the Lagrangian for a scalar field (Klein-Gordon field in relativistic quantum mechanics). Can you find the equations of motion? Here's the answer:

$$(\partial^2 + m^2)\phi = 0$$

- Just as in classical mechanics, we can also go to the Hamiltonian formalism, by defining a **momentum density**:

$$\pi \equiv \frac{\partial \mathcal{L}}{\partial \dot{\phi}}$$

  where $\dot{\phi} \equiv \partial_0 \phi$ and perform the Legendre transformation:

$$\mathcal{H}(\phi(x), \pi(x)) = \pi(x)\dot{\phi}(x) - \mathcal{L} \Rightarrow H = \int d^3x \mathcal{H} \tag{2.3.11}$$

- Notice that $\mathcal{L}$ is covariant, but $\mathcal{H}$ is not, since time becomes a special parameter via the momentum density. This is why we often prefer to use the Lagrangian formalism whenever working with relativistic mechanics. Nevertheless, this formalism is still generally useful since H is related to the energy in the field.

2.4 Noether's Theorem

A theory is said to be symmetric with respect to a transformation whenever the action doesn't change under that transformation. This implies that the Lagrangian can only change up to a total divergence (via Gauss's Theorem). To see this, notice that if the Lagrangian changes by a total divergence, we have:

$$S' = \int \mathcal{L} d^4x + \int \partial_\mu \mathcal{J}^\mu d^4x = S + \int d\Sigma_\mu \mathcal{J}^\mu$$

where the last surface integral is an integral over a surface at infinity. In order for quantities like energy to be finite, we often require that fields and their derivatives vanish on such a

surface. Thi implies that the above surface integral vanishes (or at least doesn't contribute to the Equations of Motion[3]). Thus we have

Noether's Theorem: For every symmetry in the theory, there corresponds a conserved current, and vice versa.

Proof: I will prove Noether's theorem for a scalar field, but it is true in general for any type of field. Let's consider the general transformation that acts on the field according to:

$$\phi \to \phi + \alpha\delta\phi$$

where α is a constant infinitesimal parameter[4]. Then if we insist this transformation leave the action invariant, the previous argument says that the Lagrangian can only change by a total divergence:

$$\mathcal{L} \to \mathcal{L} + \alpha\partial_\mu \mathcal{J}^\mu$$

Let's calculate the change in $\mathcal{L}$ explicitly:

$$\begin{aligned}
\alpha\delta\mathcal{L} &= \frac{\partial\mathcal{L}}{\partial\phi}(\alpha\delta\phi) + \frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\partial_\mu(\alpha\delta\phi) \\
&= \alpha\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\delta\phi\right) + \alpha\left[\frac{\partial\mathcal{L}}{\partial\phi} - \left(\frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\right)_{,\mu}\right]\delta\phi \\
&= \alpha\partial_\mu\mathcal{J}^\mu
\end{aligned}$$

To get line 1 from line 2, use the product rule. The term in brackets is zero by Equation (2.3.10), leaving only the total divergence. Then we can combine terms to get:

$$\boxed{j^\mu \equiv \left(\frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\delta\phi\right) - \mathcal{J}^\mu} \Rightarrow \partial_\mu j^\mu = 0 \tag{2.4.12}$$

hense j^μ is a conserverd current, which is precicely what we wanted to show. **QED.**

Notice that we could also write Noether's theorem by saying that for every symmetry of the theory, there is a "charge" which is conserved:

$$Q \equiv \int j^0 d^3x \Rightarrow \frac{dQ}{dt} = 0$$

There are many applications of this theorem. You can prove that isotropy of space (rotational invariance) implies angular momentum conservation, homogeneity of space (translational invariance) implies linear momentum conservation, time translation invariance implies

[3]Sometimes this integral does *not* vanish, and this leads to nontrivial "topological" effects. But this is advanced stuff, and well beyond the scope of this review. So for now let's just assume that the surface integral is irrelevant, as it often is in practice.

[4]This theorem also holds for nonconstant α - I leave the proof to you.

energy conservation, etc. You can also consider more exotic internal symmetries; for example, gauge invariance in electrodynamics implies electric charge conservation.

Chapter 3

Geometry I: Flat Space

The most important aspect of geometry is the notion of distance. What does it mean for two points in a space to be "close" to each other? This notion is what makes geometry distinct from topology and other areas of abstract mathematics. In this chapter, I want to define some important quantities that appear again and again in geometric descriptions of physics, and give some examples.

In this chapter we will always be working in flat space (no curvature). In the next chapter, we will introduce curvature and see how it affects our results.

3.1 THE Tensor: $g_{\mu\nu}$

- Like transformations, we can define distance infinitesimally and work our way up from there. Consider two points in space that are separated by infinitesimal distances dx, dy and dz. Then the total displacement (or metric) is given by the triangle rule of metric spaces (a.k.a. the Pythagorean theorem):

$$ds^2 = dx^2 + dy^2 + dz^2 \equiv g_{\mu\nu} dx^\mu dx^\nu$$

  Here, $g_{\mu\nu} = \delta_{\mu\nu}$.

- *Definition:* $g_{\mu\nu}$ is called the **metric tensor** or the **first fundamental form**. Knowing the form of $g_{\mu\nu}$ for any set of coordinates let's you know how to measure distances. The metric tensor is given by Equation (1.1.4), but in this chapter we will often cheat, since sometimes we already know what ds^2 is. But when faced with trying to calculate the metric in general, usually you have to resort to Equation (1.1.4).

- **WARNING:** Unlike the usual inner product notation, we never write $ds^2 = g^{\mu\nu} dx_\mu dx_\nu$. The dx's *always* have upper indices[1].

- In Cartesian coordinates, the metric tensor is simply the identity matrix, but in general, it could be off-diagonal, or even be a function of the coordinates.

[1]This is because $\{dx^\mu\}$ forms a basis for the covectors (1-forms); see Chapter 5. Compare this to the basis vectors of Chapter 1 $\{\vec{e}_i\}$ which always came with lower indices, so that $\langle dx^i, \vec{e}_j \rangle = \delta^i_j$. Notice that the indices work out exactly as they should for the Kroneker delta - this is, once again, not a coincidence!

3.1.1 Curvilinear Coordinates

- In spherical coordinates, the directions are given by $(dr, d\theta, d\phi)$, and the metric is given by:

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2$$

so the metric tensor is therefore:

$$[g_{\mu\nu}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}$$

- Cyllindrical coordinates (or polar coordinates) are given by the coordinates:

$$\rho = \sqrt{x^2 + y^2} \tag{3.1.1}$$
$$\phi = \tan^{-1}\frac{y}{x} \tag{3.1.2}$$
$$z = z \tag{3.1.3}$$

The metric is given by:

$$ds^2 = d\rho^2 + \rho^2 d\phi^2 + dz^2$$

and therefore:

$$[g_{\mu\nu}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- **Parabolic coordinates** are derived from the cyllindrical coordinates in the following way:

$$z = \frac{1}{2}(\xi - \eta) \tag{3.1.4}$$
$$\rho = \sqrt{\xi\eta} \tag{3.1.5}$$
$$\phi = \phi \tag{3.1.6}$$

or using the spherical radial distance:

$$\xi = r + z \qquad\qquad \eta = r - z$$

Plugging these expressions into the metric for cyllindrical coordinates, we derive the parabolic coordinate metric:

$$ds^2 = \frac{1}{4}\left[\left(1+\frac{\eta}{\xi}\right)d\xi^2 + \left(1+\frac{\xi}{\eta}\right)d\eta^2\right] + \xi\eta d\phi^2$$

therefore:

$$[g_{\mu\nu}] = \begin{pmatrix} \frac{1}{4}\left(1+\frac{\eta}{\xi}\right) & 0 & 0 \\ 0 & \frac{1}{4}\left(1+\frac{\xi}{\eta}\right) & 0 \\ 0 & 0 & \xi\eta \end{pmatrix}$$

Parabolic coordinates describe families of concentric paraboloids symmetric about the z-axis. They are very useful in problems with the appropriate symmetry, such as parabolic mirrors.

- As you might have guessed, there are also hyperbolic coordinates and elliptical coordinates, but I won't get into those here. There are even more complicated coordinate systems that are useful (toroidal coordinates, etc).

3.2 Differential Volumes and the Laplacian

3.2.1 Jacobians

- If you transform your coordinates, you will modify your differential volume element by a factor called the **Jacobian**:

$$y^\alpha = y^\alpha(x^\beta) \Rightarrow d^n x = \mathcal{J} d^n y$$

where in general, $\mathcal{J} = \det\left(\frac{\partial x^a}{\partial y^\mu}\right)$

- To find the general Jacobian, let us consider a transformation from Cartesian coordinates (x^a) to general coordinates (y^μ) and ask how the metric transforms. Remembering that the metric is a $(0,2)$ tensor and $g_{ab} = \delta_{ab}$ we have:

$$g_{\mu\nu} = \delta_{ab}\frac{\partial x^a}{\partial y^\mu}\frac{\partial x^b}{\partial y^\nu} \tag{3.2.7}$$

Now take the determinant of both sides, and defining $g \equiv \det(g_{\mu\nu})$ we find that $\mathcal{J} = \sqrt{g}$, so in general:

$$d^n x = \sqrt{g} d^n y$$

.

3.2.2 Veilbeins - a Prelude to Curvature

I will talk about curvature in the next chapter, but now would be a good time to introduce a very useful tool called the vielbein (e^μ_a):

- Consider a point p with coordinates x^μ on a manifold (generally a manifold with curvature, which we'll talk about next Chapter). At every point $p \in M$, there is a "tangent space" – this is basically the flat plane tangent to the manifold at the point p. Give coordinates on the tangent plane ξ^a such that the point p is mapped to the origin. The key point to notice is that the latin index is for flat space while the greek index is for curved space.

- Now consider a point p on the manifold, and the tangent space at p. Then there is a (linear) mapping between the points near the origin of the tangent plane (with coordinates ξ^a) to a neighborhood of p on the manifold (with coordinates x^μ). This linear mapping is given by a matrix called the **vielbein**:

  $$e^\mu_a \equiv \frac{\partial x^\mu}{\partial \xi^a}$$

  This allows us to write Equation (3.2.7) as:

  $$g_{\mu\nu} = \delta_{ab} e^a_\mu e^b_\nu$$

  So the vielbein is, in some sense, the *square root* of the metric! In particular, the above analysis shows that the Jacobian is just e, the determinant of the vielbein. There are many cases in physics where this formalism is very useful[2]. I won't mention any of it here, but at least you have seen it.

3.2.3 Laplacians

- We can find the Laplacian of a scalar field in a very clever way using the Variational principle in electrostatics. For our functional, consider that we wish to minimize the energy of the system, and in electrostatics, that is (up to constants) the integral of $\mathbf{E}^2$. So using our scalar potential, let's define the energy functional in cartesian coordinates:

  $$W[\Phi] = \int d^n x (\nabla\Phi)^2$$

  where $d^n x$ is the $n-$dimensional differential volume element.

- We can find the extremum of this functional $\delta W[\Phi] = 0$ and see what happens. At first, notice that an integration by parts lets you rewrite the integrand as $-\Phi\nabla^2\Phi$, and the variation of that is simply $-\delta\Phi\nabla^2\Phi$. Now change coordinates using the above prescription and repeat the variation. We set the two expressions equal to each other,

[2]For example, when spin is involved.

since the integral must be the same no matter what coordinate system we use. Then you get as a final expression for the Laplacian:

$$\boxed{\nabla^2 = \frac{1}{\sqrt{g}}\partial_\alpha[\sqrt{g}g^{\alpha\beta}\partial_\beta]} \tag{3.2.8}$$

Exercise: Prove this formula. It's not hard, but it's good practice.

Notice that ∂_μ is covariant, so we require the contravariant $g^{\mu\nu}$ - remember that this is the *inverse* of the metric, so be careful!

- Notice that this expression is only valid for *scalar fields*! For vector fields, it is generally more complicated. To see this, notice that at first, you might assume that $\nabla^2\mathbf{A} = (\nabla^2 A_i)$ for each component A_i. In Cartesian coordinates, this is fine. But in curvilinear coordinates, you must be extra careful. The true definition of the i^{th} vector component is $A_i = \mathbf{A}\cdot\vec{e}_i$, and in curvilinear coordinates, $\vec{e}_i$ has spacial dependence, so you have to take its derivative as well. We will take the derivative of a unit vector and discuss its meaning in the next chapter.

- *Exercise:* Find the Laplacian in several different coordinate systems using Equation (3.2.8). The usual Laplacians are on the cover of Jackson. Can you also find the Laplacian in parabolic coordinates? Here's the answer:

$$\nabla^2 = \frac{4}{\xi+\eta}\left[\xi\frac{\partial^2}{\partial\xi^2} + \frac{\partial}{\partial\xi} + \eta\frac{\partial^2}{\partial\eta^2} + \frac{\partial}{\partial\eta}\right] + \frac{1}{\xi\eta}\frac{\partial^2}{\partial\phi^2}$$

3.3 Euclidean vs Minkowskian Spaces

What is the difference between Euclidean space and Minkowski space? In all the examples we have considered so far, we have been in three dimensions, whereas Minkowski space has four. But this is no problem, since nothing above depended on the number of dimensions at all, so that generalizes very nicely (now $g_{\mu\nu}$ is a 4×4 matrix). There is one more difference, however, that must be addressed.

Minkowski space is a "hyperbolic space": the loci of points equadistant from the origin form hyperbolas. This is obvious when you look at the metric in spacetime:

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2$$

(where I am looking in only one spacial dimension for simplicity). For constant Δs, this is the equation of a hyperbola.

The key to this strange concept is the minus sign. We immediately notice that trouble will arise if we naively pretend that there is nothing wrong, since that minus sign makes g (the determinant of the metric tensor) negative, and the square root of a negative number makes no sense (you can't have immaginary distance!). To fix this problem, we will forever in Minkowski space make the substitution $g \to -g$, and all our woes will vanish! Another

way to think of this (and perhaps the better and more general way) is to replace $g \to |g|$ in all of our equations. Then there will never be a problem.

Just for completeness, the metric tensor of Minkowski spacetime is:

$$\eta_{\mu\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

This is called the "mostly-minus" prescription. Sometimes people will use the "mostly plus" prescription which is the negative of this. You must be very careful which convention is being used - particle physicists often use the minus prescription I use here, while GR people tend to use the mostly plus prescription. But sometimes they switch! Of course, physics doesn't change, but the formulas might pick up random minus signs, so be careful.

Chapter 4

Geometry II: Curved Space

Consider the planet Earth as a sphere. We know that in Euclidean geometry, if two lines are perpendicular to another line, then they are necessarely parallel to each other (they never intersect). Yet a counterexample of this is longitude lines: two lines of longitude are always perpendicular to the equator, and so they are "parallel"; yet they meet at two points: the north and south poles. Obviously, something is wrong with Euclidean geometry on a sphere.

In everything that we have been doing, we have been assuming that the space we live in was flat. Even when we were using curvilinear coordinates, we were still talking about flat space. The situation is fundamentally different when space itself is curved. It is here where differential geometry truly becomes interesting!

There is a *very* deep theorem in mathematics that states that any manifold of any dimension and any curvature can always be embedded in a flat space of some larger dimension. For example, a sphere (which is two dimensional) can be embedded into Euclidean 3-space. So in principle, we can avoid the issue of curvature completely if we think in enough dimensions. But this is not practical or necessary in general, and curvature plays a very important role in mathematical physics.

Differential geometry and curved space have their most obvious role in general relativity (GR), where gravity manifests as a warping of spacetime. However, it also has applications in several other branches of mathematical physics, including (but not limited to) optics, solid-state physics and quantum field theory. So it might be useful to see some of the basics in one place. Hence this chapter.

Differential geometry is truly a deep and profound subject, and I recommend anyone interested to study it either in a class or on your own (I took three classes in it as an undergraduate and I'm still trying to learn it!) But I do not plan to teach differential geometry here; only give an overview and a source for common equations you might come across in physics. For a deeper understanding, you have to go to a textbook. I give some recomendations in the bibliography.

4.1 Connections and the Covariant Derivative

One of the biggest problems with geometry on curved spaces is the notion of the derivative of a vector field. When you take the derivative, you are actually taking the difference of two

vectors that are separated by an infitiesimal distance:

$$d\mathbf{A} = \mathbf{A}(\mathbf{x}+\epsilon) - \mathbf{A}(\mathbf{x})$$

But this doesn't even make sense - you can only add or subtract different vectors when they have the same base point. In flat space, this is not a problem - you simply move one of the vectors over so that their base points coincide, and then you can subtract them like usual. However, as we will see in a later section, you cannot generally move a vector around arbitrarely in curved space without changing it. Hence, taking the difference of two vectors with different base points is a tricky subject, and therefore the naive derivative will not work here.

Instead, we must define a new tensor called a connection:

Definition: A **connection** on a manifold M is a map that takes two tensors to a tensor. It is denoted $\nabla_X Y$ and necessarely has the following properties:

1. $\nabla_{f(\mathbf{x})X_1+g(\mathbf{x})X_2}Y = f(\mathbf{x})\nabla_{X_1}Y + g(\mathbf{x})\nabla_{X_2}Y, \qquad f,g \in C^\infty(M).$
2. $\nabla_X(aY_1+bY_2) = a\nabla_X Y_1 + b\nabla_X Y_2, \qquad a,b$ constants.
3. $\nabla_X(f(\mathbf{x})Y) = f(\mathbf{x})\nabla_X Y + (Xf(\mathbf{x}))Y$

The first two rules give the connection linear properties, just like we expect the derivative to do. Notice that the first item applies to general *functions*, not just constants. Resorting to mathematical terminology, we say that the connection $\nabla_X Y$ is **tensorial** in X and **local** in Y. The adjective "tensorial" means that it only depends on X *at a specific point*, as opposed to a local neighborhood around the point. This is why we can factor out the functional dependence. Notice that the ordinary directional derivative in vector calculus is exactly the same. Finally, the third rule is the analogue of Leibniz's rule[1] in vector calculus. This becomes easy to see if we let $X = x^i\partial_i$, where we have chosen $\vec{\mathbf{e}}_i \equiv \partial_i$ – a standard trick in differential geometry.

Armed with this knowedge, we can ask: what is the connection of our coordinates? Let's choose a coordinate basis for our manifold ($\vec{\mathbf{e}}_i$). Then we have the following *vital* definition:

$$\nabla_{\vec{\mathbf{e}}_i}\vec{\mathbf{e}}_j \equiv \Gamma^k_{ij}\vec{\mathbf{e}}_k \tag{4.1.1}$$

(Note the Einstein notation on the index k). The n^3 numbers Γ^k_{ij} introduced here[2] are called **Christoffel symbols**, or sometimes **connection coefficients**. *They are* **not** *tensors!!* But they are tensorial in the sense described above. They have the *inhomogeneous* transformation law:

$$\boxed{\Gamma^{\sigma'}_{\mu'\nu'} = A^{\mu}_{\cdot\mu'}A^{\nu}_{\cdot\nu'}A^{\sigma'}_{\cdot\sigma}\Gamma^{\sigma}_{\mu\nu} + A^{\sigma'}_{\cdot\sigma}\frac{\partial^2 x^\sigma}{\partial x^{\mu'}x^{\nu'}}} \tag{4.1.2}$$

[1] Sometimes called the product rule.

[2] $n = dim M$

where the matrices A are the coordinate transformation matrices defined in Chapter 1.

It seems reasonable that this object might be like the generalization of the derivative. However, it turns out that we're not quite done - these three axioms define a connection, but they do not define it uniquely. If we want to define something like a derivative, we need to invoke more axioms. We usually chose the following two:

4. $\nabla_X g \equiv 0 \qquad \forall X$

5. $\Gamma^i_{jk} = \Gamma^i_{kj}$

The first of these is often put into words as: "The connection is **compatible** with the metric." It basically means that the derivative of a product makes sense. The second property is that the connection is **symmetric** or **torsionless**. These two additional properties, along with the three already included, are necessary and sufficient to define a unique connection called the **Riemann Connection**, or the **Levi-Civita Connection**. Before we proceed, notice that this is by no means the only choice I could have made. There are many theories where, for example, we drop the symmetric condition to get various other types of geometry. This is an especially popular trick in trying to quantize gravity. But this is way beyond the scope of this paper, and I will simply stick to Riemannian geometry, where these are the axioms.

Now we would like to translate this rather abstract discussion into something concrete. Consider a vector field $\mathbf{A}(\mathbf{x})$ on a manifold M. Let's take its connection with respect to a basis vector:

$$\nabla_{\vec{\mathbf{e}}_i}\mathbf{A} = \nabla_{\vec{\mathbf{e}}_i}(A^j\vec{\mathbf{e}}_j) = (\partial_i A^j)\vec{\mathbf{e}}_j + A^j\nabla_{\vec{\mathbf{e}}_i}\vec{\mathbf{e}}_j = (\partial_i A^j)\vec{\mathbf{e}}_j + A^j\Gamma^k_{ij}\vec{\mathbf{e}}_k = \left[\partial_i A^k + A^j\Gamma^k_{ij}\right]\vec{\mathbf{e}}_k$$

where I have used Leibniz's rule and renamed a dummy index in the last step ($j \to k$). If I am only interested in the coordinate form of the connection (which I usually am), I can drop the basis vectors and write[3]:

$$\boxed{D_i A^k = \partial_i A^k + A^j\Gamma^k_{ij}} \tag{4.1.3}$$

$$\boxed{D_i A_k = \partial_i A_k - A_j\Gamma^j_{ik}} \tag{4.1.4}$$

Proving Equation (4.1.4) is very straightforward - simply consider $D_i(A^jA_j) = \partial_i(A^jA_j)$ and use Equation (4.1.3). Equations (4.1.3-4.1.4) are often called the **covariant derivatives**, and replace the ordinary derivative of flat space. Notice that in the limit of flat space, there is no difference between the ordinary and covariant derivative, but in more general cases, there is a "correction" to the ordinary derivative that compensates for how you subtract the vectors with different basepoints. Physically speaking, the covariant derivative is the derivative of a vector field that you would see *on the surface*. Recall I said that you can always embed a manifold in a larger flat space, where the ordinary derivative works fine. But as long as we stay strictly on the surface, we must use the covariant derivative.

[3]When in a specific coordinate basis, we often simply write $\nabla_{\partial_i} \Leftrightarrow D_i$.

Remembering our easy notation of Chapter 1, it is sometimes easier to write the differential coordinate index with a semi-colon:

$$\boxed{A^k_{;i} = A^k_{,i} + A^j \Gamma^k_{ij}} \tag{4.1.5}$$

What if I wanted to take the covariant derivative of a tensor that is higher than rank-1? Let's consider the covariant derivative of the diadic:

$$(A_i B_j)_{;k} = A_{i;k} B_j + A_i B_{j;k} = (A_i B_j)_{,k} + (\Gamma^m_{ik} A_m) B_j + (\Gamma^m_{jk} B_m) A_i$$

where I have skipped some straightforward algebra. This, along with Equation (4.1.4), should give us enough intuition to express the covariant derivative on an arbitrary rank tensor:

$$\boxed{(T^{i_1 i_2 \cdots i_n})_{;j} = (T^{i_1 i_2 \cdots i_n})_{,j} + \sum_{\alpha=1}^{n} \Gamma^{i_\alpha}_{\beta j} T^{i_1 \cdots i_{\alpha-1} \beta i_{\alpha+1} \cdots i_n}} \tag{4.1.6}$$

In other words – for a rank-n tensor, we get a Γ term for each contravariant index. Similarly there would be a minus Γ term for each covariant index. By the way, a quick corollary to this would be:

$$\Phi_{;i} = \Phi_{,i}$$

for a scalar quantity Φ. This should be so, since the Christoffel symbols only appeared when we had basis vectors to worry about.

Now that we have the covariant derivative under our belts, we can use it to find a useful expression for the connection coefficients. Take the covariant derivative of the metric tensor three times (each with a different index), and use the property of compatability to get:

$$\boxed{\Gamma^\rho_{\nu\sigma} = \frac{1}{2} g^{\mu\rho} \left[g_{\mu\nu,\sigma} - g_{\nu\sigma,\mu} + g_{\sigma\mu,\nu} \right]} \tag{4.1.7}$$

As a final exercise, consider a parametrized curve on a manifold M, $\mathbf{w}(s) : I \to M$, where I is the unit interval [0,1]. Then we can take the parametrized covariant derivative by resorting to the chain rule:

$$\frac{Dw^j}{ds} = D_i w^j \frac{dx^i}{ds} = \frac{\partial w^j}{\partial x^i} \frac{dx^i}{ds} + (\Gamma^j_{ik} w^k) \frac{dx^i}{ds}$$

Combining terms:

$$\boxed{\frac{Dw^j}{ds} = \frac{dw^j}{ds} + \Gamma^j_{ik} w^k \frac{dx^i}{ds}} \tag{4.1.8}$$

This expression is very useful.

Exercise: Prove Equation (4.1.2). Hint: use the covariant derivative.

Exercise: Prove Equation (4.1.7) using the method suggested.

4.2 Parallel Transport and Geodesics

Let's return to our spherical planet example from the beginning of the chapter. Consider two parallel vectors tangent to the Earth at the equator and the prime meridian (zero longitude, zero lattitude). Now move one of the vectors along the equator until it comes to the international date line (180 degrees around). Now proceed to move both vectors up the longitude line until they meet at the north pole. Notice that they are no longer parallel, but antiparallel. If you were to do the same experiment, but move the vector only 90 degrees around the globe, the two vectors would be *perpendicular* at the pole[4]!

Again, we see that there is something fishy about the notion of parallel vectors in curved space. Geometers knew this, and they did their best to generalize the notion of parallel vectors by inventing an algorithm called **parallel transport**. In this section, I will talk a little about that, and then use our new-found intuition to define a geodesic.

We begin with a definition:

Definition: Consider a parametrized path $\alpha(s) : I \to M$ and a vector field along the curve: $\mathbf{w}(s) = \mathbf{w}(\alpha(s))$ all in a manifold M. Then we say that $\mathbf{w}(s)$ is a **parallel vector field** if its covariant derivative vanishes everywhere:

$$\frac{D\mathbf{w}}{ds} = 0 \qquad \forall s \in I$$

The uniquness theorem of differential equations tells us that if a vector field has a given value at some initial point ($\mathbf{w}(0) = \mathbf{w}_0$), then there is a *unique* vector $\mathbf{w}$ that is parallel to $\mathbf{w}_0$ that comes from parallel transporting $\mathbf{w}_0$ some distance *along the contour* $\alpha(s)$.

Armed with this knowedge, we can finally move onto one of the most useful objects in geometry:

Definition: A **geodesic** on a manifold M is a parametrized curve $\mathbf{x}(s) : I \to M$ whose tangent vectors form a parallel vector field. In other words, the geodesic satisfies the differential equation:

$$\frac{D\mathbf{x}'}{ds} \equiv 0$$

Using Equation (4.1.8), with $\mathbf{x}' \equiv \frac{d\mathbf{x}}{ds}$ as the vector field we're differentiating, we get a famous nonlinear differential equation:

$$\boxed{\frac{d^2x^\mu}{ds} + \Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{ds}\frac{dx^\sigma}{ds} = 0} \tag{4.2.9}$$

Going back to physics, note that a particle's trajectory $x^\mu(s)$ should ultimately be given by Newton's laws. Notice that if we're in flat space, these equations do indeed reduce to

[4]This is actually a sketch of the proof of the famous theorem from geometry/topology that states that you can never have a continuous vector field defined everywhere on a sphere that doesn't vanish somewhere. But there is no need to get into the details of that here.

Newton's laws of motion! General relativity suggests that in the absense of inertial forces, a particle moves along a geodesic, so its equation of motion is given by Equation (4.2.9). Gravity manifests itself as the curvature of space, *i.e.:* the connection coefficients. So we can think of the Γ terms in Equation (4.2.9) as source terms for the gravitational field in Newton's equations of motion. If we wished to include other inertial forces (such as an electric field, for instance), we would modify the right hand side of Equation (4.2.9).

Additionally, a geodesic is the is the path of shortest distance: this can be proved using variational calculus techniques. So this equation tells us that particles like to travel along the path the minimizes the distance travelled. In flat space (no forces), that's a straight line; but in curved space (when gravity is present, for example) then in can be more complicated.

4.3 Curvature- The Riemann Tensor

Now that we have worked out some details on parallel transport and geodesics, let's see what else the covariant derivative can give us. One question that analysts like to ask is "Are mixed partials equal?" It turns out that understanding the answer to this question gives us a rigorous way to quantify the curvature of the manifold.

Consider the commutator of two covariant derivatives acting on a vector[5]:

$$A_{\nu;\rho;\sigma} - A_{\nu;\sigma;\rho} = A_\mu R^\mu_{\ .\nu\rho\sigma}$$

where

$$\boxed{R^\mu_{\ .\nu\rho\sigma} \equiv \Gamma^\mu_{\nu\sigma,\rho} - \Gamma^\mu_{\nu\rho,\sigma} + \Gamma^\alpha_{\nu\sigma}\Gamma^\mu_{\alpha\rho} - \Gamma^\alpha_{\nu\rho}\Gamma^\mu_{\alpha\sigma}} \tag{4.3.10}$$

$R^\mu_{\ .\nu\rho\sigma}$ is a tensor, called the **Riemann-Christoffel Curvature Tensor**. We know it's a tensor from the quotient rule of tensors: *if a tensor equals another tensor contracted with another object, then that other object must be a tensor.* There are a number of symmetries[6]:

1. $R_{\beta\nu\rho\sigma} = -R_{\nu\beta\rho\sigma}$
2. $R_{\beta\nu\rho\sigma} = -R_{\beta\nu\sigma\rho}$
3. $R_{\beta\nu\rho\sigma} = +R_{\rho\sigma\beta\nu}$
4. $R^\mu_{\ .\nu\rho\sigma} + R^\mu_{\ .\rho\sigma\nu} + R^\mu_{\ .\sigma\nu\rho} = 0$
5. $R^\mu_{\ .\nu\rho\sigma;\tau} + R^\mu_{\ .\nu\sigma\tau;\rho} + R^\mu_{\ .\nu\tau\rho;\sigma} = 0$

The first four symmetries are manifest from Equation (4.3.10); the last item is known as the **Bianchi Identity**. These symmetries have the effect of putting some powerful constraits on the number of independent quantities in the curvature tensor. Without any symmetries, there would be d^4 numbers, where d is the dimension of our manifold. However, with all of these symmetries, there are:

[5] I'm not going to prove this: *you* do it!

[6] $R^\mu_{\ .\nu\rho\sigma} = g^{\mu\beta}R_{\beta\nu\rho\sigma}$

$$\frac{d^2(d^2-1)}{12} \tag{4.3.11}$$

independent quantities. Notice that the rank of the Riemann-Christoffel tensor is always 4, despite the dimension of the manifold.

The Bianchi Identity is a different sort of relation in the sense that it involves derivatives. There is a similar equation in Maxwell's electrodynamics:

$$F_{\mu\nu,\rho} + F_{\nu\rho,\mu} + F_{\rho\mu,\nu} = 0$$

where $F_{\mu\nu}$ is the field strength tensor (see Jackson or Griffiths). These equations are very closely related to the topic of "Gauge Invariance" – we will talk about that in the next chapter.

We can try to construct lower rank tensors from this monster in Equation (4.3.10), but thanks to the symmetries, there are only a few nonzero contractions, and up to sign, they all yield the same thing. We define the rank-2 tensor:

$$R_{\mu\nu} \equiv R^{\tau}_{.\mu\nu\tau} \tag{4.3.12}$$

It is called the **Ricci tensor**. It is a symmetric tensor. Any other nonzero contraction would give the same object up to a sign. You should prove this for yourself. Finally we can take the trace of Equation (4.3.12) to get the **Ricci scalar**:

$$R \equiv R^{\mu}_{\mu} \tag{4.3.13}$$

R is sometimes called the **scalar curvature**. It is a measure of the intrisic curvature of the manifold.

Finally, I would like to mention an interesting curiosity. Take a good look at Equation (4.3.10). Einstein used this tensor to describe gravity. Notice that it has derivatives of the connection coefficients, and is a tensor. It is, in a sense, a strange kind of curl of the connection coefficients. Now look at Equation (4.1.2) - this says that the connection coefficients transform under coordinate transformations with an inhomogeneous term. Does this look like anything we know? What about the vector potential of electrodynamics? That has an inhomogeneous term when you perform a gauge transformation ($\mathbf{A} \to \mathbf{A} + \nabla\Lambda$), and yet when you take its curl to get $\mathbf{B}$, the resulting field is gauge invariant. Here, the "gauge transformation" is actually a coordinate transformation (technically, a "diffeomorphism"), and the Riemann tensor transforms "covariantly" (which is a fancy word for "properly") under these transformations, even though the connection coefficients do not. You can see the parallel: you might think of the connection coefficients as "gravity vector potentials", and the curvature tensor as the "gravitational field". Then we have something like a gauge theory of gravity! Unfortunately, things are more complicated than that. However, this is getting us too far afield, so I'll stop now.

4.3.1 Special Case: $d = 2$

Before leaving differential geometry behind, I wanted to talk a little about a very special case of mathematics - two dimensional manifolds. In one dimension, the Riemann tensor

is not well-defined, and we do not need any of this formalism to discuss geometry; in fact, one dimensional differential geometry is nothing more than vector calculus! In three or more dimensions, the Riemann tensor is complicated, and it is much more difficult to say anything quantitative and general about your manifold. But in the special case of two dimensions, the Riemann tensor has only one component (see Equation (4.3.11)), and that is the Ricci scalar curvature. Here, we can talk about curvature in a *very* quantitative matter.

To understand how this works, we need to state a few definitions. A two dimensional Riemannian manifold is often simply referred to as a "surface".

Definition: Consider a curve parametrized by $\alpha(s)$ living on a surface M. The the **curvature** of the curve is given by:

$$k(s) \equiv |\alpha''(s)|$$

If $\mathbf{n}$ is the normal vector to α, and $\mathbf{N}$ is the normal to the surface, then we define the angle:

$$\cos\theta_{\alpha(s)} \equiv \langle \mathbf{n}(s), \mathbf{N}(s) \rangle$$

Then

$$k_n(s) \equiv k(s)\cos\theta_{\alpha(s)}$$

is called the **normal curvature** along the path $\alpha(s)$.

Now consider all the curves in the surface, and all of the normal curvatures. Let us call the maximum normal curvature k_1, and the minimum normal curvature k_2.[7] These two quantities are properties of the surface, not of the curves. Finally, we can define two new objects:

$$\boxed{H = \frac{1}{2}(k_1 + k_2)} \tag{4.3.14}$$

$$\boxed{K = k_1 k_2} \tag{4.3.15}$$

H is called the **mean curvature**. It is very useful in many applications, especially in engineering. However, it is not as useful for a theoretical tool because it is not isometrically invariant, where an "isometrically invariant" quantity only depends on the surface itself, not on how the surface is embedded in 3-space. K is called the **Gaussian curvature**, and is of fundamental importance. Gauss proved that it is isometrically invariant, which means that it describes the surface in a fundamental way. One can prove the relation:

$$K \propto R$$

and hense make the connection between two dimensional geometry and higher dimensional geometry.

[7] There is no difference to which one is which in the literature.

Armed with this new tool, we can state one of the most important theorems of differential geometry, and one of the key reasons why differential geometry in two dimensions is so well understood:

Theorem: For any compact, oriented Riemannian 2-manifold M:

$$\boxed{\int_M K d\Omega = 2\pi\chi(M)} \tag{4.3.16}$$

where the integral is a surface integral over the entire manifold and $\chi(M)$ is a topologically invariant quantity called the **Euler characteristic**. This theorem is called the **Gauss-Bonnet Theorem**, and has several versions for different cases of M, integrating over only part of M, etc; but this is the most general statement of the theorem.

The Euler characteristic $\chi(M)$ is of fundamental importance in topology and geometry. To calculate it for a surface, you divide your surface into a bunch of triangles; this process is called **triangulation**. You then count the total number of faces (F), edges (E) and vertices (V) and apply the following formula:

$$\chi(M) = V - E + F \tag{4.3.17}$$

$\chi(M)$ is well-defined in the sense that it is independent of how you triangulate your surface. This is not a trivial point. We call this quantity a "topological invariant" because if you have two surfaces that have the same value of χ, then they are homeomorphic; in other words, you can always deform one of the surfaces into another in a continuous manner, (*i.e.:* without tearing or gluing. Thus the Euler characteristic provides a very powerful techinque for classifying different types of surfaces.

As a very quick application of Equation (4.3.16), consider Gauss' Law from electrodynamics. If you interpret K as the "field strength" we see that the surface integral must be some topological invariant, i.e. "charge"! This kind of connection between field theory (classical and quantum) and differential geometry is very common, and very powerful!

4.3.2 Higher Dimensions

In more than two dimensions, things become considerably more complicated. In order to define the analog of Gaussian curvature, one considers all the independent two dimensional "cross sections" of the manifold and calculates each of their Gaussian curvatures. These curvatures are know as **sectional curvatures**. It can be proved that all of the sectional curvatures are contained in the Riemann tensor, which is why the Riemann-Levi-Civita formalism is so useful in higher dimensions, even if it is overkill in two dimensions.

Unfortunately, there is no such thing as a Gauss-Bonnet theorem for manifolds of more than two dimensions. First of all, the Gaussian curvature is no longer well-defined. Instead, you might use some information from the Riemann tensor, but it is not obvious how to do this. Secondly, and more importantly, *there is no known universal topological invariant in more than two dimensions!* In other words, manifolds with the same Euler characteristic

need no longer be topologically equivalent in higher dimensions, and we know of no single quantity that does classify manifolds in this way. This is way beyond the scope of this review, so I will stop now. Let me just say that this is one of the most active areas in modern mathematics research. If you are at all interested, I strongly encourage you to check it out!

Chapter 5

Differential Forms

Differential forms are, without a doubt, one of the most beautiful inventions in all of mathematics! Their elegance and simplicity are without bound. Nevertheless, like anything good, it takes a long time before they become natural and second nature to the student. Whether you realize it or not, you have been using differential forms for a long time now (gradient).

Before diving into forms, I should mention two quick things. First of all, it is likely that many of you will never need to know any of this in your life. If that is the case, then you have no reason to read this chapter (unless you're curious). I have included it anyway, since writing these notes has helped me to understand some of the more subtle details of the mathematics, and I find forms to be very useful. Differential forms are extremely useful when, for example, you wish to discuss electromagnetism, or any gauge field for that matter, in curved space (i.e.: with gravity). This is why I thought they would do well in this review.

The second point I want to make is that no matter how much you like or dislike forms, remember that they are nothing more than notation. If mathematicians had never discovered differential forms, we would not be at a serious loss; however, many calculations would become long and tedious. The idea behind this beautiful notation is that terms cancel early. If we had never heard of differential forms, one would have to carry out terms that would ultimately vanish, making the calculations nasty, at best! But let us not forget that there is nothing truly "fundamental" about forms – they are a convienience, nothing more.

That being said, let's jump in...

5.1 The Mathematician's Definition of Vector

One thing you have to be very careful of in mathematics is definitions. We have to construct a definition that both allows for all the cases that we know should exist, and also forbids any pathological examples. This is a big challenge in general mathematical research. We have been talking a lot about "vectors", where we have been using the physicist's definition involving transformation properties. But this is unfortunately not the only mathematical definition of a vector. Mathematicians have another definition that stems from the subject of linear algebra. For the moment, we are going to completely forget about the physical definition (no more transformations) and look on vectors from the mathematician's point of view. To this end we must make a new definition; the key definition of linear algebra:

Definition: A **Vector Space** (sometimes called a **Linear Space**), denoted by $\mathcal{V}$, is a set of objects, called **vectors**, and a field, called **scalars**, and two operations called vector addition and scalar multiplication that has the following properties:

1. ($\mathcal{V}$,+) is an abelian group.
2. $\mathcal{V}$ is closed, associative, and distributive under scalar multiplication, and there is a scalar identity element ($1\vec{v} = \vec{v}$).

In our case, the field of scalars will always be $\mathbb{R}$, but in general it could be something else, like $\mathbb{C}$ or some other field from abstract algebra.

Notice that this definition of vectors has absolutely nothing to do with the transformation properties that we are used to dealing with. Nevertheless, you can prove to yourself very easily that the set of n-dimensional real vectors ($\mathbb{R}^n$) is a vector space. But so are much more abstract quantities. For example, the set of all n-differentiable, real-valued functions ($C^n(\mathbb{R})$) is also a vector space, where the "vectors" are the functions themselves. Another example of a vector space is the set of all eigenfunctions in quantum mechanics, where the eigenfunctions are the vectors. This is why it's called "abstract" algebra!

Sticking to abstract linear algebra for a minute, there is a very important fact about vector spaces. Recall that a **linear transformation** is a function that operates on $\mathcal{V}$ and preserves addition and scalar multiplication:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

Then let us consider the set of all linear transformations that go from $\mathcal{V}$ to $\mathbb{R}$:

$$\Lambda^1(\mathcal{V}) = \{\phi| \;\; \phi : \mathcal{V} \to \mathbb{R}; \;\; \phi \text{ linear}\} \tag{5.1.1}$$

I'll explain the notation in a moment. This set is very intriguing - every element in this set has the property that it acts on any vector in the original vector space and returns a number. It is, in general, true that for any vector space, this special set exists. It is also true that this space also forms a vector space itself. It is so important in linear algebra, it is given a name. We call it the **dual space of** $\mathcal{V}$, and denote it as $\mathcal{V}^*$. The vectors (functions) in the space are called **covectors**, or **1-forms** (hense the 1 superscript).

Armed with this knowledge we can go back and redefine what we mean by "tensor": consider the set of all multilinear (that is, linear in each of its arguments) transformations that send k vectors and l covectors to a number:

$$T : \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{\text{k times}} \times \underbrace{\mathcal{V}^* \times \cdots \times \mathcal{V})^*}_{\text{l times}} \to \mathbb{R}$$

The set of all such transformations is called T^l_k and the elements of this set are called **tensors of rank** $\begin{pmatrix} l \\ k \end{pmatrix}$. Notice immediately that $T^0_1 = \mathcal{V}$, $T^1_0 = \mathcal{V}^*$. Also notice that this definition of tensor has nothing to do with transformations, as we defined them in Chapter 1. For

example, acording to this definition, the Christoffel symbols of differential geometry *are* $(1,2)$ tensors. So you must be sure you know which definition people are using.

That's enough abstraction for one day. Let us consider a concrete example. Consider the vector space $\mathbb{R}^n$ and the set of *multilinear* transformations that take k vectors to a number:

$$\phi : \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{\text{k times}} \to \mathbb{R} \tag{5.1.2}$$

with the additional property that they are "alternating", i.e.:

$$\phi(\ldots, \vec{x}, \vec{y}, \ldots) = -\phi(\ldots, \vec{y}, \vec{x}, \ldots) \tag{5.1.3}$$

When ϕ takes k vectors in $\mathbb{R}^n$, then ϕ is called a **k-form**. The set of all k-forms is denoted $\Lambda^k(\mathbb{R}^n)$. Note that this set is also a vector space. The dual space is an example of this, hence the notation.

In linear algebra, we know that every vector space has a basis, and the dimension of the vector space is equal to the number of elements in a basis. What is the standard basis for $\Lambda^k(\mathbb{R}^n)$? Let us start by considering the dual space to $\mathbb{R}^n$ ($k = 1$). We know the standard basis for $\mathbb{R}^n$; let's denote it by $\{\vec{e}_1, ..., \vec{e}_n\}$. Now let us *define* a 1-form dx^i with the property:

$$dx^i(\vec{e_j}) \equiv \langle dx^i, \vec{e_j} \rangle = \delta^i_j \tag{5.1.4}$$

Then the **standard basis of $\Lambda^k(\mathbb{R}^n)$** is the set of all k-forms:

$$dx^i \wedge dx^j \wedge \cdots dx^l; \quad i \neq j \neq \cdots \neq l; \quad i, j, \cdots, l \leq n \tag{5.1.5}$$

where each "wedge product" has k "dx"s[1].

Before going any further, perhaps it would be wise to go over this and make sure you understand it. Can you write down the standard basis for $\Lambda^2(\mathbb{R}^3)$? How about $\Lambda^2(\mathbb{R}^4)$? Here are the answers:

$$\begin{aligned} \Lambda^2(\mathbb{R}^3) &: \{dx^1 \wedge dx^2, dx^1 \wedge dx^3, dx^2 \wedge dx^3\} \Rightarrow D = 3 \\ \Lambda^2(\mathbb{R}^4) &: \{dx^1 \wedge dx^2, dx^1 \wedge dx^3, dx^1 \wedge dx^4, dx^2 \wedge dx^3, dx^2 \wedge dx^4, dx^3 \wedge dx^4\} \Rightarrow D = 6 \end{aligned}$$

By looking at these examples, you should be able to figure out what the dimension of $\Lambda^k(\mathbb{R}^n)$ is, in general ($k < n$):

$$\dim \Lambda^k(\mathbb{R}^n) = \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{5.1.6}$$

Notice that if $k > n$ the space is trivial – it only has the zero element.

[1] Wedge products are discussed below, but for now all you need to know of them is that they are anti-symmetric products, so $A \wedge B = -B \wedge A$ for A, B 1−forms.

5.2 Form Operations

5.2.1 Wedge Product

In order to define the standard basis for $\Lambda^k(\mathbb{R}^n)$, we needed to introduce a new operation, called the **wedge product** ($\wedge$). Let's look more carefully at this new form of multiplication. As a "pseudo-definition", consider two 1-forms $\omega, \phi \in \Lambda^1(\mathbb{R}^n)$; then let us define an element in $\Lambda^2(\mathbb{R}^n)$ by the following:

$$\omega \wedge \phi(\vec{v}, \vec{w}) = \omega(\vec{v})\phi(\vec{w}) - \omega(\vec{w})\phi(\vec{v}) \tag{5.2.7}$$

Notice that there are two immediate corollaries:

$$\omega \wedge \phi = -\phi \wedge \omega \tag{5.2.8}$$

$$\omega \wedge \omega = 0 \qquad \forall \omega \in \Lambda^k(\mathbb{R}^n) \tag{5.2.9}$$

So the wedge product defines a noncommutative product of forms. Notice that this is different from the ordinary tensor product:

$$\omega \otimes \phi(\vec{v}, \vec{w}) \equiv \omega(\vec{v})\phi(\vec{w}) \tag{5.2.10}$$

Looking at this as an antisymmetric product on forms, you might be reminded of the cross product. This is exactly right: the cross-product of two vectors is the same thing as a wedge product of two 1-forms. We will prove this explicitly soon.

Now that we have talked about the specific case of 1-forms, we can generalize to the wedge product of any forms:

Definition: Let $\omega \in \Lambda^k(\mathbb{R}^n), \phi \in \Lambda^l(\mathbb{R}^n)$. Then we define $\omega \wedge \phi \in \Lambda^{k+l}(\mathbb{R}^n)$ as the sum of all the antisymmetric combinations of wedge products. This product is

1. Distributive: $(dx^1 + dx^2) \wedge dx^3 = dx^1 \wedge dx^2 + dx^1 \wedge dx^3$
2. Associative: $(dx^1 \wedge dx^2) \wedge dx^3 = dx^1 \wedge (dx^2 \wedge dx^3) = dx^1 \wedge dx^2 \wedge dx^3$
3. Skew-commutative: $\omega \wedge \phi = (-1)^{kl}\phi \wedge \omega$

Notice that for $k + l > n$ the space is trivial, so $\phi \wedge \omega = 0$.

Before going any further, let's do some examples. Consider $\omega, \phi \in \Lambda^1(\mathbb{R}^3)$, i.e.: 1-forms in 3-space. Writing it out explicitly, we have:

$$\begin{aligned}
\omega &= a_1 dx^1 + a_2 dx^2 + a_3 dx^3 \\
\phi &= b_1 dx^1 + b_2 dx^2 + b_3 dx^3 \\
\omega \wedge \phi &= (a_1 dx^1 + a_2 dx^2 + a_3 dx^3) \wedge (b_1 dx^1 + b_2 dx^2 + b_3 dx^3) \\
&= (a_1 b_2 - a_2 b_1) dx^1 \wedge dx^2 + (a_1 b_3 - a_3 b_1) dx^1 \wedge dx^3 + (a_2 b_3 - a_3 b_2) dx^2 \wedge dx^3
\end{aligned}$$

Notice how I've been careful to keep track of minus signs. Also notice that this looks very similar to a cross product, as stated earlier. What about taking the wedge product of two 1-forms in $\mathbb{R}^4$:

$$\begin{aligned}
\xi &= a_1 dx^1 + a_2 dx^2 + a_3 dx^3 + a_4 dx^4 \\
\chi &= b_1 dx^1 + b_2 dx^2 + b_3 dx^3 + b_4 dx^4 \\
\xi \wedge \chi &= (a_1 b_2 - a_2 b_1) dx^1 \wedge dx^2 + (a_1 b_3 - a_3 b_1) dx^1 \wedge dx^3 + (a_1 b_4 - a_4 b_1) dx^1 \wedge dx^4 \\
&\quad + (a_2 b_3 - a_3 b_2) dx^2 \wedge dx^3 + (a_2 b_4 - a_4 b_2) dx^2 \wedge dx^4 + (a_3 b_4 - a_4 b_3) dx^3 \wedge dx^4
\end{aligned}$$

Note that this does not look like a "cross product"; it is a six-dimensional object in $\mathbb{R}^4$. However, if you took another wedge product with another 1-form, you would get a 3-form in $\mathbb{R}^4$ (which is four-dimensional), and that would give you something like a four-dimensional cross product. In general, taking the wedge product of $(n-1)$ 1-forms in $\mathbb{R}^n$ will give you something analogous to a cross-product. We'll get back to this later.

As a final example, what if you took the wedge product of a 1-form and a 2-form in $\mathbb{R}^3$:

$$\alpha = c_1 dx^1 + c_2 dx^2 + c_3 dx^3 \tag{5.2.11}$$

$$\beta = L_{12} dx^1 \wedge dx^2 + L_{13} dx^1 \wedge dx^3 + L_{23} dx^2 \wedge dx^3 \tag{5.2.12}$$

$$\alpha \wedge \beta = (c_1 L_{23} - c_2 L_{13} + c_3 L_{12}) dx^1 \wedge dx^2 \wedge dx^3 \tag{5.2.13}$$

Notice that if we identify $\beta = \omega \wedge \phi$ from above, then we have shown that the triple wedge product of 1-forms in $\mathbb{R}^3$ is just:

$$\omega \wedge \phi \wedge \alpha = \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} dx^1 \wedge dx^2 \wedge dx^3 \tag{5.2.14}$$

In fact, the wedge product of n 1-forms in $\mathbb{R}^n$ is always such a determinant. This is a very important fact that will come up again and again.

Next, we can ask how a general wedge product of forms acts on vectors? The answer is that the form farthest to the left acts on the vector, and then you must permute the form so that all the 1-forms act on the vector. As an example, consider a 2-form acting on a vector:

$$dx \wedge dy(\vec{v}) = dx(\vec{v}) dy - dy(\vec{v}) dx = v_x dy - v_y dx$$

The minus sign comes from flipping the dx and the dy.

Before moving on, let us conclude this section with a definition and a theorem:

Definition: A **simple k-form** is a "monomial" k-form; that is, there is no addition.

Theorem: All k-forms can be written as a linear combination of simple k-forms.

Proof: The proof of this theorem is straightforward: a simple k-form is just a basis element, possibly multiplied by a number. Then since we are dealing with a vector space, any element of $\Lambda^k(\mathbb{R}^n)$ can be written as a linear combination of these basis elements. **QED.**

5.2.2 Tilde

The tilde operator is the operator that translates us from vectors to forms, and vice versa. In general, it is an operation: (rank-k tensors)$\mapsto$(k-forms). Let's see some examples:

$$\begin{aligned} \vec{v} &= v^1\vec{e}_1 + v^2\vec{e}_2 + v^3\vec{e}_3 \in \mathbb{R}^3 \\ \tilde{v} &= v_1 dx^1 + v_2 dx^2 + v_3 dx^3 \in \Lambda^1(\mathbb{R}^3) \end{aligned}$$

Notice that the tilde changed subscript indices to superscript. This is important when we have to keep track of what is covariant and contravariant.

Let's do another less trivial example:

$$\begin{aligned} L &= \begin{pmatrix} 0 & L_{12} & L_{13} \\ -L_{12} & 0 & L_{23} \\ -L_{13} & -L_{23} & 0 \end{pmatrix} \in \mathbb{R}^3 \quad \text{(Rank-2)} \\ \tilde{L} &= \left(L_{12}dx^1 \wedge dx^2 + L_{13}dx^1 \wedge dx^3 + L_{23}dx^2 \wedge dx^3\right) \in \Lambda^2(\mathbb{R}^3) \\ &= \frac{1}{2}L_{ij}dx^i \wedge dx^j \end{aligned}$$

where in the last line I'm using Einstein notation for the indices. Notice how much simpler it is to write a 2-form as opposed to writing a rank-2 tensor. Also note that both L_{ij} and L_{ji} are taken care of, including the minus sign, by using wedge products as opposed to using tensor products - this is why we need the factor of $\frac{1}{2}$ in the last line. We are beginning to see why form notation can be so useful!

Incidentally, notice that I am only considering antisymmetric tensors when I talk about forms. This is because the form notation has antisymmetry built into it. For symmetric tensors, the concept of a form is not very useful.

Finally, let me state a pretty obvious result:

Theorem: $\tilde{\tilde{v}} = v$.

Proof: Left to the reader.

5.2.3 Hodge Star

The next tool we introduce for differential forms is the **Hodge Star** ($*$). The Hodge Star converts forms to their so-called **dual form**:

$$* : \Lambda^k(\mathbb{R}^n) \to \Lambda^{n-k}(\mathbb{R}^n) \tag{5.2.15}$$

As for many other things in linear algebra, it is sufficient to consider how the Hodge star acts on the basis elements. Let's look at some examples.

In $\mathbb{R}^2$:

$$\begin{aligned} *dx^1 &= dx^2 \\ *dx^2 &= -dx^1 \end{aligned}$$

In $\mathbb{R}^3$:

$$\begin{array}{rclrcl} *dx^1 &=& dx^2 \wedge dx^3 & *(dx^1 \wedge dx^2) &=& dx^3 \\ *dx^2 &=& dx^3 \wedge dx^1 & *(dx^2 \wedge dx^3) &=& dx^1 \\ *dx^3 &=& dx^1 \wedge dx^2 & *(dx^3 \wedge dx^1) &=& dx^2 \end{array}$$

In $\mathbb{R}^4$:

$$\begin{array}{rclrclrcl} *dx^1 &=& +dx^2 \wedge dx^3 \wedge dx^4 & *(dx^1 \wedge dx^2) &=& +dx^3 \wedge dx^4 & *(dx^1 \wedge dx^2 \wedge dx^3) &=& +dx^4 \\ *dx^2 &=& -dx^3 \wedge dx^4 \wedge dx^1 & *(dx^1 \wedge dx^3) &=& -dx^2 \wedge dx^4 & *(dx^1 \wedge dx^2 \wedge dx^4) &=& -dx^3 \\ *dx^3 &=& +dx^4 \wedge dx^1 \wedge dx^2 & *(dx^1 \wedge dx^4) &=& +dx^2 \wedge dx^3 & *(dx^1 \wedge dx^3 \wedge dx^4) &=& +dx^2 \\ *dx^4 &=& -dx^1 \wedge dx^2 \wedge dx^3 & *(dx^2 \wedge dx^3) &=& +dx^1 \wedge dx^4 & *(dx^2 \wedge dx^3 \wedge dx^4) &=& -dx^1 \\ & & & *(dx^2 \wedge dx^4) &=& -dx^1 \wedge dx^3 & & & \\ & & & *(dx^3 \wedge dx^4) &=& +dx^1 \wedge dx^2 & & & \end{array}$$

We can see a pattern here if we remember how the Levi-Civita simbol works. For any form in $\Lambda^k(\mathbb{R}^n)$:

$$\Phi = \phi_{\mu_1 \ldots \mu_k} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k} \Rightarrow$$

$$\boxed{*\Phi = \frac{1}{(n-k)!} \epsilon_{\nu_1 \ldots \nu_n} \phi^{\nu_1 \ldots \nu_k} dx^{\nu_{k+1}} \wedge \cdots \wedge dx^{\nu_n}} \tag{5.2.16}$$

or

$$\boxed{*\phi_{\mu_1 \ldots \mu_{n-k}} = \frac{1}{(n-k)!} \epsilon_{\nu_1 \ldots \nu_k \mu_1 \ldots \mu_{n-k}} \phi^{\nu_1 \ldots \nu_k}} \tag{5.2.17}$$

From the above, we can now present a theorem:

Theorem: For any k-form $\omega \in \Lambda^k(\mathbb{R}^n)$, $** \omega = (-1)^{n-1}\omega$.

The proof follows from using Equation (5.2.16) and the properties of the Levi-Civita tensor. Notice that the double-Hodge star does *not* depend on k at all!

Note that in all cases, we get the general result:

$$dx^i \wedge *dx^i = \bigwedge_{j=1}^{n} dx^j \equiv d^n x \in \Lambda^n(\mathbb{R}^n) \tag{5.2.18}$$

$\Lambda^n(\mathbb{R}^n)$ is one-dimensional; it's single basis element is the wedge product of all the dx^j. This product is called the **volume form of** $\mathbb{R}^n$. With this in mind, we can reinterpret Equation (5.2.14). Given three vectors in $\mathbb{R}^3$, take their corresponding forms with the tilde and then take the wedge product of the result. The answer you get is the determinant of the matrix formed by these three vectors, multiplied by the volume form of $\mathbb{R}^3$. What is this determinant? It is simply the Jacobian of the transformation from the standard basis to the

basis of the three vectors you chose! In other words, differential forms automatically give us the vector calculus result:

$$d^n\vec{x}' = \mathcal{J} d^n\vec{x} \tag{5.2.19}$$

5.2.4 Evaluating k-Forms

We have done a lot of defining, but not too much calculating. You might be asking at this point: OK, so we have all of this machinery- now what?

The appearance of determinants should give you a clue. Whenever we need to evaluate a k-form, we construct a matrix according to the prescription the form gives us. Finally, we take the determinant of the matrix.

To construct the matrix, we simply chose whatever elements of our vectors get picked out by the simple form.

Examples: evalutate the following forms:

1. $$dx \wedge dy \left[\begin{pmatrix} a \\ b \\ c \end{pmatrix}, \begin{pmatrix} d \\ e \\ f \end{pmatrix} \right]$$

2. $$dx \wedge dz \left[\begin{pmatrix} a \\ b \\ c \end{pmatrix}, \begin{pmatrix} d \\ e \\ f \end{pmatrix} \right]$$

3. $$dx \wedge dy \wedge dz \left[\begin{pmatrix} a \\ b \\ c \end{pmatrix}, \begin{pmatrix} d \\ e \\ f \end{pmatrix}, \begin{pmatrix} g \\ h \\ i \end{pmatrix} \right]$$

Answers:

1. Construct the matrix by picking out the "x" and "y" components of the vectors, and take the determinant:
$$\det \begin{pmatrix} a & d \\ b & e \end{pmatrix}$$

2. This is similar, but now pick out the "x" and "z" components:
$$\det \begin{pmatrix} a & d \\ c & f \end{pmatrix}$$

3. For this case, it is simply what we had before:
$$\det \begin{pmatrix} a & d & g \\ b & e & h \\ c & f & i \end{pmatrix}$$

Hopefully you get the hang of it by now.

5.2.5 Generalized Cross Product

Eariler, I mentioned that we can define a generalized cross product in $\mathbb{R}^n$ by wedging together $(n-1)$ 1-forms. Let's see this explicitly:

In $\mathbb{R}^n$, take $n-1$ vectors, and tilde them so they are now 1-forms: $\{\tilde{v}_1, \ldots, \tilde{v}_{n-1}\}$. Now wedge these forms together to give you an $(n-1)$-form, and take the Hodge star of the product to give us a 1-form. Finally tilde the whole thing to give us an n-vector, and we have defined a general product of vectors:

$$\boxed{\vec{v}_1 \times \cdots \times \vec{v}_{n-1} \equiv [*(\tilde{v}_1 \wedge \widetilde{\cdots \wedge} \, \tilde{v}_{n-1})] \in \mathbb{R}^n} \tag{5.2.20}$$

5.3 Exterior Calculus

A **k-form field** is a k-form whose coefficients depend on the coordinates. This is exactly analogous to a vector field in vector calculus. So, if we have a vector field in $\mathbb{R}^3$:

$$\begin{aligned} \vec{v}(\vec{x}) &= v^1(\vec{x})\vec{e}_1 + v^2(\vec{x})\vec{e}_2 + v^3(\vec{x})\vec{e}_3 \\ \tilde{v}(\vec{x}) &= v_1(\vec{x})dx^1 + v_2(\vec{x})dx^2 + v_3(\vec{x})dx^3 \end{aligned}$$

5.3.1 Exterior Derivative

Now that we have a notion of "function", let's see what we can do in the way of calculus. We can define an operator

$$d : \Lambda^k(\mathbb{R}^n) \to \Lambda^{k+1}(\mathbb{R}^n)$$

with the all-important property:

$$d^2 = dd \equiv 0 \tag{5.3.21}$$

By construction we will let d act on forms in the following way:

$$d(a(x)dx) = da(x) \wedge dx \tag{5.3.22}$$

Let's look at some examples.

$$f(\vec{x}) \in \Lambda^0(\mathbb{R}^3) \to df(\vec{x}) = \left(\frac{\partial f}{\partial x^1}\right) dx^1 + \left(\frac{\partial f}{\partial x^2}\right) dx^2 + \left(\frac{\partial f}{\partial x^3}\right) dx^3 \in \Lambda^1(\mathbb{R}^3)$$

This is just the gradient of a function! How about the curl?

$$
\begin{aligned}
\omega &= a_1 dx^1 + a_2 dx^2 + a_3 dx^3 \rightarrow \\
d\omega &= \left[\left(\frac{\partial a_1}{\partial x^1} \right) dx^1 + \left(\frac{\partial a_1}{\partial x^2} \right) dx^2 + \left(\frac{\partial a_1}{\partial x^3} \right) dx^3 \right] \wedge dx^1 \\
&\quad + \left[\left(\frac{\partial a_2}{\partial x^1} \right) dx^1 + \left(\frac{\partial a_2}{\partial x^2} \right) dx^2 + \left(\frac{\partial a_2}{\partial x^3} \right) dx^3 \right] \wedge dx^2 \\
&\quad + \left[\left(\frac{\partial a_3}{\partial x^1} \right) dx^1 + \left(\frac{\partial a_3}{\partial x^2} \right) dx^2 + \left(\frac{\partial a_3}{\partial x^3} \right) dx^3 \right] \wedge dx^3 \\
&= \left(\frac{\partial a_2}{\partial x^1} - \frac{\partial a_1}{\partial x^2} \right) dx^1 \wedge dx^2 + \left(\frac{\partial a_3}{\partial x^1} - \frac{\partial a_1}{\partial x^3} \right) dx^1 \wedge dx^3 + \left(\frac{\partial a_3}{\partial x^2} - \frac{\partial a_2}{\partial x^3} \right) dx^2 \wedge dx^3
\end{aligned}
$$

Notice that terms like $\left(\frac{\partial a_1}{\partial x^1}\right) dx^1 \wedge dx^1$ vanish immediately because of the wedge product. Again we see how the beauty of differential forms notation pays off!

5.3.2 Formulas from Vector Calculus

At this point we can immediately write down some expressions. Some of them I have shown. Can you prove the rest of them? In all cases, f is any (differentiable) function, and $\vec{v}$ is a vector field.

$$
\begin{aligned}
\widetilde{df} &= \nabla f && (5.3.23) \\
\widetilde{*d\tilde{v}} &= \nabla \times \vec{v} \quad (\text{in } \mathbb{R}^3) && (5.3.24) \\
d\tilde{v} &= \nabla \cdot \vec{v} && (5.3.25) \\
*d*df &= \nabla^2 f && (5.3.26)
\end{aligned}
$$

Notice that even though $dd = 0$, $d * d \neq 0$ in general!

5.3.3 Orthonormal Coordinates

Differential forms allow us to do calculations without ever refering to a coordinate system. However, sooner or later we will want to get our answers back into a coordinate frame. This could be tricky.

The key point to remember is that each of these dx^i is an *orthonormal* coordinate[2]. Therefore we must make sure to translate them into orthonormal coordinates in our coordinate frame; otherwise, the Hodge star will not work.

Let's consider a simple example by looking at the two-dimensional Laplacian. From above, we know that $\nabla^2 = *d*d$. Cartesian coorindates $(x^1 = x, x^2 = y)$ are orthonormal, so we can easily plug into our formula:

[2]That means that $< dx^i, \vec{e}_j >= \delta^i_j$ and $< \vec{e}_i, \vec{e}_j >= \delta_{ij}$.

$$
\begin{aligned}
df &= \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \\
*df &= \frac{\partial f}{\partial x}dy - \frac{\partial f}{\partial y}dx \\
d*df &= \frac{\partial^2 f}{\partial x^2}dx\wedge dy - \frac{\partial^2 f}{\partial y^2}dy\wedge dx \\
&= \left[\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right] dx\wedge dy \\
*d*df &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}
\end{aligned}
$$

All is well with the world. But what if I wanted to do my work in polar coordinates? Then my coordinate system is $(x^1 = r, x^2 = \theta)$, but the *orthonormal* polar coordinate basis vectors are $\vec{\mathbf{e}}_1 = \hat{\mathbf{r}}, \vec{\mathbf{e}}_2 = \frac{\hat{\theta}}{r}$, and the dual basis is $(dr, rd\theta)$ (see previous footnote). So:

$$
\begin{aligned}
df &= \frac{\partial f}{\partial r}dr + \frac{\partial f}{\partial \theta}d\theta\left(\frac{r}{r}\right) = \frac{\partial f}{\partial r}dr + \frac{1}{r}\frac{\partial f}{\partial \theta}(rd\theta) \\
*df &= \frac{\partial f}{\partial r}(rd\theta) - \frac{1}{r}\frac{\partial f}{\partial \theta}dr \\
d*df &= \left\{\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right)dr\wedge d\theta - \frac{\partial}{\partial \theta}\left(\frac{1}{r}\frac{\partial f}{\partial \theta}\right)d\theta\wedge dr\right\}\left(\frac{r}{r}\right) \\
&= \left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2}\right] dr\wedge(rd\theta) \\
*d*df &= \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} = \nabla^2 f \qquad \text{Wow!}
\end{aligned}
$$

where I have introduced $(\frac{r}{r})$ in the first and third line in order to keep the coordinates orthonormal before using the Hodge star. So, as long as we are careful to use orthonormal coordinates when using the Hodge star, our formulas work for *any* coordinate system! This is a vital reason why differential forms are so useful when working with general relativity, where you want to prove things about the physics independent of your reference frame or coordinate basis.

5.4 Integration

We have discussed how to differentiate k-forms; how about integrating them? First I will discuss how to formally evaluate forms, and then I will present a formula for integrating them. With this definition in mind, we will be able to derive Stokes' Theorem- possibly the most important theorem differential forms will provide us with.

5.4.1 Evaluating k-form Fields

First, a definition:

Definition: An **Oriented k-Parallelogram**, denoted by $\pm P^0_{\vec{x}}\{\vec{v}_i\}$, $(i = 1, \ldots, k)$, is a k-parallelogram spanned by the k n-vectors $\{\vec{v}_i\}$ with basepoint $\vec{x}$. $P^0_{\vec{x}}\{\vec{v}_i\}$ is antisymmetric under $\vec{v}_i$ exchange.

Oriented k-Parallelograms allow you to think geometrically about evaluating k-forms, but as far as paperwork goes, they are just a way to keep track of what vectors are what. We use them explicitly to evaluate k-form fields at a point. Let's see how that works:

Let's evaluate the 2-form field $\phi = \cos(xz)dx \wedge dy \in \Lambda^2(\mathbb{R}^3)$. We'll evaluate it on the oriented 2-parallelogram spanned by the 3-vectors $\vec{v}_1 = (1, 0, 1)$ and $\vec{v}_2 = (2, 2, 3)$. We'll evaluate it at two points: $(1, 2, \pi)$ and $(\frac{1}{2}, 2, \pi)$:

1.
$$\phi\left[P^0_{\begin{pmatrix}1\\2\\\pi\end{pmatrix}}\left\{\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}2\\2\\3\end{pmatrix}\right\}\right] = \cos(1\cdot\pi)\det\begin{pmatrix}1 & 2\\0 & 2\end{pmatrix} = -2$$

2.
$$\phi\left[P^0_{\begin{pmatrix}\frac{1}{2}\\2\\\pi\end{pmatrix}}\left\{\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}2\\2\\3\end{pmatrix}\right\}\right] = \cos(\frac{1}{2}\cdot\pi)\det\begin{pmatrix}1 & 2\\0 & 2\end{pmatrix} = 0$$

5.4.2 Integrating k-form Fields

Now that we have a formal way to evaluate k-form fields, we can talk about integrating them. We *define* what the integral of a k-form field is:

Definition: Let $\phi \in \Lambda^k(\mathbb{R}^n)$ be a k-form field, $A \subset \mathbb{R}^k$ be a pavable set, and $\gamma : A \to \mathbb{R}^n$ be a (vector-valued) differentiable map. Then we define the integral of ϕ over $\gamma(A)$ as:

$$\boxed{\int_{\gamma(A)} \phi \equiv \int_A \phi\left[P^0_{\gamma(\vec{u})}\left\{\left.\frac{\partial\vec{\gamma}}{\partial x^1}\right|_{\vec{u}}, \left.\frac{\partial\vec{\gamma}}{\partial x^2}\right|_{\vec{u}}, \ldots, \left.\frac{\partial\vec{\gamma}}{\partial x^k}\right|_{\vec{u}}\right\}\right] d^k\vec{u}} \tag{5.4.27}$$

where $d^k\vec{u}$ is the k-volume form.

Like in vector calculus, we can define the integral formally in terms of Riemann sums - yuck! Let me just say that this is an equivalent definition; the more curious students can go prove it.

Let's do two examples to get the hang of it:

1. Consider $\phi \in \Lambda^1(\mathbb{R}^2)$ and integrate over the map:

$$\gamma(u) = \begin{pmatrix} R\cos u \\ R\sin u \end{pmatrix}$$

over the region $A = [0, \alpha]$, $(\alpha > 0)$. Let $\phi = xdy - ydx$:

$$\begin{aligned}
\int_{\gamma(A)} \phi &= \int_{[0,\alpha]} (xdy - ydx) \left[P^0 \begin{pmatrix} R\cos u \\ R\sin u \end{pmatrix} \begin{pmatrix} -R\sin u \\ R\cos u \end{pmatrix} \right] du \\
&= \int_0^\alpha du[(R\cos u)(R\cos u) - (R\sin u)(-R\sin u)] \\
&= \int_0^\alpha du R^2[\cos^2 u + \sin^2 u] \\
&= \alpha R^2
\end{aligned}$$

2. Consider $\phi = dx \wedge dy + ydx \wedge dz \in \Lambda^2(\mathbb{R}^3)$, and the map:

$$\gamma \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} s+t \\ s^2 \\ t^2 \end{pmatrix}$$

over the region $C = \{ \begin{pmatrix} s \\ t \end{pmatrix} | 0 \le s \le 1, 0 \le t \le 1 \}$:

$$\begin{aligned}
\int_{\gamma(C)} \phi &= \int_0^1 \int_0^1 (dx \wedge dy + ydx \wedge dz) \left[P^0 \begin{pmatrix} s+t \\ s^2 \\ t^2 \end{pmatrix} \left\{ \begin{pmatrix} 1 \\ 2s \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2t \end{pmatrix} \right\} \right] dsdt \\
&= \int_0^1 \int_0^1 \left[\det \begin{pmatrix} 1 & 1 \\ 2s & 0 \end{pmatrix} + s^2 \det \begin{pmatrix} 1 & 1 \\ 0 & 2t \end{pmatrix} \right] dsdt \\
&= \int_0^1 \int_0^1 (-2s + 2s^2 t) dsdt \\
&= \int_0^1 ds(-2s + s^2) = -\frac{2}{3}
\end{aligned}$$

5.4.3 Stokes' Theorem

Now we can move on to present one of if not the most important theorems that differential forms has to offer- Stokes' Theorem. I will present it correctly; do not be overly concerned with all the hypotheses; it suffices that the area you are integrating over has to be appropriately "nice".

Theorem: Let X be a compact, piece-with-boundary of a (k+1)-dimensional oriented manifold M$\subset \mathbb{R}^n$. Give the boundary of X (denoted by ∂X) the proper orientation, and consider a k-form field $\phi \in \Lambda^k(\mathbb{R}^n)$ defined on a neighborhood of X. Then:

$$\boxed{\int_{\partial X} \phi = \int_X d\phi} \tag{5.4.28}$$

What is this theorem saying? My old calculus professor used to call it the "Jumping-d theorem", since the "d" jumps from the manifold to the form. In words, this theorem says that the integral of a form over the boundary of a sufficiently nice manifold is the same thing as the integral of the derivative of the form over the whole mainfold itself.

You have used this theorem many times before. Let's rewrite it in more familiar notation, for the case of $\mathbb{R}^3$:

k	$d\phi$	X	$\int_X d\phi = \int_{\partial X} \phi$	Theorem Name
0	$\nabla f \cdot d\vec{x}$	Path from $\vec{a}$ to $\vec{b}$	$\int_{\vec{a}}^{\vec{b}} \nabla f \cdot d\vec{x} = f(\vec{b}) - f(\vec{a})$	FTOC
1	$(\nabla \times \vec{f}) \cdot d\vec{S}$	Surface (Ω)	$\int_\Omega (\nabla \times \vec{f}) \cdot d\vec{S} = \oint_{\partial\Omega} \vec{f} \cdot d\vec{x}$	Stokes Theorem
2	$(\nabla \cdot \vec{f}) d^3x$	Volume (V)	$\int_V (\nabla \cdot \vec{f}) d3x = \oint_{\partial V} \vec{f} \cdot d\vec{S}$	Gauss Theorem

Here, I am using vector notation (even though technically I am supposed to be working with forms) and for the case of $\mathbb{R}^3$, I've taken advantage of the following notations:

$$\begin{aligned} d\vec{x} &= (dx, dy, dz) \\ d\vec{S} &= (dy \wedge dz, dz \wedge dx, dx \wedge dy) \\ d^3x &= dx \wedge dy \wedge dz \end{aligned}$$

As you can see, all of the theorems of vector calculus in three dimensions are reproduced as specific cases of this generalized Stokes Theorem. However, in the forms notation, we are not limited to three dimensions!

Before leaving differential forms behind, I should mention one important point that I have gone out of my way to avoid: the issue of orientation. For the sake of this introduction, let me just say that all the manifolds we consider must be "orientable with acceptable boundary". What does this mean? It means that the manifold must have a sense of "up" (no Möbius strips). "Acceptable boundary" basically means that we can, in a consistent and smooth way, "straighten" the boundary (no fractals). These are technical issues that I have purposely left out, but they are important if you want to be consistent.

5.5 Forms and Electrodynamics

As a finale for differential forms, I thought it would be nice to summarize briefly how one uses forms in theories such as E&M. Recall that in covariant electrodynamics, we have an antisymmetric, rank-2 4-tensor known as the "Field Strength" tensor:

$$\mathbf{F} \equiv F_{\mu\nu} dx^\mu \otimes dx^\nu \tag{5.5.29}$$

As we have seen, it is always possible to use 2-forms instead of (antisymmetric) tensors, and we can rewrite the above tensor as a differential form in $\Lambda^2(M_4)$[3]:

$$F \equiv \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu \in \Lambda^2(M_4) \tag{5.5.30}$$

Notice that the antisymmetry of $F_{\mu\nu}$ is immediate in form notation. Also, this is a 2-form in four dimensions, which means that its dual is also a 2-form. Now using Equation (5.5.30), I can write the Lorentz Force Law for a particle with charge e moving with a velocity vector $\vec{u}$ as:

$$\dot{\vec{p}} = eF(\vec{u}) \in \Lambda^1(M_4) \tag{5.5.31}$$

Let's consider an example. Suppose we have a magnetic field in the $\hat{\mathbf{x}}$ direction, so $F = B_x dy \wedge dz$. Then the force on the particle is:

$$\begin{aligned} \dot{\vec{p}} = eF(\vec{u}) &= eB_x \langle dy \wedge dz, \vec{u} \rangle \\ &= eB_x [dy \langle dz, \vec{u} \rangle - dz \langle dy, \vec{u} \rangle] \\ &= eB_x [u_z dy - u_y dz] \Rightarrow \dot{\vec{p}} = eB_x (u_z \hat{\mathbf{y}} - u_y \hat{\mathbf{z}}) \end{aligned}$$

This is exactly what you would have gotten if you had used your vector-form of Lorentz's Law, with cross products.

Another vector that is important in E&M is the 4-current (J). We will again think of J as a 1-form in M_4; therefore, it's dual is a 3-form.

Armed with all we need, we can write down Maxwell's Equations:

$$dF = 0 \tag{5.5.32}$$

$$d * F = 4\pi * J \tag{5.5.33}$$

To interpret these equations, one can think of F ($*F$) as an object that represents "tubes of force" flowing through space-time.

We can take advantage of Stokes Theorem (just like in regular vector calculus) to put these equations in integral form:

[3] M_4 is Minkowski space-time. In general, I can go to any space I want; for now, I'm sticking to flat space (i.e.: ignoring gravity). But notice that this analysis is totally general, and gravity can be easily included at this stage.

$$\int_\Sigma dF = \oint_{\partial\Sigma} F = 0 \tag{5.5.34}$$

$$\int_\Sigma d*F = \oint_{\partial\Sigma} *F = 4\pi(\text{charge}) \tag{5.5.35}$$

The first of these equations says that the total flux of F through a closed region of space-time is zero; the second equation says that the total flux of $*F$ through a closed region of space-time is proportional to the amount of charge in that region. Notice that this description never mentions coordinate systems: once again, differential forms has allowed us to describe physics and geometry without ever referring to a coordinate system! Notice the similarity of this equation with the Gauss-Bonnet Theorem – in gauge theories, you may think of the field strength as a curvature to some space!

The final step in studying electrodynamics is to notice that Equation (5.5.32) suggests something. Recall that if a differential form was the exterior derivative of some other form, then it's exterior derivative was necessarely zero, via Equation (5.3.21). Is the converse true? Namely, if you have a form whose exterior derivative is zero, can it be written as the exterior derivative of some other form?

Your first impulse might be to say "yes": surely if the form is an exterior derivative, your condition is satisfied. But it turns out that the question is a little more subtle than that. For I can give you a form which is not an exterior derivative of another form, and yet still has vanishing exterior derivative! This is a famous case of "necessity" versus "sufficiency" in mathematics. That a form is an exterior derivative is sufficient for its exterior derivative to vanish, but not necessary.

The key point is that this property depends not on the differential forms themselves, but on the *global properties of the space* they live in! It turns out that M_4 does indeed have the property of necessity; specifically, it is simply connected. Therefore it is safe to assume a la Equation (5.5.32) that we can write:

$$F = dA \tag{5.5.36}$$

for some 1-form A. So indeed, electrodynamics in flat space can be described by a 4-vector potential. But do be careful not to jump to any conclusions before you know the properties of the universe you are describing!

Finally, we can consider what happens if we let $A \to A + d\lambda$, where λ is any (reasonably well-behaved) function. Then by just plugging into Equation (5.5.36), with the help of Equation (5.3.21), we see that F remains the same. Hense, the 4-vector potential is only defined up to a 4-gradient; this is exactly what gauge invariance tells us should happen!

This gives a very powerful geometric intuition for our field-strength, and therefore for electricity and magnetism. You can do similar analyses for any gauge field, and by adjusting the manifold, you can alter your theory to include gravity, extra dimensions, strings, or whatever you want! This is one of the biggest reasons why differential forms are so useful to physicists.

Chapter 6

Complex Analysis

In this chapter, I would like to present some key results that are used in analytic work. There is of course much more material that what I present here, but hopefully this small review will be useful.

Complex analysis is concerned with the behavior of functions of complex variables. It turns out that these functions have a much different behavior than functions of real variables. The techniques of complex analysis, pioneered in the eighteenth and nineteenth centuries, have become very important in almost all branches of applied mathematics. I am assuming that you are familiar with the basics of complex numbers ($z = x + iy \in \mathbb{C}$) and will immediately move on.

In my experience, applications of complex numbers fall into two general catagories: Cauchy's Integral Theorem and Conformal Mapping. This review will consider the calculus of complex variables. I will not mention too much about conformal mapping, except to define it and give an example or two.

Complex numbers are written as z. Then a function of a complex variable has the form:

$$f(z) = f(x + iy) = u(x, y) + iv(x, y)$$

where u, v are both real valued functions of the real variables x, y. You can also write a complex number in polar coordinates $z = re^{i\theta}$, where $r \equiv |z|$ and θ is called the **argment** of z, sometimes denoted as $\arg z$.

6.1 Analytic Functions

- *Definition:* A complex function $f(z)$ is **analytic at a point** z_0 if it has a Taylor expansion that converges in a neighborhood of that point:

$$f(z) = \sum_{i=0}^{\infty} a_i (z - z_0)^i$$

If $f(z)$ is analytic at all points inside some domain $D \subset \mathbb{C}$ then it is said to be analytic in D. If it is analytic for all finite complex numbers, it is said to be **entire**. If it is analytic for all complex numbers, including infinity, then $f(z)$ =constant!

- *Theorem:* The real and imaginary parts of an analytic function $f(z) = u(x, y)+iv(x, y)$ satisfy the **Cauchy-Riemann (CR) Equations**:

$$u_x = v_y \tag{6.1.1}$$
$$v_x = -u_y \tag{6.1.2}$$

  where u_x means the partial derivative of u with respect to x, etc. From the CR equations, it can be seen that both u and v satisfy Laplace's equation:

$$u_{xx} + u_{yy} = v_{xx} + v_{yy} = 0$$

  Functions that satisfy Laplace's equation are called **harmonic functions**. If $u(x, y)$ is a harmonic function, then up to an additive constant, there is a unique function $v(x, y)$ that satisfies the CR equations with $u(x, y)$. This function is called the **harmonic conjugate** of $u(x, y)$. Notice that this relation is not reflexive in general; in other words, if v is the harmonic conjugate of u, then it is *not* generally true that u is the harmonic conjugate of v. When it is true, it is possible to show that u and v are constants.

- Let's do an example:

$$f(z) = z^2 = (x^2 - y^2) + i(2xy)$$

  This function is analytic everywhere in the finite complex plane (it is entire): its Taylor series about $z = 0$ is trivial: $a_2 = 1$, all others zero. It also satisfies the CR equations. Note that the function $v(x, y) = 2xy$ is the harmonic conjugate to $u(x, y) = x^2 - y^2$, but as we expect from the above note, the converse is not true (prove it!).

- Let's see how one might construct the harmonic conjugate of a real-valued harmonic function. Consider the function $u(x, y) = y^3-3x^2y$. This function is certainly harmonic (prove it!), so up to an additive constant, there should be a unique function $v(x, y)$ that is the harmonic conjugate of $u(x, y)$. To find it, note first of all that by the first CR equation, $v_y = u_x = -6xy$. Integrating this equation with respect to y gives $v(x, y) = -3xy^2+\phi(x)$ where $\phi(x)$ is some arbitrary function of x alone. Plugging into the second CR equation gives $\phi'(x) = 3x^2$. Integrating this gives $\phi(x) = x^3+C$ with an arbitrary constant C, and therefore $v(x, y) = x^3 - 3xy^2 + C$ is the harmonic conjugate to $u(x, y)$. Therefore $f(x, y) = u(x, y)+iv(x, y)$ must be an analytic function. *Exercise: show that $f(x, y)$ can be rewritten as $f(z) = i(z^3 + C)$, which is certainly analytic.*

- Consider two analytic functions $f_1(z), f_2(z)$ defined on domains D_1, D_2 respectively, where $D_1 \cap D_2 \neq \emptyset$. Also assume that the two functions agree for all points in the intersection. Then $f_2(z)$ is called the **analytic continuation** of $f_1(z)$. When an analytic continuation exists, then it is unique. The proof of this is not hard, but the key is in the analytic. This means that there is only one way to "extend" a function analytically out of its original domain. Notice, however, that if $f_3(z)$ is another analytic continuation of $f_2(z)$ into another domain D_3, it need not be a continuation of $f_1(z)$ into that domain!

- As an example, let's look at our old example of $f(z) = z^2$. This function agrees with the function $f(x) = x^2$ in the domain of real numbers ($y = 0$). It is also analytic, therefore it is the analytic continuation of the real-valued function $f(x) = x^2$. Further, it is unique: there is *no* other function that is analytic and matches with our original function!

- That was an obvious one. What about a more complicated example? Consider the factorial function defined on $\mathbb{N}$: $f(n) = n!$ I will now define a function on the real numbers:

$$\Gamma(x) = \int_0^\infty dss^{x-1}e^{-s} \tag{6.1.3}$$

  Integrating by parts shows that when x is a natural number, $\Gamma(x) = (x-1)!$ Furthermore, $\Gamma(x)$ is in fact an analytic function of x, so long as x is not a negative integer or zero. Hense, $\Gamma(x)$ is the (unique) analytic continuation of the factorial function to the real numbers (without the negative integers or zero). Indeed, we can go further: by replacing $x \to z$ (as long as z is not a negative integer or zero – *why not?*) this extends to all of $\mathbb{C}$.

6.2 Singularities and Residues

- *Definition:* A **Laurent Expansion** of a function $f(z)$ about a point z_0 is an expansion in both positive and negative powers:

$$f(z) = \sum_{n=-\infty}^{\infty} c_n(z-z_0)^n = \sum_{n=1}^{\infty} \frac{b_{-n}}{(z-z_0)^n} + \sum_{n=0}^{\infty} a_n(z-z_0)^n \tag{6.2.4}$$

- As you can see, a Laurent expansion is a generalization of a Taylor expansion. Notice that the function is not defined at the point z_0 unless all of the $b_n = 0$. If this is the case, $f(z)$ is analytic at z_0. A Taylor expansion typically exists in an open disk about the point z_0. The generalization to the Laurent expansion converges inside an *annulus* $R_1 < |z - z_0| < R_2$.

- A **singularity** of a function is a point where the function is not defined. If z_0 is a singularity of $f(z)$ and you can define a neighborhood (open set) about that point $U(z_0)$ such that $f(z)$ is analytic in the domain $U(z_0) - \{z_0\}$ (so $R_1 \to 0$ above), then z_0 is called an **isolated singularity**. There are three general types of isolated singularities:

  1. **Removable singularities** are the most mild form of singularity, where all the $b_{-n} = 0$. An example of this is the function $\frac{\sin x}{x}$. This function is not defined at the origin, but it does have a well-defined limit there. You can see this by performing a Taylor expansion of the sine function and dividing out by x, leaving a finite value plus terms that are $\mathcal{O}(x)$ and hense vanish in the limit. Removable singularities are trivial, and I will not talk anymore about them.

2. If a function has a Laurent expansion about z_0 but there exists an integer $m > 0$ such that all the $b_{-n} = 0$ for every $n > m$ (but with $b_{-m} \neq 0$), then the function is said to have a **Pole of order m** at z_0. In the special case of $m = 1$, the function is said to have a **simple pole**. An example of a function with a simple pole is $\frac{\cos x}{x}$ at the origin, where it behaves as $1/x$. Poles are very useful and important in complex analysis, as we will see.
3. If a function has a Laurent expansion about z_0 but there is *not* an integer $m > 0$ such that all the $b_{-n} = 0$ for every $n > m$ (so the denominator terms do not stop), then the function is said to have an **Essential Singularity** at z_0. An example of this is the function $f(z) = e^{1/z}$ at the origin. This singularity is a true nightmare, and there is very little to be done for it!

- If a function $f(z)$ has a pole of order m at z_0 (not necessarely a simple pole), then the quantity b_{-1} is called the **Residue** of the pole, also denoted by $\text{Res}_{z=z_0}\{f(z)\}$. This quantity plays a vital role in the calculus of complex-valued functions. A function that is analytic in a domain, up to maybe a countable number of poles (so it still has a Laurent expansion), is said to be **meromorphic** in that domain.

- Residues are so important, I want to list various methods of finding them. I will state several methods here without proof:

  1. If $p(z_0) \neq 0$ and $q(z)$ has a zero of order m at z_0[1] then $\frac{p(z)}{q(z)}$ has a pole of order m at z_0.
  2. If $f(z)$ has a pole of order m at z_0, it can be written as $f(z) = \frac{p(z)}{(z-z_0)^m}$ where $p(z_0) \neq 0$. Then the residue of $f(z)$ at z_0 is $b_{-1} = \frac{p^{(m-1)}(z_0)}{(m-1)!}$
  3. If $q(z)$ has a simple zero at z_0 and $p(z_0) \neq 0$, then the residue of $\frac{p(z)}{q(z)}$ at z_0 is given by $b_{-1} = \frac{p(z_0)}{q'(z_0)}$

- Before leaving this section, there is one more type of singularity that comes up. This is *not* an isolated singularity, and therefore is not defined through its Laurant expansion. Rather this singularity comes from an ambiguity in *multiply defined* functions. For example, consider the complex Logarithmic function:

$$\log z = \log r + i\theta$$

  where I have used the famous result $z = re^{i\theta}$. The imaginary part of this function is a problem, since θ is only defined up to an arbitrary factor of 2π. Therefore, this expression has an infinite number of solutions for a single value of z. In real analysis, we would throw such functions away, but not in complex analysis. The solution is to define a *cut* where we explicitly say what values of θ we allow. For example, we can say that we will only let $-\pi < \theta < +\pi$. This is called the **principle branch** of the logarithmic function, and is sometimes denoted Log z.

[1] *i.e.:* $q(z) = (z - z_0)^m r(z)$, where $r(z_0) \neq 0$.

- **Branch cuts** are lines in the complex plane that tell you where to cut off the definition of your function. For example, the branch cut for the logarithmic function is a line that connects the origin to infinity. For the principle branch, that line extends along the negative real axis. It could have gone in any direction, which would have corresponded to a different branch of the logarithm. In addition, it does not even have to be a straight line. Notice that the function is discontinuous as you cross the branch cut. In the case of the Logarithm: $\log(-x + i\epsilon) - \log(-x - i\epsilon) = 2\pi i$, where x is a positive real number. Branch cuts are discontinuities in the function.

- The branch cut you make is not unique, but what *is* true is that all branch cuts connect two points, and those two points are unique. In the case of the Logarithm, those points were the origin and infinity. Since those points are common to all branch cuts, you cannot ever define the logarithmic function there. Therefore they are singularities. They are called **branch points**.

- Aside from the logarithm, the most common function with branch point singularities is the function $f(z) = x^a$ where $0 < \mathrm{Re}(a) < 1$. This function is multiply defined since $f(z) = f(ze^{i2\pi n/a})$, for integer n. It too has branch points at $z = 0, \infty$.

6.3 Complex Calculus

When doing calculus on a function of complex variables, a whole new world of possibilities opens up! At first, you can think of the complex plane as a special representation of $\mathbb{R}^2$, and you might think that the complex calculus is the same as real vector calculus. But this is not true! There are a variety of theorems that make calculus (especially integral calculus) on the complex plane a rich and fascinating subject, both theoretically as well as practically.

Complex analysis is a deep subject, but since this is supposed to be a practical review for physics grad students, I will skip over all the proofs and just present the most commonly-used results. I strongly encourage you to read up on the theory behind this stuff; it's really an amazing subject in and of itself.

- A general function of a complex variable can be written as $f(z, z^*)$. Such a function is analytic if and only if

$$\frac{df}{dz^*} \equiv 0 \tag{6.3.5}$$

  so that $f = f(z)$. Such a function is sometimes called **holomorphic**, but this is just a synonym for analytic.

- Just like in vector calculus, you can define contour integrals in the complex plane. if C is a contour described by a path in the complex plane $z(t)$ where $t \in [a, b]$ is a real parameter, we can define the contour integral of an analytic function in terms of a one-dimensional Riemann integral:

$$\int_C f(z)dz \equiv \int_a^b f(z(t))z'(t)dt \tag{6.3.6}$$

- *Cauchy-Goursat Theorem*: Let C be a closed contour in the complex plane and $f(z)$ an analytic function on the domain interior to and on C. Then:

$$\oint_C f(z)dz \equiv 0 \tag{6.3.7}$$

  This is a generalization of the fundamental theorem of calculus. A powerful consequence of this theorem is that a contour integral is invariant to deformations of the contour, as long as the deformation does not pass through a singularity of $f(z)$.

- *Cauchy's Integral Formula:* Let $f(z)$ be analytic everywhere within and on a simple, closed, positively oriented contour[2] C, and let z_0 be any point on the interior to C. Then:

$$f(z_0) = \frac{1}{2\pi i}\oint_C \frac{f(z)dz}{z - z_0} \tag{6.3.8}$$

- Using the Cauchy integral formula and some complex analysis, we can actually derive a very powerful formula for the n-th derivative of an analytic function:

$$f^{(n)}(z) = \frac{n!}{2\pi i}\oint_C \frac{f(z)dz}{(z - z_0)^{n+1}} \tag{6.3.9}$$

  From this, we come to a rather amazing result: if a function is analytic (which is the same as saying $\frac{df}{dz}$ exists), then *all* of its derivatives exist and are given by the above integral! This is very different from real analysis, where a function that has a first derivative need not have higher derivatives. For example, $x^{3/2}$ has a first derivative at $x = 0$ but it does not have a second derivative there. This kind of thing can never happen in complex analysis. *Exercise: why doesn't $z^{3/2}$ contradict this result?*

- Notice immediately the marvelous corollary to the above formulas: for all $n \in \mathbb{Z}$,

$$\oint_C \frac{dz}{(z - z_0)^n} = 2\pi i\delta_{n1} \tag{6.3.10}$$

  as long as C encloses z_0. This follows for $n > 1$ since the derivative of a constant is zero, and it follows for $n \leq 0$ since such terms are analytic inside and on the contour and hence the integral vanishes by the Cauchy-Goursat theorem. Finally, the result for $n = 1$ follows directly from the Cauchy integral formula, or by direct integration.

[2] A *positively oriented contour* means you travel around the contour in a counter-clockwise direction; a *negatively oriented contour* travels around the contour in a clockwise direction.

- Using this result, we come to a remarkable and vital theorem used throughout all areas of applied and abstract mathematics. Consider a general, meromorphic function on the complex plane. Such a function has a Laurent expansion. Now let's do a closed contour integral, where we make sure the contour does not pass through any singularities of $f(z)$. Some of the isolated singularities of $f(z)$ are on the interior of C (call this set $Z = \{z_0, z_1, \ldots, z_n\}$) and some are on the exterior, where Z can be a finite or infinite set. Then we have the amazing and most important result:

$$\boxed{\oint_C f(z)dz = 2\pi i \sum_{z_k \in Z} \text{Res}_{z=z_k}\{f(z)\}} \tag{6.3.11}$$

  This result is usually called the **Residue Theorem**. This is why it's so important to know how to compute residues of meromorphic functions.

- To see how vitally useful the residue theorem is, consider the following real integral:

$$I = \int_{-\infty}^{+\infty} \frac{\cos x dx}{(x^2 + a^2)^3} \tag{6.3.12}$$

  Try doing this integral directly! Before you shoot yourself, let's see how to do it in just a few steps. First of all, we analytically continue the integrand to the complex plane by replacing x by z (check that this is, indeed, the analytic continuation of the integrand). Now in order to make sense of the cosine of a complex variable, we use the definition $\cos z \equiv (e^{iz} + e^{-iz})/2$. We split the integral into two terms $I = I_+ + I_-$, where

$$I_\pm = \frac{1}{2}\int_{-R}^{+R} \frac{e^{\pm iz}dz}{(z^2 + a^2)^3}$$

  where we will take the limit $R \to \infty$ in the end. Let's just consider I_+ for now. We can rewrite the denominator as $(z + ia)^3(z - ia)^3$, so by rule 2 in the residues section, the integrand has a third-order pole at both of these points and the residue at these points are given by

$$\frac{1}{2!}\frac{d^2}{dz^2}\left.\frac{e^{iz}}{(z \pm ia)^3}\right|_{z=\pm ia} = \mp i\frac{e^{\mp a}}{16a^5}(3 + a(a \pm 3))$$

  So far, we cannot apply the residue theorem since the contour is not closed. However, notice that when we go into the complex plane, the exponential in the numerator picks up a factor of $e^{-\text{Im}(z)}$. So in particular, if we add to I_+ a term that is the contour integral over the semicircle of radius R going through the upper-half of the complex plane ($\text{Im}(z) > 0$), such a term will go like e^{-R} and hence vanish as $R \to \infty$. If we include this term, we don't change the value of I_+ but we do close the contour, and hence can apply the residue theorem. Since we only surround the pole at $z = +ia$ we only keep that residue, and we get:

$$\lim_{R\to\infty} I_+ = \frac{\pi e^{-a}}{16a^5}(3 + a(a+3))$$

To compute I_-, the story is very similar, but there are a few differences to take into account. First of all, with the exponential being negative, this will change our formula for the residue slightly. Also, we will need to close the integral in the lower half of the complex plane rather than the upper half, in order to make sure the added contribution vanishes. Finally, since the new contour is oriented clockwise rather than counterclockwise, there is an extra negative sign. Putting this all together gives us that $I_- = I_+$ so the final integral is

$$\int_{-\infty}^{+\infty} \frac{\cos x dx}{(x^2+a^2)^3} = \frac{\pi e^{-a}}{8a^5}(3 + a(a+3))$$

Try doing that integral without the help of the residue theorem – I wouldn't even know where to begin!!

- The residue theorem gives us a way to integrate (or differentiate) complicated meromorphic functions which are analytic on the contour of integration. But what about functions that have singularities *on* the contour? Such integrals by themselves are not well-defined, but we can still make sense of them by means of a *regulator*. The final answer we get will not be unique since it will depend on the regularization scheme used. But once we specify this scheme, we can proceed to make well-defined statements.

- It turns out that there is a particular regularization scheme that is a very natural choice, called the **Principle Value**. As an example, consider the integral

$$\int_{-\infty}^{\infty} x dx$$

This integral is indeterminant until we chose precisely how to take the limit to infinity. The Principle Value of this integral is defined as the symmetric limit:

$$\mathcal{P}\int_{-\infty}^{\infty} x dx = \lim_{R\to\infty}\int_{-R}^{R} x dx = 0$$

Notice that this choice of regulator is by no means unique. For example, we could have defined the integral as

$$\lim_{R\to\infty}\int_{-R-\frac{a}{R}}^{R+\frac{b}{R}} x dx = b - a$$

- There is nothing special about any of these choices. However, there is good reason to use the principle value. Let $f(x)$ be an analytic real-valued function. Then we have the following vital result:

$$\boxed{\lim_{\epsilon\to 0}\int_{-\infty}^{\infty}\frac{f(x)}{x-x_0\pm i\epsilon}=\mathcal{P}\int_{-\infty}^{\infty}\frac{f(x)}{x-x_0}\mp i\pi f(x_0)} \tag{6.3.13}$$

where the principle value on the right hand side should be understood as

$$\mathcal{P}\int_{-\infty}^{\infty}\frac{f(x)}{x-x_0}\equiv\lim_{R\to\infty}\lim_{\rho\to 0}\left(\int_{-R}^{-\rho}\frac{f(x)}{x-x_0}+\int_{\rho}^{R}\frac{f(x)}{x-x_0}\right) \tag{6.3.14}$$

This theorem goes by many names, but I like to call it the **Principle Value Theorem**, although you might find it under something else in your favorite textbook.

- I will not prove the Principle Value theorem (it's not very difficult to prove – try it yourself), but I will use it in a simple example. Let's calculate the integral

$$\int_{-\infty}^{\infty}\frac{\cos x}{x+i\epsilon}dx$$

As $\epsilon\to 0$, this integral is ill-definined, due to its pole on the real axis. Nonetheless, the Principle Value Theorem gives us a way to solve the problem:

$$\int_{-\infty}^{\infty}\frac{\cos x}{x+i\epsilon}=\mathcal{P}\int_{-\infty}^{\infty}\frac{\cos x}{x}-i\pi\cos(0)$$

If the universe is fair, the first integral on the right-hand side should vanish due to the symmetry of the integrand, but that will only happen if the divergences are under control. Notice that near the origin, where the integrand has a simple pole, the principle value tells us how to cancel the divergence there:

$$\lim_{R\to\infty}\lim_{\rho\to 0}\left(\int_{-R}^{-\rho}\frac{dx}{x}+\int_{\rho}^{R}\frac{dx}{x}\right)=\lim_{R\to\infty}\lim_{\rho\to 0}\left(-\int_{\rho}^{R}\frac{du}{u}+\int_{\rho}^{R}\frac{dx}{x}\right)=0$$

where I set $u=-x$ in the first integral. Thus the principle value has tamed the divergence. So this integral does exist, and it equals $-i\pi$.

- Notice in the above example, if we had changed the sign of ϵ (placing the pole above the real axis rather than below it) we would have found a different answer $+i\pi$, even though we take the limit $\epsilon\to 0$, and you might have thought that we should get the same answer in the end. This is *not* a coincidence! It is due to the fact that the integral (without the ϵ) is not well defined. So this begs the question: which sign should we use? This is where mathematics ends and physics begins. You will find that the choice of sign typically comes from boundary conditions, and is sensitive to the particular problem at hand.

- As a final example of how complex calculus is so very useful, let's consider how to integrate a function with a branch-point singularity. This is a little different since a branch point is not an isolated singularity, so the residue theorem doesn't help by itself. But we can still do pretty well.

  When there is a branch-point singularity, you have to specify a branch cut that connects the singularities. So for example, let's consider the following integral:

$$\int_0^\infty \frac{x^{-a}dx}{x+1}$$

  where $0 < a < 1$. This integrand is multiply defined, so we must choose a branch cut to define it. This is a line connecting the two branch points, in this case the origin and infinity. This comes from the fact that we can write the numerator as $e^{-a\log x}$. Also, notice the simple pole at $x = -1$, so we shouldn't choose our branch cut to overlap with that pole. We will make our cut along the positive real axis. This integral is the same as

$$\int_{\mathbb{R}_+} \frac{z^{-a}dz}{z+1}$$

  where the integral is along the positive real axis in the complex plane.

  We will extend this contour: first, the contour goes from $\rho + i\epsilon$ ($\rho < 1$) to $R + i\epsilon$ ($R > 1$) on the real axis (L_+). Then the contour goes counterclockwise around a circle of radius R until we get back to just below the real axis $R - i\epsilon$ (C_R). Then we travel back down the real axis until we get to $\rho - i\epsilon$ (L_-). Finally, we go clockwise around a circle of radius ρ until we get back to the starting point at $\rho + i\epsilon$. The ϵ is there to avoid the branch cut - we can let $\epsilon \to 0$ as long as we keep in mind how to define the exponent (see below). With our choice of ρ and R this contour encircles the pole, so the residue theorem applies. So we have:

$$\left(\int_{L_+} + \int_{C_R} + \int_{L_-} + \int_{C_\rho}\right) \frac{z^{-a}dz}{z+1} = 2\pi i e^{-ia\pi}$$

  The two circle contour integrals vanish. To see that, we need the fact that $|z_1 + z_2| \geq ||z_1| - |z_2||$. Then for the small circle:

$$\left|\int_{C_\rho} \frac{z^{-a}dz}{z+1}\right| \leq \frac{\rho^{-a}(2\pi\rho)}{||z|-1|} = \frac{2\pi}{1-\rho}\rho^{1-a} \xrightarrow{\rho\to 0} 0$$

  and a similar analysis can be done with the large circle. The interesting part comes from the remaining two integrals. The integral across L_+ is above the branch cut, so $\log z = \log r + i0$ as $\epsilon \to 0$, while the integral along L_- is below the branch cut, so $\log z = \log r + i2\pi$ as $\epsilon \to 0$. Plugging this in:

$$\int_{L_+} \frac{z^{-a}dz}{z+1} = \int_\rho^R \frac{r^{-a}dr}{r+1} \qquad \int_{L_-} \frac{z^{-a}dz}{z+1} = \int_R^\rho \frac{r^{-a}e^{2ia\pi}dr}{r+1}$$

Putting all of this together and taking the limits $\rho \to 0$ and $R \to \infty$ gives

$$\int_0^\infty \frac{r^{-a}dr}{r+1} = \frac{2\pi i e^{-ia\pi}}{1-e^{2ia\pi}} = \frac{\pi}{\sin a\pi}$$

6.4 Conformal Mapping

- Conformal mapping is one of those tools that, in my experience, takes a great deal of time and practice to get comfortable with. It is one of those things in mathematics which is almost as much art as science. Whole textbooks have been written on this rich and amazing subject; perhaps it is overly ambitious for me to try and write this section. But I've always loved a challenge.

  I cannot begin to write a complete review of this subject – it would take way too long and I am not an expert on the subject anyway. So in this section I only intend to present some of the basic theorems and one or two examples. Maybe one day I'll do more than that, but not today.

- Conformal mapping is a technique that is useful in solving partial differential equations in two dimensions with certain boundary conditions. The applications to such problems range everywhere from fluid dynamics to electrodynamics to financial forecasting! The basic idea is to transform complicated boundary conditions into simpler ones, and thus make solving the problem much easier.

- Simply put, a conformal transformation is a transformation that preserves angles. Technically, consider a contour in the complex plane parametrized by a function $z(t)$, and *a map $f(z)$ that is analytic and has a nonvanishing first derivative;* such maps are called **conformal maps**. Let $w(t) = f(z(t))$ be the image of the contour under the map f. We will call the complex w-plane $\mathbb{C}_w$ and the complex z-plane $\mathbb{C}_z$, so $f : \mathbb{C}_z \to \mathbb{C}_w$.

- So what does this have to do with angles? The tangent vector in $\mathbb{C}_w$ is given by

  $$w'(t) = f'(z(t))z'(t)$$

  From this it follows that

  $$\arg w'(t) = \arg f'(z) + \arg z'(t) \tag{6.4.15}$$

  Now let's imagine two curves described by $z_1(t)$ and $z_2(t)$ that intersect at a point $z_1(t_0) = z_2(t_0) = z_0 \in \mathbb{C}_z$. Then from the above formula, we have the result:

$$\arg w_1'(t) - \arg w_2'(t) = \arg z_1'(t) - \arg z_2'(t) \tag{6.4.16}$$

So indeed we find that conformal transformations do preserve angles between contours.

- Before going further, let me point out a very important fact without proof: let $f(z) = u(x,y) + iv(x,y)$ be an analytic function, so that $u(x,y)$ and $v(x,y)$ are both harmonic, and v is the harmonic conjugate to u. Let $\mathbf{T}_u(x,y) = \nabla u(x,y)$ be the tangent vector of u at the point (x,y), and similarly for $\mathbf{T}_v(x,y)$, and let (x_0,y_0) be a point such that $u(x_0,y_0) = v(x_0,y_0)$. Then $\mathbf{T}_u(x_0,y_0) \cdot \mathbf{T}_v(x_0,y_0) = 0$; that is, harmonic conjugates have orthogonal tangent vectors at points of intersection. I leave the proof of this to you (*Hint: use the CR Equations*). This will be very useful in what follows.

- Another important fact: the definition of a conformal transformation turns out to provide a necessary and sufficient condition for an inverse function to exist, at least locally; this is a result of the infamous "Inverse Function Theorem". Thus, if we find a conformal transformation that simplifies our boundary conditions in $\mathbb{C}_w$, we can always invert the transformation to get the solution on $\mathbb{C}_z$, which is, of course, what we really want.

- The differential equation we are interested in is Laplace's equation. Remember that the real and imaginary parts of an analytic function are harmonic functions; that is, solutions to Laplace's equation in two dimensions. We are interested in two types of boundary conditions for a harmonic function $h(x,y)$ on a boundary described by a curve C in the complex plane $z = x + iy$:

  1. *Dirichlet Boundary Conditions:* $h|_C = h_0$ for any constant h_0.
  2. *Neumann Boundary Conditions:* $\vec{n} \cdot \nabla h|_C = 0$, where $\vec{n}$ is a unit vector normal to the boundary.

  It turns out that for these two types of boundary conditions are preserved under conformal transformations. Any other type of boundary condition will generally not be preserved. for example, $\frac{dh}{dn}\big|_C$ =constant$\neq 0$ is *not* preserved by conformal transformations. Therefore, whenever faced with trying to solve Laplace's equation with either Dirichlet or Neumann boundary conditions in two dimensions, conformal mapping is an available method of finding the solution. As mentioned above, this problem comes up many times and in many places, so this is a very powerful technique.

- As an example, let's consider the problem of trying to find the electrostatic scalar potential inside an electrical wire. The wire has a circular cross section, where half of the cylinder is kept at $V = 0$ and the other half is kept at $V = 1$. The cylindrical symmetry of the problem means the potential cannot depend on z, so this is a two-dimensional problem. If we place the cross section of the wire on a unit circle in the complex plane $\mathbb{C}_z$, we can look for a harmonic function that agrees with the boundary conditions. This, by uniqueness of solutions, is the final answer. This is a Dirichlet boundary condition problem with circular boundary. Such boundaries are complicated, but we can use conformal mapping to turn this problem into an easier one.

Consider the mapping:

$$w = i\frac{1-z}{1+z} \tag{6.4.17}$$

This is an example of a *Möbius Transformation*, also known as a *linear fractional transformation* . It has the effect of mapping the interior of the unit circle in $\mathbb{C}_z$ to the entire upper-half plane of $\mathbb{C}_w$, where the upper half of the circle is mapped to the positive real w axis while the lower half of the circle is mapped to the negative real w axis. I leave it to you to prove that this map is, indeed, a conformal map.

Now the next step is to find a map on $\mathbb{C}_w$ that is harmonic and satisfies the boundary conditions. Consider the function

$$\frac{1}{\pi}\log w = \frac{1}{\pi}\left(\log|w| + i\arg(w)\right) \tag{6.4.18}$$

which is analytic and spans the upper half plane of $\mathbb{C}_w$ for $0 \leq \arg w \leq \pi$. Notice that the imaginary part of this function satisfies the boundary conditions we need. Since it is the imaginary component of an analytic function is harmonic, so we found our solution.

By writing $w = u + iv$, we have $\arg(w) = \tan^{-1}\left(\frac{v}{u}\right)$. By expanding Equation (6.4.17) to compute $u(x, y)$ and $v(x, y)$, we find

$$V(\vec{r}) = \frac{1}{\pi}\tan^{-1}\left(\frac{1-x^2-y^2}{2y}\right) = \frac{1}{\pi}\tan^{-1}\left(\frac{1-r^2}{2r\sin\theta}\right) \tag{6.4.19}$$

where we take the branch of the arctangent function $0 \leq \tan^{-1} t \leq \pi$. It gets better: the harmonic conjugate of the potential is another harmonic function whose tangent vector is always orthogonal to the potential's tangent vector. But the tangent vector to the electrostatic potential is the electric field. And the lines perpendicular to the field are just the field flux lines. From Equation (6.4.18) we know right away that these flux lines are then given by the level curves of the imaginary part of $-(i/\pi)\log w$, or upon using Equation (6.4.17):

$$\mathcal{E}(\vec{r}) = \frac{1}{2\pi}\log\left(\frac{(1-x)^2+y^2}{(1+x)^2+y^2}\right) = \frac{1}{2\pi}\log\left(\frac{1+r^2-2r\cos\theta}{1+r^2+2r\cos\theta}\right)$$

- Perhaps this relatively straightforward example has been able to show just how much art goes into solving problems this way. First, we had to find the conformal transformation that simplifies the problem. Believe it or not, this is not as bad as it sounds – there are big, fat tables online and in textbooks cataloging hundreds of conformal maps. So you can go through those and look for one that helps. The next thing we had to do is find the analytic function whose harmonic component gives us our solution. In this case, that was relatively straightforward. There are a whole bunch of analytic functions that you can try – polynomials and roots, trig functions, logs and exponents,

etc. When you've done problem after problem, you begin to develop an intuition for finding the right function. But it is something of an art, and like all forms of art, it takes practice to get good at it. But the rewards are worth it! Once you get good at finding the conformal map and the analytic function, you have completely solved the problem. You even get the field flux lines for free.