Mathematical Models in Physics: Relativistic Electrodynamics and Differential Forms

James Cook
MA 430
Department of Mathematics at NCSU

Let me try to summarize the physical and mathematical motivations for covering the hodge-podge of topics that appear in these notes.

The physical motivations of the course stem from a single question, "why are Maxwell's equations relativistic ?". In order to answer the question we must first understand it. Addressing this question is largely the goal of Part II of the notes. We devote a chapter each to Newtonian mechanics, Maxwell's equations and special relativity. Our approach is to educate the student to the general idea of the theory, however we will not get terribly deep in problem solving. We simply have not the time for it. This course cuts a wide spectrum of physics, so I don't think it reasonable to ask difficult physical problems. On the other hand I think it is perfectly reasonable to ask you to duplicate a particular physical argument I make in lecture (given fair warning of course).

The mathematical motivations of the course stem from a simple question, "what are differential forms ?". Addressing this question is largely the goal of Part III. To answer this question we must discuss multilinear algebra and tensors. Once the idea of a tensor is settled we introduce the "wedge" product. As a mathematical aside we'll derive some familiar formulas for determinants via the wedge product. Then we consider differential forms, Hodge Duality, and the correspondence to ordinary vector calculus. In the course of these considerations we will rephrase Maxwell's equations in terms of differential forms, this will give us an easy and elegant answer to our primary physical question. Lastly we see how the generalized Stokes' theorem for integrating differential forms unifies ordinary integration involving vector fields. I intend for Part III to occupy about half of our time, but be patient please its the last half.

Part I is largely a review of prerequisite material and we will not cover it in its entirety. We will have some homework from it to get the course in motion and also to familiarize you to the Einstein index convention. I'm hoping that will soften the blow of later topics. The repeated index notation is a little confusing at first, but once you get the hang of it you'll find it an incredible tool. Well, at least I do.

Contents

I Vector Calculus **4**

1 Vectors at a Point and their Algebraic Structure **5**
1.1 vectors based at a point . . . 5
1.2 vector operations in $T_P\mathbb{R}^3$ . . . 6
1.3 Newton's universal law of gravitation . . . 6
1.4 electrostatic repulsion, Coulomb's law . . . 6
1.5 cross product . . . 6
1.6 Lorentz force law . . . 7
1.7 dot product . . . 7
1.8 lengths and angles . . . 7
1.9 orthogonality . . . 8
1.10 tiny work from a tiny displacement . . . 8
1.11 power expended during a tiny displacement . . . 8
1.12 Einstein summation notation . . . 8
1.13 Kronecker delta and the Levi Civita symbol . . . 10
1.14 translating vector algebra into Einstein's notation . . . 10
1.15 tricks of the trade . . . 11
1.16 applying Einstein's notation . . . 11
1.17 work done by magnetic field . . . 12
1.18 work done by electric field . . . 12

2 Vector Fields on $\mathbb{R}^n$ **13**
2.1 frames . . . 13
2.2 spherical and coordinates and their frame . . . 14
2.3 cylindrical coordinates and their frame . . . 15

3 Differential Calculus on Vector Fields in $\mathbb{R}^n$ **16**
3.1 The ∇ operator . . . 16
3.2 grad, curl and div in Cartesians . . . 16
3.3 properties of the gradient operator . . . 17

4 Integral Vector Calculus **18**
4.1 line integrals . . . 18
4.2 surface and volume integrals . . . 21
4.3 Stokes's and Gauss's theorems . . . 24

II Geometry and Classical Physics **25**

5 Maxwell's Equations **26**
5.1 basics of electricity and magnetism . . . 26
5.2 Gauss law . . . 27
5.3 Ampere's law . . . 28
5.4 Faradays law . . . 29
5.5 Maxwell's correction term . . . 29
5.6 no magnetic monopoles . . . 30
5.7 overview of potentials in electromagnetism . . . 31
5.8 summary and apology . . . 32

6 Euclidean Structure and Newton's Laws **33**
6.1 Euclidean geometry . . . 34
6.2 noninertial frames, a case study of circular motion . . . 42
6.3 a lonely example . . . 44

7 Special Relativity **45**
7.1 Maxwell's equations verses Galilean relativity . . . 45
7.2 a short history of Ether . . . 46
7.3 Einstein's postulates . . . 47
7.4 Lorentz transformations . . . 47
7.5 the Minkowski metric . . . 50
7.6 general Lorentz transformations . . . 52
7.7 rapidity and velocity addition rule in special relativity . . . 54
7.7.1 rapidities are additive . . . 56
7.7.2 velocities are not additive . . . 57
7.8 translations and the Poincaire group . . . 58
7.9 vectors in Minkowski space . . . 59
7.9.1 additional consistency checks . . . 61
7.10 relativistic mechanics . . . 62
7.10.1 Energy . . . 63
7.10.2 more on the relativistic force law . . . 64

8 Matrix Lie groups and Lie algebras **66**
8.1 matrix Lie groups . . . 66
8.2 matrix Lie algebras . . . 68
8.3 exponentiation . . . 69

III Multilinear Algebra and Differential Forms in $\mathbb{R}^n$ **71**

9 Tensors and Forms on Vector Space **72**
9.1 multilinear maps . . . 72
9.2 multilinear maps on V and tensor products of dual vectors . . . 75
9.2.1 constructing bilinear maps via the tensor product of dual vectors . . . 75
9.2.2 constructing trilinear maps via the tensor product of dual vectors . . . 76

9.2.3 multilinear maps from tensor products . . . 77
9.3 multilinear maps on V^* and tensor products of vectors . . . 78
9.4 tensors on a vector space V . . . 80
9.5 raising and lowering indices on tensors . . . 82
9.6 coordinate change and tensors . . . 85

10 The Exterior Algebra of Forms **87**
10.1 exterior algebra from the wedge product . . . 87
10.1.1 wedge product verses cross product . . . 89
10.1.2 theory of determinants via the exterior algebra . . . 90
10.2 the wedge product constructed from the tensor product . . . 91
10.3 Hodge duality . . . 93

11 Differential Forms **95**
11.1 basics and the definition of differential form . . . 95
11.2 differential forms in $\mathbb{R}^3$ . . . 96
11.3 differential forms in Minkowski space . . . 100
11.4 exterior derivative . . . 104
11.4.1 coderivatives and comparing to Griffith's relativitic E & M . . . 107
11.5 Maxwell's equations are relativistically covariant . . . 108
11.6 Poincaire's Lemma and $d^2 = 0$ . . . 110

12 Integration of forms in $\mathbb{R}^n$ **113**
12.1 definitions . . . 113
12.2 Generalized Stokes Theorem . . . 116
12.3 Electrostatics in Five dimensions . . . 119

A Digressions and Comments **124**
A.1 speed of light invariant under Lorentz transformation . . . 124
A.2 Campbell-Baker-Hausdorff Relation . . . 124
A.3 Quaternions . . . 124
A.4 Coordinate change and Hodge Duality . . . 125
A.5 Dimension of the exterior algebra . . . 125

Part I

Vector Calculus

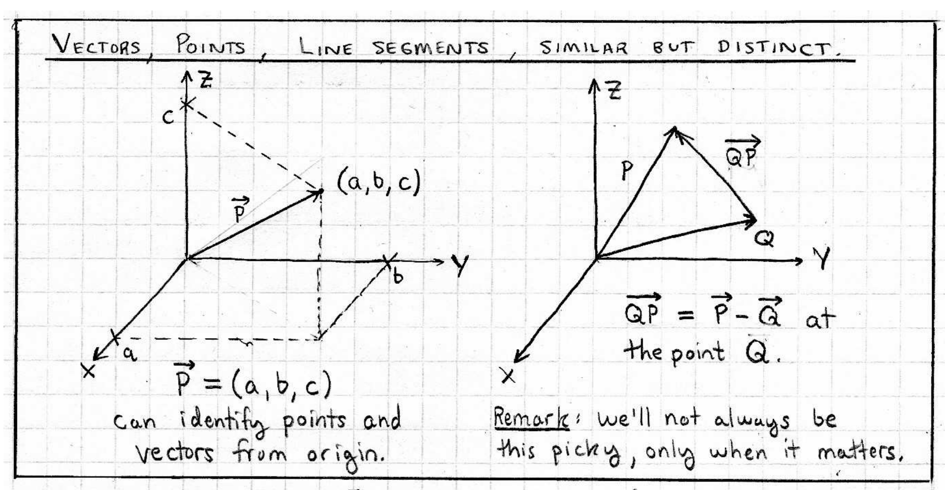
Chapter 1

Vectors at a Point and their Algebraic Structure

For the majority of this chapter we focus on the three dimensional case, however as we note at the conclusion of this chapter many of these ideas generalize to an arbitrary finite dimension. This chapter is intended as a lightning review of vector analysis, it's main purpose is to introduce notation and begin our discussion of mechanics and electromagnetism.

1.1 vectors based at a point

Let us begin by recalling that a point $P \in \mathbb{R}^3$ can be identified by the it's Cartesian coordinates (P_1, P_2, P_3). If we are given two points, say P and Q then the directed line segment from P to Q is called a *vector based at P* and is denoted $\vec{PQ}$. For a vector $\vec{PQ}$ we say the point P is the *tail* and the point Q is the *tip*. Vectors are drawn from tail to tip with an arrow customarily drawn at the tip to indicate the direction.



If the vector $\vec{PQ}$ has tail $P = (0,0,0)$ then we can identify the vector with it's tip $\vec{PQ} = \vec{Q}$. In practice this is also done when the tail is not at the origin, we usually imagine transporting the vector to the origin so that we can identify it with the point that it reaches. There is nothing wrong with this unless you care about where the tail of the vector resides. Thus, to be careful if we move a vector $\vec{PQ} = \vec{A}$ to the origin to figure out it's *components with respect to the standard basis* then we should make a note to the side this is a vector at P. We write,

$$\vec{A} = (A_1, A_2, A_3) \qquad \text{based at } P.$$

Of course if we use the directed line segment notation this would be redundant, when we say $\vec{PQ} =< A_1, A_2, A_3 >$ it is already implicit that the vector is based at P. The numbers A_i, $i = 1, 2, 3$ are called the components of $\vec{A}$ with respect to the standard basis. We may also write

$$\vec{A} = A_1\hat{\mathbf{i}} + A_2\hat{\mathbf{j}} + A_3\hat{\mathbf{k}} \qquad \text{based at } P.$$

As you may recall,

$$\hat{\mathbf{i}} = (1, 0, 0) \equiv e_1 \qquad \hat{\mathbf{j}} = (0, 1, 0) \equiv e_2 \qquad \hat{\mathbf{k}} = (0, 0, 1) \equiv e_3.$$

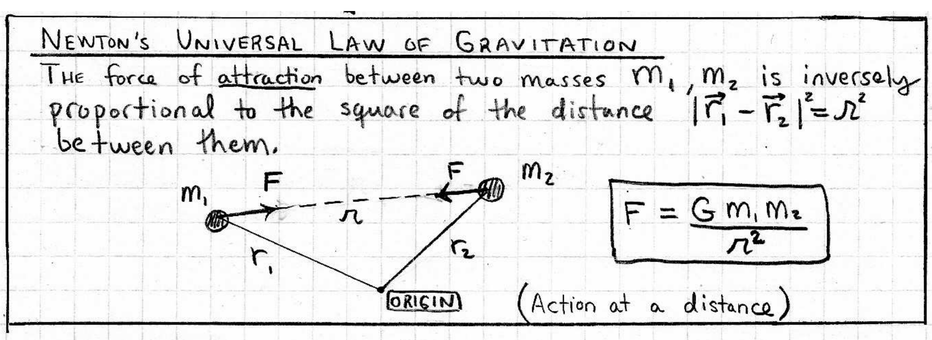
where the notation e_i is the standard basis (notation borrowed from linear algebra). For our future convenience let us denote the set off all vectors at P in $\mathbb{R}^3$ by $T_P\mathbb{R}^3$ which is called the *tangent space to* $\mathbb{R}^3$ *at* P. The fact that $\mathbb{R}^3$ can be identified with $T_P\mathbb{R}^3$ is very special, in general when the space is curved we cannot identify the space and the tangent space.

1.2 vector operations in $T_P\mathbb{R}^3$

Let $\vec{A}, \vec{B}$ be vectors based at a point $P \in \mathbb{R}^3$ and $c \in \mathbb{R}$ then recall we can add, substract and scale vectors to obtain new vectors at P,

$$\begin{aligned} \vec{A} + \vec{B} &= (A_1 + B_1, A_2 + B_2, A_3 + B_3) \\ \vec{A} - \vec{B} &= (A_1 - B_1, A_2 - B_2, A_3 - B_3) \\ c\vec{A} &= (cA_1, cA_2, cA_3). \end{aligned} \tag{1.1}$$

1.3 Newton's universal law of gravitation

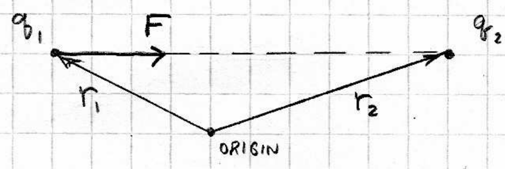


1.4 electrostatic repulsion, Coulomb's law

Electrostatic interactions involve charges that have fixed positions.

COULOMBS FORCE LAW

The force of attraction (for unlike charges) or repulsion (for like charges) is proportional to the inverse of the square of the distance seperating q_1 and q_2



$$\boxed{F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}}$$

we have drawn the attractive case, note the similarity with Newton's Gravity Law. In contrast, we will eventually learn that this Law only appears to indicate action at a distance, in fact the electric field communicates the Coulomb force. In the case of gravity it was only with the advent of general relativity that the apparent non-locality of Newtonian Gravity was recast as a subcase of a local (no action at a distance) classical field theory. ("classical" in physics means no quantum mechanics in play.)

1.5 cross product

We define the *cross-product* of $\vec{A} \in T_P\mathbb{R}^3$ with $\vec{B} \in T_P\mathbb{R}^3$ according to the rule,

$$\vec{A} \times \vec{B} = (A_2B_3 - A_3B_2, A_3B_1 - A_1B_3, A_1B_2 - A_2B_1). \tag{1.2}$$

The output of the cross product is again a vector based at P. One way to remember the formula above is that the i^{th} slot does not contain the i^{th} components of $\vec{A}$ or $\vec{B}$ and the indices in the order $\{1, 2, 3, 1\}$ get a positive sign whereas indices in the order $\{3, 2, 1, 3\}$ obtain a minus sign. Notice the cross product has the properties,

$$\begin{array}{ll} \vec{A} \times \vec{B} = -\vec{B} \times \vec{A} & \text{skewsymmetry} \\ (\vec{A} + \vec{B}) \times \vec{C} = \vec{A} \times \vec{C} + \vec{B} \times \vec{C} & \\ \vec{A} \times (c\vec{B}) = c(\vec{A} \times \vec{B}) & \end{array} \tag{1.3}$$

for all $\vec{A}, \vec{B}, \vec{C} \in T_P\mathbb{R}^3$ and $c \in \mathbb{R}$. It is straightforward to verify that

$$\begin{array}{lll} \hat{\mathbf{i}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}} & \hat{\mathbf{j}} \times \hat{\mathbf{k}} = \hat{\mathbf{i}} & \hat{\mathbf{k}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}} \\ \hat{\mathbf{i}} \times \hat{\mathbf{i}} = 0 & \hat{\mathbf{j}} \times \hat{\mathbf{j}} = 0 & \hat{\mathbf{k}} \times \hat{\mathbf{k}} = 0 \end{array} \tag{1.4}$$

Finally, we note the following is a useful heuristic for remembering the cross-product,

$$\vec{A} \times \vec{B} = \det \begin{pmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{pmatrix} \equiv \hat{\mathbf{i}} \det \begin{pmatrix} A_2 & A_3 \\ B_2 & B_3 \end{pmatrix} - \hat{\mathbf{j}} \det \begin{pmatrix} A_1 & A_3 \\ B_1 & B_3 \end{pmatrix} + \hat{\mathbf{k}} \det \begin{pmatrix} A_1 & A_2 \\ B_1 & B_2 \end{pmatrix}. \tag{1.5}$$

the expression above is heuristic because the determinant is typically defined only for matrices filled with numbers, not vectors like $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$.

1.6 Lorentz force law

LORENTZ FORCE LAW

Given electric and magnetic fields $\vec{E}$ and $\vec{B}$ measured in a particular frame, say S. If we have a charge q moving at $\vec{V}$ relative to S then the force it experiences in S is

$$\vec{F} = q(\vec{E} + \vec{V} \times \vec{B})$$

1.7 dot product

Let $\vec{A}, \vec{B} \in T_P\mathbb{R}^3$ then define the *dot product of $\vec{A}$ with $\vec{B}$*,

$$\vec{A} \cdot \vec{B} = A_1 B_1 + A_2 B_2 + A_3 B_3. \tag{1.6}$$

The output of the dot product is a number. It is simple to show the dot product has the following properties,

$$\begin{aligned} &\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A} \\ &(\vec{A} + \vec{B}) \cdot \vec{C} = \vec{A} \cdot \vec{C} + \vec{B} \cdot \vec{C} \\ &\vec{A} \cdot (c\vec{B}) = c(\vec{A} \times \vec{B}) \end{aligned} \tag{1.7}$$

for all $\vec{A}, \vec{B}, \vec{C} \in T_P\mathbb{R}^3$ and $c \in \mathbb{R}$. It is straightforward to verify that

$$\begin{array}{ccc} \hat{\mathbf{i}} \cdot \hat{\mathbf{j}} = 0 & \hat{\mathbf{j}} \cdot \hat{\mathbf{k}} = 0 & \hat{\mathbf{k}} \cdot \hat{\mathbf{i}} = 0 \\ \hat{\mathbf{i}} \cdot \hat{\mathbf{i}} = 1 & \hat{\mathbf{j}} \cdot \hat{\mathbf{j}} = 1 & \hat{\mathbf{k}} \cdot \hat{\mathbf{k}} = 1 \end{array} \tag{1.8}$$

1.8 lengths and angles

The length of a vector $\vec{A} \in T_P\mathbb{R}^3$ is denoted $||\vec{A}||$ and is defined to be the length from the origin to the point which corresponds to $\vec{A}$, that is,

$$||\vec{A}|| = \sqrt{(A_1)^2 + (A_2)^2 + (A_3)^2}. \tag{1.9}$$

Notice that the length function $||\ .\ || : T_P\mathbb{R}^3 \to \mathbb{R}$ forms a *norm* on $T_P\mathbb{R}^3$. A *norm* is defined to be a mapping from a vector space to scalars such that,

$$\begin{array}{ll} ||\vec{A}|| \geq 0 & \text{non-negative} \\ ||\vec{A}|| = 0 \iff A = 0 & \text{positive definite} \\ ||\vec{A} + \vec{B}|| \leq ||\vec{A}|| + ||\vec{B}|| & \text{triangle inequality} \\ ||c\vec{A}|| = |c|\ ||\vec{A}|| & \text{pull out absolute value of scalars} \end{array} \tag{1.10}$$

for all $\vec{A}, \vec{B} \in T_P\mathbb{R}^3$ and $c \in \mathbb{R}$.

The *angle* θ between two non-zero vectors $\vec{A}, \vec{B} \in T_P\mathbb{R}^3$ with lengths A, B respectively is defined to be

$$\theta = \cos^{-1}\left[\frac{\vec{A}\cdot\vec{B}}{AB}\right]. \tag{1.11}$$

Unit vectors are denoted by hats, if $\vec{A}$ has length A then define $\vec{A} \equiv A\hat{A}$. A unit vector for $\vec{A}$ is vector of length one that points in the same direction as $\vec{A}$.

We have the following results for non-zero vectors $\vec{A} = A\hat{A}$ and $\vec{B} = B\hat{B}$ separated by angle θ,

$$\begin{aligned} \vec{A}\cdot\vec{B} &= AB\cos(\theta) \\ \vec{A}\times\vec{B} &= AB\sin(\theta)\hat{A}\times\hat{B}. \end{aligned} \tag{1.12}$$

The first of these relations is trivial given our definition of angle, the second requires some thought. Finally notice that,

$$||\vec{A}||^2 = \vec{A}\cdot\vec{A}.$$

1.9 orthogonality

In three dimensions the when vectors are perpendicular this means the angle between them is $\pi/2$ radians. Since $\cos(\pi/2) = 0$ it follows that the dot product of such vectors is zero. Orthogonality is simply then generalization of the concept of perpendicularity to arbitrary dimensions, we define orthogonality with respect to the dot product on $\mathbb{R}^n$,

$$\vec{A} \text{ is orthogonal to } \vec{B} \text{ iff } \vec{A}\cdot\vec{B} = 0. \tag{1.13}$$

Moreover we say a set of vectors is orthogonal if each pair of distinct vectors in the set is orthogonal. A set of vectors which are orthogonal and all of length one is called *orthonormal*. For example, the standard basis on $\mathbb{R}^n$ is orthonormal since $e_i \cdot e_j = 0$ if $i \neq j$ and they all have length one.

1.10 tiny work from a tiny displacement

INFINITESIMAL WORK dW

Suppose a force $\vec{F}$ propels a particle of mass m a small displacement $d\vec{l}$ then the work done on m by $\vec{F}$ is defined to be

$$dW = \vec{F}\cdot d\vec{l}$$

Notice parallell forces do the most work, whereas perpendicular forces do none.

1.11 power expended during a tiny displacement

POWER OUTPUT BY FORCE DOING INFINITESIMAL WORK.

Power is the change in energy with respect to the change in time; $P = dE/dt$. Now here the energy is realeased doing work dW then

$$P = \frac{dW}{dt} = \frac{d}{dt}\left(\vec{F}\cdot d\vec{l}\right) = \vec{F}\cdot\frac{d\vec{l}}{dt} = \boxed{\vec{F}\cdot\vec{V} = P}$$

Here I have pulled $\vec{F}$ out since it cannot change during the infinitesimal time it takes to move m $d\vec{l}$. If we have a conservative force then $\exists\, U$ such that

$$\boxed{F = -\nabla U} \qquad U = \text{potential energy}$$

then there is a nice clean derivation, $U = U(x,y,z)$, not time.

$$\begin{aligned}
\frac{dE}{dt} = -\frac{dU}{dt} &= -\left(\frac{\partial U}{\partial x}\frac{dx}{dt} + \frac{\partial U}{\partial y}\frac{dy}{dt} + \frac{\partial U}{\partial z}\frac{dz}{dt}\right) \\
&= (-\nabla U)\cdot\vec{V} \\
&= \boxed{\vec{F}\cdot\vec{V} = P}
\end{aligned}$$

The minus stems from the conservation of energy theorem, the work done by a conservative force is minus the change in potential energy. Since the work done by a force F is equal to the change in Kinetic energy this gives $K_i + U_i = K_f + U_f$.

1.12 Einstein summation notation

We adopt the following convention, Latin indices (like i,j,k,...) which are repeated in the same term are to be summed over their values. In $\mathbb{R}^3$ we have indices ranging over $\{1,2,3\}$ so,

$$A_iB_i \equiv A_1B_1 + A_2B_2 + A_3B_3. \tag{1.14}$$

When we do not wish to sum over repeated indices we will explicitly indicate "no sum". For example,

$$A_iB_i \equiv \sum_{i=1}^{3} A_iB_i \qquad \text{no summation on i.} \tag{1.15}$$

In contrast if we stick to our repeated index convention then we'd get

$$\begin{aligned}
A_iB_i &= \textstyle\sum_{i=1}^{3} A_iB_i \\
&= \textstyle\sum_{i=1}^{3}(A_1B_1 + A_2B_2 + A_3B_3) \\
&= 3(A_1B_1 + A_2B_2 + A_3B_3)
\end{aligned} \tag{1.16}$$

that's why you need to write "no sum on i" if you want ordinary sums to mean what they usually do.

If the index is repeated in different terms then it is taken to be free,

$$A_i + B_i = 3 \tag{1.17}$$

would mean that $A_1 + B_1 = 3$ and $A_2 + B_2 = 3$ and $A_3 + B_3 = 3$ since, unless we say otherwise, an index which is not summed over is allowed to take all it's values. If there are other indices hanging around then they just ride along, for example,

$$A_{ij}B_{ik} \equiv A_{1j}B_{1k} + A_{2j}B_{2k} + A_{3j}B_{3k} \tag{1.18}$$

here the indices j and k are free which means that there are actually nine equations hiding in the statement above. If there are several repeated indices then that indicates a double or triple summation, for example

$$A_{ij}B_{ij} \equiv \sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}B_{ij} \qquad \text{no summation on i or j.} \tag{1.19}$$

Remark 1.12.1. Ignore this for now if you've never studied special relativity *Although this comment in inappropriate at this juncture I need to make it for those of you who know more than you should. We will later find that $A_i = A^i$ because we will use the 4-metric that is mostly positive. More precisely, we will consider generalized dot-product of 4 dimensional contravariant vectors $(A^\mu) = (A^0, A^1, A^2, A^3) = (A^0, \vec{A})$ and $(B^\mu) = (B^0, B^1, B^2, B^3) = (B^0, \vec{B})$, the Minkowski dot product which we will define as $\bar{A}\cdot\bar{B} = -A^0B^0 + A^1B^1 + A^2B^2 + A^3B^3$ which can be written with the help of the covariant version of $\bar{B}$ which is $(B_\mu) = (B_0, B_1, B_2, B_3)$ as $\bar{A}\cdot\bar{B} = A^0B_0 + A^1B_1 + A^2B_2 + A^3B_3 = A^\mu B_\mu$ invoking the repeated index convention for 4 dimensions where we will use $\mu = 0, 1, 2, 3$. The covariant and contravariant components of a particular vector are almost identical modulo the time component,*

$$B_0 = -B^0 \qquad\qquad B_i = B^i \qquad i = 1, 2, 3.$$

With our conventions we cannot distinguish at the level of components a covariant verses a contravariant spatial component. Also when we have $A^iB_i = \vec{A}\cdot\vec{B}$, if we used the other popular metric which is mostly negative then instead we would have $A^iB_i = -\vec{A}\cdot\vec{B}$, neither is wrong of course. I have chosen our conventions so that they correlate with Griffith's "Introduction to Electrodynamics" and also the usual component notation in ordinary calculus texts which would never bother to write indices up. So we can begin the course without worrying about indices up or down. Later we will put our indices in their proper place and everything we've done here will still be correct since there is no difference; $B_i = B^i$ (if you use the other metric then all manner of funny signs enter the discussion at some point)

1.13 Kronecker delta and the Levi Civita symbol

Definition 1.13.1. *The Kronecker delta function δ_{ij} is defined as follows,*

$$\delta_{ij} \equiv \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}.$$

Definition 1.13.2. *The Levi Civita symbol* ϵ_{ijk} *is defined as follows*

$$\epsilon_{ijk} \equiv \begin{cases} 1, & \{i,j,k\} \text{ cyclic permutation of } \{1,2,3\} \\ -1, & \{i,j,k\} \text{ cyclic permutation of } \{3,2,1\} \\ 0, & \text{any pair of indices repeated} \end{cases}$$

this is also known as the completely antisymmetric symbol because exchanging any pair of indices will result in generating a minus sign. Just to be explicit we list it's non-zero values,

$$\begin{aligned} &\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1 \\ &\epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1. \end{aligned} \tag{1.20}$$

Thus the symbol is invariant upto cyclic exchanges of its indices,

$$\epsilon_{ijk} = \epsilon_{jki} = \epsilon_{kij}. \tag{1.21}$$

The following identities are often useful for calculations,

$$\begin{aligned} &\epsilon_{ijk}\epsilon_{mjk} = 2\delta_{im} \\ &\epsilon_{ijk}\epsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{jl}\delta_{im}. \\ &\delta_{kk} = 3 \end{aligned} \tag{1.22}$$

The first and third identities hold **only for three dimensions**, they are multiplied by different constants otherwise. In fact if n is a positive integer then

$$\begin{aligned} &\epsilon_{ii_2i_3\ldots i_n}\epsilon_{ji_2i_3\ldots i_n} = (n-1)!\delta_{ij} \\ &\delta_{kk} = n. \end{aligned} \tag{1.23}$$

Although we have given the definition for the antisymmetric symbol in three dimensions with three indices it should be clear that we can construct a similar object with n-indices in n-dimensions, simply define that $\epsilon_{12\ldots n} = 1$ and the symbol is antisymmetric with respect to the exchange of any two indices.

1.14 translating vector algebra into Einstein's notation

Now let us restate some earlier results in terms of the Einstein repeated index conventions, let $\vec{A}, \vec{B} \in T_P\mathbb{R}^3$ and $c \in \mathbb{R}$ then

$$\begin{array}{ll} \vec{A} = A_k e_k & \text{basis expansion} \\ e_i \cdot e_j = \delta_{ij} & \text{orthonormal basis} \\ (\vec{A}+\vec{B})_i = \vec{A}_i + \vec{B}_i & \text{vector addition} \\ (\vec{A}-\vec{B})_i = \vec{A}_i - \vec{B}_i & \text{vector subtraction} \\ (c\vec{A})_i = c\vec{A}_i & \text{scalar multiplication} \\ \vec{A}\cdot\vec{B} = A_kB_k & \text{dot product} \\ (\vec{A}\times\vec{B})_k = \epsilon_{ijk}A_iB_j & \text{cross product.} \end{array} \tag{1.24}$$

All but the last of the above are readily generalized to dimensions other than three by simply increasing the number of components. The cross product is special to three dimensions, we will see why as we go on. I can't emphasize enough that the formulas given above for the dot and cross products are much easier to utilize for abstract calculations. For specific vectors with given numerical entries the formulas you learned in multivariate calculus will do just fine, but as we go on in this course we will deal with vectors with arbitrary entries so an abstract language allows us to focus on what we know without undue numerical clutter.

1.15 tricks of the trade

The Einstein notation is more a hindrance then a help if you don't know the tricks. I'll now make a list of common "tricks", most of them are simple,

$(i.)\ A_iB_j = B_jA_i$	components are numbers!
$(ii.)\ A_iB_i\delta_{ij} = A_jB_j$	no sum over j
$(iii.)\ A_{ij}B_{ij} = A_{ji}B_{ji} = A_{mp}B_{mp}$	switching dummies
$(iv.)$ If $S_{ij} = S_{ji}$ and $A_{ij} = -A_{ji}$ then $A_{ij}B_{ij} = 0$	symmetric kills antisymmetric

Part (i.) is obviously true. Part (ii.) is not hard to understand, the index j is a fixed (but arbitrary) index and the terms in the sum are all zero except when $i = j$ as a consequence of what δ_{ij} means. Part (iii.) is also simple, the index of summation is just notation, wether I use i or j or m or p the summation includes the same terms. A word of caution on (iii.) is that when an index is free (like j in (i.)) we cannot just change it to something else. Part (iv.) deserves a small proof which we give now, assuming the conditions of (iv.),

$$\begin{aligned} A_{ij}S_{ij} &= A_{mp}S_{mp} && \text{by (iii.)} \\ &= -A_{pm}S_{pm} && \text{using our hypothesis} \\ &= -A_{ij}S_{ij} && \text{by (iii.)} \end{aligned}$$

therefore $2A_{ij}S_{ij} = 0$ thus $A_{ij}S_{ij} = 0$ as claimed.

There are other tricks, but these should do for now. It should be fairly clear these ideas are not particular to three dimensions. The Einstein notation is quite general.

1.16 applying Einstein's notation

We now give an example of how we can implement Einstein's notation to prove an otherwise cumbersome identity. (if you don't believe me that it is cumbersome feel free to try to prove it in Cartesian components).

Proposition 1.16.1.

$$\begin{aligned} &(i.)\ \vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B}) \\ &(ii.)\ \vec{A} \cdot (\vec{B} \times \vec{C}) = \vec{B} \cdot (\vec{C} \times \vec{A}) = \vec{C} \cdot (\vec{A} \times \vec{B}) \end{aligned} \tag{1.25}$$

Proof: The proof of (i.) hinges on eq. [1.21]. Let's look at the k^{th} component of $\vec{A} \times (\vec{B} \times \vec{C})$,

$$\begin{aligned} [\vec{A} \times (\vec{B} \times \vec{C})]_k &= \epsilon_{ijk}A_i(\vec{B} \times \vec{C})_j \\ &= \epsilon_{ijk}A_i\epsilon_{mpj}B_mC_p \\ &= -\epsilon_{ikj}\epsilon_{mpj}A_iB_mC_p \\ &= -(\delta_{im}\delta_{kp} - \delta_{ip}\delta_{km})A_iB_mC_p \\ &= \delta_{ip}\delta_{km}A_iB_mC_p - \delta_{im}\delta_{kp}A_iB_mC_p \\ &= A_iB_kC_i - A_iB_iC_k \\ &= B_k(\vec{C} \cdot \vec{A}) - C_k(\vec{A} \cdot \vec{B}). \end{aligned} \tag{1.26}$$

Now this equation holds for each value of k thus the vector equation in question is true. Now let us attack (ii.),

$$
\begin{aligned}
\vec{A}\cdot(\vec{B}\times\vec{C}) &= A_k(\vec{B}\times\vec{C})_k \\
&= A_k\epsilon_{ijk}B_iC_j && \text{bob} \\
&= B_i\epsilon_{ijk}C_jA_k && \text{components are numbers} \\
&= B_i\epsilon_{jki}C_jA_k && \text{since } \epsilon_{ijk} = -\epsilon_{jik} = \epsilon_{jki} \\
&= \vec{B}\cdot(\vec{C}\times\vec{A}) && \text{half done.} \\
&= C_j\epsilon_{ijk}A_kB_i && \text{going back to bob and commuting numbers} \\
&= C_j\epsilon_{kij}A_kB_i && \text{since } \epsilon_{ijk} = -\epsilon_{ikj} = \epsilon_{kij} \\
&= \vec{C}\cdot(\vec{A}\times\vec{B}) && \text{and finished.}
\end{aligned}
\tag{1.27}
$$

Thus we see the cyclicity of ϵ_{ijk} has translated into the cyclicity of the triple product identity (ii.). Hence the proposition is true, and again I emphasize the Einstein notation has saved us much labor in these calculations.

1.17 work done by magnetic field

We'll use repeated index notation to show that the magnetic field does no work.

WORK DONE BY MAGNETIC FIELD IS ZERO

Consider a charge q with velocity $\vec{V}$ then place it in a magnetic field $\vec{B}$ and find if any work is done. Lets look at the power instead,

$$
\begin{aligned}
P = \frac{dW}{dt} = \vec{F}\cdot\vec{V} &= q(\vec{V}\times\vec{B})\cdot\vec{V} \\
&= q\,\epsilon_{ijk}V_iB_jV_k \\
&= q\,\epsilon_{ijk}V_iV_kB_j \\
&= -q\,\epsilon_{kji}V_iV_kB_j \\
&= -q\,\epsilon_{kji}V_kV_iB_j \\
&= -q(V\times\vec{B})\cdot\vec{V} \implies P = 0 \implies \boxed{dW = 0}
\end{aligned}
$$

1.18 work done by electric field

We recall a few concepts from freshman physics in this example, sorry we don't have more time to develop them properly.

Work Done by Electric Field and the Electric Potential

Although we'll not completely introduce the line integral till a later chapter, its nice to use it here to discuss the work done by a force on a particle moving a finite distance.

$$W_{ab} \equiv \int_a^b \vec{F} \cdot d\vec{l}$$

Now for the electric force $\vec{F} = q\vec{E}$ thus

$$W_{ab} = \int_a^b q\vec{E} \cdot d\vec{l}$$

Then the potential energy function will be some U such that $\vec{F} = -\nabla U$ where $U = U(x,y,z)$ but not time. For example the Coulomb force on q due to Q is

$$\vec{F} = \frac{1}{4\pi\epsilon_0}\frac{qQ}{r^2}\hat{r}$$

where we have placed the "source" charge Q at the origin for convenience. Notice the potential energy function U is

$$U = \frac{1}{4\pi\epsilon_0}\frac{qQ}{r} \quad \text{since,} \quad \vec{F} = -\frac{\partial}{\partial r}\left(\frac{1}{4\pi\epsilon_0}\frac{qQ}{r}\right)\hat{r} = \frac{qQ}{4\pi\epsilon_0 r^2}\hat{r}.$$

Then the work done by the Coulomb force along a path will be easily calculated

$$\begin{aligned} W_{ab} &= \int_a^b \vec{F} \cdot d\vec{l} \\ &= \int_a^b -\nabla U \cdot d\vec{l} \\ &= U(a) - U(b) \\ &= q\frac{Q}{4\pi\epsilon_0}\left(\frac{1}{a} - \frac{1}{b}\right) \end{aligned}$$

WARNING: I'm assuming you recall some ideas here. We will cover them again later so this example is a bit out of order, sorry.

For reasons that may become clearer by the end of this course the Electric Potential or "Scalar Potential" is of more use to us, it is simply the potential energy per unit charge.

$$V = \frac{U}{q} = \frac{Q}{4\pi\epsilon_0 r} = \text{potential due to } Q.$$

V is taken with respect to some ref. point, we used $r = \infty$. We defer the general def^{n} of V till later, for now just Coulomb.

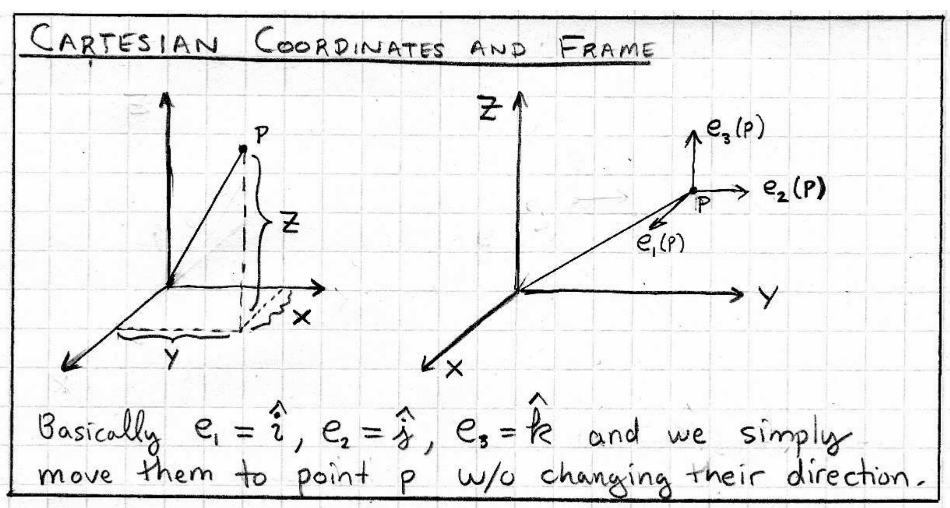
Chapter 2

Vector Fields on $\mathbb{R}^n$

A vector field is a function from $\mathbb{R}^n \to T_P\mathbb{R}^n$, it is an assignment of a vector $\vec{A}$ to each point $P \in \mathbb{R}^n$. Sometimes we will allow that assignment will depend on time.

2.1 frames

A *frame* is an assignment of an ordered basis to $T_P\mathbb{R}^n$ for each $P \in \mathbb{R}^n$. We define the *Cartesian coordinate frame* for $\mathbb{R}^n$ to be $\{e_1(p), e_2(p), \ldots, e_n(p)\}$ where we are using the notation $e_i(p)$ to denote the i^{th} standard basis vector paralell transported to the point p.



It is trivial to verify that it forms a basis for $T_P\mathbb{R}^n$. We could say that a frame forms a basis for vector fields, however we should be careful to realize that this would not be a basis in the sense of linear algebra. An arbitrary vector field $\vec{X}$ in $\mathbb{R}^n$ would have the form

$$\vec{X}(P) = X^i(P)e_i(P) \tag{2.1}$$

where the components $X^i(P)$ are *functions* of position, not just numbers. This means that the set of all vector fields on $\mathbb{R}^n$ is a *module* over functions from $\mathbb{R}^n \to \mathbb{R}$. A module is basically a vector space except the "scalars" are taken from a ring instead of a field. But, this is a topic for a different course.

Although $X^i(P)$ is better than $X_i(P)$ conceptually we will let the indices stay down in this chapter to keep expressions pretty. There are other conventions.

Definition 2.1.1. *An orthonormal frame is a frame that assigns a orthonormal basis at each point.*

The frame $\{e_1(p), e_2(p), \dots, e_n(p)\}$ is an orthonormal frame, as are the spherical and cylindrical coordinate frames which we will soon discuss.

2.2 spherical and coordinates and their frame

Cartesian coordinates are great for basic formalism but they are horrible to use on many common physical problems. Often a physical problem has the property that from a certain point all directions look the same, this is called *spherical symmetry*. In such cases using *spherical coordinates* will dramatically reduce the tedium of calculation.

We recall spherical coordinates relate to Cartesian coordinates as follows (I use physics conventions (r, θ, ϕ) instead of the usual math version (ρ, ϕ, θ) sorry, but it's for my own good. When I refer to Colley you'll need to translate me $\mapsto$ Colley as $r \mapsto \rho$, $\theta \mapsto \phi$ and $\phi \mapsto \theta$ if you try to compare these notes to conventional math books)

$$\begin{aligned} x &= r\cos(\phi)\sin(\theta) \\ y &= r\sin(\phi)\sin(\theta) \\ z &= r\cos(\theta) \end{aligned} \tag{2.2}$$

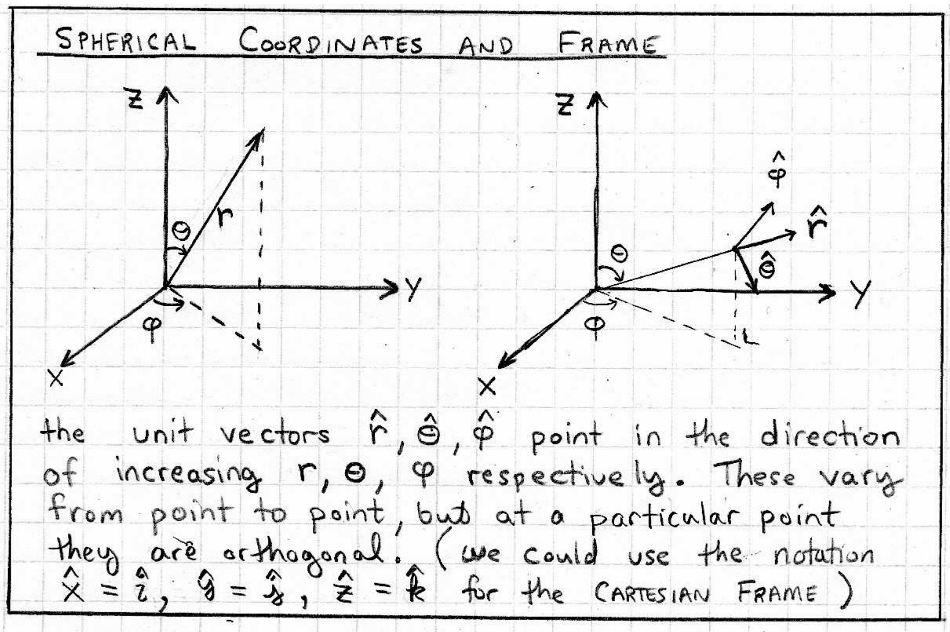
where $r > 0$ and $0 \le \theta \le \pi$ and $0 \le \phi \le 2\pi$. We can derive,

$$\begin{aligned} r^2 &= x^2 + y^2 + z^2 \\ \tan(\theta) &= \sqrt{x^2 + y^2}/z \\ \tan(\phi) &= y/x \end{aligned} \tag{2.3}$$

Let us denote unit vectors in the direction of increasing r, θ, ϕ by e_r, e_θ, e_ϕ respectively, then it can be shown geometrically (see pages 76-77 of Colley) that,

$$\begin{aligned} \hat{r} = e_r &= \sin(\theta)\cos(\phi)\hat{\mathbf{i}} + \sin(\theta)\sin(\phi)\hat{\mathbf{j}} + \cos(\theta)\hat{\mathbf{k}} \\ \hat{\theta} = e_\theta &= -\cos(\theta)\cos(\phi)\hat{\mathbf{i}} - \cos(\theta)\sin(\phi)\hat{\mathbf{j}} + \sin(\theta)\hat{\mathbf{k}} \\ \hat{\phi} = e_\phi &= -\sin(\phi)\hat{\mathbf{i}} + \cos(\phi)\hat{\mathbf{j}}. \end{aligned} \tag{2.4}$$

the notation $\hat{r}, \hat{\theta}, \hat{\phi}$ matches Griffith's notation.



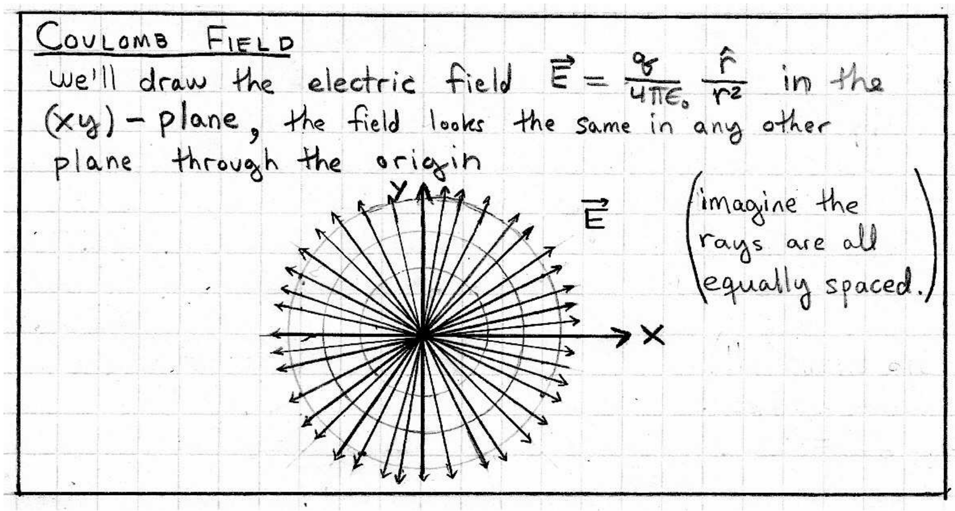
This is also an orthonormal frame on part of $\mathbb{R}^n$. Although I will not belabor such points much in this course it is important to note that spherical coordinates are ill-defined at certain points. (a pretty detailed discussion about some of the dangers can be found in pages 67-77 of Colley). In contrast, Cartesian coordinates are well defined everywhere.

We give the standard order to the basis at each point, $\{e_r, e_\theta, e_\phi\}$. The ordering is made so that the cross product of vectors matches the usual pattern,

$$\begin{array}{lll} \hat{\mathbf{i}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}} & \hat{\mathbf{j}} \times \hat{\mathbf{k}} = \hat{\mathbf{i}} & \hat{\mathbf{k}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}} \\ e_r \times e_\theta = e_\phi & e_\theta \times e_\phi = e_r & e_\phi \times e_r = e_\theta \end{array} \tag{2.5}$$

We also have the notation $\hat{\mathbf{i}} = e_1$, $\hat{\mathbf{j}} = e_2$, $\hat{\mathbf{k}} = e_3$ which allows us to compactly state the first line of the above as $e_i \times e_j = \epsilon_{ijk} e_k$. Thus if we define $f_1 = e_r$, $f_2 = e_\theta$, $f_3 = e_\phi$ we will likewise obtain $f_i \times f_j = \epsilon_{ijk} f_k$.

A good example of a physical problem that is best described by spherical coordinates is the point charge. Pictured below is the electric field due to a point charge, it is called the *Coulomb field*



2.3 cylindrical coordinates and their frame

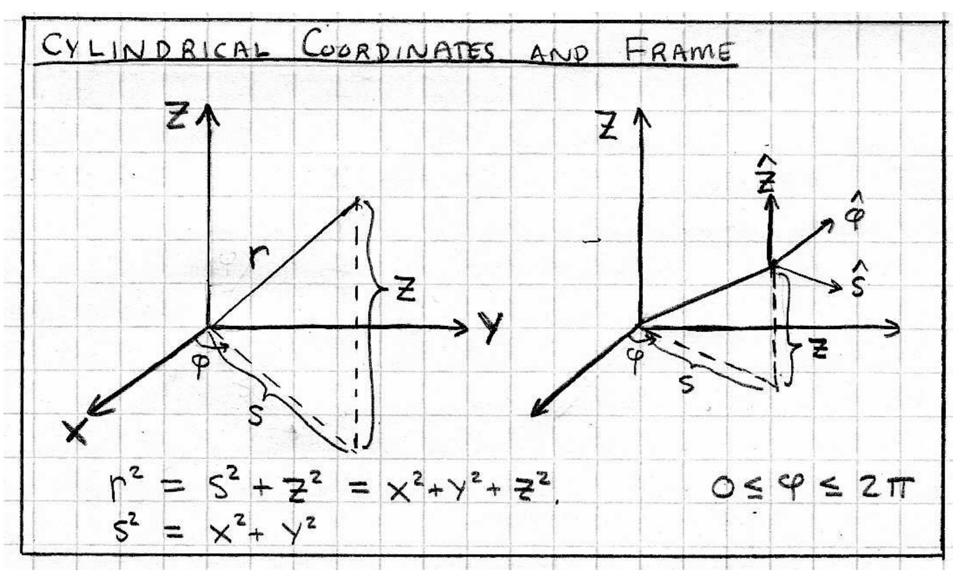
Cylindrical coordinates are useful for problems that have a cylindrical symmetry, you can guess what that means. Anyway, cylindrical coordinates (s, ϕ, z) are related to Cartesian coordinates as follows,

$$\begin{aligned} x &= s\cos(\phi) \\ y &= s\sin(\phi) \\ z &= z \end{aligned} \tag{2.6}$$

where ϕ is same as in spherical coordinates; $0 \leq \phi \leq 2\pi$. We can derive,

$$\begin{aligned} s^2 &= x^2 + y^2 \\ \tan(\phi) &= y/x \\ z &= z. \end{aligned} \tag{2.7}$$

Graphically,

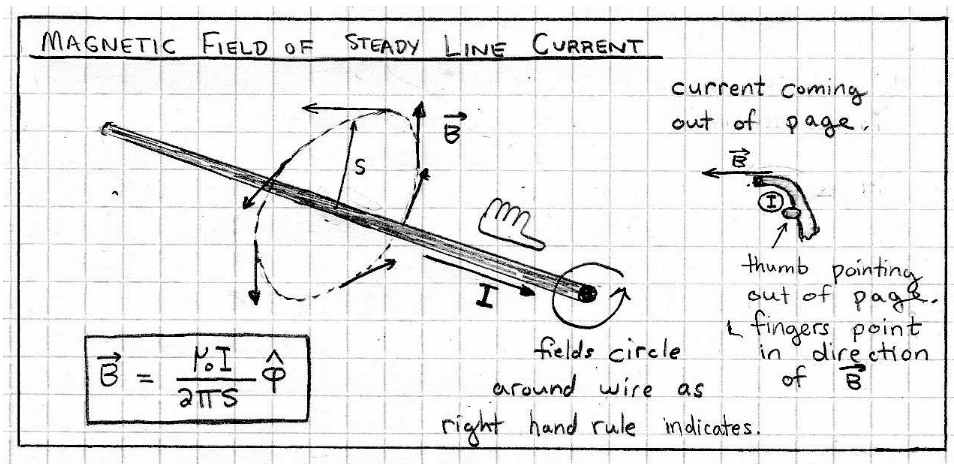


Let us denote unit vectors in the direction of increasing s, ϕ, z by e_s, e_ϕ, e_z respectively, then it can be shown geometrically (see pages 76-77 of Colley) that,

$$\begin{aligned} \hat{s} &= e_s = \cos(\phi)\hat{\mathbf{i}} + \sin(\phi)\hat{\mathbf{j}} \\ \hat{\phi} &= e_\phi = -\sin(\phi)\hat{\mathbf{i}} + \cos(\phi)\hat{\mathbf{j}} \\ \hat{z} &= e_z = \hat{\mathbf{k}}. \end{aligned} \tag{2.8}$$

It can be shown that $\{e_s, e_\phi, e_z\}$ form an orthonormal frame. Indeed if we define $h_1 = e_s$ and $h_2 = e_\phi$ and $h_3 = e_z$ we could verify that $h_k = \epsilon_{ijk} h_i h_j$.

The magnetic field due to a long linear current is best described by cylindrical coordinates. The right hand rule helps us maintain a consistent system or orientations, I've tried to illustrate it...



Chapter 3

Differential Calculus on Vector Fields in $\mathbb{R}^n$

Our goal here is to review all the basic notions of differentiation that involve vector fields. We will use the ∇ operator to phrase the gradient, curl and divergence.

3.1 The ∇ operator

We define the ∇ operator in Cartesian coordinates,

$$\nabla = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z} = e_i\partial_i \tag{3.1}$$

where we have introduced the notation ∂_i for $\frac{\partial}{\partial x^i}$. Admittably ∇ is a strange object, a vector of operators. Much like the determinant formula for the cross product this provides a convenient mnemonic for the vector calculus.

3.2 grad, curl and div in Cartesians

Let $U \subset \mathbb{R}^n$ and f a function from U to $\mathbb{R}$. The *gradient* of f is a vector field on U defined by,

$$\mathbf{grad}(f) = \nabla f = \hat{\mathbf{i}}\frac{\partial f}{\partial x} + \hat{\mathbf{j}}\frac{\partial f}{\partial y} + \hat{\mathbf{k}}\frac{\partial f}{\partial z} = e_i\partial_i f \tag{3.2}$$

Let $\vec{F} = F^i e_i$ be a vector field on U. The *curl* of $\vec{F}$ is defined by,

$$\mathbf{curl}(\vec{F}) = \nabla \times \vec{F} = \hat{\mathbf{i}}(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}) + \hat{\mathbf{j}}(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}) + \hat{\mathbf{k}}(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}) = \epsilon_{ijk}(\partial_i F_j)e_k. \tag{3.3}$$

Let $\vec{G} = G^i e_i$ be a vector field on U. The *divergence* of $\vec{G}$ is defined by,

$$\mathbf{div}(\vec{G}) = \nabla \cdot \vec{G} = \frac{\partial G_1}{\partial x} + \frac{\partial G_2}{\partial y} + \frac{\partial G_3}{\partial z} = \partial_i G_i. \tag{3.4}$$

All the operations above are only defined for suitable functions and vector fields, we must be able to perform the partial differentiation that is required. I have listed the definition in each of the popular notations and with the less popular (among mathematicians anyway) repeated index notation. Given a particular task you should choose the notation wisely.

3.3 properties of the gradient operator

It is fascinating how many of the properties of ordinary differentiation generalize to the case of vector calculus. The main difference is that we now must take more care to not commute things that don't commute or confuse functions with vector fields. For example, while it is certainly true that $\vec{A} \cdot \vec{B} = \vec{B} \cdot \vec{A}$ it is not even sensible to ask the question does $\nabla \cdot \vec{A} = \vec{A} \cdot \nabla$? Notice $\nabla \cdot \vec{A}$ is a function while $\vec{A} \cdot \nabla$ is an operator, apples and oranges.

Proposition 3.3.1. *Let f, g, h be real valued functions on $\mathbb{R}^n$ and $\vec{F}, \vec{G}, \vec{H}$ be vector fields on $\mathbb{R}^n$ then (assuming all the partials are well defined)*

$$\begin{aligned}
&(i.)\ \nabla(f+g) = \nabla f + \nabla g \\
&(ii.)\ \nabla \cdot (\vec{F}+\vec{G}) = \nabla \cdot \vec{F} + \nabla \cdot \vec{G} \\
&(iii.)\ \nabla \times (\vec{F}+\vec{G}) = \nabla \times \vec{F} + \nabla \times \vec{G} \\
&(iv.)\ \nabla(fg) = (\nabla f)g + f(\nabla g) \\
&(v.)\ \nabla \cdot (f\vec{F}) = (\nabla f) \cdot \vec{F} + f\nabla \cdot \vec{F} \\
&(vi.)\ \nabla \times (f\vec{F}) = \nabla f \times \vec{F} + f\nabla \times \vec{F} \\
&(vii.)\ \nabla \cdot (\vec{F} \times \vec{G}) = \vec{G} \cdot (\nabla \times \vec{F}) - \vec{F} \cdot (\nabla \times \vec{G}) \\
&(viii.)\ \nabla(\vec{F} \cdot \vec{G}) = \vec{F} \times (\nabla \times \vec{G}) + \vec{G} \times (\nabla \times \vec{F}) + (\vec{F} \cdot \nabla)\vec{G} + (\vec{G} \cdot \nabla)\vec{F} \\
&(ix.)\ \nabla \times (\vec{F} \times \vec{G}) = (\vec{G} \cdot \nabla)\vec{F} - (\vec{F} \cdot \nabla)\vec{G} + \vec{F}(\nabla \cdot \vec{G}) - \vec{G}(\nabla \cdot \vec{F})
\end{aligned} \tag{3.5}$$

Proof: The proofs of (i.),(ii.) and (iii.) are easy, we begin with (iv.),

$$\begin{aligned}
\nabla(fg) &= e_i \partial_i (fg) \\
&= e_i[(\partial_i f)g + f\partial_i g] \qquad \text{ordinary product rule} \\
&= (e_i \partial_i f)g + f(e_i \partial_i g) \\
&= (\nabla f)g + f(\nabla g).
\end{aligned} \tag{3.6}$$

Now consider (vii.), let $\vec{F} = F_i e_i$ and $\vec{G} = G_i e_i$ as usual,

$$\begin{aligned}
\nabla \cdot (\vec{F} \times \vec{G}) &= \partial_k[(\vec{F} \times \vec{G})_k] \\
&= \partial_k[\epsilon_{ijk} F_i G_j] \\
&= \epsilon_{ijk}[(\partial_k F_i)G_j + F_i(\partial_k G_j)] \\
&= \epsilon_{ijk}(\partial_k F_i)G_j - F_i \epsilon_{ikj}(\partial_k G_j) \\
&= \epsilon_{kij}(\partial_k F_i)G_j - F_i \epsilon_{kji}(\partial_k G_j) \\
&= (\nabla \times \vec{F})_j G_j - F_i(\nabla \times \vec{G})_i \\
&= (\nabla \times \vec{F}) \cdot \vec{G} - \vec{F} \cdot (\nabla \times \vec{G}).
\end{aligned} \tag{3.7}$$

The proof of the other parts of this proposition can be handled similarly, although parts (viii) and (ix) require some thought so I'll let you do those for homework.

Proposition 3.3.2. *Let f be a real valued function on $\mathbb{R}^n$ and $\vec{F}$ be a vector field on $\mathbb{R}^n$, both with well defined partial derivatives, then*

$$\begin{aligned}
&\nabla \cdot (\nabla \times \vec{F}) = 0 \\
&\nabla \times \nabla f = 0 \\
&\nabla \times (\nabla \times \vec{F}) = \nabla(\nabla \cdot \vec{F}) - \nabla^2 \vec{F}
\end{aligned} \tag{3.8}$$

Again the proof is similar to those already given in this section and I'll let you prove it in the homework.

Chapter 4

Integral Vector Calculus

If we only dealt with a finite collection of point particles then, for electromagnetism, we would not need the results of this section. However, in nature we do not just concern ourselves with a small number of localized particles. Classical electromagnetism describes how charges and currents that are smeared out effect other objects. Charge and current *densities* replace point charges and currents along a line (with no size). Intuitively we do imagine that these densities are assembled from point-like charges, but often a macroscopic description is more interesting to the macroscopic world of everyday experience. So we will learn that the laws of electromagnetism are phrased in terms of performing integrations of charge and current density. Of course the integral vector calculus has many other applications and we actually include this chapter more for the sake of understanding integration of differential forms later in the course.

4.1 line integrals

Let me give you a brief qualitative run-down of where we are going here: Before we can integrate over a curve we need to know what a curve is. Curves are understood in terms of paths, but then there are two ways to make the path go. For certain types of integrals it matters which direction the path progresses so we will need to insure the path goes in a certain sense. A curve that goes in a certain sense is called an oriented curve. All of these ideas will be required to properly understand line integrals.

Definition 4.1.1. *A* **path** *in $\mathbb{R}^3$ is a continuous function $\phi : I \subset \mathbb{R} \to \mathbb{R}^3$. If $I = [a, b]$ then we say that $\phi(a)$ and $\phi(b)$ are the endpoints of the path ϕ. When ϕ has continuous derivatives of all orders we say it is a smooth path (of class C^∞), if it has at least one continuous derivative we say it is a differentiable path(of class C^1).*

More often than not, when $t \in I \subset \mathbb{R}$ we think of t as time. This is not essential though, we can also use arclength or something else as the parameter for the path. Also, we need not take $I = [a, b]$, it could be some other connected subset of $\mathbb{R}$ like $(0, \infty)$ or (a, b).

Definition 4.1.2. *Let $\phi : [a, b] \to \mathbb{R}^n$ be a differentiable path then the* **length of the path** ϕ *is denoted L_ϕ, defined by,*

$$L_\phi = \int_a^b || \ \phi'(t) \ || dt.$$

The arclength parameter is s which is a function of t,

$$s(t) = \int_a^t || \ \phi'(u) \ ||du.$$

It is the length of ϕ from $\phi(a)$ upto $\phi(t)$.

By the fundamental theorem of calculus we can deduce that,

$$\dot{s} \equiv \frac{ds}{dt} = || \ \phi'(t) \ ||$$

where we have introduced the "$\cdot$ " notation for the time derivative. You can see that if $t = s$ then we will have $ds/dt = 1$, this is why a path parametrized by arclength is called a *unit speed curve.*

Definition 4.1.3. *Let $\phi : [a, b] \to \mathbb{R}^3$ be a differentiable path and suppose that $\phi(I) \subset dom(f)$ for a continuous function $f : dom(f) \to \mathbb{R}$ then the* **scalar line integral of f along** ϕ *is*

$$\int_\phi f \ ds \equiv \int_a^b f(\phi(t)) \ ||\phi'(t)|| \ dt.$$

We can also calculate the scalar line integral of f along some curve which is made of finitely many differentiable segments, we simply calculate each segment's contribution and sum them together. Just like calculating the integral of a piecewise continuous function with a finite number of jump-discontinuities, you break it into pieces.

Notice that if we calculate the scalar line integral of the constant function $f = 1$ then we will obtain the arclength of the curve. More generally the scalar line integral calculates the weighted sum of the values that the function f takes over the path ϕ. If we divide the result by the length of ϕ then we would have the average of f over ϕ.

Definition 4.1.4. *Let $\phi : [a, b] \to \mathbb{R}^3$ be a differentiable path and suppose that $\phi(I) \subset dom(\vec{F})$ for a continuous vector field $\vec{F}$ on $\mathbb{R}^3$ then the* **vector line integral of $\vec{F}$ along** ϕ *is*

$$\int_\phi \vec{F} \cdot d\vec{s} \equiv \int_a^b \vec{F}(\phi(t)) \cdot \phi'(t)dt = \int_\phi (\vec{F} \cdot \vec{T})ds.$$

As the last equality indicates, the vector line integral of $\vec{F}$ is given by the scalar line integral of the tangential component $\vec{F} \cdot \vec{T}$. Thus the vector line integral of $\vec{F}$ along ϕ gives us a measure of how much $\vec{F}$ points in the same direction as the path ϕ. If the vector field always cuts the path perpendicularly (if it was normal to the path) then the vector line integral would be zero.

Work done by force $\vec{F}$

Suppose a particle travels along a path $\phi : [a, b] \to \mathbb{R}^3$ from $\phi(a)$ to $\phi(b)$ then the work done by a force $\vec{F}$ on the particle is defined

$$W = \int_\phi \vec{F} \cdot d\vec{l}$$

Definition 4.1.5. *Let $\phi : [a,b] \to \mathbb{R}^3$ be a differentiable path. We say another differentiable path $\gamma : [c,d] \to \mathbb{R}^3$ is a* **reparametrization** *of ϕ if there exists a bijective (one-one and onto), differentiable function $u : [a,b] \to [c,d]$ with differentiable inverse $u^{-1} : [c,d] \to [a,b]$ such that $\phi(t) = \gamma(u(t))$ for all $t \in [a,b]$.*

Theorem 4.1.6. *Let $\phi : [a,b] \to \mathbb{R}^3$ be a differentiable path and $f : dom(f) \to \mathbb{R}$ a continuous function with $\phi([a,b]) \subset dom(f)$. If $\gamma : [c,d] \to \mathbb{R}^3$ is a reparametrization of ϕ then*

$$\int_\phi f ds = \int_\gamma f ds.$$

If $\vec{F}$ is a continuous vector field with $\phi([a,b]) \subset dom(\vec{F})$ then there are two possibilities,

$$\int_\phi \vec{F} \cdot d\vec{s} = \int_\gamma \vec{F} \cdot d\vec{s} \qquad \gamma \textit{ is } \textbf{orientation preserving}$$

$$\int_\phi \vec{F} \cdot d\vec{s} = -\int_\gamma \vec{F} \cdot d\vec{s} \qquad \gamma \textit{ is } \textbf{orientation reversing}$$

This theorem shows us that we cannot define vector line integrals on just any old set of points, when we wish to disassociate the set of points $\phi[a,b]$ from the particular parametrization we refer to the point set as a **curve** . We see that depending on the type of parametrization we could differ by a sign. This suggests we refine our idea of a curve.

Definition 4.1.7. *A* **simple curve** *$C \subset \mathbb{R}^3$ has no self-intersections. If ϕ is one-one and $C = \phi[a,b]$ for some path ϕ then we say that ϕ parametrizes C. If C is parametrized by $\phi : [a,b] \to \mathbb{R}^3$ and $\phi(a) = \phi(b)$ then it is said to be* **closed***. If we choose an orientation for the curve then it is said to be* **oriented***. Choosing an orientation means we give the curve a direction.*

It's not hard to see that there are only two possible orientations for a simple close curve, clockwise and counterclockwise. However, the terminology "clockwise" and "counterclockwise" is ambiguous without further conventions. Given a particular situation it will be clear.

Definition 4.1.8. *Given a simple curve C with parametrization ϕ and a continuous function $f : dom(f) \to \mathbb{R}$ such that $C \subset dom(f)$ then the* **scalar line integral of f along C** *is defined,*

$$\int_C f ds = \int_\phi f ds.$$

When the curve C is closed we indicate that by replacing $\int$ with $\int$.

Notice that we did not necessarily need that C was oriented, however the condition of non-selfintersection was important as it avoids double counting.

Definition 4.1.9. *Given an oriented simple curve C with a orientation preserving parametrization ϕ and suppose that $\vec{F}$ is a continuous vector field with $C \subset dom(\vec{F})$ then the* **vector line integral of $\vec{F}$ along C** *is defined,*

$$\int_C \vec{F} \cdot d\vec{s} = \int_\phi \vec{F} \cdot d\vec{s}.$$

When the curve C is closed we indicate that by replacing $\int$ with $\int$.

Notice that it is essential to have an oriented curve if we are to calculate the vector line integral of some vector field along the curve. If we did not have an orientation then we could not tell if the vector field is pointing with or against the direction of the curve. We will almost always consider vector line integrals along oriented curves, these are the interesting ones for physics and also the generalized Stokes theorem.

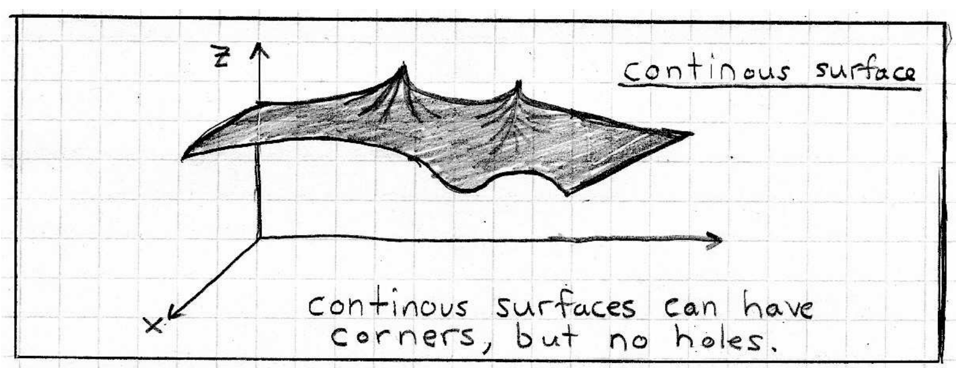
4.2 surface and volume integrals

Before we can integrate over a surface we need to know what in the world a surface is. Much like the previous section we will proceed by a series of refinements till finally arriving at the notion of an oriented surface which is the kind we are most interested in.

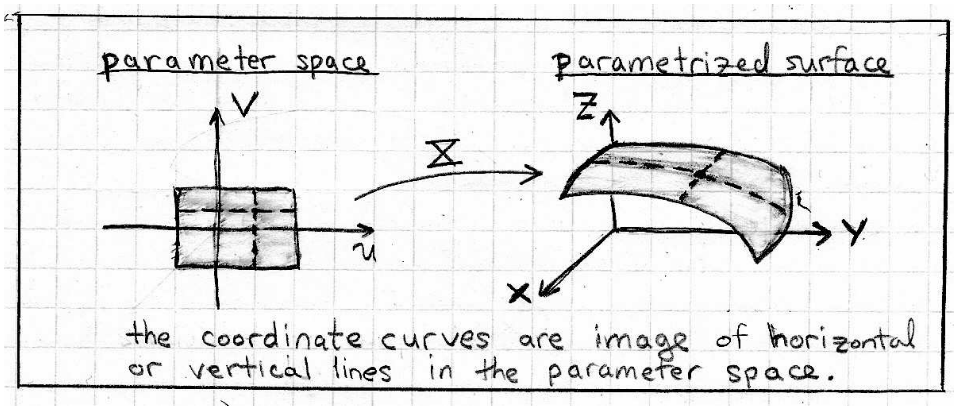
Definition 4.2.1. *Let D be a region in $\mathbb{R}^2$ consisting of an open set and its boundary or partial boundary. A* **parametrized surface** *in $\mathbb{R}^3$ is a continuous function $X : D \subset \mathbb{R}^2 \to \mathbb{R}^3$ that is one-one on D except possibly along the boundary ∂D (such a function is said to be "nearly" one-one). The image $S = X(D)$ is the underlying surface of X denoted by S. We usually employ the following notation for the parametrization for each $(u, v) \in D$,*

$$X(u, v) = (x(u, v), y(u, v), z(u, v))$$

we call $x, y, z : D \to \mathbb{R}$ the coordinate functions of X. We refer to u, v as the parameters and D as the parameter space.



Definition 4.2.2. *The* **coordinate curves through** *$X(a, b) \in X(D)$ are as follows,*
1.) The map $u \mapsto X(u, b)$ gives the **u-coordinate curve** *on S through $X(a, b)$*
2.) The map $v \mapsto X(a, v)$ gives the **v-coordinate curve** *on S through $X(a, b)$*
where we assume that X is a parametrized surface with parameters u, v.



Upto now we have only demanded continuity for the map X, this will allow for ridges and kinks and cusps and such in S. We usually wish for most of our parametrized surface to be free from such defects, so to that end we define the notion of locally smooth next.

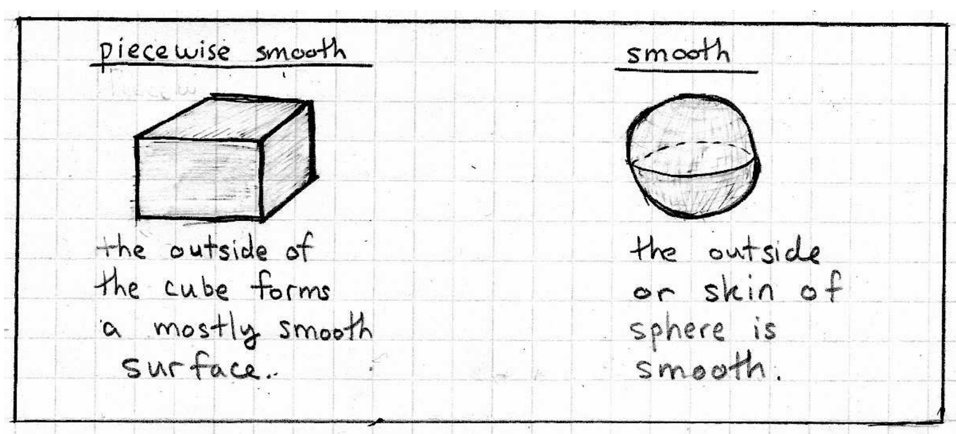
Definition 4.2.3. *The parametrized surface $S = X(D)$ is* **smooth at** $X(a, b)$ *if X is differentiable near (a, b) and if the normal vector $\vec{N}(a, b) \neq 0$ meaning,*

$$\vec{N}(a, b) = \vec{T}_u(a, b) \times \vec{T}_v(a, b)$$

where we have denoted the tangent vectors along the u and v coordinate curves by $\vec{T}_u$ and $\vec{T}_v$ respectively; that is $\vec{T}_u = \partial X/\partial u$ and $\vec{T}_v = \partial X/\partial v$. If S is smooth at every point $X(a, b) \in S$ then S is a **smooth parametrized surface**. *If S is a smooth parametrized surface the (nonzero) vector $\vec{N} = \vec{T}_u \times \vec{T}_v$ is the standard normal vector arising from the parametrization X.*

We do not need smoothness for the whole surface, so long as we have it in the bulk that will do. To be more precise, typical physical examples involve *piecewise smooth* surfaces.

Definition 4.2.4. *A* **piecewise smooth parametrized surface** *is the union of finitely many parametrized surfaces $X_i : D_i \to \mathbb{R}^3$ where $i = 1, 2, ...m < \infty$ and*
(1.) Each D_i is an open set with some of its boundary points
(2.) Each X_i is differentiable and one-one except possibly along ∂D_i
(3.) Each $S_i = X_i(D_i)$ (no sum on i) is smooth except possibly at ∂D_i.



Definition 4.2.5. *The* **surface area** *of a parametrized surface is given by*

$$Area(S) = \int\int_D ||\vec{T}_u \times \vec{T}_v||\ dudv.$$

Given a piecewise smooth surface we sum the areas of the pieces to find the total area.

Definition 4.2.6. *Let $X : D \to \mathbb{R}^3$ be a smooth parametrized surface where $D \subset \mathbb{R}^2$ is a bounded region. Let f be a continuous function with $S = X(D) \subset dom(f)$. Then the* **scalar surface integral of f** *along X is*

$$\int_X f\ dS \equiv \int\int_D f(X(u,v))||\vec{T}_u \times \vec{T}_v||\ dudv.$$

The scalar surface integral over a piecewise smooth surface is found by calculating the contribution of each piece then adding them all together.

Observe that in the case $f = 1$ we recover the surface area. This integral gives a weighted average of the values of f over the surface, if we wanted the average of f on the surface we could divide by the surface area. Incidentally, the following proposition reveals the nuts and bolts of how to actually compute such an integral.

Proposition 4.2.7. *If $X(u,v) = (x(u,v), y(u,v), z(u,v))$ and if we introduce the notation*

$$\frac{\partial(f,g)}{\partial(u,v)} \equiv \frac{\partial f}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial f}{\partial v}\frac{\partial g}{\partial u}$$

Then $\vec{N}(u,v) = \frac{\partial(y,z)}{\partial(u,v)}\hat{\mathbf{i}} - \frac{\partial(x,z)}{\partial(u,v)}\hat{\mathbf{j}} + \frac{\partial(x,y)}{\partial(u,v)}\hat{\mathbf{k}}$ thus

$$\int_X f\ dS = \int\int_D f(X(u,v))\sqrt{\left[\frac{\partial(y,z)}{\partial(u,v)}\right]^2 + \left[\frac{\partial(z,x)}{\partial(u,v)}\right]^2 + \left[\frac{\partial(x,y)}{\partial(u,v)}\right]^2}\ dudv.$$

Definition 4.2.8. *Let $X : D \to \mathbb{R}^3$ be a smooth parametrized surface where $D \subset \mathbb{R}^2$ is a bounded region. Let $\vec{F}$ be a continuous vector field with $S = X(D) \subset dom(\vec{F})$. Then the* **vector surface integral of $\vec{F}$** *along X is*

$$\int_X \vec{F} \cdot d\vec{S} \equiv \int\int_D \vec{F}(X(u,v)) \cdot \vec{N}(u,v)\ dudv.$$

The vector surface integral over a piecewise smooth surface is found by calculating the contribution of each piece then adding them all together.

Let us introduce $\hat{\mathbf{n}}(u,v)$ the unit normal at $X(u,v)$,

$$\hat{\mathbf{n}}(u,v) = \frac{\vec{N}(u,v)}{||\vec{N}(u,v)||}.$$

Observe that the vector surface integral is related to a particular scalar surface integral since $\vec{T}_u \times \vec{T}_v = \vec{N} = ||\vec{N}||\hat{\mathbf{n}} = ||\vec{T}_u \times \vec{T}_v||\hat{\mathbf{n}}$ it should be clear that

$$\int_X \vec{F} \cdot d\vec{S} = \int_X (\vec{F} \cdot \hat{\mathbf{n}})dS.$$

Thus we can see that the vector surface integral takes the weighted sum of the component of the vector field that is normal to the surface. In other words, the vector surface integral gives us a measure of how a vector field cuts through a surface, this quantity is known as the *flux of* $\vec{F}$ *through the surface*

Now we have done everything with respect to a particular parametrization X for the surface $S = X(D)$, but intuitively we should hope to find results that do not depend on the parametrization. It would seem that X is just one picture of S, there could be other parametrizations.

Definition 4.2.9. *Let $X : D_1 \to \mathbb{R}^3$ and $Y : D_2 \to \mathbb{R}^3$ be parametrizations of some surface $S = X(D_1) = Y(D_2)$. We say Y is a* **reparametrization** *of X if there exists a bijection (one-one and onto) function $H : D_2 \to D_1$ such that $Y = X \circ H$. If X and Y are smooth and H is smooth then we say that Y is a* **smooth reparametrization of X***.*

Theorem 4.2.10. *Let $X : D_1 \to \mathbb{R}^3$ and $Y : D_2 \to \mathbb{R}^3$ be parametrizations of some surface S and suppose that f is a continuous function with $S \subset dom(f)$ then,*

$$\int_Y f \ dS = \int_X f \ dS.$$

However if $\vec{F}$ is a continuous vector field with $S \subset dom(\vec{F})$ then two possibilities exist,

$$\int_Y \vec{F} \cdot d\vec{S} = \pm \int_X \vec{F} \cdot d\vec{S}.$$

If we obtain (+) then we say that Y is **orientation preserving***, but if we obtain (-) then we call Y* **orientation reversing***.*

Therefore, just as with curves, we find the integral will not be well defined for just any old surface. We need some system to insure that we pick the correct type of parametrization. Otherwise our answer will depend on the choice of parametrization we made. To avoid this dilemma we define a *two sided surface.*

Definition 4.2.11. *A smooth connected surface S is* **two sided** *(or orientable) if it is possible to define a single unit normal at each point of S so that these vary continuously over S. If S is a orientable surface where a particular choice of unit normal has been established then we say that S* **is an oriented surface***.*

Given this concept we naturally extend the idea of surface integration to oriented surfaces.

Definition 4.2.12. *If S is a smooth oriented surface with unit normal field $\hat{\mathbf{n}}$ on S then if $\vec{F}$ is a continuous vector field with $S \subset dom(\vec{F})$ then define*

$$\int_S \vec{F} \cdot d\vec{S} \equiv \int_S (\vec{F} \cdot \hat{\mathbf{n}}) dS \equiv \int_X (\vec{F} \cdot \hat{\mathbf{n}}) dS.$$

for any parametrization X of S that respects the orientation of S, meaning the normal vectors generated from X emanate from the same side of S as the orienting unit normal field (which we assumed by hypothesis).

We can be certain that this definition is sensible given the theorem above. Because we have assumed the surface is oriented we avoid the $\pm$ ambiguity.

Definition 4.2.13. *Let S be a bounded, piecewise smooth, oriented surface in $\mathbb{R}^3$. Let C be any simple closed curve lying on S. Consider the unit normal $\hat{\mathbf{n}}$ given by the orientation of S, use $\hat{\mathbf{n}}$ to orient the curve by the right hand rule. If the above is true then we say that C* **has orientation induced from the orientation of** *S. Now suppose that the boundary ∂S of S consists of finitely many simple closed curves. Then we say that ∂S* **is oriented consistently** *if each of those simple closed curves has an orientation induced from S.*

4.3 Stokes's and Gauss's theorems

We defer the proof of these theorems to Colley.

Theorem 4.3.1. (Stokes's) *Let S be a bounded, piecewise smooth, oriented surface in $\mathbb{R}^3$. Further suppose that the boundary ∂S of S consists of finitely many piecewise differentiable simple closed curves oriented consistently with S. Let $\vec{F}$ be a differentiable vector field with $S \subset dom(\vec{F})$ then,*

$$\int_S (\nabla \times \vec{F}) \cdot d\vec{S} = \int_{\partial S} \vec{F} \cdot d\vec{S}$$

and,

Theorem 4.3.2. (Gauss's) *Let D be a bounded solid region in $\mathbb{R}^3$ whose boundary ∂D consists of fintely many piecewise smooth, closed, orientable surfaces each of which is oriented with unit normals pointing away from D. Let $\vec{F}$ be a differentiable vector field with $D \subset dom(\vec{F})$ then*

$$\int_D (\nabla \cdot \vec{F}) dV = \int_{\partial D} \vec{F} \cdot d\vec{S}$$

where dV denotes the volume element in D.

hopefully we'll have time to study the generalized Stokes theorem, that theorem includes both of these as subcases and much much more...

Part II

Geometry and Classical Physics

Chapter 5

Maxwell's Equations

The goal of this chapter is to make Maxwell's equations a bit less mysterious to those of you who have not seen them before. Basically this chapter is a brief survey of electromagnetism. For most of this course we will use the *equations* electromagnetism as our primary example. In contrast to a physics course we are not going to see how to solve those equations, rather we will study the general properties and see how to restate the mathematics in terms of differential forms. If I were not to include this chapter then this course would be much more formal, our starting point would just be Maxwell's equations. Hopefully by the end of this chapter you will have some idea of where Maxwell's equations came from and also what they are for.

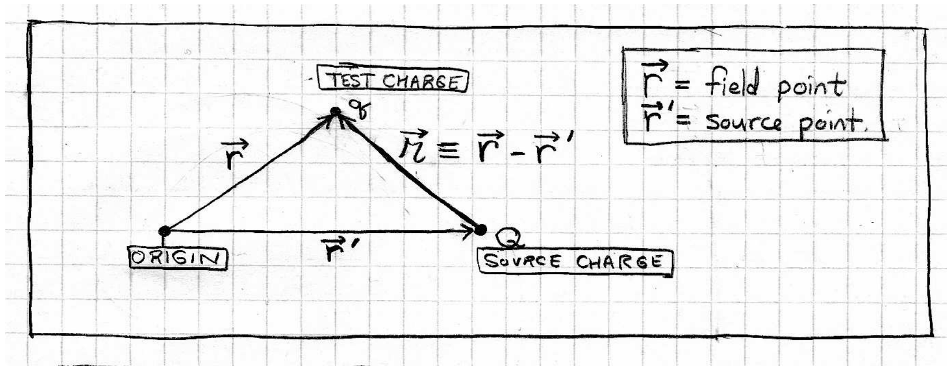
5.1 basics of electricity and magnetism

To begin we will consider the fate of a lonely point charge q in empty space. Suppose that the charge is subject to the influence of an electric field $\vec{E}$ and a magnetic field $\vec{B}$ but it does not contribute to the fields. Such a charge is called a *test charge*. The force imparted on the test charge is

$$\boxed{\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})} \tag{5.1}$$

where $\vec{v}$ is the velocity of the test charge in some inertial coordinate system. We also suppose that the force, electric field and magnetic field are taken with respect to that same inertial frame.

We know that once the force is given we can in principle calculate the trajectory the charge will follow, but to know the force we must find how to determine the electric and magnetic fields. The fields arise from the presence of other charges which we will call *source charges*. Some of the source charges might be in motion, then those would give a *source current*. The basic picture is simply



Given a collection of source charges with certain nice geometrical features we may calculate $\vec{E}$ and $\vec{B}$ according to one of the following laws,

$$\begin{array}{ll} \text{Gauss' Law} & Q_{enc}/\epsilon_o = \Phi_E \\ \text{Ampere's Law} & \mu_o I_{enc} = \int \vec{B} \cdot d\vec{l} \\ \text{Faraday's Law} & \int \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} \end{array} \tag{5.2}$$

where $\Phi_E = \int \vec{E} \cdot d\vec{A}$ is the electric flux and $\Phi_B = \int \vec{B} \cdot d\vec{A}$ is the magnetic flux, both taken over the surface in consideration. We will expand on the precise meaning of these statements in upcoming examples. Our central observation of this chapter is that it can be shown how each of the integral relations derives from one of the following partial differential equations,

$$\begin{array}{ll} \text{Gauss Law} & \nabla \cdot \vec{E} = \rho/\epsilon_o \\ \text{Ampere's Law} & \nabla \times \vec{B} = \mu_o \vec{J} + \mu_o \epsilon_o \frac{\partial \vec{E}}{\partial t} \\ \text{Faradays Law} & \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \\ \text{no magnetic monopoles} & \nabla \cdot \vec{B} = 0 \end{array} \tag{5.3}$$

where ρ is the charge per unit volume, $\vec{J}$ is the current density a.k.a the current per unit area with normal $\frac{1}{|\vec{J}|}\vec{J}$, ϵ_o, μ_o are the permitivity and permeability of free space which are known constants of nature. These equations are known as Maxwell's equations because he was the first to understand them collectively. In fact I should really say that Ampere's Law is just $\nabla \times \vec{B} = \mu_o \vec{J}$ since the other term is due to Maxwell. That other term has profound significance which we will elaborate on soon.

5.2 Gauss law

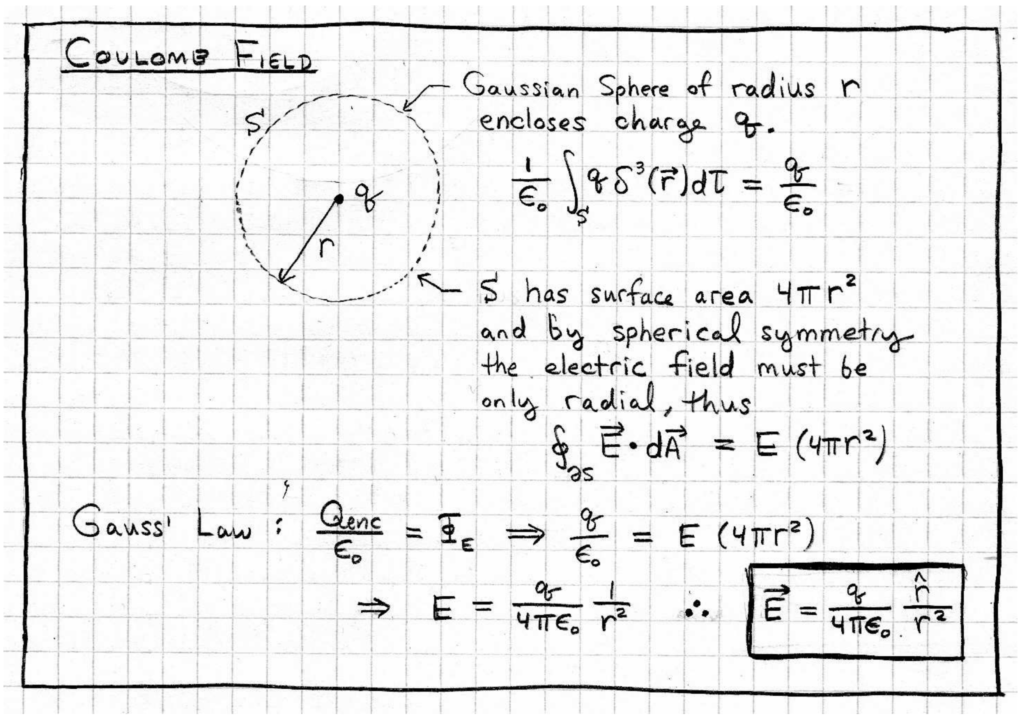
Remark 5.2.1. Gauss' Law: *The integrations below are taken over some arbitrary connected finite region of space where we know that Gauss' theorem of integral vector calculus will apply,*

$$\begin{array}{lll} \nabla \cdot \vec{E} = \rho/\epsilon_o & \Longrightarrow & \frac{1}{\epsilon_o} \int \rho d\tau = \int (\nabla \cdot \vec{E}) d\tau \\ & \Longrightarrow & Q_{enc}/\epsilon_o = \int \vec{E} \cdot d\vec{A} \equiv \Phi_E \end{array} \tag{5.4}$$

Notice that Q_{enc} is simply the total charge enclosed in the region, when we integrate a charge density we get the total charge. The quantity Φ_E is the flux of the electric field through the *boundary of the region. This short calculation establishes that Gauss' equation as a partial differential equation yields the more common expression we used in freshman level physics.*

The example to follow is basic physically speaking, but the mathematics it entails is not really so common to most mathematicians,

Example 5.2.2. *Consider a point charge q positioned at the origin. The charge density is given in terms of the three dimensional Dirac delta function, $\delta^3(\vec{r})$ which is the unusual gadget the satisfies $\int f(\vec{r})\delta^3(\vec{r})d^3\vec{r} = f(\vec{0})$ basically it just evaluates the integrand wherever its argument is zero. Technically it is not a function but rather a distribution. Dirac delta functions have been rigorized by mathematicians to some extent, see the work of Schwarz for example.*



Remark 5.2.3. *The electric field we just found is the Coulomb or monopole field for a static (motionless) point charge at the origin. Notice that if we go back to Gauss' equation $\nabla\cdot\vec{E} = \rho/\epsilon_o$ we'll find an important identity for the delta function as follows,*

$$\nabla \cdot \left(\frac{q}{4\pi\epsilon_o}\frac{\hat{r}}{r^2}\right) = \frac{q}{\epsilon_o}\delta^3(\vec{r})$$

Thus we find,

$$\nabla \cdot \frac{\hat{r}}{r^2} = 4\pi\delta^3(\vec{r}) \tag{5.5}$$

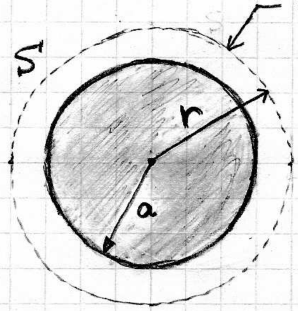
Well use this identity when we investigate the existence of magnetic monopoles.

Example 5.2.4. *Suppose that instead of a point charge we have the same charge smeared out uniformly over some tiny spherical region of radius $a > 0$,*

$$\rho \equiv \begin{cases} q/(\frac{4}{3}\pi a^3), & 0 \leq r \leq a \\ 0, & r > a \end{cases}.$$

Then to find the electric field we apply Gauss law twice, once for inside the sphere and once for outside the sphere. The dotted lines indicate the shells of the so-called Gaussian spheres.

CASE I. $r > a$



Gaussian sphere of radius r. Again by spherical symmetry

$$\Phi_E = 4\pi r^2 E$$

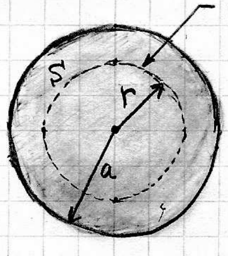
The charge enclosed is the total charge in this case,

$$Q_{enc} = \frac{4}{3}\pi a^3 \cdot \frac{q}{\frac{4}{3}\pi a^3} = q.$$

Therefore by Gauss' Law $4\pi r^2 E = q/\epsilon_o$ thus

$$\boxed{\vec{E} = \frac{q}{4\pi\epsilon_o}\frac{\hat{r}}{r^2} \quad \text{for } r > a}$$

CASE II. $r \leq a$



Gaussian sphere of radius r. Again by spherical symmetry

$$\Phi_E = 4\pi r^2 E$$

The charge enclosed is

$$Q_{enc} = \int_S \frac{q}{\frac{4}{3}\pi a^3} d\tau = \frac{q}{\frac{4}{3}\pi a^3}\frac{4}{3}\pi r^3 = q\frac{r^3}{a^3}.$$

Therefore by Gauss' Law $4\pi r^2 E = qr^3/\epsilon_o a^3$ thus

$$\boxed{\vec{E} = \frac{q}{4\pi\epsilon_o}\frac{r}{a^3}\hat{r} \quad \text{for } r \leq a}$$

We observe that the electric fields of a true idealized point charge and the smeared out point charge considered here look identical for $r \geq a$. In contrast, when $r < a$ the field considered in this example smoothly goes to zero at the origin whereas the Coulomb field diverges. That divergence is why the Dirac delta function comes into play, on the flip-side one could avoid delta functions by insisting that point charges are simply an idealization and that in fact all charges have some finite size. If you choose to insist on finite size then experiment forces you to make that size extremely small. Mathematically the current physical theories treat elementary particles as if they are at a point. Alternative viewpoints exist, string theories give strings a finite size and electrons come from vibrations of those strings, or noncommutative geometry says spacetime itself is granular at some scale so there are no points. It should be emphasized that neither of these have yet to contacted experiment conclusively.

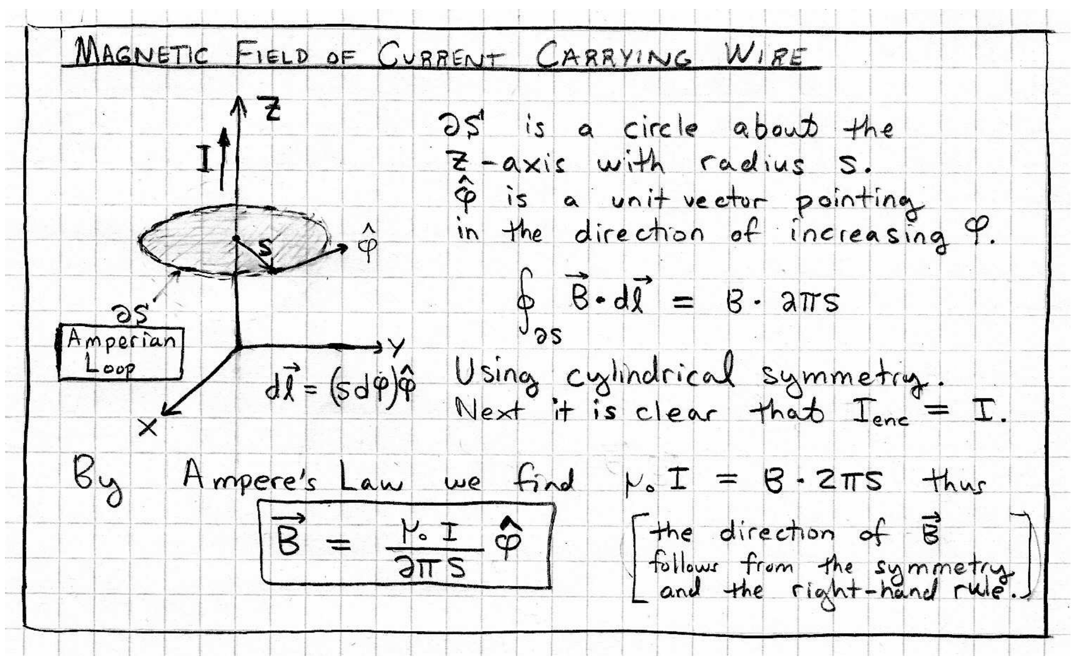
5.3 Ampere's law

Remark 5.3.1. Ampere's Law for steady currents: *The integrations below are taken over some arbitrary closed surface S in space where we know that Stokes theorem of integral vector calculus will apply,*

$$\begin{aligned} \nabla \times \vec{B} = \mu_o \vec{J} \quad &\Longrightarrow \quad \textstyle\int_S (\nabla \times \vec{B}) \cdot d\vec{A} = \mu_o \int_S (\vec{J}) \cdot d\vec{A} \\ &\Longrightarrow \quad \textstyle\int_{\partial S} \vec{B} \cdot d\vec{l} = \mu_o I_{enc} \end{aligned} \tag{5.6}$$

here I_{enc} is the current that cuts through the surface S. It should be emphasized that this rule only applied to steady currents, when currents change we will see things need some modification. This short calculation establishes that Ampere's equation as a partial differential equation yields the more common expression we used in freshman level physics.

Example 5.3.2. *Consider a wire which is very long and carries current in the $z-$ direction along the z-axis. We apply Ampere's Law to obtain the magnetic field at a radial distance s from the z-axis (we use cylindrical coordinates as they are natural for this geometry)*



If we had more time we could uncover the fact that there is a delta function hidden in the current density $\vec{J}$. That fact is reflected in the singular nature of the magnetic field at $s = 0$. We could again replace the idealized wire with a wire which has a finite size and we would find that outside the wire the same magnetic field results, but inside the wire the magnetic field is well-defined everywhere. As a general rule, whenever one considers point charges or line-currents one finds singular fields where those source charges reside.

5.4 Faradays law

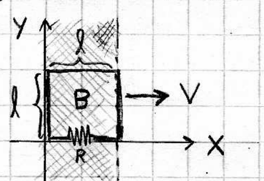
Remark 5.4.1. Faradays Law: *The integrations below are taken over some arbitrary closed* *surface S in space where we know that Stokes theorem of integral vector calculus will apply,*

$$\begin{aligned} \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad &\implies \quad \int_S (\nabla \times \vec{E}) \cdot d\vec{A} = \int_S \left(-\frac{\partial \vec{B}}{\partial t}\right) \cdot d\vec{A} \\ &\implies \quad \int_{\partial S} \vec{E} \cdot d\vec{l} = -\frac{\partial}{\partial t} \int_S \vec{B} \cdot d\vec{A} \\ &\implies \quad \textit{Voltage around } \partial S = -\frac{\partial \Phi_B}{\partial t} \end{aligned} \tag{5.7}$$

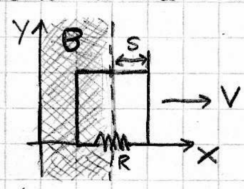
Note that ∂S is the boundary of S. This short calculation establishes that Faradays equation as a partial differential equation yields the more common expression we used in freshman level physics.

Example 5.4.2. *Consider a fixed magnetic field coming out of the page, then let a metallic loop with a resistor with resistance R start completely within the magnetic field at time zero. Then suppose the loop is pulled out of the field as shown with a constant velocity v.*

Suppose there is a constant magnetic field $\vec{B} = B\hat{k}$ for $0 \le x \le l$ then $\vec{B} = 0$ elsewhere. Further suppose a $l \times l$ loop is intially completely immersed in the B-field. At time zero pull the loop out at velocity $\vec{V} = V\hat{i}$. If the loop has a resistor with resistance R and is otherwise a perfect conductor find the current in the loop



(intially)



(at time t)

- Area of loop with B-field is $l(l-s)$. $\therefore$ $\boxed{\Phi_B = Bl(l-s)}$

Notice that $\frac{ds}{dt} = V$ thus $\frac{d\Phi_B}{dt} = \frac{d}{dt}(Bl^2 - Bls) = \boxed{-BlV = \frac{d\Phi_B}{dt}}$

We have taken the normal of the area to be out of the page thus the $(-)$ sign indicates the net B-field out of the page cutting through the loop is decreasing. Lenz' Law says nature abhors a change in flux thus there will be an induced current that increases the flux out of the page, such a current will go in the $\circlearrowleft$ counterclockwise direction, to see what the magnitude of the current is we consult Ampere's Law

$$\mathcal{V} = \text{voltage in loop} = \oint \vec{E} \cdot d\vec{l} = -\frac{d\Phi_B}{dt} = BlV = \mathcal{V}$$

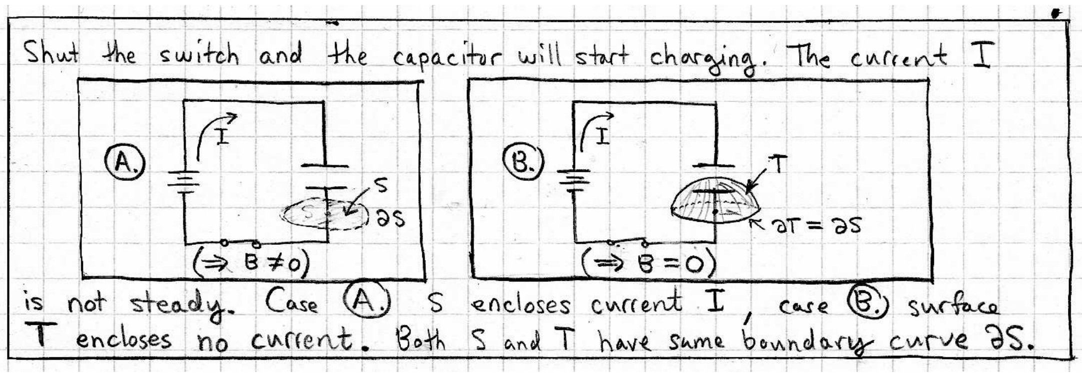
Ohm's Law says $\mathcal{V} = IR$ thus $\boxed{I = \frac{BlV}{R}}$

(we assumed $t < V/l$ so the loop is still partially in field)

Alternatively one could let the loop sit still and vary the magnetic field. This is why one should not put metallic objects in a microwave, the changing magnetic fields induce currents so *large in the metal they will often spark and cause a fire. In a metal loop with no resistor the value of R is very small so the current I is consequently large.*

5.5 Maxwell's correction term

Trouble with naive extension of Ampere's Law Naively one might hope that Ampere's Law held for non-steady currents. But, if we consider the example of a charging capacitor we can quickly deduce that something is missing,



Let us examine how Maxwell framed this inconsistency as a theoretical problem. Maxwell considered the conservation of charge an important principle, let us state it as a partial differential equation,

$$\text{conservation of charge} \qquad \Longleftrightarrow \qquad \nabla \cdot \vec{J} = -\frac{\partial \rho}{\partial t} \tag{5.8}$$

this equation says the charge that leaves some tiny region is equal to the time rate of change of the charge density in that tiny region. Consider that Gauss law indicates that

$$\begin{aligned} \nabla \cdot \vec{J} &= -\tfrac{\partial}{\partial t}(\rho) \\ &= -\epsilon_o \tfrac{\partial}{\partial t}(\nabla \cdot \vec{E}) \\ &= -\epsilon_o \nabla \cdot \tfrac{\partial \vec{E}}{\partial t} \end{aligned} \tag{5.9}$$

But, Ampere's law without Maxwell's correction says $\nabla \times B = \mu_o \vec{J}$. But recall you proved in homework that the divergence of a curl is zero thus,

$$0 = \nabla \cdot (\nabla \times \vec{B}) = \nabla \cdot (\mu_o \vec{J}) \quad \Longrightarrow \quad \nabla \cdot \vec{J} = 0. \tag{5.10}$$

This is an unacceptable result, we cannot demand that $\frac{\partial \rho}{\partial t} = 0$ in all situations. For example, in the charging capacitor the charge is bunching up on the plates as it charges, this means the charge density changes with time.
Maxwell realized that the fix to the problem was to add what he called a displacement current of $J_d = \epsilon_o \frac{\partial E}{\partial t}$ to Ampere's Law,

$$\nabla \times \vec{B} = \mu_o(\vec{J} + \vec{J}_d)$$

although technically speaking $\vec{J}_d$ is not really a current it is sometimes useful for problem solving to think of it that way. Anyway nomenclature aside lets see why charge is conserved now that

Maxwell's correction is added,

$$\begin{aligned} 0 = \nabla \cdot (\nabla \times B) &= \mu_o \nabla \cdot (\vec{J} + \vec{J}_d) \\ &= \mu_o \big(\nabla \cdot \vec{J} + \epsilon_o \nabla \cdot (\tfrac{\partial E}{\partial t})\big) \\ &= \mu_o \big(\nabla \cdot \vec{J} + \epsilon_o \tfrac{\partial}{\partial t}(\nabla \cdot \vec{E})\big) \\ &= \mu_o (\nabla \cdot \vec{J} + \tfrac{\partial \rho}{\partial t}). \end{aligned} \tag{5.11}$$

Thus $\nabla \cdot \vec{J} = -\frac{\partial \rho}{\partial t}$, charge is conserved. The genius of Maxwell in suggesting this modification is somewhat obscured by our modern vector notation. Maxwell did not have the luxury of arguing in the easy compact fashion we have here, only around 1900 did Oliver Heaviside popularize the vector notation we have taken for granted. Maxwell had to argue at the level of components. Maxwell's correction unified the theory of electricity and the theory of magnetism into a single unified field theory which we call electromagnetism. As we will expose later in the course one should not really think of just an electric field or just a magnetic field, they are intertwined and actually manifestations of a single entity. Personally, I find Maxwell's lasting success in resolving a *theoretical inconsistency* one of the better arguments in favor of theoretical research. I mean, experiment has its place but without individuals with creative vision like Maxwell physics wouldn't be nearly as much fun.

5.6 no magnetic monopoles

We have one last equation to explain, $\nabla \cdot \vec{B} = 0$. Let us suppose we have a magnetic monopole. We know from our experience with the Coulomb field it would follow an inverse square law, (this is really a definition of magnetic monopole)

$$\vec{B} = k\frac{\hat{r}}{r^2}$$

Now well show that $\vec{B} = 0$. Observe,

$$0 = \nabla \cdot \vec{B} = k \nabla \cdot \frac{\hat{r}}{r^2} = 4\pi k \delta^3(\vec{r}) \tag{5.12}$$

then integrate over any region containing $\vec{r} = 0$ to see that $4\pi k = 0$ hence there are no nontrivial magnetic monopole fields. This is why sometimes the equation $\nabla \cdot \vec{B} = 0$ is termed the no magnetic monopoles equation. In contrast the Gauss equation $\nabla \cdot \vec{E} = \rho/\epsilon_o$ allowed the electric monopole as a solution.

Remark 5.6.1. *Magnetic monopoles are fascinating objects. Many deep advances is modern field theory are in some way tied to the magnetic monopole. For example in the first half of the 20^{th} century, Dirac treated magnetic monopoles by using strange objects called Dirac Strings (no direct relation to the strings in string theory if you're wondering), one famous result of his approach was that he could explain why magnetic and electric charge came in certain increments. If just one magnetic monopole existed somewhere in the universe one could argue that charge was quantized. By the end of the 1970s it became clear that Dirac's strings were just a weird way of talking about principle fiber bundles. With that viewpoint in place it became possible to rephrase Diracs physical ideas as statements about topology. From this correspondence and its various generalizations many new facts about topology were gleaned from quantum field theory. In other words, people found a way to do math using physics. Many of string theory's most important intellectual contributions to date have this characteristic.*

5.7 overview of potentials in electromagnetism

Maxwell's equations prove intractable to solve directly for many systems, the potential formulation proves more tractable for many problems. One replaces the problem of finding $\vec{E}$ and $\vec{B}$ with the alternate problem of finding a **scalar potential** V and a **vector potential** $\vec{A}$ which are defined such that they give electric and magnetic fields according to the following prescription,

$$\begin{aligned} \vec{E} &= -\nabla V - \tfrac{\partial \vec{A}}{\partial t} \\ \vec{B} &= \nabla \times \vec{A} \end{aligned} \tag{5.13}$$

The potentials are calculated for electrostatics and magnetostatics according to the following formulas,

$$V(\vec{r}) = \frac{1}{4\pi\epsilon_o}\int \frac{\rho(\vec{r})}{\mathfrak{r}} d\tau' \qquad\qquad \vec{A}(\vec{r}) = \frac{\mu_o}{4\pi}\int \frac{\vec{J}(\vec{r})}{\mathfrak{r}} d\tau' \tag{5.14}$$

where $\vec{\mathfrak{r}} = \vec{r} - \vec{r}$ is the vector connecting the source point $\vec{r}$ to the field point $\vec{r}$ and the integrations are to be taken over all space. The formulas above really only work for time-independent source configurations, that is charges that are fixed or moving in steady currents.

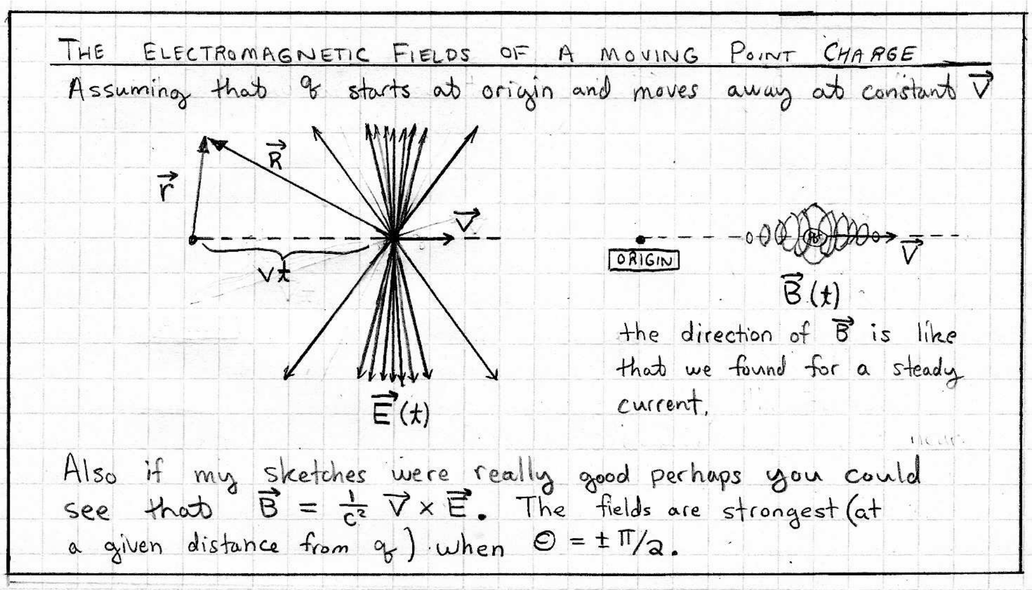
Generally, when the sources are moving one can calculate the potentials by integrating the charge and current densities at *retarded time* t_r which is defined by the time it takes a light-signal travel from the source point to the field point, that is $t_r \equiv t - \mathfrak{r}/c$ and much as before,

$$V(\vec{r},t) = \frac{1}{4\pi\epsilon_o}\int \frac{\rho(\vec{r}\,',\ t_r)}{\mathfrak{r}}\ d\tau' \qquad\qquad \vec{A}(\vec{r},t) = \frac{\mu_o}{4\pi}\int \frac{\vec{J}(\vec{r}\,',\ t_r)}{\mathfrak{r}}\ d\tau' \tag{5.15}$$

These are not simple to work with. If you consult Griffith's you'll see it takes several pages of careful reasoning just to show that the electric field of a charge q moving with constant velocity $\vec{v}$ is,

$$\vec{E}(\vec{r},t) = \frac{q}{4\pi\epsilon_o}\frac{1-v^2/c^2}{[1 - v^2 sin^2(\theta)/c^2]^{3/2}}\frac{\hat{R}}{R^2} \tag{5.16}$$

assuming the charge was at the origin at time zero and $\vec{R} = \vec{r} - t\vec{v}$ is the vector from the current location of the particle to the field point $\vec{r}$ and θ is the angle between $\vec{R}$ and $\vec{v}$.



It is odd that although the integration is based on the retarded time we see no mention of it in the formula above. Somehow the strength of the electric field depends on the current position of the moving charge, relativity could give us an explanation for this. Also it can be shown that the magnetic field due to the same moving charge is,

$$\vec{B} = \frac{1}{c^2}(\vec{v} \times \vec{E}). \tag{5.17}$$

Finally, if $v^2 << c^2$ we recover

$$\vec{E}(r,t) = \frac{q}{4\pi\epsilon_o}\frac{\hat{R}}{R^2} \qquad\qquad \vec{B}(r,t) = \frac{\mu_o q}{4\pi}\frac{1}{R^2}(\vec{v} \times \vec{R}). \tag{5.18}$$

what we can take from these formulas is that a moving charge generates both an electric and a magnetic field. Moreover the faster the charge moves the greater the magnetic field strength.

5.8 summary and apology

So, in these few brief pages I have assaulted you with the combined efforts of hundreds of physicists over about 150 years thought. Most physicists spend at least 5 full courses just thinking about how to solve Maxwell's equations in various contexts, it is not a small matter. Many hours or suffering are probably required to really appreciate the intricacies of these equations. If you wish to learn more about how to solve Maxwell's equations for particular physical situations I recommend taking the junior level sequence out of Griffith's text, *if you have time* . To my taste, electromagnetism has always been he most beautiful subject in physics. It turns out that electromagnetism is the quintessential classical field theory, it stems easily from the principle of locality and symmetry, perhaps well have time to discuss more about that later (that approach is called gauge theory). Anyway all I really want for you to take from this chapter is the names of the equations and some cursory understanding of what they describe. My hope is that this will make later portions of this course a little more down to earth.

Chapter 6

Euclidean Structure and Newton's Laws

This chapter is borrowed from Dr. Fulp's notes word for word in large part, we change some of his notation just a bit, but certainly we are indebted to him for the overall logic employed here.

Although much was known about the physical world prior to Newton that knowledge was highly unorganized and formulated in such a way that is was difficult to use and understand. The advent of Newton changed all that. In 1665-1666 Newton transformed the way people thought about the physical world, years later he published his many ideas in "Principia mathematica philosphiae naturalia" (1686). His contribution was to formulate three basic laws or principles which along with his universal law of gravitation would prove sufficient to derive and explain all mechanical systems both on earth and in the heavens known at the time. These basic laws may be stated as follows:

1. Every particle persists in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by impressed forces.

2. The rate of change of motion is proportional to the motive force impressed; and is made in the direction of the straight line in which that force is impressed.

3. To every action there is an equal reaction; or the mutual actions of two bodies upon each other are always equal but oppositely directed.

Until the early part of this century Newton's laws proved adequate. We now know, however that they are only accurate within prescribed limits. They do not apply for things that are very small like an atom or for things that are very fast like cosmic rays or light itself. Nevertheless Newton's laws are valid for the majority of our common macroscopic experiences in everyday life.

It is implicitly presumed in the formulation of Newton's laws that we have a concept of a straight line, of uniform motion, of force and the like. Newton realized that Euclidean geometry was a necessity in his model of the physical world. In a more critical formulation of Newtonian mechanics one must address the issues implicit in the above formulation of Newton's laws.

This is what we attempt in this chapter, we seek to craft a mathematically rigorous systematic statement of Newtonian mechanics.

6.1 Euclidean geometry

Nowadays Euclidean geometry is imposed on a vector space via an inner product structure. Let $x_1, x_2, x_3, y_1, y_2, y_3, c \in \mathbb{R}$. As we discussed $\mathbb{R}^3$ is the set of 3-tuples and it is a vector space with respect to the operations,

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3)$$

$$c(x_1, x_2, x_3) = (cx_1, cx_2, cx_3)$$

where $x_1, x_2, x_3, y_1, y_2, y_3, c \in \mathbb{R}$. Also we have the dot-product,

$$(x_1, x_2, x_3) \cdot (y_1, y_2, y_3) = x_1y_1 + x_2y_2 + x_3y_3$$

from which the *length* of a vector $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ can be calculated,

$$|x| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + x_3^2}$$

meaning $|x|^2 = x \cdot x$. Also if $x, y \in \mathbb{R}^3$ are nonzero vectors then the angle between them is defined by the formula,

$$\theta = \cos^{-1}\left(\frac{x \cdot y}{|x||y|}\right)$$

In particular nonzero vectors x and y are *perpendicular or orthogonal* iff $\theta = 90^o$ which is so iff $\cos(\theta) = 0$ which is turn true iff $x \cdot y = 0$.

Definition 6.1.1. *A function $L : \mathbb{R}^3 \to \mathbb{R}^3$ is said to be a* **linear transformation** *if and only if there is a 3×3 matrix A such that $L(x) = Ax$ for all $x \in \mathbb{R}^3$. Here Ax indicates multiplication by the matrix A on the column vector x (alternatively could also formulate everything in terms of rows)*

Definition 6.1.2. *An* **orthogonal transformation** *is a linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^3$ which satisfies*

$$L(x) \cdot L(y) = x \cdot y$$

for all $x, y \in \mathbb{R}^3$. Such a transformation is also called an **linear isometry of the Euclidean metric**.

The term *isometry* means the same measure, you can see why that's appropriate from the following,

$$|L(x)|^2 = L(x) \cdot L(x) = x \cdot x = |x|^2$$

for all $x \in \mathbb{R}^3$. Taking the square root of both sides yields $|L(x)| = |x|$; an orthogonal transformation preserves the lengths of vectors in $\mathbb{R}^3$. Using what we just learned its easy to show orthogonal transformations preserve angles as well,

$$\cos(\theta_L) = \frac{L(x) \cdot L(y)}{|L(x)||L(y)|} = \frac{x \cdot y}{|x||y|} = \cos(\theta)$$

Hence taking the inverse cosine of each side reveals that the angle θ_L between $L(x)$ and $L(y)$ is equal to the angle θ between x and y; $\theta_L = \theta$. Orthogonal transformations preserve angles.

Proposition 6.1.3. *Let l be a line in $\mathbb{R}^3$; that is there exist $a, v \in \mathbb{R}^3$ so that*

$$l = \{x \in \mathbb{R}^2 \mid x = a + tv \quad t \in \mathbb{R}\} \qquad \qquad \textit{definition of a line in } \mathbb{R}^3.$$

If L is an orthonormal transformation then $L(l)$ is also a line in $\mathbb{R}^3$.

To prove this we simply need to find new a' and v' in $\mathbb{R}^3$ to demonstrate that $L(l)$ is a line. Take a point on the line, $x \in l$

$$\begin{aligned} L(x) &= L(a + tv) \\ &= L(a) + tL(v) \end{aligned} \tag{6.1}$$

thus $L(x)$ is on a line described by $x = L(a) + tL(v)$, so we can choose $a' = L(a)$ and $v' = L(v)$ it turns out; $L(l) = \{x \in \mathbb{R}^3 \mid x = a' + tv'\}$.

If one has a coordinate system with unit vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ along three mutually orthogonal axes then an orthogonal transformation will create three new mutually orthogonal unit vectors $L(\hat{\mathbf{i}}) = \hat{\mathbf{i}}', L(\hat{\mathbf{j}}) = \hat{\mathbf{j}}', L(\hat{\mathbf{k}}) = \hat{\mathbf{k}}'$ upon which one could lay out new coordinate axes. In this way orthogonal transformations give us a way of constructing new "rotated" coordinate systems from a given coordinate system. Moreover, it turns out that Newton's laws are preserved (have the same form) under orthogonal transformations. Transformations which are not orthogonal can greatly distort the form of Newton's laws.

Remark 6.1.4. *If we view vectors in $\mathbb{R}^3$ as column vectors then the dot-product of x with y can be written as $x \cdot y = x^t y$ for all $x, y \in \mathbb{R}^3$. Recall that x^t is the transpose of x, it changes the column vector x to the corresponding row vector x^t.*

Let us consider an orthogonal transformation $L : \mathbb{R}^3 \to \mathbb{R}^3$ where $L(x) = Ax$. What condition on the matrix A follows from the the L being an orthogonal transformation ?

$$\begin{aligned} L(x) \cdot L(y) = x \cdot y \quad &\iff (Ax)^t(Ay) = x^t y \\ &\iff x^t(A^tA)y = x^t y \\ &\iff x^t(A^tA)y = x^t I y \\ &\iff x^t(A^tA - I)y = 0. \end{aligned} \tag{6.2}$$

But $x^t(A^tA - I)y = 0$ for *all* $x, y \in \mathbb{R}^3$ iff $A^tA - I = 0$ or $A^tA = I$. Thus L is orthogonal iff its matrix A satisfies $A^tA = I$. This is in turn equivalent to A having an inverse and $A^{-1} = A^t$.

Claim 6.1.5. *The set of* **orthogonal transformations** *on $\mathbb{R}^3$ is denoted O(3). The operation of function composition on O(3) makes it a group. Likewise we also denote the set of all* **orthogonal matrices** *by O(3),*

$$O(3) = \{A \in \mathbb{R}^{3\times 3} \mid A^tA = I\}$$

it is also a group under matrix multiplication.

Usually we will mean the matrix version, it should be clear from the context, it's really just a question of notation since we know that L and A contain the same information thanks to linear algebra. Recall that every linear transformation L on a finite dimensional vector space can be represented by matrix multiplication of some matrix A.

Claim 6.1.6. *The set of* **special orthogonal matrices** *on* $\mathbb{R}^3$ *is denoted SO(3),*

$$SO(3) = \{A \in \mathbb{R}^{3\times 3} \mid A^t A = I \text{ and } det(A) = 1\}$$

it is also a group under matrix multiplication and thus it is a subgroup of O(3). It is shown in standard linear algebra course that every special orthogonal matrix rotates $\mathbb{R}^3$ *about some line. Thus, we will often refer to SO(3) as the* **group of rotations**.

There are other transformations that do not change the geometry of $\mathbb{R}^3$.

Definition 6.1.7. *A* **translation** *is a function* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *defined by* $T(x) = x + v$ *where* v *is some fixed vector in* $\mathbb{R}^3$ *and* x *is allowed to vary over* $\mathbb{R}^3$.

Clearly translations do not change the distance between two points $x, y \in \mathbb{R}^3$,

$$|T(x) - T(y)| = |x + v - (y - v)| = |x - y| = \text{ distance between } x \text{ and } y.$$

Also if x, y, z are points in $\mathbb{R}^3$ and θ is the angle between $y - x$ and $z - x$ then θ is also the angle between $T(y) - T(x)$ and $T(z) - T(x)$. Geometrically this is trivial, if we shift all points by the same vector then the difference vectors between points are unchanged thus the lengths and angles between vectors connecting points in $\mathbb{R}^3$ are unchanged.

Definition 6.1.8. *A function* $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ *is called a* **rigid motion** *if there exists a vector* $r \in \mathbb{R}^3$ *and a rotation matrix* $A \in SO(3)$ *such that* $\phi(x) = Ax + r$.

A rigid motion is the composite of a translation and a rotation therefore it will clearly preserve lengths and angles in $\mathbb{R}^3$. So rigid motions are precisely those transformations which preserve Euclidean geometry and consequently they are the transformations which will preserve Newton's laws. If Newton's laws hold in one coordinate system then we will find Newton's laws are also valid in a new coordinate system iff it is related to the original coordinate system by a rigid motion. We now proceed to provide a careful exposition of the ingredients needed to give a rigorous formulation of Newton's laws.

Definition 6.1.9. *We say that* $\mathcal{E}$ *is an* **Euclidean structure** *on a set* S *iff* $\mathcal{E}$ *is a family of bijections from* S *onto* $\mathbb{R}^3$ *such that,*
(1.) $\mathcal{X}, \mathcal{Y} \in \mathcal{E}$ *then* $\mathcal{X} \circ \mathcal{Y}^{-1}$ *is a rigid motion.*
(2.) if $\mathcal{X} \in \mathcal{E}$ *and* ϕ *is a rigid motion then* $\phi \circ \mathcal{X} \in \mathcal{E}$.
Also a **Newtonian space** *is an ordered pair* $(S, \mathcal{E})$ *where* S *is a set and* $\mathcal{E}$ *is an Euclidean structure on* S.

Notice that if $\mathcal{X}, \mathcal{Y} \in \mathcal{E}$ then there exists an $A \in SO(3)$ and a vector $r \in \mathbb{R}^3$ so that we have $\mathcal{X}(p) = A\mathcal{Y}(p) + r$ for every $p \in S$. Explicitly in cartesian coordinates on $\mathbb{R}^3$ this means,

$$[\mathcal{X}_1(p), \mathcal{X}_2(p), \mathcal{X}_3(p)]^t = A[\mathcal{Y}_1(p), \mathcal{Y}_2(p), \mathcal{Y}_3(p)]^t + [r_1, r_2, r_3]^t.$$

Newtonian space is the mathematical model of space which is needed in order to properly formulate Newtonian mechanics. The first of Newton's laws states that an object which is subject to no forces must move along a straight line. This means that some observer should be able to show that the object moves along a line in space. We take this to mean that the observer chooses an inertial frame and makes measurements to decide wether or not the object executes straight line motion in the coordinates defined by that frame. If the observations are to be frame independent then the notion of a straight line in space should be independent of which inertial coordinate system is used to make the measurements. We intend to identify inertial coordinate systems as precisely those elements of $\mathcal{E}$. Thus we need to show that if l is a line as measured by $\mathcal{X} \in \mathcal{E}$ then l is also a line as measured by $\mathcal{Y} \in \mathcal{E}$.

Definition 6.1.10. *Let $(S, \mathcal{E})$ be a Newtonian space. A subset l of S is said to be* **a line in S** *iff $\mathcal{X}(l)$ is a line in $\mathbb{R}^3$ for some choice of $\mathcal{X} \in \mathcal{E}$.*

The theorem below shows us that the choice made in the definition above is not special. In fact our definition of a line in S is coordinate independent. Mathematicians almost always work towards formulating geometry in a way which is independent of the coordinates employed, this is known as the coordinate free approach. Physicists in contrast almost always work in coordinates.

Theorem 6.1.11. *If l is a line in a Newtonian space $(S, \mathcal{E})$ then $\mathcal{Y}(l)$ is a line in $\mathbb{R}^3$ for* **every** *$\mathcal{Y} \in \mathcal{E}$.*

Proof: Because l is a line in the S we know there exists $\mathcal{X} \in \mathcal{E}$ and $\mathcal{X}(l)$ is a line in $\mathbb{R}^3$. Let $\mathcal{Y} \in \mathcal{E}$ observe that,

$$\mathcal{Y}(l) = (\mathcal{Y} \circ \mathcal{X}^{-1} \circ \mathcal{X})(l) = (\mathcal{Y} \circ \mathcal{X}^{-1})(\mathcal{X}(l)).$$

Now since $\mathcal{X}, \mathcal{Y} \in \mathcal{E}$ we have that $\mathcal{Y} \circ \mathcal{X}^{-1}$ is a rigid motion on $\mathbb{R}^3$. Thus if we can show that rigid motions take lines to lines in $\mathbb{R}^3$ the proof will be complete. We know that there exist $A \in SO(3)$ and $r \in \mathbb{R}^3$ such that $(\mathcal{Y} \circ \mathcal{X}^{-1})(x) = Ax + r$. Let $x \in \mathcal{X}(l) = \{x \in \mathbb{R}^3 \mid x = p + tq \;\; t \in \mathbb{R}$ and p,q are fixed vectors in $\mathbb{R}^3\}$, consider

$$\begin{aligned}
(\mathcal{Y} \circ \mathcal{X}^{-1})(x) &= Ax + r \\
&= A(p + tq) + r \\
&= (Ap + r) + tAq \\
&= p' + tq' \qquad\qquad \text{letting } p' = Ap + r \text{ and } q' = Aq.
\end{aligned} \tag{6.3}$$

The above hold for all $x \in \mathcal{X}(l)$, clearly we can see the line has mapped to a new line $\mathcal{Y}(l) = \{x \in \mathbb{R}^3 \mid x = p' + tq' \;\;, t \in \mathbb{R}\}$. Thus we find what we had hoped for, lines are independent of the frame chosen from $\mathcal{E}$ in the sense that a line is always a line no matter which element of $\mathcal{E}$ describes it.

Definition 6.1.12. *An* **observer** *is a function from an interval $I \subset \mathbb{R}$ into $\mathcal{E}$. We think of such a function $\mathcal{X} : I \to \mathcal{E}$ as being a time-varying coordinate system on S. For each $t \in I$ we denote $\mathcal{X}(t)$ by $\mathcal{X}_t$; thus $\mathcal{X}_t : S \to \mathbb{R}^3$ for each $t \in I$ and $\mathcal{X}_t(p) = [\mathcal{X}_{t1}(p), \mathcal{X}_{t2}(p), \mathcal{X}_{t3}(p)]$ for all $p \in S$.*

Assume that a material particle or more generally a "point particle" moves in space S in such a way that at time t the particle is centered at the point $\gamma(t)$. Then the mapping $\gamma : I \to S$ will be called the **trajectory** of the particle.

Definition 6.1.13. *Let us consider a particle with trajectory $\gamma : I \to S$. Further assume we have an observer $\mathcal{X} : I \to \mathcal{E}$ with $t \mapsto \mathcal{X}_t$ then:*

(1.) $\mathcal{X}_t(\gamma(t))$ *is the* **position vector** *of the particle at time $t \in I$ relative to the observer $\mathcal{X}$.*

(2.) $\frac{d}{dt}\big[\mathcal{X}_t(\gamma(t))\big]|_{t=t_o}$ *is called the* **velocity** *of the particle at time* $t_o \in I$ *relative to the observer* $\mathcal{X}$*, it is denoted* $v_{\mathcal{X}}(t_o)$*.*

(3.) $\frac{d^2}{dt^2}\big[\mathcal{X}_t(\gamma(t))\big]|_{t=t_o}$ *is called the* **acceleration** *of the particle at time* $t_o \in I$ *relative to the observer* $\mathcal{X}$*, it is denoted* $a_{\mathcal{X}}(t_o)$*.*

Notice that position, velocity and acceleration are only defined with respect to an observer. We now will calculate how position, velocity and acceleration of a particle with trajectory $\gamma : I \to S$ relative to observer $\mathcal{Y} : I \to \mathcal{E}$ compare to those of another observer $\mathcal{X} : I \to \mathcal{E}$. To begin we note that each particular $t \in I$ we have $\mathcal{X}_t, \mathcal{Y}_t \in \mathcal{E}$ thus there exists a rotation matrix $A(t) \in SO(3)$ and a vector $v(t) \in \mathbb{R}^3$ such that,

$$\mathcal{Y}_t(p) = A(t)\mathcal{X}_t(p) + r(t)$$

for all $p \in S$. As we let t vary we will in general find that $A(t)$ and $r(t)$ vary, in other words we have A a matrix-valued function of time given by $t \mapsto A(t)$ and r a vector-valued function of time given by $t \mapsto r(t)$. Also note that the *origin* of the coordinate coordinate system $\mathcal{X}(p) = 0$ moves to $\mathcal{Y}(p) = r(t)$, this shows that the correct interpretation of $r(t)$ is that it is the position of the old coordinate's origin in the new coordinate system. Consider then $p = \gamma(t)$,

$$\boxed{\mathcal{Y}_t(\gamma(t)) = A(t)\mathcal{X}_t(\gamma(t)) + r(t)} \tag{6.4}$$

this equation shows how the position of the particle in $\mathcal{X}$ coordinates transforms to the new position in $\mathcal{Y}$ coordinates. We should not think that the particle has moved under this transformation, rather we have just changed our viewpoint of where the particle resides. Now move on to the transformation of velocity, (we assume the reader is familiar with differentiating matrix valued functions of a real variable, in short we just differentiate component-wise)

$$\begin{aligned} v_{\mathcal{Y}}(t) &= \tfrac{d}{dt}\big[\mathcal{Y}(\gamma(t))\big] \\ &= \tfrac{d}{dt}\big[A(t)\mathcal{X}_t(\gamma(t)) + r(t)\big] \\ &= \tfrac{d}{dt}[A(t)]\mathcal{X}_t(\gamma(t)) + A(t)\tfrac{d}{dt}[\mathcal{X}_t(\gamma(t))] + \tfrac{d}{dt}[r(t)] \\ &= A'(t)\mathcal{X}_t(\gamma(t)) + A(t)v_{\mathcal{X}}(t) + r'(t). \end{aligned} \tag{6.5}$$

Recalling the dot notation for time derivatives and introducing $\gamma_{\mathcal{X}} = \mathcal{X} \circ \gamma$,

$$\boxed{v_{\mathcal{Y}} = \dot{A}\gamma_{\mathcal{X}} + Av_{\mathcal{X}} + \dot{r}.} \tag{6.6}$$

We observe that the velocity according to various observes depends not only on the trajectory itself, but also the time evolution of the observer itself. The case $A = I$ is more familiar, since $\dot{A} = 0$ we have,

$$v_{\mathcal{Y}} = Iv_{\mathcal{X}} + \dot{r} = v_{\mathcal{X}} + \dot{r}. \tag{6.7}$$

The velocity according to the observer $\mathcal{Y}$ moving with velocity $\dot{r}$ relative to $\mathcal{X}$ is the sum of the velocity according to $\mathcal{X}$ and the velocity of the observer $\mathcal{Y}$. Obviously when $A \neq I$ the story is more complicated, but the case $A = I$ should be familiar from freshman mechanics.
Now calculate how the accelerations are connected,

$$\begin{aligned} a_{\mathcal{Y}}(t) &= \tfrac{d^2}{dt^2}\big[\mathcal{Y}(\gamma(t))\big] \\ &= \tfrac{d}{dt}\big[A'(t)\mathcal{X}_t(\gamma(t)) + A(t)v_{\mathcal{X}}(t) + r'(t)\big] \\ &= A''(t)\mathcal{X}_t(\gamma(t)) + A'(t)\tfrac{d}{dt}[\mathcal{X}_t(\gamma(t))] + A'(t)v_{\mathcal{X}}(t) + A(t)\tfrac{d}{dt}[v_{\mathcal{X}}(t)] + r''(t) \\ &= A''(t)\mathcal{X}_t(\gamma(t)) + 2A'(t)v_{\mathcal{X}}(t) + +A(t)a_{\mathcal{X}}(t) + r''(t) \end{aligned} \tag{6.8}$$

Therefore we relate acceleration in $\mathcal{X}$ to the acceleration in $\mathcal{Y}$ as follows,

$$\boxed{a_{\mathcal{Y}} = Aa_{\mathcal{X}} + \ddot{r} + \ddot{A}\gamma_{\mathcal{X}} + 2\dot{A}v_{\mathcal{X}}.} \tag{6.9}$$

The equation above explains many things, if you take the junior level classical mechanics course you'll see what those things are. This equation does not look like the one used in mechanics for noninertial frames, it is nevertheless the same and if you're interested I'll show you.

Definition 6.1.14. *If $\gamma : I \to S$ is the trajectory of a particle then we say the particle and $\mathcal{X} : I \to \mathcal{E}$ is an observer. We say the particle is in a* **state of rest** *relative to the observer $\mathcal{X}$ iff $v_{\mathcal{X}} = \frac{d}{dt}[\mathcal{X}_t(\gamma(t))] = 0$. We say the particle experiences* **uniform rectilinear motion** *relative to the observer $\mathcal{X}$ iff $t \mapsto \mathcal{X}_t(\gamma(t))$ is a straight line in $\mathbb{R}^3$ with velocity vector some nonzero constant vector.*

We now give a rigorous definition for the existence of *force*, a little later we'll say how to calculate it.

Definition 6.1.15. *A particle* **experiences a force** *relative to an observer $\mathcal{X}$ iff the particle is neither in a state of rest nor is it in uniform rectilinear motion relative to $\mathcal{X}$. Otherwise we say the particle experiences no force relative to $\mathcal{X}$.*

Definition 6.1.16. *An observer $\mathcal{X} : I \to \mathcal{E}$ is said to be an* **inertial observer** *iff there exists $\mathcal{X}_o \in \mathcal{E}$, $A \in SO(3)$, $v, w \in \mathbb{R}^3$ such that $\mathcal{X}_t = A\mathcal{X}_o + tv + w$ for all $t \in I$. A particle is called a* **free particle** *iff it experiences no acceleration relative to an inertial observer.*

Observe that a constant mapping into $\mathcal{E}$ is an inertial observer and that general inertial observers are observers which are in motion relative to a "stationary observer" but the motion is "constant velocity" motion. We will refer to a constant mapping $\mathcal{X} : I \to \mathcal{E}$ as a **stationary observer**.

Theorem 6.1.17. *If $\mathcal{X} : I \to \mathcal{E}$ and $\mathcal{Y} : I \to \mathcal{E}$ are inertial observers then there exists $A \in SO(3)$, $v, w \in \mathbb{R}^3$ such that $\mathcal{Y}_t = A\mathcal{X}_t + tv + w$ for all $t \in I$. Moreover if a particle experiences no acceleration relative to $\mathcal{X}$ then it experiences no acceleration relative to $\mathcal{Y}$.*

Proof: Since $\mathcal{X}$ and $\mathcal{Y}$ are inertial we have that there exist $\mathcal{X}_o$ and $\mathcal{Y}_o$ in $\mathcal{E}$ and fixed vectors $v_x, w_x, v_y, w_y \in \mathbb{R}^3$ and particular rotation matrices $A_x, A_y \in SO(3)$ such that

$$\mathcal{X}_t = A_x\mathcal{X}_o + tv_x + w_x \qquad\qquad \mathcal{Y}_t = A_y\mathcal{Y}_o + tv_y + w_y.$$

Further note that since $\mathcal{X}_o, \mathcal{Y}_o \in \mathcal{E}$ there exists fixed $Q \in SO(3)$ and $u \in \mathbb{R}^3$ such that $\mathcal{Y}_o = Q\mathcal{X}_o + u$. Thus, noting that $\mathcal{X}_o = A_x^{-1}(\mathcal{X}_t - tv_x - w_x)$ for the fourth line,

$$\begin{aligned} \mathcal{Y}_t &= A_y\mathcal{Y}_o + tv_y + w_y \\ &= A_y(Q\mathcal{X}_o + u) + tv_y + w_y \\ &= A_yQ\mathcal{X}_o + A_yu + tv_y + w_y \\ &= A_yQA_x^{-1}(\mathcal{X}_t - tv_x - w_x) + tv_y + A_yu + w_y \\ &= A_yQA_x^{-1}\mathcal{X}_t + t[v_y - A_yQA_x^{-1}v_x] - A_yQA_x^{-1}w_x + A_yu + w_y \end{aligned} \tag{6.10}$$

Thus define $A = A_yQA_x^{-1} \in SO(3)$, $v = v_y - A_yQA_x^{-1}v_x$, and $w = -A_yQA_x^{-1}w_x + A_yu + w_y$. Clearly $v, w \in \mathbb{R}^3$ and it is a short calculation to show that $A \in SO(3)$, we've left it as an exercise to the reader but it follows immediately if we already know that $SO(3)$ is a group under matrix multiplication (we have not proved this yet). Collecting our thoughts we have established the first half of the theorem, there exist $A \in SO(3)$ and $v, w \in \mathbb{R}^3$ such that,

$$\mathcal{Y}_t = A\mathcal{X}_t + tv + w$$

Now to complete the theorem consider a particle with trajectory $\gamma : I \to S$ such that $a_{\mathcal{X}} = 0$. Then by eqn.[6.9] we find, using our construction of A, v, w above,

$$\begin{aligned} a_{\mathcal{Y}} &= Aa_{\mathcal{X}} + \ddot{r} + \ddot{A}\gamma_{\mathcal{X}} + 2\dot{A}v_{\mathcal{X}} \\ &= A0 + 0 + 0\gamma_{\mathcal{X}} + 2(0)v_{\mathcal{X}} \\ &= 0. \end{aligned} \tag{6.11}$$

Therefore if the acceleration is zero relative to a particular inertial frame then it is zero for **all** inertial frames.

Consider that if a particle is either in a state of rest or uniform rectilinear motion then we can express it's trajectory γ relative to an observer $\mathcal{X} : I \to S$ by

$$\mathcal{X}_t(\gamma(t)) = tv + w$$

for all $t \in I$ and fixed $v, w \in \mathbb{R}^3$. In fact if $v = 0$ the particle is in a state of rest, whereas if $v \neq 0$ the particle is in a state of uniform rectilinear motion. Moreover,

$$\gamma_{\mathcal{X}}(t) = tv + w \iff v_{\mathcal{X}} = v \iff a_{\mathcal{X}} = 0.$$

Therefore we have shown that according to any inertial frame a particle that has zero acceleration necessarily travels in rectilinear motion or stays at rest.

Let us again ponder Newton's laws.

1. **Newton's First Law** Every particle persists in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by impressed forces.

2. **Newton's Second Law** The rate of change of motion is proportional to the motive force impressed; and is made in the direction of the straight line in which that force is impressed.

3. **Newton's Third Law** To every action there is an equal reaction; or the mutual actions of two bodies upon each other are always equal but oppositely directed.

It is easy to see that if the first law holds relative to one observer then it does not hold relative to another observer which is rotating relative to the first observer. So a more precise formulation of the first law would be that it holds relative to *some* observer, or some class of observers, but not relative to all observers. *We have just shown that if $\mathcal{X}$ is an inertial observer then a particle is either in a state of rest or uniform rectilinear motion relative to $\mathcal{X}$ iff its acceleration is zero.* If γ is the trajectory of the particle the second law says that the force F acting on the body is proportional to $m(dv_{\mathcal{X}}/dt) = ma_{\mathcal{X}}$. Thus the second law says that a body has zero acceleration iff the force acting on the body is zero (assuming $m \neq 0$). It seems to follow that the first law is a consequence of the second law. What then *does* the first law say that is *not* contained in the second law ?

The answer is that the *first law is not a mathematical axiom but a physical principle.* It says it should be *possible to physically construct, at least in principle, a set* of coordinate systems at each instant of time which may be modeled by the mathematical construct we have been calling an inertial observer. Thus the *first law can be reformulated* to read:

There exists an inertial observer

The second law is also subject to criticism. When one speaks of the force on a body what is it that one is describing? Intuitively we think of a force as something which pushes or pulls the particle off its natural course.

The truth is that a course which seems natural to one observer may not appear natural to another. One usually models forces as vectors. These vectors provide the push or pull. The components of a vector in this context are observer dependent. The second law could almost be relegated to a definition. The force on a particle at time t would be defined to be $ma_{\mathcal{X}}(t)$ relative to the observer $\mathcal{X}$. Generally physicists require that the second law hold **only** for inertial observers. One reason for this is that if $F_{\mathcal{X}}$ is the force on a particle according to an inertial observer $\mathcal{X}$ and $F_{\mathcal{Y}}$ is the force on the same particle measured relative to the inertial observer $\mathcal{Y}$ then we claim $F_{\mathcal{Y}} = AF_{\mathcal{X}}$ where $\mathcal{X}$ and $\mathcal{Y}$ are related by

$$\mathcal{Y}_t = A\mathcal{X}_t + tv + w$$

for $v, w \in \mathbb{R}^3$ and $A \in SO(3)$ and for all t. Consider a particle traveling the trajectory γ we find it's accelerations as measured by $\mathcal{X}$ and $\mathcal{Y}$ are related by,

$$a_{\mathcal{Y}} = Aa_{\mathcal{X}}$$

where we have used eqn.[6.9] for the special case that A is a fixed rotation matrix and $r = tv+w$. Multiply by the mass to obtain that $ma_{\mathcal{Y}} = A(ma_{\mathcal{X}})$ thus $F_{\mathcal{Y}} = AF_{\mathcal{X}}$. Thus the form of Newton's law is maintained under admissible transformations of observer.

Remark 6.1.18. *The invariance of the form of Newton's laws in any inertial frame is known as the Galilean relativity principle. It states that no inertial frame is preferred in the sense that the physical laws are the same no matter which inertial frame you take observations from. This claim is limited to mechanical or electrostatic forces. The force between to moving charges due to a magnetic field does not act along the straight line connecting those charges. This exception was important to Einstein conceptually. Notice that if no frame is preferred then we can never, taking observations solely within an inertial frame, deduce the velocity of that frame. Rather we only can deduce relative velocities by comparing observations from different frames.*

In contrast, if one defines the force relative to one observer $\mathcal{Z}$ which is rotating relative to $\mathcal{X}$ by $F_{\mathcal{Z}} = ma_{\mathcal{Z}}$ then one obtains a much more complex relation between $F_{\mathcal{X}}$ and $F_{\mathcal{Z}}$ which involves the force on the particle due to rotation. Such forces are called *fictitious forces* as they arise from the choice of noninertial coordinates, not a genuine force.

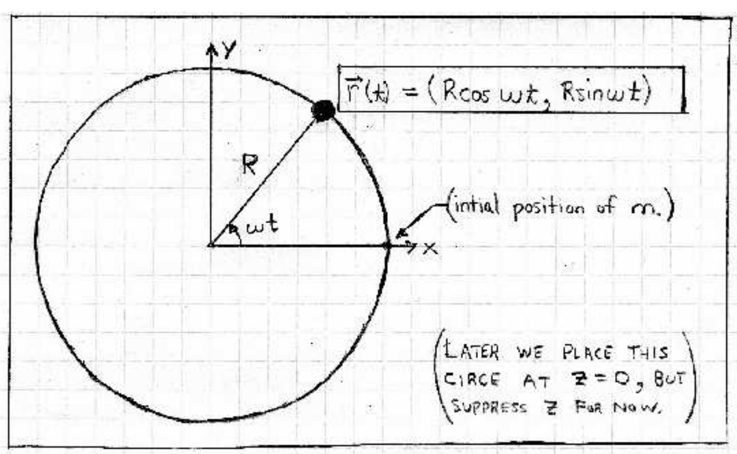
6.2 noninertial frames, a case study of circular motion

Some argue that any force proportional to mass may be viewed as a fictitious force, for example Hooke's law is F=kx, so you can see that the spring force is genuine. On the other hand gravity looks like $F = mg$ near the surface of the earth so some would argue that it is fictitious, however the conclusion of that thought takes us outside the realm of classical mechanics and the mathematics of this course. Anyway, if you are in a noninertial frame then for all intents and purposes fictitious forces are very real. The most familiar of these is probably the centrifugal force. Most introductory physics texts cast aspersion on the concept of centrifugal force (radially outward directed) because it is not a force observed from an inertial frame, rather it is a force due to noninertial motion. They say the centripetal (center seeking) force is really what maintains the motion and that there is no such thing as centrifugal force. I doubt most people are convinced by such arguments because it really feels like there is a force that wants to throw you out of a car when you take a hard turn. If there is no force then how can we feel it ? The desire of some to declare this force to be "fictional" stems from there belief that everything should be understood from the perspective of an inertial frame. Mathematically that is a convenient belief, but it certainly doesn't fit with everday experience. Ok, enough semantics. Lets examine circular motion in some depth.

For notational simplicity let us take $\mathbb{R}^3$ to be physical space and the identity mapping $\mathcal{X} = id$ to give us a stationary coordinate system on $\mathbb{R}^3$. Consider then the motion of a particle moving in a circle of radius R about the origin at a constant angular velocity of ω in the counterclockwise direction in the xy-plane. We will drop the third dimension for the most part throughout since it does not enter the calculations. If we assume that the particle begins at $(R, 0)$ at time zero then it follows that we can parametrize its path via the equations,

$$\begin{aligned} x(t) &= Rcos(\omega t) \\ y(t) &= Rsin(\omega t) \end{aligned} \tag{6.12}$$

this parametrization is geometric in nature and follows from the picture below, remember we took ω constant so that $\theta = \omega t$



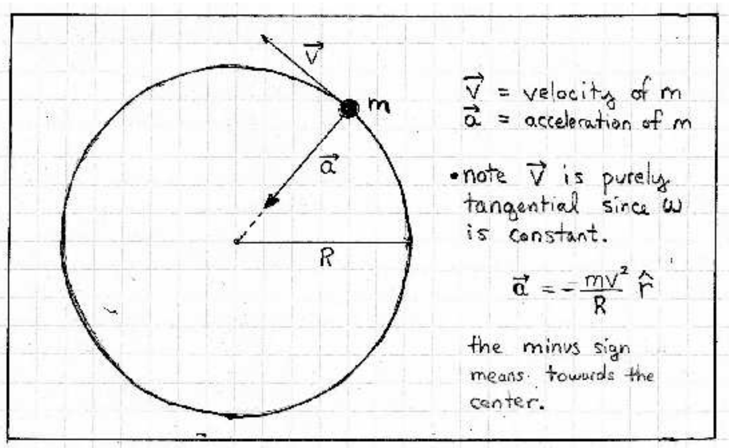
Now it is convenient to write $\vec{r}(t) = (x(t), y(t))$. Let us derive what the acceleration is for the particle, differentiate twice to obtain

$$\begin{aligned}
\vec{r}''(t) &= (x''(t), y''(t)) \\
&= (-R\omega^2 cos(\omega t), -R\omega^2 sin(\omega t)) \\
&= -\omega^2 \vec{r}(t)
\end{aligned}$$

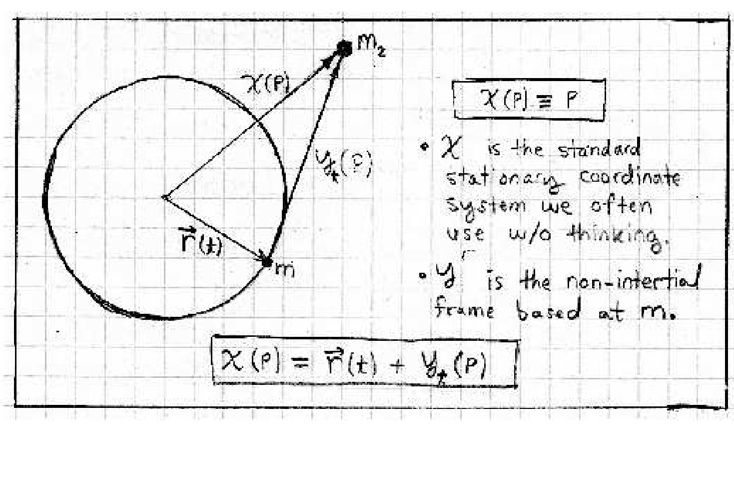
Now for pure circular motion the tangential velocity v is related to the angular velocity ω by $v = \omega R$. In other words $\omega = v/R$, radians per second is given by the length per second divided by the length of a radius. Substituting that into the last equation yields that,

$$\vec{a}(t) = \vec{r}''(t) = -\frac{v^2}{R^2} r(t) \tag{6.13}$$

The picture below summarizes our findings thus far.



Now define a second coordinate system that has its origin based at the rotating particle. We'll call this new frame $\mathcal{Y}$ whereas we have labeled the standard frame $\mathcal{X}$. Let $p \in \mathbb{R}^3$ be an arbitrary point then the following picture reveals how the descriptions of $\mathcal{X}$ and $\mathcal{Y}$ are related.



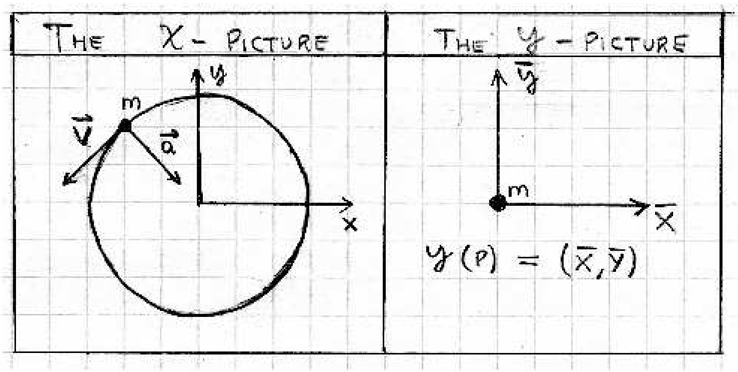
Clearly we find,

$$\mathcal{X}(p) = \mathcal{Y}(p) + \vec{r}(t) \tag{6.14}$$

note that the frames $\mathcal{X}$ and $\mathcal{Y}_t$ are not related by an rigid motion since $\vec{r}$ is not a constant function. Suppose that γ is the trajectory of a particle in $\mathbb{R}^3$, lets compare the acceleration of γ in frame $\mathcal{X}$ to that of it in $\mathcal{Y}_t$.

$$\begin{aligned} &\mathcal{X}(\gamma(t)) = \mathcal{Y}_t(\gamma(t)) + \vec{r}(t) \\ &\implies a_{\mathcal{X}}(t) = \gamma''(t) = a_{\mathcal{Y}_t}(t) + \vec{r}''(t) \end{aligned} \tag{6.15}$$

If we consider the special case of $\gamma(t) = r(t)$ we find the curious but trivial result that $\mathcal{Y}_t(r(t)) = 0$ and consequently $a_{\mathcal{Y}_t}(t) = 0$. Perhaps a picture is helpful,

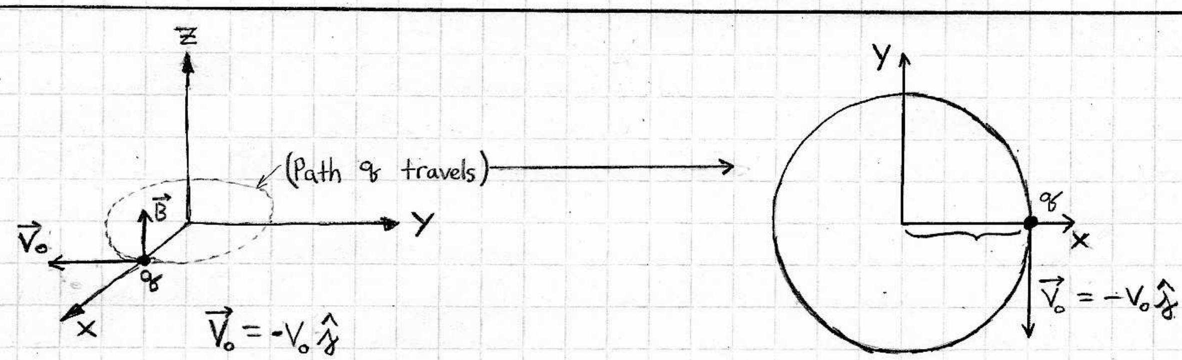


We have radically different pictures of the motion of the rotating particle, in the $\mathcal{X}$ picture the particle is accelerated and using our earlier calculation,

$$a_{\mathcal{X}} = \vec{r}''(t) = \frac{-v^2}{R}\hat{r}$$

on the other hand in the $\mathcal{Y}_t$ frame the mass just sits at the origin with $a_{Ycalt} = 0$. Since $F = ma$ we would conclude (ignoring our restriction to inertial frames for a moment) that the particle has an external force on it in the $\mathcal{X}$ frame but not in the $\mathcal{Y}$ frame. This clearly throws a wrench in the universality of the force concept, it is for this reason that we must restrict to inertial frames if we are to make nice broad sweeping statements as we have been able to in earlier sections. If we allowed noninertial frames in the basic set-up then it would be difficult to ever figure out what if any forces were in fact genuine. Dwelling on these matters actually led Einstein to his theory of general relativity where noninertial frames play a central role in the theory.
Anyway, lets think more about the circle. The relation we found in the $\mathcal{X}$ frame does not tell us *how* the particle is remaining in circular motion, rather only that if it is then it must have an acceleration which points towards the center of the circle with precisely the magnitude mv^2/R. I believe we have all worked problems based on this basic relation. An obvious question remains, which force makes the particle go in a circle? Well, we have not said enough about the particle yet to give a definitive answer to that question. In fact many forces could accomplish the task. You might imagine the particle is tethered by a string to the central point, or perhaps it is stuck in a circular contraption and the contact forces with the walls of the contraption are providing the force. A more interesting possibility for us is that the particle carries a charge and it is

subject to a magnetic field in the z-direction. Further let us assume that the initial position of the charge q is $(mv/qB, 0, 0)$ and the initial velocity of the charged particle is v in the negative y-direction. I'll work this one out one paper because I can.



We assume that we have a charge q with mass m that has intial position $(mV_o/qB, 0, 0)$ with intial velocity $\vec{V}_o = (0, -V_o, 0)$ then the charge is subject to a constant background magnetic field of $\vec{B} = (0, 0, B)$ where $B > 0$. We show that the charge travels a circle of radius $R = mV_o/qB$. We begin with Newton's 2^{nd} law paired with the Coulomb force law,

$$m\frac{d^2\vec{r}}{dt^2} = q\vec{V}\times\vec{B} = qB\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right)\times(0,0,1)$$

$$\frac{d^2\vec{r}}{dt^2} = \frac{qB}{m}\left(\frac{dy}{dt}\hat{i} - \frac{dx}{dt}\hat{j}\right) = \frac{d^2x}{dt^2}\hat{i} + \frac{d^2y}{dt^2}\hat{j} + \frac{d^2z}{dt^2}\hat{k}$$

This gives us three 2^{nd} order ODEs to solve, let $\alpha \equiv qB/m$.

$$\frac{d^2z}{dt^2} = 0 \text{ has sol}^{\underline{n}}: \; z(t) = c_1 + c_2 t \Rightarrow z(t) = 0$$
(using intial conditions)

$$\left\{\begin{aligned} \frac{d^2y}{dt^2} &= -\alpha\frac{dx}{dt} \\ \frac{d^2x}{dt^2} &= \alpha\frac{dy}{dt} \end{aligned}\right\} \text{ \underline{Coupled}.}$$

Notice that $\frac{dy}{dt} = \frac{1}{\alpha}\frac{d^2x}{dt^2} \Rightarrow \frac{d^2y}{dt^2} = \frac{1}{\alpha}\frac{d^3x}{dt^3} = -\alpha\frac{dx}{dt}$.

Thus we need to solve $x''' = -\alpha^2 x'$, introduce $W \equiv x'$ then our DEq$^{\underline{n}}$ becomes $W'' = -\alpha^2 W$ this has well known sol$^{\underline{n}}$,

$$W = \boxed{\frac{dx}{dt} = c_1\cos(\alpha t) + c_2\sin(\alpha t)}$$

Then we can calculate everything else from this

$$\frac{dy}{dt} = \frac{1}{\alpha}\frac{d^2x}{dt^2} = \frac{1}{\alpha}\left(-c_1\alpha\sin(\alpha t) + c_2\alpha\cos(\alpha t)\right)$$

$$\Rightarrow \boxed{\frac{dy}{dt} = -c_1\sin(\alpha t) + c_2\cos(\alpha t)}$$

Continuing,

Continuing, we found for $\alpha = qB/m$ that

$$\boxed{\begin{aligned} \frac{dx}{dt} &= C_1 \cos(\alpha t) + C_2 \sin(\alpha t) \\ \frac{dy}{dt} &= -C_1 \sin(\alpha t) + C_2 \cos(\alpha t) \end{aligned}}$$

Now integrate both to find $x(t)$ & $y(t)$

$$\boxed{\begin{aligned} x(t) &= \frac{1}{\alpha}\left(C_1 \sin(\alpha t) - C_2 \cos(\alpha t)\right) + C_3 \\ y(t) &= \frac{1}{\alpha}\left(C_1 \cos(\alpha t) + C_2 \sin(\alpha t)\right) + C_4 \end{aligned}}$$

Finally apply intial conditions,

$$\begin{aligned} x'(0) &= 0 = C_1 \Rightarrow \boxed{C_1 = 0} \\ y'(0) &= -V_o = C_2 \Rightarrow \boxed{C_2 = -V_o} \\ x(0) &= mV_o/qB = V_o/\alpha = \frac{1}{\alpha}(V_o) + C_3 \Rightarrow \boxed{C_3 = 0} \\ y(0) &= 0 = \frac{1}{\alpha}(0) + C_4 \Rightarrow \boxed{C_4 = 0} \end{aligned}$$

Thus collecting our results we find, since $C_2/\alpha = -V_o/\alpha = \frac{-mV_o}{qB}$

$$\boxed{\begin{aligned} x(t) &= (mV_o/qB) \cos(qBt/m) \\ y(t) &= -(mV_o/qB) \sin(qBt/m) \\ z(t) &= 0 \end{aligned}}$$

This is the circle $x^2 + y^2 = R^2$ with radius $R = mV_o/qB$. lying in the $z = 0$ plane.

It is curious that magnetic forces cannot be included in the Galilean relativity. For if the velocity of a charge is zero in one frame but not zero in another then does that mean that the particle has a non-zero force or no force? In the rest frame of the constant velocity charge apparently there is no magnetic force, yet in another inertially related frame where the charge is in motion there would be a magnetic force. How can this be? The problem with our thinking is we have not asked how the magnetic field transforms for one thing, but more fundamentally we will find that you cannot separate the magnetic force from the electric force. Later we'll come to a better understanding of this, there is no nice way of addressing it in Newtonian mechanics that I know of. It is an inherently relativistic problem, and Einstein attributes it as one of his motivating factors in dreaming up his special relativity.

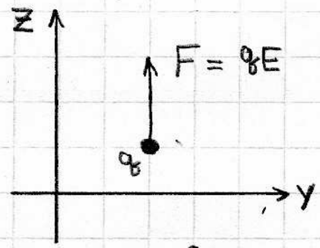
> "What led me more or less directly to the special theory of relativity was the conviction that the electromotive force acting on a body in motion in a magnetic field was nothing else but an electric field"

Albert Einstein, 1952.

6.3 a lonely example

Assume that we have a constant electric field. Let a particle of charge q start at the origin. Find its velocity as a function of time, what if any is the maximum speed it attains?

Consider a charge q with mass m. Place it at rest in a uniform electric field $\vec{E} = E\hat{k}$ find the resulting equation of motion and velocity



$$m\frac{d^2\vec{r}}{dt^2} = qE\hat{k}$$

thus $z'' = qE/m$, $y'' = 0$, $x'' = 0$ suppose that q is at (x_o, y_o, z_o) at time zero then we may integrate from $t = 0$ to time t and obtain,

$$z'(t) - z'(0) = \int_0^t \frac{qE}{m}\,d\bar{t} = \frac{qE}{m}t = z'(t)$$

$$z(t) - z(0) = \int_0^t \frac{qE}{m}\bar{t}\,d\bar{t} = \frac{qE}{2m}t^2 \Rightarrow z(t) = z_o + \frac{qE}{2m}t^2$$

thus,

$$\boxed{\begin{aligned} \vec{r}(t) &= \left(x_o,\ y_o,\ z_o + \frac{1}{2}\frac{qE}{m}t^2\right) \\ \vec{r}\,'(t) &= \left(0,\ 0,\ \frac{qE}{m}t\right) \end{aligned}}$$

Notice the speed $|r'(t)| = qEt/m$ is <u>unbounded</u> it may reach arbitrarily large values as $t \to \infty$.

Chapter 7

Special Relativity

In the last chapter we saw how Euclidean geometry and Newtonian physics go hand in hand. In this chapter we will see why electromagnetism is at odds with Newtonian physics. After giving a short history of the dilemma we will examine the solution given by Einstein. His axioms replace Newtonian mechanics with a new system of mechanics called *special relativity*. The geometry implicit within special relativity is *hyperbolic geometry* since the rotations in this new geometry are parametrized by hyperbolic angles. Euclidean space is replaced with Minkowski space and rotations are generalized to Lorentz transformations. There are many fascinating non-intuitive features of special relativity, but we will not dwell on those matters. Many good books are available to ponder the paradoxes (Taylor and Wheeler, Rindler, Resnick,French...). Our focus will be on the overall motivations and mathematical structure and we will use linear algebra throughout. We have borrowed some physical arguments from Resnick, I find them clearer than most.

7.1 Maxwell's equations verses Galilean relativity

Why is it that Galilean relativity does not apply to electrodynamic situations ? If most forces are to be treated equally by all inertial frames then why not magnetic forces as well? Let us derive the wave equation for an electromagnetic wave (light) in vacuum in order to arrive at a more concrete manifestation of the problem. Maxwell's equations in "empty" space are

$$\begin{array}{ll} \nabla \cdot \vec{E} = 0 & \nabla \cdot \vec{B} = 0 \\ \nabla \times \vec{E} = -\partial_t \vec{B} & \nabla \times \vec{B} = \mu_o \epsilon_o \partial_t \vec{E} = 0 \end{array} \tag{7.1}$$

where μ_o, ϵ_o are the permeability and permitivity of free space. Thus applying eqn.3.8 to Maxwell's equations we find,

$$\begin{gathered} \nabla \times (\nabla \times \vec{E}) = \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = -\nabla^2 \vec{E} \\ \nabla \times (\nabla \times \vec{E}) = \nabla \times (-\partial_t \vec{B}) = -\partial_t(\nabla \times \vec{B}) = -\partial_t(\mu_o \epsilon_o \partial_t \vec{E}) \end{gathered} \tag{7.2}$$

Comparing the equations above and performing similar calculation for $\vec{B}$ we find,

$$\begin{aligned} \nabla^2 \vec{E} &= \mu_o \epsilon_o \frac{\partial^2 \vec{E}}{\partial t^2} \\ \nabla^2 \vec{B} &= \mu_o \epsilon_o \frac{\partial^2 \vec{B}}{\partial t^2}. \end{aligned} \tag{7.3}$$

The equation for a three dimensional wave has the form $\nabla^2 f = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}$ where the speed of the wave is v. Given this we can identify that the eqn.7.3 says that light is a wave with velocity $v = 1/\sqrt{\mu_o \epsilon_o} \equiv c$. Question: where does the inertial observer enter these calculations ? Intuitively it seems that the speed of light should depend on who is observing it, but the speed we just derived depends only on the characteristics of empty space itself. Given the success of Newtonian mechanics it was natural to suppose that velocity addition was not wrong, so to explain this universal speed it was posited that light propagated relative to a substance called *Ether*. This turns out to be wrong for reasons I will elaborate on next.

7.2 a short history of Ether

Light has been interpreted as an electromagnetic wave since the discovery of Maxwell's equations in the mid-nineteenth century. In the classical physics all waves travel through some media. Moreover when the media moves then the wave moves with it and consequently moves faster or slower. So it was only natural to assume that since light was also a wave it should propagate at various speeds depending on the motion of its media. But what media does light propogate in? It must be everywhere otherwise we could not see distant starlight. Also this media must be very unusual as it does not effect the orbits of planets and such, that indicates it is very transparent both optically and mechanically. It would be a substance that we could not feel or see except in through its propagation of light. This media was called *Ether*, the necessity for this immaterial substance was predicated on the belief of physicists that light could not just propogate through empty space and also the wave equation argument for the universal speed of light. Once one excepts the existence of the Ether one has a preferred frame, the frame of the Ether. Since light propagates with respect to the Ether it follows under unmodified Newtonian mechanics that light should be observed to go faster or slower from frames moving relative to the Ether. This is something one could check experimentally. The famous Michelson Morely experiment concluded a null result. Light had the same velocity relative to completely different frames.

Thus there was a puzzle to explain. If there was no Ether then what ? Why was the velocity of light always the same ? There were a number of attempts made to explain the apparently constant of the speed of light. I'll list a few to give a flavor

1.) *Emission Theory* : the speed of a light ray is relative to the speed of its emitter.

2.) *Ether Drag* : material particles drag the ether along with them so that the ether is always stationary relative to the frame, hence the speed of light would be the same for various observers.

3.) *Modifications to Newtonian Mechanics* : Lorentz and others advocated ad-hoc modifications of Newtonian mechanics to match what was found by the experiment. Particularly he advocated that distances were shrunk in the direction of motion in order to explain the constant speed of light.

Each of the attempts above fails in one way or another. I recommend reading Resnick's account of these matters if you are interested in more details.

7.3 Einstein's postulates

Curiously it seems that Einstein knew only pieces of the above history at the time he provided a solution. Basically, he found the unique characteristics of light strange in the view of Newtonian mechanics. The existence of the Ether frame seemed to fly in the face of most of physics, other phenomena follow the same rules in every inertial frame, why should light be different ? His objections were theoretical rather than experimental. After much thought Einstein came to the following two postulates for mechanics,

> 1.) The laws of physics are the same for all inertial observers. There is no preferred frame of reference.
> 2.) The speed of light in free space has the same value c according to all inertial observers.

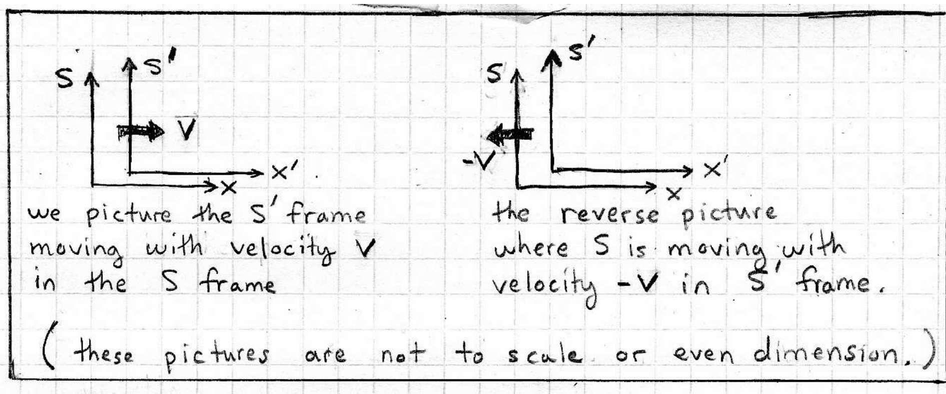
Notice that the first postulate extends the relativity principle to electrodynamics, Einsteinian relativity is more general than Galilean relativity, it holds for all forces not just the straight line directed ones. The second postulate seems like cheating at first glance, I mean sure it explains the experiments but what is the power of such a starting point? Remember though Einstein's motivation was not so much to explain the Michelson Morely experiment, rather it was to reconcile electrodynamics with mechanics. It turns out that these two postulates are consistent with Maxwell's equations, thus in a system of mechanics built over these postulates we will find that Maxwell's equations have the same form in all inertial frames.

Special relativity contains precisely the system of mechanics which is consistent with Maxwell's equations. The cost of reconciling mechanics and electromagnetism was common sense. Once one accepts that special relativity describes nature one must admit that things are much stranger than everyday experience indicates, time is not universal, the length of a given object is not the same according to different inertial observers, electric fields in one frame can be magnetic fields in another, energy is tied to mass, velocities do not add, classical momentum is not conserved, ... the list goes on. All of these questions are answered in the standard special relativity course so we will focus on just a select few as a mathematical exercise, we again recommend Resnick for discussions of those standard topics.

Please note that we have not *yet* proved that Maxwell's equations are consistent with special relativity. We will find an easy proof once we have translated Maxwell's equations into the language of differential forms. If you wish to be contrary you could argue this is unnecessary, afterall the proof in conventional vector notation can be found in Resnick on pgs. 178-181. That is true, but it wouldn't be as much fun. Not to mention our treatment will generalize to higher dimensions which is cool.

7.4 Lorentz transformations

To be precise we should emphasize that the definition of "inertial" observer must change for special relativity. We should examine what kind of transformations of space and time are allowed by Einstein's postulates. Our derivation here will be slightly heuristic, but its conclusion is not. To simplify the derivation let us consider $\mathbb{R}^2$ with one space and one time dimension.



Assume that (t, x) is an inertial coordinate system S. Further suppose that (t', x') is another inertial coordinate system S' that is related to the original system as follows,

$$\begin{aligned} x' &= ax + bt \\ x &= a'x' + b't' \end{aligned} \tag{7.4}$$

Let us find the velocity the point $x' = 0$ with respect to S,

$$0 = ax + bt \quad \Longrightarrow \quad a\frac{dx}{dt} + b = 0 \quad \Longrightarrow \quad v = \frac{dx}{dt} = \frac{-b}{a}. \tag{7.5}$$

This tells us that the S' frame has velocity $v = -b/a$ in the S frame. Likewise find the velocity of the point x=0 in the S' frame,

$$0 = a'x' + b't' \quad \Longrightarrow \quad a'\frac{dx'}{dt'} + b' = 0 \quad \Longrightarrow \quad -v = \frac{dx'}{dt'} = \frac{-b'}{a'}. \tag{7.6}$$

We have noted that the velocity of the S' frame with respect to S must be equal and opposite to the velocity of the S frame with respect to S'. Rindler calls this the "v-reversal" symmetry. There are other coordinate exchange symmetries that we could use to more formally derive the Lorentz transformation. We defer a pickier derivation to the subtle book of Rindler. Let us return to our derivation, note that equations [7.5] and [7.6] reveal,

$$\begin{aligned} x' &= a(x - vt) \\ x &= a'(x' + vt'). \end{aligned} \tag{7.7}$$

Since there is no preferred frame it follows that $a = a'$ thus,

$$\begin{aligned} x' &= a(x - vt) \\ x &= a(x' + vt'). \end{aligned} \tag{7.8}$$

Upto now we have really only used symmetries that derive from the non-existence of a preferred frame of reference. Next consider emitting a photon from the S frame when the frames coincide. Let $\phi(t)$ be its position in the (t, x) frame and $\psi(t')$ be its position in the (t', x') frame. Applying equation [7.8] to the photon yields $\psi(t') = a(\phi(t) - vt)$ and $\phi(t) = a(\psi(t') + vt')$. We know by the second postulate that the speed of light is c thus,

$$c = \frac{d\psi}{dt'} = \frac{d}{dt'}\Big(a(\phi(t) - vt)\Big) = a\Big(\frac{d\phi}{dt}\frac{dt}{dt'} - v\frac{dt}{dt'}\Big) = a(c - v)\frac{dt}{dt'} \tag{7.9}$$

likewise

$$c = \frac{d\phi}{dt} = \frac{d}{dt}\Big(a(\psi(t') + vt')\Big) = a\left(\frac{d\psi}{dt'}\frac{dt'}{dt} + v\frac{dt'}{dt}\right) = a(c+v)\frac{dt'}{dt}. \tag{7.10}$$

From the two equations just above we find,

$$c^2 = a(c-v)\frac{dt}{dt'}a(c+v)\frac{dt'}{dt} = a^2(c^2 - v^2) \tag{7.11}$$

where we have used that $\frac{dt}{dt'}\frac{dt'}{dt} = \frac{dt}{dt} = 1$. Therefore we find that the constant a depends on the velocity of the frame and the speed of light, we give it a new name following physics tradition,

$$a = \gamma \equiv \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{c}{\sqrt{c^2 - v^2}}.$$

Another popular notation is $\beta = v/c$ yielding,

$$\gamma \equiv \frac{1}{\sqrt{1 - \beta^2}}.$$

Finally, note that the $domain(\gamma) = (-c, c)$, thus the transformation will only be defined when $|v| < c$. If a velocity is smaller in magnitude than the speed of light then we say that it is a *subluminal velocity*. We find that inertial frames are in constant subluminal velocity motion relative to each other. To summarize we find that inertial frames S and S' are related by the Lorentz Transformations below,

$$\begin{aligned} x' &= \gamma(x - vt) \\ x &= \gamma(x' + vt') \end{aligned} \tag{7.12}$$

from which it follows that

$$\begin{aligned} x' &= \gamma(x - \beta ct) \\ ct' &= \gamma(ct - \beta x). \end{aligned} \tag{7.13}$$

Thus to make things more symmetrical it is nice to rescale the time coordinate by a factor of c, customarily one defines $x^0 = ct$. **That said, we will for the remainder of these notes take $c = 1$ to avoid clutter**, this means that for us $x^0 = t$. Then in matrix notation,

$$\begin{pmatrix} t' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} \tag{7.14}$$

If you use results in these notes with other books you should remember that some factors of c must be put in to be correct (You may compare our equations with those of Griffith's to see where those factors of c belong).

By almost the same derivation we find in $\mathbb{R}^4$ that if the frame S', with coordinates (t', x', y', z') has velocity v with respect to S, with coordinates (t, x, y, z), then provided they share the same spacetime origin (meaning the origins match up; $(t, x, y, z) = (0, 0, 0, 0) = (t', x', y', z')$) we find,

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \tag{7.15}$$

This transformation is called an **x-boost**, it is just the previous two dimensional Lorentz transformation paired with $y' = y$ and $z' = z$. We make no claim that this is the only Lorentz transformation in $\mathbb{R}^4$; unlike the two-dimensional case there are many other transformations which respect Einstein's postulates. That will be the topic of the next section.

7.5 the Minkowski metric

Given the important place of the dot-product in Euclidean geometry it is natural to seek out an analogue for spacetime. We wish to find an invariant quantity that characterizes the geometry of special relativity. Since in the limit of $\beta \to 0$ we should recover Euclidean geometry it is reasonable to suppose we should have the dot-product inside this invariant, but we know time must also be included. Thus consider for an x-boost,

$$\begin{aligned} -(t')^2 + (x')^2 + (y')^2 + (z')^2 &= -\gamma^2(t - vx)^2 + \gamma^2(x - vt)^2 + y^2 + z^2 \\ &= \gamma^2(-(t^2 - 2vxt + v^2x^2) + (x^2 - 2vxt + v^2t^2) + y^2 + z^2 \\ &= \gamma^2(1 - v^2)(-t^2 + x^2) + y^2 + z^2 \\ &= -t^2 + x^2 + y^2 + z^2. \end{aligned}$$

Let us define a generalized dot-product the Minkowski product in view of this interesting calculation,

Definition 7.5.1. *Let* $v = (v^0, v^1, v^2, v^3)^t, w = (w^0, w^1, w^2, w^3)^t \in \mathbb{R}^4$ *then the* **Minkowski metric** $< \ , \ >: \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ *is defined by*

$$< v, w >= -v^0w^0 + v^1w^1 + v^2w^2 + v^3w^3.$$

Equivalently, encode the minus sign by $(w_\mu) = (-w^0, w^1, w^2, w^3)$ *and write*

$$< v, w >= v^\mu w_\mu$$

where the μ *is summed over its values* $0, 1, 2, 3$. *The components* v^μ *are called* **contravariant** *components while the related components* v_μ *are called the* **covariant components**.

It might be more apt to say the Minkowski "metric" since technically it is **not** a metric in the traditional mathematical sense. To be more precise we should call the Minkowski metric a *psuedo-metric*, other authors term it a *Lorentzian inner product*. Recall that $d : V \times V \to \mathbb{R}$ is said to be a metric on V if it is symmetric $d(x, y) = d(y, x)$, positive definite $d(x, x) = 0 \iff x = 0$, and satisfies the triangle inequality $z = x + y \implies d(x, z) \leq d(x, y) + d(y, z)$. For example, the dot-product on $\mathbb{R}^3$ is a metric. So why is the Minkowski metric not a metric ?

$$< (a, a, 0, 0)^t, (a, a, 0, 0)^t >= -a^2 + a^2 = 0$$

but $(a, a, 0, 0)^t$ is not the zero vector, this demonstrates the Minkowski metric is not positive definite. It can also be shown that the triangle inequality fails for certain vectors.

The Minkowski metric does share much in common with the dot product and other genuine metrics, in particular,

Proposition 7.5.2. *The Minkowski product* $<,>: \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ *satisfies:*
(1.) $<x,y>=<y,x>$ *for all* $x,y \in \mathbb{R}^4$
(2.) $<cx,y>=c<x,y>$ *for all* $x,y \in \mathbb{R}^4$ *and* $c \in \mathbb{R}$
(3.) $<x+y,z>=<x,z>+<y,z>$ *for all* $x,y,z \in \mathbb{R}^4$
(4.) $<x,y>=0$ *for* **all** $y \in \mathbb{R}^4 \implies x=0$

Proof: left to the reader as an exercise.

Notice the phrase **all** in (4.) is not necessary in the case of a true metric, for a genuine metric $<x,x>=0$ implies $x=0$. But for the Minkowski product we have nontrivial null-vectors, just as we saw before the proposition. Physically these vectors are important, they describe paths that light may possibly travel. The definition below helps us to make distinctions between physically distinct directions in Minkowski space.

Definition 7.5.3. *Let* $v=(v^0,v^1,v^2,v^3)^t \in \mathbb{R}^4$ *then we say*
(1.) v *is a timelike vector if* $<v,v> < 0$
(2.) v *is a lightlike vector if* $<v,v> = 0$
(3.) v *is a spacelike vector if* $<v,v> > 0$

If we consider the trajectory of a massive particle in $\mathbb{R}^4$ that begins at the origin then at any later time the trajectory will be located at a timelike vector. If we consider a light beam emitted from the origin then at any future time it will located at the tip of a lightlike vector. Finally, spacelike vectors point to points in $\mathbb{R}^4$ which cannot be reached by the motion of physical particles that pass throughout the origin. We say that massive particles are confined within their light cones, this means that they are always located at timelike vectors relative to their current position in space time. You have a homework that elaborates on these ideas and their connection to the constant speed of light.

It is useful to write the Minkowski product in terms of a matrix multiplication. Observe that for $x,y \in \mathbb{R}^4$,

$$<x,y>=-x^0y^0+x^1y^1+x^2y^2+x^3y^3=\begin{pmatrix} x^0 & x^1 & x^2 & x^3 \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} y^0 \\ y^1 \\ y^2 \\ y^3 \end{pmatrix} \equiv x^t \eta y$$

where we have introduced η the matrix of the Minkowski product. I find the component versions of the statements above to be more useful in practice. The components of η are simply $(\eta_{\mu\nu})$ whereas the components of η^{-1} are $(\eta^{\mu\nu})$ so that $\eta_{\mu\beta}\eta^{\beta\nu}=\delta^\nu_\mu$. Let me make a list of useful identities,

$$\begin{aligned} &v_\mu = \eta_{\mu\nu}v^\nu \\ &v^\mu = \eta^{\mu\nu}v_\nu \\ &v^\mu v_\mu = \eta^{\mu\nu}v_\nu v_\mu = \eta_{\mu\nu}v^\mu v^\nu \end{aligned} \tag{7.16}$$

We defer the proper discussion of the true mathematical meaning of these formulas till a later section.

Remark 7.5.4. *Property (1.) above says the Minkowski metric is symmetric. If one studies Symplectic geometry the object that plays the role of the metric in that geometry is not even symmetric. So when you are reading physics and someone uses the term "metric" be careful it means what you think it means.*

Lets settle the terminology here on out,

Definition 7.5.5. *A* **metric** *on a vector space V is a symmetric, nondegenerate bilinear form on V.*

this is what you should understand me to mean when I mention a metric in later chapters.

7.6 general Lorentz transformations

In this section we present other transformations that relate relativistic inertial frames, then we give an elegant coordinate free definition for a Lorentz transformation.
Notice our derivation of an x-boost just as well follows for y or z playing the role of x, such Lorentz transformations are **y-boosts** and **z-boosts** respectively. Explicitly,

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & -\gamma\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\gamma\beta & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \qquad \begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \tag{7.17}$$

Another transformation that does not violate Einstein's postulates is a spatial rotation, that is fix the time variable ($t' = t$) and rotate the spatial coordinates as we did in the Euclidean case. For example a rotation about the z-axis would be,

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & & 0 \\ 0 & cos(\theta) & sin(\theta) & 0 \\ 0 & -sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \tag{7.18}$$

More generally if $R \in O(3)$ then the following will preserve the invariant quantity we discovered,

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & & 0 \\ 0 & R_{11} & R_{12} & R_{13} \\ 0 & R_{21} & R_{22} & R_{23} \\ 0 & R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \tag{7.19}$$

Recall that we must take $R \in SO(3)$ if we wish to insure that the new spatial coordinates are of the same "handedness" as the old.

Now that we have seen some examples of Lorentz transformations it is desirable to give a compact and coordinate free characterization of the idea.

Definition 7.6.1. *A* **Lorentz Transformation** *on $\mathbb{R}^4$ is a linear transformation $L : \mathbb{R}^4 \to \mathbb{R}^4$ such that $< L(v), L(w) >=< v, w >$ for all $v, w \in \mathbb{R}^4$. The group of all such transformations is the* **Lorentz group** *which is denoted by $\mathcal{L}$. If we restrict to the subgroup of timelike vector preserving transformations which maintain the right-hand rule then we call that the* **orthochronous Lorentz group** *and denote it by $\mathcal{L}^{\uparrow}_{+}$.*

This definition of Lorentz transformation makes it clear that the Minkowski product is the same in all inertial frames of reference. In principle one might have worried that the definition for $<,>$ was coordinate that dependent, afterall it is defined explicitly in components relative to some frame. However now that we have refined what a Lorentz transformation truly is, it is obvious that the Minkowksi product of two vectors is the same value in any two frames related by a Lorentz transformation. Mathematically it is cleaner to just begin with the definition above and then we actually could *derive* the speed of light is the same for all inertial frames. You can find a proof in the last chapter, its not difficult if you approach the question in the right way.

Proposition 7.6.2. *Let L be a linear transformation on $\mathbb{R}^4$ and A be the matrix of L so that $L(x) = Ax$ for all $x \in \mathbb{R}^4$ then,*

$$A^t \eta A = \eta \qquad \Longleftrightarrow \qquad L \in \mathcal{L}.$$

We call A a Lorentz transformation if it is such a matrix.

Proof: Let $x, y \in \mathbb{R}^4$ and let L be a linear transformation with $L(x) = Ax$, Recall from the last section that,

$$< x, y >= x^t \eta y. \tag{7.20}$$

Next consider that

$$\begin{aligned} < L(x), L(y) > &=< Ax, Ay > \\ &= (Ax)^t \eta Ay \\ &= x^t A^t \eta Ay. \end{aligned} \tag{7.21}$$

Thus,

$$< L(x), L(y) >=< x, y > \Longleftrightarrow x^t A^t \eta A y = x^t \eta y \tag{7.22}$$

for all $x, y \in \mathbb{R}^4$. Therefore, since the equation above holds for all vectors,

$$L \in \mathcal{L} \iff A^t \eta A = \eta. \tag{7.23}$$

The proof is complete.

Remark 7.6.3. *We refer the reader to Wu ki Tung's "Group Theory in Physics" for further definitions and discussion of the discrete spacetime symmetries of charge, parity and time-reversal which all have fascinating physical implications to representation theory. Also it can be shown that topologically $\mathcal{L}$ is disconnected into four connected components, the component containing the identity is $\mathcal{L}^{\uparrow}_{+}$. The proof of these statements is involved and we refer the reader to David Bleeker's "Gauge Theory and Variational Principles" for the proof.*

7.7 rapidity and velocity addition rule in special relativity

Every rotation in $SO(3)$ can be written as a rotation about some axis. In practice we like to think about rotations about the principle coordinate axes, in particular

$$
\begin{aligned}
R_z(\theta) &= \begin{pmatrix} cos(\theta) & sin(\theta) & 0 \\ -sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} && \text{rotates about the z-axis} \\
R_y(\theta) &= \begin{pmatrix} cos(\theta) & 0 & sin(\theta) \\ 0 & 1 & 0 \\ -sin(\theta) & 0 & cos(\theta) \end{pmatrix} && \text{rotates about the y-axis} \\
R_x(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\theta) & sin(\theta) \\ 0 & -sin(\theta) & cos(\theta) \end{pmatrix} && \text{rotates about the x-axis}
\end{aligned}
$$

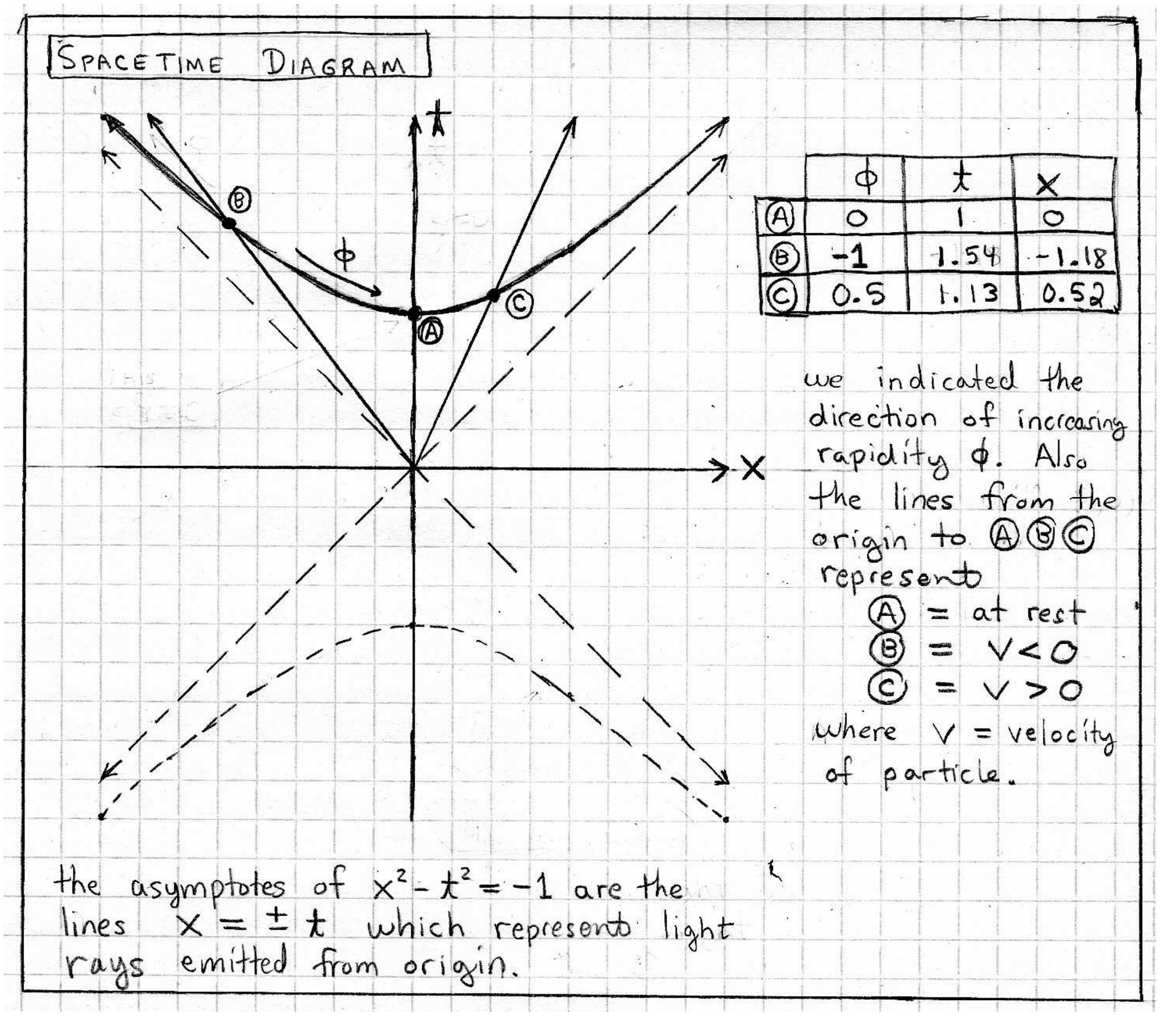
We can elevate these to rotations in $\mathbb{R}^4$ by adjoining a 1 as in equation 7.19. We now explore how we can view the x,y,z-boosts as *hyperbolic rotations*. In Euclidean geometry we study sine and cosine because they form ratios of the lengths of sides of a triangle inscribed in a circle, it should be familiar that we can parametrize a circle using sine and cosine,

$$x^2 + y^2 = 1 \qquad x = cos(\theta) \;\; y = sin(\theta) \;\; 0 < \theta \leq 2\pi$$

The analogue of a circle here is a locus of equidistant points relative to the Minkowski product. Choosing the locus of all timelike points with interval -1 from the origin gives a hyperbola which can be parametrized by hyperbolic sine and cosine,

$$-t^2 + r^2 = -1 \qquad t = cosh(\phi) \;\; r = sinh(\phi) \;\; -\infty < \phi < \infty$$

here we have let $r^2 = x^2 + y^2 + z^2$ and for simplicity of discussion we'll suppress two of the spatial directions as you may otherwise justly complain we really have a hyperboloid. As you can see in the diagram below as we trace out the hyperbola the straight line connecting the point and the origin represents the motion of a particle traveling at various velocities away from x=0. When $\phi = 0$ we get a particle at rest, whereas for $\pm\phi >> 0$ we approach the asymptotes of the hyperbola $r = \pm t$ which are the equations of light-rays emitted from the origin.



Thus we see that velocity and rapidity are in a one-one correspondence.

Definition 7.7.1. *The hyperbolic angle the parametrizes a boost in the k^{th} spatial coordinate direction is defined by*

$$tanh(\phi_k) = \beta_k$$

where β_k is the velocity in the k-th direction. We call ϕ_k the **rapidity** *of a $x_k -$ boost.*

Before going any further we stop to remind the reader everything we should have learned in calculus about hyperbolic functions,

Proposition 7.7.2. *We begin by recalling the definitions of "cosh" and "sinh",*

$$cosh(x) = \frac{1}{2}(e^x + e^{-x}) \qquad sinh(x) = \frac{1}{2}(e^x - e^{-x})$$

Then the hyperbolic tangent is "tanh" defined by

$$tanh(x) = \frac{sinh(x)}{cosh(x)}$$

Recalling the imaginary exponential form of sine and cosine

$$cos(x) = \frac{1}{2}(e^{ix} + e^{-ix}) \qquad sin(x) = \frac{1}{2i}(e^{ix} - e^{-ix})$$

you can see the analogy, moreover you can easily derive that,

$$cosh(ix) = cos(x) \qquad sinh(ix) = isin(x)$$

this means you can translate trigonometric identities into hyperbolic identities,

$$\begin{array}{ll} (1.) & cosh^2(x) - sinh^2(x) = 1 \\ (2.) & cosh(a+b) = cosh(a)cosh(b) + sinh(a)sinh(b) \\ (3.) & sinh(a+b) = sinh(a)cosh(b) + sinh(b)cosh(a) \\ (4.) & tanh(a+b) = \frac{tanh(a)+tanh(b)}{1+tanh(a)tanh(b)} \end{array} \tag{7.24}$$

Finally note the derivatives,

$$\frac{d}{dx}[cosh(x)] = sinh(x) \qquad \frac{d}{dx}[sinh(x)] = cosh(x) \qquad \frac{d}{dx}[tanh(x)] = sech^2(x)$$

where $sech^2(x) = 1/cosh^2(x)$.

Proof: Parts (1.),(2.),(3.) are easily verified direction via calculations involving products of exponentials or by translating trigonometric identities as hinted at in the theorem. We'll give an explicit account of how (4.) is true assuming the previous parts of the proposition,

$$\begin{aligned} tanh(a+b) &= \frac{sinh(a+b)}{cosh(a+b)} \\ &= \frac{sinh(a)cosh(b) + sinh(b)cosh(a)}{cosh(a)cosh(b) + sinh(a)sinh(b)} \\ &= \frac{tanh(a)cosh(b) + sinh(b)}{cosh(b) + sinh(b)tanh(a)} \\ &= \frac{tanh(a) + tanh(b)}{1 + tanh(a)tanh(b)}. \end{aligned}$$

Finally the derivatives follow quickly from differentiating the definitions of the hyperbolic functions.

Proposition 7.7.3. *If* $tanh(\phi) = \beta$ *then* $cosh(\phi) = \gamma$ *where* $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ *as before.*

Proof: Lets calculate,

$$tanh(\phi) = \frac{sinh(\phi)}{cosh(\phi)} = \frac{sinh(\phi)}{\sqrt{1 + sinh^2(\phi)}} = \beta \tag{7.25}$$

solving for $sinh(\phi)$ yields,

$$sinh(\phi) = \frac{\beta}{\sqrt{1-\beta^2}} = \beta\gamma. \tag{7.26}$$

Therefore,

$$cosh(\phi) = \frac{sinh(\phi)}{tanh(\phi)} = \frac{\beta\gamma}{\beta} = \gamma. \tag{7.27}$$

this completes the proof.

Thus we may parametrize the x,y,z-boosts in terms of rapidities corresponding to the x,y,z velocity of the boost. Using the notation $B_i(\phi)$ for a boost in the i-direction by rapidity ϕ we have,

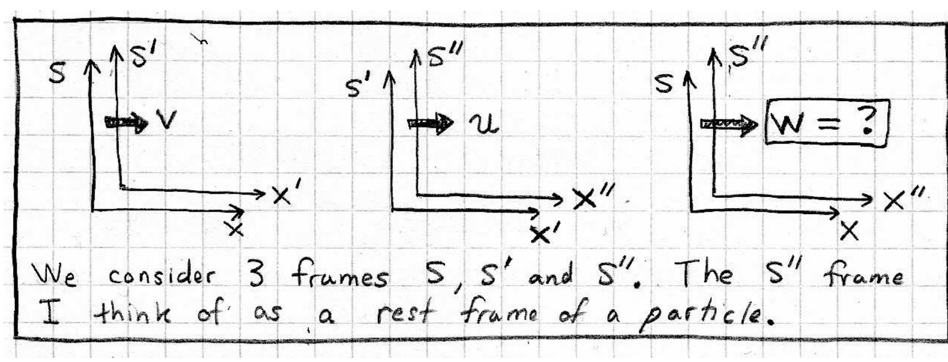
$$B_x(\phi) = \begin{pmatrix} cosh(\phi) & -sinh(\phi) & 0 & 0 \\ -sinh(\phi) & cosh(\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{boost by rapidity } \phi \text{ in x-direction}$$

$$B_y(\phi) = \begin{pmatrix} cosh(\phi) & 0 & -sinh(\phi) & 0 \\ 0 & 1 & 0 & 0 \\ -sinh(\phi) & 0 & cosh(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{boost by rapidity } \phi \text{ in y-direction}$$

$$B_z(\phi) = \begin{pmatrix} cosh(\phi) & 0 & 0 & -sinh(\phi) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -sinh(\phi) & 0 & 0 & cosh(\phi) \end{pmatrix} \quad \text{boost by rapidity } \phi \text{ in z-direction}$$

7.7.1 rapidities are additive

Let us consider three inertial frames S, S' and S''. Suppose that all three frames are alligned at $t = t' = t'' = 0$. Further suppose that frame S' travels at velocity v in the x-direction of the S frame. Also suppose that frame S'' travels at velocity u in the x'-direction of the S' frame.



In terms of boosts we may restate our assumptions as follows,

$$x' = B_x(a)x \qquad x'' = B_{x'}(b)x'$$

where $tanh(a) = v$ and $tanh(b) = u$. Let us find what the rapidity of S'' is with respect to the S frame. Notice that $x'' = B_{x'}(b)x' = B_{x'}(b)B_x(a)x$. We calculate $B_{x'}(b)B_x(a)$,

$$\begin{pmatrix} cosh(b)cosh(a) + sinh(b)sinh(a) & -sinh(b)cosh(a) - cosh(b)sinh(a) & 0 & 0 \\ -sinh(b)cosh(a) - cosh(b)sinh(a) & cosh(b)cosh(a) + sinh(b)sinh(a) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{7.28}$$

Apply the hyperbolic trig-identities to find

$$B_{x'}(b)B_x(a) = \begin{pmatrix} cosh(a+b) & -sinh(a+b) & 0 & 0 \\ -sinh(a+b) & cosh(a+b) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = B_x(a+b). \tag{7.29}$$

Which says that a boost by rapidity a followed by another boost of rapidity b is the same as a single boost by rapidity $a+b$, assuming that the boosts are in the same direction its fairly clear that the calculation above will be similar for the y or z boosts.

7.7.2 velocities are not additive

Let us return to the question we began the last section with and rephrase it in terms of velocity. What is the velocity of S'' with respect to the S frame ? Let us define that velocity to be w, note

$$\begin{aligned} w &= tanh(a+b) \\ &= \frac{tanh(a) + tanh(b)}{1 + tanh(a)tanh(b)} \\ &= \frac{u+v}{1+uv}. \end{aligned}$$

Breaking from our usual convention of omitting c let us put it in for a moment,

$$w = \frac{u+v}{1+uv/c^2} \tag{7.30}$$

It is clear that if $|v| << c$ then we get back the Newtonian velocity addition rule, simply $w = u+v$. However, if $|v|$ is not small then the rule for adding relative velocities is appreciably modified from the common-sense rule of Newtonian mechanics.

Remark 7.7.4. *I tried the same approach for a perpendicularly moving S'' frame but I could not see how to get the known velocity addition rule in that case. Probably a better way to derive the relative velocity addition rules is differentiating the Lorentz transformations. The derivation above is somewhat novel and I really don't recommend trying it for other situations. If you wish to find a slick mathematically questionable way of deriving these things take a look at Rindler, sorry I'm a bit suspicious of his way of dividing the differentials.*

7.8 translations and the Poincaire group

A simple generalization of a linear transformation is an *affine* transformation.

Definition 7.8.1. *An affine transformation on $\mathbb{R}^4$ is a mapping $F : \mathbb{R}^4 \to \mathbb{R}^4$ such that $F(x) = Ax + b$ for some matrix A and vector b, when $b = 0$ it is a linear transformation, when $A = 0$ then we define $F(x) = x + b$ to be a translation.*

Let us consider two points in $\mathbb{R}^4$, say $x, y \in \mathbb{R}^4$ then we define the interval between x and y as

$$I(x, y) \equiv < x - y, x - y > . \tag{7.31}$$

We define the Poincaire group $\mathcal{P}$ to be the set of all affine transformations that leave the interval $I(x, y)$ invariant for all $x, y \in \mathbb{R}^4$. Notice that if $L \in \mathcal{L}$ then,

$$\begin{aligned} I(L(x), L(y)) &= < L(x) - L(y), L(x) - L(y) > \\ &= < L(x), L(x) > - < L(x), L(y) > - < L(y), L(x) > + < L(y), L(y) > \\ &= < x, x > -2 < x, y > + < y, y > \\ &= < x - y, x - y > \\ &= I(x, y) \end{aligned}$$

thus Lorentz transformations are in the Poincaire group. Also we can quickly verify that spacetime translations are in the Poincaire group, let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be defined by $T(x) = x + b$,

$$\begin{aligned} I(T(x), T(y)) &= < T(x) - T(y), T(x) - T(y) > \\ &= < x + b - (y + b), x + b - (y + b) > \\ &= < x, y > \\ &= I(x, y) \end{aligned}$$

thus spacetime translations are in the Poincaire group.

Definition 7.8.2. *Let ϕ be a* **Poincaire transformation** *on $\mathbb{R}^4$ then it is an affine mapping $\phi : \mathbb{R}^4 \to \mathbb{R}^4$ such that there exist $L \in \mathcal{L}$ and $b \in \mathbb{R}^4$ with*

$$\phi(x) = L(x) + b.$$

In other words a Poincaire transformation is generally a composition of a Lorentz transformation and a spacetime translation.

Proposition 7.8.3. *Every affine transformation of the Poincaire group $\mathcal{P}$ is a Poincaire transformation.*

Proof: left to the reader as an exercise.

Notice Poincaire transformations are analogous to the rigid motions of Euclidean geometry. Thus we proceed to define an analogue to the Euclidean structure, we will call it a Minkowski structure in honor of Minkowski. It was Minkowski who was largely responsible for supplying the mathematics to complement Einstein's physical genius. Without these mathematical refinements of special relativity it is doubtful that general relativity would have been found as quickly as it was (1905-1916 approximately).

Definition 7.8.4. *A* **Minkowski structure** *on a set $\mathcal{M}$ is a family of bijections $\mathcal{H}$ from $\mathbb{R}^4$ onto $\mathcal{M}$ such that*
(1.) $\mathcal{X}, \mathcal{Y} \in \mathcal{H}$ then $\mathcal{X} \circ \mathcal{Y}^{-1} \in \mathcal{H}$
(2.) If $\mathcal{X} \in \mathcal{H}$ and $\phi \in \mathcal{P}$ then $\mathcal{X} \circ \phi \in \mathcal{H}$.
A **Minkowskian Space** *is a pair $(\mathcal{M}, \mathcal{H})$ where $\mathcal{M}$ is a set with a Minkowski structure $\mathcal{H}$.*

Often we will just say $\mathcal{M}$ is a Minkowskian space and when we say that $\mathcal{M}$ is Minkowski Space we are identifying $\mathcal{M}$ with $\mathbb{R}^4$. We introduce this definition to bring out the analogy with Euclidean space, however pragmatically we will always identify Minkowski space with $\mathbb{R}^4$ in this course in order to avoid extra notation. This identification is much like the identification of ordinary spatial dimensions with $\mathbb{R}^3$, conceptually it is useful, but at a more basic level it is incorrect strictly speaking since $\mathbb{R}^3$ is not physical space.

If $\mathcal{X}$ is an inertial frame then $\mathcal{Y}$ is also an inertial frame if $\mathcal{X}^{-1} \circ \mathcal{Y} \in \mathcal{P}$. Thus in the context of special relativity inertial frames are related by Poincaire transformations. Thus to prove a theory is consistent with special relativity it suffices to demonstrate that the defining equations of the theory have the same form in all inertial frames.

7.9 vectors in Minkowski space

We have thus far avoided the problem of covariant and contravariant indices in this course, except in the definition of the Minkowksi metric. Let us continue that discussion now. In earlier chapters we wrote matrices with indices down, but now that there is a difference between indices being up and down it will become important to distinguish them. Heuristically you can think of it as a conservation of up and down indices. Consider a Lorentz transformation $L \in \mathcal{L}$ with matrix $\Lambda = (\Lambda^{\mu}_{\nu})$ then $L(x) = \Lambda x = \bar{x}$ explicitly entails,

$$\bar{x}^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu}$$

where the repeated ν is understood to be summed over its values $0, 1, 2, 3$. Notice that in order to "conserve indices" we need one up and one down index.

Given the Lorentz transformation as above it follows that if v is a vector in spacetime then in the x^{μ}-coordinates we have $v = v^{\mu} e_{\mu}$ whereas in the $\bar{x}^{\mu}$-coordinates we have $v = \bar{v}^{\nu} \bar{e}_{\nu}$. The vector v is in one-one correspondence to a point in spacetime and when we change coordinates we do not move the points themselves, rather our description of those points is what changes. That is why we write v in both systems. Before addressing the question of how the different pictures of v are related we will find it useful to consider a point $P \in \mathbb{R}^4$,

$$P = P^{\mu} e_{\mu} = \bar{P}^{\mu} \bar{e}_{\mu} = \Lambda^{\mu}_{\nu} P^{\nu} \bar{e}_{\mu} \tag{7.32}$$

Thus as the point P in the equation above is arbitrary it follows

$$e_{\mu} = \Lambda^{\mu}_{\nu} \bar{e}_{\nu} \tag{7.33}$$

Then multiply by the inverse of Λ to obtain

$$\boxed{\bar{e}_{\nu} = (\Lambda^{-1})^{\mu}_{\nu} e_{\mu}.} \tag{7.34}$$

Now that we know how the basis $\{e_\mu\}$ "rotates" under a Lorentz transformation we can figure out how the components v^μ must in turn change. Consider that

$$\begin{aligned} v &= v^\mu e_\mu \\ &= \bar{v}^\mu \bar{e}_\mu \\ &= \bar{v}^\mu (\Lambda^{-1})^\mu_\nu e_\mu. \end{aligned}$$

Thus as $\{e_\mu\}$ is a basis we can read off the transformation of the components,

$$v^\mu = \bar{v}^\mu (\Lambda^{-1})^\mu_\nu.$$

Multiply by the Lorentz transformation to cancel the inverse and find,

$$\boxed{\bar{v}^\mu = \Lambda^\nu_\mu v^\nu.} \tag{7.35}$$

Components v^μ that transforms to $\bar{v}^\mu = \Lambda^\mu_\nu v^\nu$ under a Lorentz transformation $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$ are called **contravariant components**. Contravariant components are the components of the vector v, however the covariant components will actually be the components of a corresponding *dual vector* or *covector*. Recall the idea of the *dual space* from linear algebra,

Definition 7.9.1. *Let V be a vector space then $V^* = \{\alpha : V \to \mathbb{R} \mid \alpha a \text{ linear map}\}$. Given an ordered basis $\{e_i\}$ for V then the* **dual basis** *on V^* is denoted $\{e^i\}$ and $e^j(e_i) = \delta^j_i$. We also say that $\{e^i\}$ is the basis on V^* dual to $\{e_i\}$ on V. Additionally, elements of V^* are called* **covectors or dual vectors**.

Now let us apply the general idea of a dual space to the specific case of interest to us, namely $\mathbb{R}^4 = V$. If $\alpha \in V^*$ then $\alpha = \alpha_\mu e^\mu$. Let us find how the components of a covector must change under a Lorentz transformation. We'll proceed much as we did for vectors, first we'll find how the dual basis "rotates" under a Lorentz transformation, then we'll use that to pin down the transformation property we must require for the components of a dual vector.

Let us consider the basis $\{e_\mu\}$ and the Lorentz transformed $\{\bar{e}_\mu\}$ where the Lorentz transformation we have in mind is the same as before $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$. Now consider the dual bases to $\{e_\mu\}$ and $\{\bar{e}_\mu\}$ which we denote $\{e^\mu\}$ and $\{\bar{e}^\mu\}$ respectively. Consider,

$$\begin{aligned} \bar{e}^\alpha(e_\beta) &= \bar{e}^\alpha(\Lambda^\gamma_\beta \bar{e}_\gamma) \\ &= \Lambda^\gamma_\beta \bar{e}^\alpha(\bar{e}_\gamma) \\ &= \Lambda^\gamma_\beta \delta^\alpha_\gamma \\ &= \Lambda^\alpha_\beta \\ &= \Lambda^\alpha_\mu \delta^\mu_\beta \\ &= \Lambda^\alpha_\mu e^\mu(e_\beta) \end{aligned}$$

Recall from linear algebra that a linear operator is defined by its action on a basis, thus we may read from the equation above

$$\boxed{\bar{e}^\alpha = \Lambda^\alpha_\mu e^\mu.} \tag{7.36}$$

Therefore the dual basis transforms inversely to the basis. Let's check the consistency of this result,

$$\begin{aligned}
\delta^\alpha_\beta &= \bar{e}^\alpha(\bar{e}_\beta) \\
&= \Lambda^\alpha_\mu e^\mu((\Lambda^{-1})^\nu_\beta e_\nu) \\
&= \Lambda^\alpha_\mu (\Lambda^{-1})^\nu_\beta e^\mu(e_\nu) \\
&= \Lambda^\alpha_\mu (\Lambda^{-1})^\nu_\beta \delta^\mu_\nu \\
&= (\Lambda^{-1})^\mu_\beta \Lambda^\alpha_\mu \\
&= \delta^\alpha_\beta.
\end{aligned}$$

The components of a dual vector $b = b_\mu \bar{e}^\mu$ are the numbers b_μ, these are called **covariant components**. Let us determine how they change under the Lorentz transformation,

$$\begin{aligned}
b &= b_\mu e^\mu \\
&= \bar{b}_\alpha \bar{e}^\alpha \\
&= \bar{b}_\mu \Lambda^\alpha_\mu e^\mu.
\end{aligned}$$

Since $\{e^\mu\}$ is a basis we can read from the above that $b_\mu = \bar{b}_\mu \Lambda^\alpha_\mu$ which upon multiplication by the inverse Lorentz transformation yields,

$$\boxed{\bar{b}_\mu = (\Lambda^{-1})^\nu_\mu b_\nu} \tag{7.37}$$

7.9.1 additional consistency checks

The calculations that follow are not strictly speaking necessary. However, I myself have at times gotten extremely confused trying to make a coherent whole of the various interlocking ideas here. Our goal here is to show that the metric is coordinate invariant, that is it has the same values in one frame as it does in any other frame related by a Lorentz transformation. Of course the very definition of the Lorentz transformation insures that the Minkowski metric is coordinate invariant, $L \in \mathcal{L}$ implies $< L(x), L(y) >=< x, y >$. I will now show how to verify this at the level of components. We can attempt this as we have just determined how the covariant components transform.

To begin we note some alternative characterizations of a Lorentz transformation in components. We assume that A is the matrix of the linear transformation L.

$$\begin{aligned}
L \in \mathcal{L} \quad &\Longleftrightarrow A^t \eta A = \eta \\
&\Longleftrightarrow (A^t)^\alpha_\mu \eta_{\alpha\beta} A^\beta_\nu = \eta_{\mu\nu} \\
&\Longleftrightarrow (A)^\mu_\alpha \eta_{\alpha\beta} A^\beta_\nu = \eta_{\mu\nu} \\
&\Longleftrightarrow \eta^{\sigma\mu}(A)^\mu_\alpha \eta_{\alpha\beta} A^\beta_\nu = \eta^{\sigma\mu}\eta_{\mu\nu} \\
&\Longleftrightarrow \eta^{\sigma\mu}(A)^\mu_\alpha \eta_{\alpha\beta} A^\beta_\nu = \delta^\sigma_\nu \\
&\Longleftrightarrow (A^{-1})^\sigma_\beta = \eta^{\sigma\mu}(A)^\mu_\alpha \eta_{\alpha\beta} \\
&\Longleftrightarrow (A^{-1})^\sigma_\beta = \eta_{\alpha\beta}(A)^\mu_\alpha \eta^{\sigma\mu}
\end{aligned} \tag{7.38}$$

Let v, w be vectors in Minkowski space where their Minkowski product $< v, w >= v^\mu w_\mu$ with respect to the coordinates x^μ. Change to new coordinates $\bar{x}^\mu$ where $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$ so that $\bar{v}^\mu = \Lambda^\mu_\nu v^\nu$ and $\bar{v}_\mu = (\Lambda^{-1})^\mu_\nu v_\nu$. We now seek to demonstrate the coordinate invariance of the Minkowski product directly in terms of components. Observe,

$$\begin{aligned}
v^\mu w_\mu =< v, w > &= < L(v), L(w) > \\
&= < \Lambda^\mu_\nu v^\nu e_\mu, \Lambda^\alpha_\beta w^\beta e_\alpha > \\
&= \Lambda^\mu_\nu v^\nu \Lambda^\alpha_\beta w^\beta < e_\mu, e_\alpha > \\
&= \Lambda^\mu_\nu v^\nu \Lambda^\alpha_\beta w^\beta \eta_{\mu\alpha} \\
&= (\Lambda^\mu_\nu v^\nu)(\eta_{\mu\alpha} \Lambda^\alpha_\beta w^\beta) \\
&= (\Lambda^\mu_\nu v^\nu)(\eta_{\mu\alpha} \Lambda^\alpha_\beta \eta^{\beta\alpha} w_\alpha) \\
&= (\Lambda^\mu_\nu v^\nu)((\Lambda^{-1})^\alpha_\mu w_\alpha) \\
&= (\bar{v}^\mu)(\bar{w}_\mu).
\end{aligned}$$

Again I emphasize this is doing things the hard way, this result is actually immediate from our definition of Lorentz transformations. The only reason I've kept this calculation is to give you an example of a more advanced index calculation.

7.10 relativistic mechanics

Upto now we have just discussed the geometry of special relativity, the next thing to address is how we must modify mechanics to fit the postulates of special relativity. In the course of this endeavor we will try to return to our example of circular motion to contrast the relativistic verses the Newtonian picture. This section is in large part adapted from chapter 3 of Resnick. I will not *prove* that the definitions below are consistent with special relativity, in the interest of time you'll just have to trust me.

Definition 7.10.1. *A particle's* **rest frame** *is the frame which is comoving with the particle. The* **rest mass** *of the particle is the mass of the particle measured in the rest frame, we denote it m_o. If the particle travels with velocity u in some frame then we define the* **relativistic mass** *of the particle to be*

$$m = m_o \gamma(u) = \frac{m_o}{\sqrt{1 - u^2/c^2}}$$

Notice that $m(u = 0) = m_o \gamma(0) = m_o$ so the relativistic mass measured in the rest frame is simply the rest mass. Next we define relativistic three-momentum,

Definition 7.10.2. **Relativistic momentum** *of a particle traveling with velocity $\vec{u}$ is denoted $\vec{p}$ and for massive particles it is defined $\vec{p} = m\vec{u}$ where m is the relativistic mass.*

In Newtonian mechanics we had the force law $\vec{F} = d\vec{p}/dt$. The relativistic force law is defined by the same rule except we replace Newtonian momentum with relativistic momentum,

Definition 7.10.3. *If $\vec{F}$ is the* **relativistic force** *on a particle then*

$$\vec{F} = \frac{d\vec{p}}{dt} = \frac{d}{dt}\left(\frac{m_o \vec{u}}{\sqrt{1 - u^2/c^2}}\right)$$

When you think about Newton's second law there are two parts, half of the law is that $\vec{F} = d\vec{p}/dt$, the other half is what $\vec{F}$ actually is. In principle when we generalize to the relativistic case we might have to modify both "halves" of the second law. For certain forces that is the case, but for the Lorentz force only $d\vec{p}/dt$ needs modification. The force on a particle of rest mass m_o and charge q with velocity $\vec{u}$ is,

$$q(\vec{E} + \vec{u} \times \vec{B}) = \frac{d}{dt}\left(\frac{m_o\vec{u}}{\sqrt{1 - u^2/c^2}}\right) \tag{7.39}$$

Remark 7.10.4. *Newtonian momentum $m_o u$ is not conserved relativistically, however relativistic momentum is conserved,*

$$\vec{F} = \frac{d}{dt}\left(m_o\gamma(u)\vec{u}\right) = 0 \implies \quad p = m_o\gamma(u)\vec{u} = constant$$

this means in a collision the net relativistic momentum is the same before and after the collision, according to a particular frame of reference. It is not the case that the net momentum is the same in all frames, that is not even the case in Newtonian mechanics. A particle has zero momentum in its rest frame in relativistic or ordinary mechanics.

7.10.1 Energy

One version of the work-energy theorem says that the kinetic energy of a particle accelerated by a force $\vec{F}$ from rest is equal to the work done by the force,

$$K = \int_{u=0}^{u=u} \vec{F} \cdot d\vec{l}$$

For example in the one dimensional case we recover the usual formula for kinetic energy by a simple calculation,

$$K = \int F dx = \int m_o \frac{du}{dt} dx = \int m_o \frac{dx}{dt}\frac{du}{dx} dx = \int m_o u du = \frac{1}{2} m_o u^2 \tag{7.40}$$

all the integrations above technically should have bounds corresponding to an upper bound of velocity u and a lower bound of velocity zero. Given this viewpoint and the fact we have already defined the relativistic force we can derive the relativistic kinetic energy, again we'll focus on the one-dimensional case to reduce clutter, the arguments are easily generalized for more dimensions,

$$\begin{aligned} K &= \int F dx \\ &= \int \frac{d}{dt}[mu] dx \\ &= \int (\frac{dm}{dt}u + m\frac{du}{dt}) dx \\ &= \int (u^2 dm + mudu) \end{aligned}$$

Now we should note that dm and du are implicitly related according to the definition of m,

$$\begin{aligned} m = \frac{m_o}{\sqrt{1-u^2/c^2}} \quad &\Longrightarrow \quad m^2c^2 - m^2u^2 = m_o^2 \\ &\Longrightarrow \quad 2mc^2dm - 2mu^2dm - 2m^2udu = 0 \\ &\Longrightarrow \quad c^2dm = u^2dm + mudu \end{aligned}$$

substituting what we just learned we find

$$K = \int_{u=0}^{u=u} Fdx = \int_{u=0}^{u=u} c^2dm = mc^2|_{m_o}^{m(u)} = mc^2 - m_oc^2 \tag{7.41}$$

Therefore,

$$\boxed{K = m_oc^2(\gamma - 1)} \tag{7.42}$$

Moreover we may define the *total energy* of a free particle (no external forces) by

$$\boxed{E = \gamma m_oc^2} \tag{7.43}$$

Then $E = K + E_o$ where the quantity $E_o = m_oc^2$ is the *rest energy*. Evidently free particles in relativity have more than just Kinetic energy, they also have a rest energy. This is the famed mass-energy correspondence of relativity. A particle at rest has an enormous quantity of energy, this energy is what powers nuclear fusion and fission.

7.10.2 more on the relativistic force law

Proposition 7.10.5. *The relativistic force is only partially proportional to the acceleration in general,*

$$\boxed{\vec{F} = \frac{1}{c^2}(\vec{F}\cdot\vec{u})\vec{u} + m\frac{d\vec{u}}{dt}}$$

Proof: Remember that m is a function of u thus,

$$\vec{F} = \frac{d\vec{p}}{dt} = \frac{d}{dt}[mu] = \frac{dm}{dt}u + m\frac{du}{dt}$$

However, we also know that $E = mc^2$ thus $m = E/c^2$ hence,

$$\frac{dm}{dt} = \frac{1}{c^2}\frac{dE}{dt} = \frac{1}{c^2}\frac{dK}{dt}.$$

Differentiating $K = \int \vec{F}\cdot d\vec{l}$ with respect to time yields that

$$\frac{dK}{dt} = \vec{F}\cdot\vec{u}$$

Putting it together we get,

$$\frac{dm}{dt} = \frac{1}{c^2}(\vec{F}\cdot\vec{u})$$

and the proposition follows.

Proposition 7.10.6. *The relativistic force reduces as follows in the special case that (1.) the $\vec{a}, \vec{F}, \vec{u}$ are paralell and (2.) the force is perpendicular to $\vec{u}$, a.k.a $\vec{F} \cdot \vec{u} = 0$.*

$$(1.) \quad \vec{F}_{||} = m_o \gamma^3(u) \frac{d\vec{u}}{dt} \qquad\qquad (2.) \quad \vec{F}_{\perp} = m_o \gamma \frac{d\vec{u}}{dt} \tag{7.44}$$

Proof: I have made this a homework. It should be easy.

Proposition 7.10.7. Relativistic Charged Particle in circular motion: *Let us suppose we have a charge q moving with intial velocity $(0, -v_o, 0)$ subject to the constant magnetic field $\vec{B} = (0, 0, B)$. Everything is the same as before except now let the intial position be $(mv_o/qB, 0, 0)$ where $m = \gamma(v_o)m_o$ is the relativistic mass. It can be shown that the particle travels in a circle of radius $R = \gamma(v_o)m_o v_o/qB$ centered at the origin lying in the $z = 0$ plane.*

Proof: I have made this a homework.

We see that charged particles moving at relativistic speeds through perpendicular magnetic fields also travel in circles relative to the frame where the magnetic field is perpendicular to the velocity. In the course of proving the proposition above you will notice that if the velocity is initially perpendicular to the magnetic field then it remains as such in the absence of other influences. One might wonder if the particle will travel in a circle in the view of other frames of reference. The analysis of that question must wait till we find how the magnetic field transforms when we change frames.

Remark 7.10.8. *We have introduced the concepts of relativistic force and momenta in the context of a single frame of reference. To make the treatment complete we ought to show that these concepts are coherent with Einstein's axiom that the laws of physics are the same in all inertial frames. The nice way to show that is to group momentum and energy together into one "4-vector" and likewise the force and power into another "4-vector", it can then be shown that they transform as covariant vectors in Minkowski space. I'd love to spend a few days telling you about how to use 4-vectors and energy-momentum conservation but we must go on.*

Chapter 8

Matrix Lie groups and Lie algebras

This chapter is a slight digression (we may not cover it depending on time, obviously I'll only assign the homeworks referenced in this chapter if I lecture on it). I include it only because it might be good to see where the orthogonal and Lorentz transformations fit into more abstract math. Although the math in this chapter is not terribly abstract, believe it or not.

8.1 matrix Lie groups

It is the case that all the examples given in this section are *Lie groups*. A Lie group is a group on which the group operations are *smooth*. We will not explain what we mean by *smooth*, if you'd like to know I'll tell you in office hours. Recall that a group is a set with an operation,

Definition 8.1.1. *Let G be a set with an operation from $G \times G \to G$ namely $(a, b) \mapsto ab$ for all $a, b \in G$ then we say G is a group if is has an identity, the operation is an associative operation, the operation closes on G, and every element is invertible. That is*
(1.) there exists $e \in G$ such that $xe = ex = x$ for all $x \in G$
(2.) if $a, b, c \in G$ then $(ab)c = a(bc)$
(3.) $a, b \in G$ implies $ab \in G$
(4.) If $x \in G$ then there exists $x^{-1} \in G$ such that $xx^{-1} = e = x^{-1}x$
when $ab = ba$ for all $a, b \in G$ we say that G is **Abelian** *otherwise we say G is* **nonabelian***.*

I'd like to just consider *matrix groups* in this section, typically they are nonabelian.

Example 8.1.2. *The general linear group $GL(n)$ is the set $n \times n$ matrices with nonzero determinant.*

$$GL(n) = \{A \in \mathbb{R}^{n\times n} \mid det(A) \neq 0\}$$

The group operation is simply matrix multiplication, we know from linear algebra that it is an associative operation. The identity of the group is just the $n \times n$ identity matrix. If $A, B \in Gl(n)$ then $0 \neq det(A)$ and $0 \neq det(B)$ then as $det(AB) = det(A)det(B)$ we find $0 \neq det(AB)$ thus the product of invertible matrices is invertible. Finally, from linear algebra we know that matrices with nonzero determinant are nonsingular, the inverse exists.

All the other matrix groups live in $GL(n)$ as subgroups. We will restrict our attention to matrices with real entries but this is not a necessary restriction, complex matrix groups as just as interesting, maybe more.

Definition 8.1.3. *Let H be a subset of G then H is a subgroup of G if it is also a group under the group operations of G.*

Recall that we do not have to prove all the group axioms when checking if a subgroup is really a subgroup,

Proposition 8.1.4. *Let $H \subseteq G$ then H is a subgroup of G if*
(1.) $e \in H$
(2.) $a, b \in H$ implies $ab \in H$.

Example 8.1.5. *Let $H = SL(n)$ be the set of matrices in $GL(n)$ with determinant one,*

$$SL(n) = \{A \in GL(n) \mid det(A) = 1\}.$$

this forms a subgroup of $GL(n)$ since clearly $det(I) = 1$ thus $I \in SL(n)$ and if $A, B \in SL(n)$ then $det(AB) = det(A)det(B) = 1$ thus $SL(n)$ is closed under matrix multiplication. Thus $SL(n)$ is a subgroup of $GL(n)$. We call $SL(n)$ the **special linear group**

Orthogonal matrices form another subgroup of $O(n)$,

Example 8.1.6. *Let $H = O(n)$ be the subset of $GL(n)$ defined by,*

$$O(n) = \{A \in GL(n) \mid A^t A = I\}$$

Then since $I^t I = I$ it follows $I \in O(n)$. Moreover if $A, B \in O(n)$ then

$$(AB)^t(AB) = B^t A^t AB = B^t IB = I$$

where we have used that $A^t A = I$ in the second equality and $B^t IB = B^t B = I$ in the third. Thus $O(n)$ is closed under matrix multiplication and hence is a subgroup. We call $O(n)$ the **group of orthogonal matrices**

One can also think about *special orthogonal matrices*

Example 8.1.7. *Define the set of* **special orthogonal matrices** *to be*

$$SO(n) = \{A \in O(n) \mid det(A) = 1\}$$

this is a subgroup of $GL(n)$, $SL(n)$ and $O(n)$. I'll let you prove that in a homework.

The Lorentz matrices also form a group,

Example 8.1.8. *We denote the set of Lorentz matrices $O(1, 3)$ they are a subset of $GL(4)$ defined by*

$$O(1, 3) = \{A \in GL(4) \mid A^t \eta A = \eta\}$$

These from a group, again I'll let you show it in homework.

We can restrict to the Lorentz matrices corresponding to the orthochronous Lorentz transformations,

Example 8.1.9. *We denote the set of special Lorentz matrices $SO(1,3)$ they are a subset of $O(1,3)$ defined by*

$$SO(1,3) = \{A \in O(1,3) \ | det(A) = 1\}$$

These from a group, again I'll let you show it in homework.

We can include many of these examples in the following example,

Example 8.1.10. *Let $J \in GL(n)$ be a fixed matrix. Define*

$$H_J = \{A \in GL(n) \mid A^t JA = J\}$$

when $J = I$ we recover $O(n)$ whereas when $J = \eta$ and $n = 4$ we recover $0(1,3)$. Again I'll let you show that this is a group in homework.

8.2 matrix Lie algebras

A Lie algebra is a vector space paired with an operation called a bracket. For matrices we can form Lie algebras using the "commutator". The commutator of $a, b \in \mathbb{R}^{n\times n}$ is $[a, b] \equiv ab - ba$. It is an easy algebra exercise to verify the following identities for the commutator,

$$\begin{aligned}
&[\lambda a, b] = \lambda[a, b] \\
&[a + b, c] = [a, c] + [b, c] \\
&[a, b] = -[b, a] \\
&[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0
\end{aligned} \tag{8.1}$$

for all $a, b, c \in \mathbb{R}^{n\times n}$ and $\lambda \in \mathbb{R}$. These properties makes $\mathbb{R}^{n\times n}$ paired with the commutator a *Lie algebra* which we denote $gl(n)$. The connection between Lie algebras and Lie groups is that the Lie algebra appears as the tangent space to the identity of the Lie group. What this means is that if you take a curve *of matrices* in the matrix Lie group that passes through I then the tangent to that curve is in the Lie algebra. The collection of all such tangents forms the Lie algebra of the Lie group.

Remark 8.2.1. *The notation is confusing if you are caught unaware of the difference between the group and the algebra. There is a big difference between $gl(n)$ and $GL(n)$. Generally I try to follow the notation capital letter for the group like G and lower Germanic letter for the algebra $\mathfrak{g}$.*

Example 8.2.2. *The Lie algebra of $GL(n)$ is $gl(n)$. Let $\gamma : \mathbb{R} \to GL(n)$ be a curve such that $\gamma(0) = I$. What can we say about the tangents to this curve? It is completely arbitrary, depending on what γ is we could have $\gamma'(0)$ be most anything. All we really know here is that $\gamma'(0) \in gl(n) = \mathbb{R}^{n\times n}$.*

Just as the other matrix Lie groups are subsets of $GL(n)$ we will find that the Lie algebras corresponding to the sub Lie groups give subalgebras of $gl(n)$ (you can guess the definition of subalgebra)

Example 8.2.3. *The Lie algebra of $SL(n)$ is $sl(n)$, where $sl(n) \equiv \{a \in gl(n) \mid trace(a) = 0\}$. There is a nice identity involving the matrix exponential, trace, determinant and ordinary exponential*

$$det(exp(a)) = exp(trace(a))$$

this is easy to prove when a is diagonalizable, and it even holds when a is not. I know a nice proof of it if you are interested ask me. Observe that we can write a curve through I in $SL(n)$ in terms of the matrix exponential, $\gamma(t) = exp(at)$ clearly $\gamma(0) = exp(0) = I$ and the curve will lie in $SL(n)$ provided we demand that $trace(a) = 0$ since in that case,

$$det(exp(ta)) = exp(trace(ta)) = exp(ttrace(a)) = exp(0) = 1.$$

Finally note that $\frac{d}{dt}exp(ta) = aexp(ta)$ so $\gamma'(0) = a$ and we have shown that the set of traceless matrices forms the Lie algebra of the the Lie group $SL(n)$

We could go on and derive the Lie algebra $o(n)$ of $O(n)$ or the Lie algebra $so(n)$ of $SO(n)$ or even the Lie algebra $\mathfrak{h}_J$ of H_J. In each case we would take a curve through the identity and deduce what algebraic constraint characterized the Lie algebra.

Historically the correspondence between the Lie algebra and group probably gave rise to much of the interest in Lie algebras. However, Lie algebras are fascinating even without the group. Lie algebras and its various infinite dimensional generalizations are still a very active area of algebraic research today. For the physicist the correspondence between the Lie algebra and group is central to understanding Quantum Mechanics. I highly recommend the book *Symmetries and Quantum Mechanics* by Greiner if you'd like to better understand how quantum numbers and symmetry groups are connected. It makes Quantum Mechanics much less ad-hoc in my estimation.

8.3 exponentiation

The process of generating a Lie group from its Lie algebra is called *exponentiation*. Well to be more careful I should mention we cannot usually get the whole group, but rather just part of the part of the group which is connected to the identity.

Definition 8.3.1. Matrix Exponential: *let $A \in gl(n)$ then define,*

$$exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$

The importance of the matrix expontial to systems of ordinary differential equations is easily summarized. Given **any** system of ordinary differential equations with the form $dx/dt = Ax$ (where $x = (x_1, x_2, \dots, x_n)^t$ is a vector of functions of t and A is a constant matrix) the general solution is $x = exp(tA)c$ where $c = (c_1, c_2, \dots, c_n)^t$ is a vector of arbitrary constants. One uses generalized eigenvectors to assemble the solution in an acessible form. You should have seen some of these things in the differential equations course. Anyway, from the perspective of matrix Lie group/algebras the matrix exponential is important because it satisfies the following identity known as the Campbell-Baker Hausdorff relation,

$$\boxed{exp(A)exp(B) = exp(A + B + \frac{1}{2}[A, B] + \frac{1}{12}[[A, B], B] - \frac{1}{12}[[B, A], A] + \cdots)} \tag{8.2}$$

where the higher order terms are all formed in terms of nested commutators. You are probably familar with the most simple case of this identity, if $[A, B] = 0$ then all the commutator terms vanish leaving

$$exp(A)exp(B) = exp(A + B).$$

What equation 8.2 tells us is that if we know how all the commutators work in the algebra we can contruct the products in the group. Some mathematicians define the group in this way, they assume that the Baker-Campbell-Hausdorff relation holds and define the group multiplication in view of that (these are the so-called "formal groups" of Serre). Anyway lets find what the exponentiation of $gl(n)$ forms.

Example 8.3.2. *$exp(gl(n)) \subset GL^+(n)$: Let $A \in gl(n)$ then exponentiate A and calculate its determinant, we'll use the identity I mentioned before,*

$$det(exp(A)) = exp(trace(A)).$$

Now we know that the ordinary exponential has $exp : \mathbb{R} \to (0, \infty)$ so then $exp(trace(A)) > 0$. We define

$$GL^+(n) = \{B \in gl(n) \mid det(B) > 0 \}$$

clearly then $exp(gl(n)) \subset GL^+(n)$. In fact if we take products of matrix exponentials then we can cover $GL^+(n)$ but that requires more thought. I leave that for you. In topology you will learn that if a set can be divided up into a disjoint finite union then that is called a "seperation" of the set. If a topological space has no open seperation then it is said to be connected, otherwise it is disconnected. Notice that $GL(n)$ is disconnected because we can seperate it into positive and negative determinant matrices, defining $GL^-(n) = \{ B \in gl(n) \mid det(B) < 0 \}$ we see that

$$GL(n) = GL^+(n) \cup GL^-(n).$$

What has happened is that the exponential mapping has missed half of the group, we only obtained the part of the group which is connected to the identity (and not even that unless we take products of exponentials as well).

Remark 8.3.3. *I have only attempted to give you a brief introduction to matrix Lie groups and algebras. Both areas have a rather beautiful theoretical underpinning, I have only shown you a few examples. The bracket can be much more abstract than the commutator and the group does not usually lie inside the algebra as in our examples here.*

Part III

Multilinear Algebra and Differential Forms in $\mathbb{R}^n$

Chapter 9

Tensors and Forms on Vector Space

In this chapter we will explore the algebraic foundations of tensors. To begin we will study multilinear maps on V. We will then ease into the concept of the tensor product, it will provide us a basis for the multilinear maps to begin with. Then once the concept of a multilinear map is exhausted we will add mappings on the dual space as well. Those will mimic much of what we did to begin with, except the tensor product will be defined a little differently. Finally with all the special cases settled in earlier sections we will define tensors on V. This will include everything in earlier sections plus some new mixed cases. Again the tensor product will induce a basis on the vector space of tensors on V. Finally we will add a metric to the discussion in the last section. The metric will give isomorphisms which allow us to convert tensor type, that is to raise and lower indices. Throughout this chapter we will try to understand the interplay between mappings and components. In this chapter we are thinking of everything at a point, this means the components are really just numbers. Later on the components will become functions, but the algebra we develop in this chapter will still be very much relevant so we focus on it to begin.

9.1 multilinear maps

A multilinear map is a natural extension of the concept of a linear mapping.

Definition 9.1.1. *A* **multilinear map** *on a vector space V to a vector space W is a mapping $L : V \times V \times \cdots \times V \to W$ that is linear in each slot, meaning for all $x_1, x_2, \ldots, x_p, y \in V$ and $c \in \mathbb{R}$,*

$$L(x_1, x_2, \ldots, x_k + cy, \ldots, x_p) = L(x_1, x_2, \ldots, x_k, \ldots, x_p) + cL(x_1, x_2, \ldots, y, \ldots, x_p)$$

for $k = 1, 2, \ldots p$. When $p = 1$ we say it is a linear mapping, when $p = 2$ we say it is a bilinear mapping, in general we say it is a p-linear mapping on V to W. Also we may say that L is a W-valued multilinear map on V.

In fact, you have already had experience with linear and bilinear maps.

Example 9.1.2. *Dot product on $V = \mathbb{R}^3$ is a real-valued bilinear mapping,*

$$\begin{aligned} (\vec{x} + c\vec{y}) \cdot \vec{z} &= \vec{x} \cdot \vec{z} + c\vec{y} \cdot \vec{z} \\ \vec{x} \cdot (\vec{y} + c\vec{z}) &= \vec{x} \cdot \vec{y} + c\vec{x} \cdot \vec{z} \end{aligned} \tag{9.1}$$

for all $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^3$ and $c \in \mathbb{R}$.

Bilinear maps like the one above have a special property, they are said to by symmetric. But first we should define some combinatorial notations to help with discussing *permutations* of indices,

Definition 9.1.3. *A permutation on $\{1, 2, \dots p\}$ is a bijection onto $\{1, 2, \dots p\}$. We define the set of permutations on $\{1, 2, \dots p\}$ to be Σ_p. Further, define the sign of a permutation to be $sgn(\sigma) = 1$ if σ is the product of an even number of transpositions whereas $sgn(\sigma) = -1$ if σ is the product of a odd number transpositions.*

Let us consider the set of permutations on $\{1, 2, 3, \dots n\}$, this is called S_n the symmetric group, its order is $n!$ if you were wondering. Let me remind you how the cycle notation works since it allows us to explicitly present the number of transpositions contained in a permutation,

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 5 & 4 & 6 & 3 \end{pmatrix} \quad \Longleftrightarrow \quad \sigma = (12)(356) = (12)(36)(35) \tag{9.2}$$

recall the cycle notation is to be read right to left. If we think about inputing 5 we can read from the matrix notation that we ought to find $5 \mapsto 6$. Clearly that is the case for the first version of α written in cycle notation; (356) indicates that $5 \mapsto 6$ and nothing else messes with 6 after that. Then consider feeding 5 into the version of α written with just two-cycles (a.k.a. transpositions), first we note (35) indicates $5 \mapsto 3$, then that 3 hits (36) which means $3 \mapsto 6$, finally the cycle (12) doesn't care about 6 so we again have that $\alpha(5) = 6$. Finally we note that $sgn(\sigma) = -1$ since it is made of 3 transpositions.

It is always possible to write any permutation as a product of transpositions, such a decomposition is not unique. However, if the number of transpositions is even then it will remain so no matter how we rewrite the permutation. Likewise if the permutation is an product of an odd number of transpositions then any other decomposition into transpositions is also comprised of an odd number of transpositions. This is why we can **define** an **even** permutation is a permutation comprised by an even number of transpositions and an **odd** permutation is one comprised of an odd number of transpositions.

Example 9.1.4. Sample cycle calculations: *we rewrite as product of transpositions to determin if the given permutation is even or odd,*

$$\sigma = (12)(134)(152) = (12)(14)(13)(12)(15) \quad \Longrightarrow \quad sgn(\sigma) = -1$$

$$\lambda = (1243)(3521) = (13)(14)(12)(31)(32)(35) \quad \Longrightarrow \quad sgn(\lambda) = 1$$

$$\gamma = (123)(45678) = (13)(12)(48)(47)(46)(45) \quad \Longrightarrow \quad sgn(\gamma) = 1$$

We will not actually write down permutations as I've done in the preceding discussion, instead we will think about moving the indices around as we have from the beginning of this course. I have recalled the cycle notation for two reasons. First, it allows us to rigorously define symmetric and antisymmetric in a nice compact form. Second, probably some of you like modern algebra so these calculations bring a calm nostalgic feel to this chapter, its your happy place. If you have no idea how to do cycle calculations don't worry about it, so long as you understand what I mean by "symmetric" and "antisymmetric" you should be ok. (Modern algebra's not a prerequisite for this course)

Now we are ready to define symmetric and antisymmetric.

Definition 9.1.5. *A p-linear mapping $L : V \times V \times \cdots \times V$ is completely symmetric if*

$$L(x_1, \ldots, x, \ldots, y, \ldots, x_p) = L(x_1, \ldots, y, \ldots, x, \ldots, x_p)$$

for all possible pairs (x,y). Conversely, if a mapping has

$$L(x_1, \ldots, x, \ldots, y, \ldots, x_p) = -L(x_1, \ldots, y, \ldots, x, \ldots, x_p)$$

for all possible pairs (x,y) then it is said to be completely antisymmetric or alternating. Equivalently a p-linear mapping L is alternating if for all $\pi \in \Sigma_p$

$$L(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_p}) = sgn(\pi) L(x_1, x_2, \ldots, x_p)$$

.

Example 9.1.6. *The* **Minkowski product** *on $V = \mathbb{R}^4$ is a symmetric bilinear mapping.*

$$\begin{aligned} &< x, y + cz > = < x, y > + c < x, z > \\ &< x, y > = < y, x > \end{aligned} \tag{9.3}$$

for all $x, y, z \in \mathbb{R}^4$ and $c \in \mathbb{R}$. Notice that when we have a symmetric mapping it is sufficient to know it is linear in one slot. Once we know that we can use symmetry to find linearity in all the other slots, consider

$$\begin{aligned} < y + cz, x > \; &= < x, y + cz > \\ &= < x, y > + c < x, z > \\ &= < y, x > + c < z, x > \end{aligned} \tag{9.4}$$

thus linearity in the 2nd slot has given us linearity in the 1st slot thanks to the symmetric property of $<, >$.

Example 9.1.7. *The* **cross product** *on $V = \mathbb{R}^3$ is an antisymmetric vector-valued bilinear mapping.*

$$\begin{aligned} &\vec{A} \times \vec{B} = -\vec{B} \times \vec{A} \\ &(\vec{A} + c\vec{B}) \times \vec{C} = \vec{A} \times \vec{C} + c\vec{B} \times \vec{C} \end{aligned} \tag{9.5}$$

for all $\vec{A}, \vec{B}, \vec{C} \in \mathbb{R}^3$ and $c \in \mathbb{R}$. We also sometimes use the term "skewsymmetry" as an equivalent term for alternating or antisymmetric. Notice again it was enough to know it is antisymmetric and linear in one of the slots.

$$\begin{aligned} \vec{C} \times (\vec{A} + c\vec{B}) \; &= -(\vec{A} + c\vec{B}) \times \vec{C} \\ &= -(\vec{A} \times \vec{C} + c\vec{B} \times \vec{C}) \\ &= \vec{C} \times \vec{A} + c\vec{C} \times \vec{B} \end{aligned} \tag{9.6}$$

thus linearity in the 1st slot has given us linearity in the 2nd slot thanks to the antisymmetric property of the cross-product.

Example 9.1.8. *Define $h(u, v, w) \equiv u \cdot (v \times w)$ this will be an antisymmetric multilinear map from $\mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$. The proof follows quickly from the identity given in proposition* 1.25.

Example 9.1.9. *Let $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$ then define $L(v_1, v_2, \ldots, v_n) = A[v_1|v_2|\cdots|v_n]$ for some fixed $n \times n$ matrix A, then L is a multilinear map from $\mathbb{R}^n$ to the vector space of matrices which can be identified with $\mathbb{R}^{n^2}$ if you wish. Unless a special choice of A is made this mapping does not necessarily have any special properties except of course multilinearity which itself is a pretty stringent condition when you think about all the possible maps one could imagine.*

9.2 multilinear maps on V and tensor products of dual vectors

You may recall from linear algebra that the set of all linear transformations from V to W forms a vector space under pointwise addition of the maps. Equivalently those maps can be viewed as $dim(W)$ by $dim(V)$ matrices which do form a vector space under matrix addition and scalar multiplication. Moreover, we could even write a convenient basis for the vector space of matrices. Remember it was E_{ij} where that is the matrix with all zeros except the $(i,j)-th$ slot. Natural question to ask here is what is the analogue for the multilinear maps ? Does the set of all multilinear maps form a vector space and if so what is the basis ? It is true that the set of all p-multilinear mappings forms a vector space, I have left the verification of that fact as a homework problem.

To begin, notice we already have a basis for the case $p = 1$ and $W = \mathbb{R}$. The set of all linear mappings from $V \to \mathbb{R}$ is simply the dual space V^*. If $V = span\{e_1, \ldots e_n\}$ then $V^* = span\{e^1, \ldots e^n\}$ where $e^i(e_j) = \delta^i_j$.

Thus one should expect that the basis for multilinear maps on V is built from the dual basis. The method of construction is called the *tensor product*.

9.2.1 constructing bilinear maps via the tensor product of dual vectors

Let us define the tensor product of two dual vectors,

Definition 9.2.1. *Then tensor product of $\alpha \in V^*$ with $\beta \in V^*$ is denoted $\alpha \otimes \beta$ and is defined by*

$$(\alpha \otimes \beta)(x,y) = \alpha(x)\beta(y) \qquad \text{for all } x, y \in V$$

It is simple to verify that $\alpha \otimes \beta$ is a bilinear map on V.

Proposition 9.2.2. *Given dual vectors $\alpha, \beta, \gamma \in V^*$ and $a, b, c \in \mathbb{R}$ the tensor product satisfies*
(1.) $\alpha \otimes (c\beta) = (c\alpha) \otimes \beta = c(\alpha \otimes \beta)$
(2.) $\alpha \otimes (a+b)\beta = a\alpha \otimes \beta + b\alpha \otimes \beta$
(3.) $(\alpha + \beta) \otimes \gamma = \alpha \otimes \gamma + \beta \otimes \gamma$
(4.) $\alpha \otimes (\beta + \gamma) = \alpha \otimes \beta + \alpha \otimes \gamma$
(5.) $\alpha \otimes 0 = 0$

Some of the comments above are admittably redundant. Further notice

$$\begin{aligned} (\alpha \otimes \beta)(x,y) &= \alpha(x)\beta(y) \\ &= (\alpha_i e^i)(x)(\beta_j e^j)(y) \\ &= \alpha_i \beta_j e^i(x) e^j(y) \\ &= \alpha_i \beta_j (e^i \otimes e^j)(x,y). \end{aligned} \tag{9.7}$$

for all $x, y \in V$. Thus $\alpha \otimes \beta = \alpha_i \beta_j e^i \otimes e^j$.

Proposition 9.2.3. *The set $\{e^i \otimes e^j \mid 1 \leq i, j \leq n\}$ forms a basis for the set of all bilinear maps on the n-dimensional space V. Any bilinear map $B : V \times V \to \mathbb{R}$ can be written $B = B_{ij} e^i \otimes e^j$.*

In some sense the tensor product $e^i \otimes e^j$ is a place-holder in the vector space of all bilinear maps. The real information is contained in the numbers that multiply $e^i \otimes e^j$.

Definition 9.2.4. *If $B : V \times V \to \mathbb{R}$ can be written $B = B_{ij}e^i \otimes e^j$ then the numbers B_{ij} are the* **components** *of B. Moreover the components are said to be symmetric if $B_{ij} = B_{ji}$ or antisymmetric if $B_{ij} = -B_{ji}$.*

Physicists often omit the basis $e^i \otimes e^j$ from their analysis and focus exclusively on the components. Thats not necessarily a bad thing until you ask certain questions, like why do the components transform as they do. We'll see later on the presence of the basis is essential to gain a clear understanding of the coordinate change rule for bilinear forms. These comments are not limited to bilinear forms, in fact this contrast between mathematicians and physicists holds for pretty much any *tensor*.

Proposition 9.2.5. *Symmetric bilinear mappings have symmetric components. Antisymmetric bilinear maps have antisymmetric components.*

The proof is left as an exercise for the reader.

Example 9.2.6. *The Minkowski metric g is the bilinear map $g = \eta_{\mu\nu}e^\mu \otimes e^\nu$ on $\mathbb{R}^4$. Notice that g is symmetric just as $\eta_{\mu\nu} = \eta_{\nu\mu}$.*

Example 9.2.7. *Let consider a bilinear mapping $B : V \times V \to \mathbb{R}$ then we can always write B as a the sum of a symmetric and antisymmetric mapping,*

$$B(x, y) = \frac{1}{2}\Big[B(x, y) + B(y, x)\Big] + \frac{1}{2}\Big[B(x, y) - B(y, x)\Big]$$

At the level of components the same thought becomes,

$$B_{ij} = \frac{1}{2}(B_{ij} + B_{ji}) + \frac{1}{2}(B_{ij} - B_{ji})$$

In both of the equations we see that the arbitrary bilinear mapping can be decomposed into a purely symmetric and purely antisymmetric part. This is not the case for higher orders.

9.2.2 constructing trilinear maps via the tensor product of dual vectors

You could probably guess the definition to follow given our discussion thus far.

Definition 9.2.8. *Let $B : V \times V \to \mathbb{R}$ be a bilinear mapping and $\gamma \in V^*$ then the tensor products of B and γ are defined by*

$$(B \otimes \gamma)(x, y, z) = B(x, y)\gamma(z) \qquad (\gamma \otimes B)(x, y, z) = \gamma(x)B(y, z)$$

for all $x, y, z \in V$.

The tensor product is an associative product,

Proposition 9.2.9. *Let $\alpha, \beta, \gamma \in V^*$ then*

$$(\alpha \otimes \beta) \otimes \gamma = \alpha \otimes (\beta \otimes \gamma)$$

This means we may omit the parentheses above without danger of confusion. Indeed we can say all the same things that we did for the tensor product of two dual vectors,we can verify that $\alpha \otimes \beta \otimes \gamma$ is a trilinear map on V.

Proposition 9.2.10. *Given dual vectors $\alpha, \beta, \gamma, \sigma \in V^*$ and $c \in \mathbb{R}$ the tensor product satisfies*
(1.) $\alpha \otimes \beta \otimes (c\gamma) = \alpha \otimes (c\beta) \otimes \gamma = c(\alpha \otimes \beta \otimes \gamma)$
(2.) $(\alpha + \beta) \otimes \gamma \otimes \sigma = \alpha \otimes \gamma \otimes \sigma + \beta \otimes \gamma \otimes \sigma$
(3.) $\alpha \otimes (\beta + \gamma) \otimes \sigma = \alpha \otimes \beta \otimes \sigma + \alpha \otimes \gamma \otimes \sigma$
(4.) $\alpha \otimes \beta \otimes (\gamma + \sigma) = \alpha \otimes \beta \otimes \gamma + \alpha \otimes \beta \otimes \sigma$
(5.) $0 \otimes \beta \otimes \gamma = \alpha \otimes 0 \otimes \gamma = \alpha \otimes \beta \otimes 0 = 0$

In (5.) the O on the LHS's are the zero dual vectors, whereas the 0 on the RHS is the O mapping on $V \times V \times V$. Lets consider a generic trilinear mapping $T : V \times V \times V \to \mathbb{R}$. Observe

$$\begin{aligned} T(x, y, z) &= T(x^i e_i, y^j e_j, z^k e_k) \\ &= x^i y^j z^k T(e_i, e_j, e_k) \\ &= x^i y^j z^k T_{ijk} \\ &= e^i(x) e^j(y) e^k(z) T_{ijk} \\ &= T_{ijk}(e^i \otimes e^j \otimes e^k)(x, y, z) \end{aligned} \tag{9.8}$$

for all $x, y, z \in V$ where we have **defined** the components of T by $T(e_i, e_j, e_k) = T_{ijk}$. It should be clear that the calculation we have just completed verifies the following proposition,

Proposition 9.2.11. *The set $\{e^i \otimes e^j \otimes e^k \mid 1 \leq i, j, k \leq n\}$ forms a basis for the set of all trilinear maps on the n-dimensional space V. Any trilinear map $T : V \times V \times V \to \mathbb{R}$ can be written $T = T_{ijk} e^i \otimes e^j \otimes e^k$.*

Definition 9.2.12. *If $T : V \times V \times V \to \mathbb{R}$ can be written $T = T_{ijk} e^i \otimes e^j \otimes e^k$ then the numbers T_{ijk} are the* **components** *of T. The components are said to be symmetric if its values are identical under the exchange of any pair of indices; $T_{ijk} = T_{jik} = T_{ikj} = T_{kji}$ for all $i, j, k \in \{1, 2, \dots n\}$. The components are antisymmetric if*

$$T_{ijk} = T_{jki} = T_{kij} = -T_{kji} = -T_{jik} = -T_{ikj}$$

for all $i, j, k \in \{1, 2, \dots n\}$. Equivalently if

$$T_{i_1 i_2 i_3} = sgn(\pi) T_{i_{\pi(1)} i_{\pi(2)} i_{\pi(3)}}$$

for all $\pi \in \Sigma_3$ then T_{ijk} are antisymmetric components.

Proposition 9.2.13. *Symmetric trilinear mappings have symmetric components. Antisymmetric trilinear maps have antisymmetric components.*

The proof is left as an exercise for the reader.

9.2.3 multilinear maps from tensor products

By now the following statements should hopefully not be to surprising,

Definition 9.2.14. *Then tensor product of $\alpha_1, \alpha_2, \dots \alpha_p \in V^*$ is denoted $\alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_p$ and is defined by*

$$(\alpha_1 \otimes \alpha_2 \otimes \cdots \otimes \alpha_p)(x_1, x_2, \dots, x_p) = \alpha_1(x_1)\alpha_2(x_2) \cdots \alpha_p(x_p)$$

for all $x_1, x_2, \dots, x_p \in V$.

The *components* with respect to the basis $\{e_i\}$ of V of a p-multilinear map are given by acting on the basis elements.

Definition 9.2.15. *Let $T : V \times V \times \cdots \times V \to \mathbb{R}$ then the* **components** *of T are defined to be*

$$T(e_{i_1}, e_{i_2}, \dots, e_{i_p}) = T_{i_1 i_2 \dots i_p}.$$

If $T_{i_1 i_2 \dots i_p} = T_{i_{\pi(1)} i_{\pi(2)} \dots i_{\pi(p)}}$ for all $\pi \in \Sigma_p$ then $T_{i_1 i_2 \dots i_p}$ are **symmetric***.*
If $T_{i_1 i_2 \dots i_p} = sgn(\pi) T_{i_{\pi(1)} i_{\pi(2)} \dots i_{\pi(p)}}$ for all $\pi \in \Sigma_p$ then $T_{i_1 i_2 \dots i_p}$ are **antisymmetric***.*

As before if V has the basis $\{e_i\}$ then the tensor product induces a basis for the p-multilinear maps on V as follows, the components we just defined are simply the coordinates of the multilinear maps with respect to the induced tensor basis.

Proposition 9.2.16. *Let $T : V \times V \times \cdots \times V \to \mathbb{R}$ then*

$$T = T_{i_1 i_2 \dots i_p} e^{i_1} \otimes e^{i_2} \otimes \cdots \otimes e^{i_p}.$$

Thus $\{e^{i_1} \otimes e^{i_2} \otimes \cdots \otimes e^{i_p}\}$ with $1 \leq i_1, i_2, \dots, i_p \leq n$ is a basis on the vector space of all p-multilinear maps on V.

Remark 9.2.17. *The dimension of a vector space is the number of vectors in basis. A basis is a linearly independent spanning set. For the vector space of all multilinear mappings on an n-dimensional vector space V we have just mentioned that $\{e^{i_1} \otimes e^{i_2} \otimes \cdots \otimes e^{i_p}\}$ is a basis. Thus we can calculate the dimension by counting, it is n^p.*

9.3 multilinear maps on V^* and tensor products of vectors

For the most part in this section we will just follow the same terminology as in the case of multilinear maps on V. We take V to be a vector space throughout this section. Also although we could study mappings into another vector space W we will drop that from the beginning as the case $W = \mathbb{R}$ is truly what interests us.

Definition 9.3.1. *A* **multilinear map** *on the dual space V^* is a mapping $L : V^* \times V^* \times \cdots \times V^* \to \mathbb{R}$ that is linear in for all $\alpha, \beta \in V^*$ each slot, meaning for all $\alpha_1, \alpha_2, \dots, \alpha_p, \beta \in V^*$ and $c \in \mathbb{R}$,*

$$L(\alpha_1, \alpha_2, \dots, \alpha_k + c\beta, \dots, \alpha_p) = L(\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_p) + cL(\alpha_1, \alpha_2, \dots, \beta, \dots, \alpha_p)$$

for $k = 1, 2, \dots p$. When $p = 1$ we say it is a linear mapping on V^*, when $p = 2$ we say it is a bilinear mapping on V^*, in general we say it is a p-linear mapping on V^*. Also we may say that L is a $\mathbb{R}$-valued multilinear map on V^*.

Lets think about the simple case $p = 1$ to begin. Recall that a linear mapping from a vector space to $\mathbb{R}$ is called a dual vector. Now V^* is itself a vector space thus the space of all real-valued linear maps on V^* is V^{**} the **double dual to** V. It turns out for *finite dimensional* vector spaces there is a *natural* isomorphism $\Phi : V \to V^{**}$, defined as follows

$$\Phi(v)(\alpha) \equiv \alpha(v)$$

for all $\alpha \in V^*$. It is straightforward to check this is an isomorphism of vector spaces. We will identify v with $\Phi(v)$ throughout the rest of the course. We include these comments here so that you can better understand what is meant when we act on a dual vector by a vector. We are not really using the vector, rather its double dual.

While we are thinking about isomorphisms its worth mentioning that V and V^* are also isomorphic, however there is no *natural* isomorphism intrinsic within just the vector space structure itself. If we have a the additional structure of a metric then we can find a *natural* isomorphism. We will explore such isomorphisms at the conclusion of this chapter.

Definition 9.3.2. *Then tensor product of $v \in V$ with $w \in V$ is denoted $v \otimes w$ and is defined by*

$$(v \otimes w)(\alpha, \beta) = \alpha(v)\beta(w) \qquad \textit{for all } \alpha, \beta \in V^*$$

We can demonstrate that $v \otimes w : V^* \times V^* \to \mathbb{R}$ is a bilinear mapping on V^*. Observe,

$$\begin{aligned}(v \otimes w)(\alpha, \beta + c\gamma) &= \alpha(v)(\beta + c\gamma)(w) \\ &= \alpha(v)(\beta(w) + c\gamma(w)) \\ &= \alpha(v)\beta(w) + c\alpha(v)\gamma(w) \\ &= (v \otimes w)(\alpha, \beta) + c(v \otimes w)(\alpha, \gamma).\end{aligned} \tag{9.9}$$

The linearity in the first slot falls out from a similar calculation. Let $v, w \in V$, observe

$$\begin{aligned}(v \otimes w)(\alpha, \beta) &= \alpha(v)\beta(w) \\ &= (\alpha_i e^i)(v)(\beta_j e^j)(w) \\ &= \alpha_i \beta_j e^i(v) e^j(w) \\ &= \alpha_i \beta_j v^i w^j \\ &= v^i w^j \alpha(e_i)\beta(e_j) \\ &= v^i w^j (e_i \otimes e_j)(\alpha, \beta) \\ &= (v^i w^j e_i \otimes e_j)(\alpha, \beta).\end{aligned} \tag{9.10}$$

for all $\alpha, \beta \in V^*$. Not every bilinear mapping on V^* is the tensor product of two vectors, this is a special case. Lets complete the thought,

Definition 9.3.3. *If $B : V^* \times V^* \to \mathbb{R}$ can be written $B = B^{ij} e_i \otimes e_j$ then the numbers B^{ij} are the* **components** *of B. Moreover the components are said to be symmetric if $B^{ij} = B^{ji}$ or antisymmetric if $B^{ij} = -B^{ji}$.*

Proposition 9.3.4. *The set $\{e_i \otimes e_j \mid 1 \leq i, j \leq n\}$ forms a basis for the set of all bilinear maps on the n-dimensional dual space V^*. Any bilinear map $B : V^* \times V^* \to \mathbb{R}$ can be written $B = B^{ij} e_i \otimes e_j$.*

Thus any bilinear map on V^* can be written as a linear combination of tensor products of the basis for V.
Lets present the general definitions without further ado,

Definition 9.3.5. *Then tensor product of $v_1, v_2, \cdots, v_p \in V$ is denoted $v_1 \otimes v_2 \otimes \cdots \otimes v_p$ and is defined by*

$$(v_1 \otimes v_2 \otimes \cdots \otimes v_p)(\alpha_1, \alpha_2, \ldots \alpha_p) = \alpha_1(v_1)\alpha_2(v_2) \cdots \alpha_p(v_p)$$

for all $\alpha_1, \alpha_2, \ldots \alpha_p \in V^*$.

The *components* with respect to the basis $\{e^i\}$ of V^* of a p-linear map on the dual space V^* are given by acting on the basis elements of the dual space.

Definition 9.3.6. *Let $T : V^* \times V^* \times \cdots \times V^* \to \mathbb{R}$ then the* **components** *of T are defined to be*

$$T(e^{i_1}, e^{i_2}, \ldots, e^{i_p}) = T^{i_1 i_2 \ldots i_p}.$$

If $T^{i_1 i_2 \ldots i_p} = T^{i_{\pi(1)} i_{\pi(2)} \cdots i_{\pi(p)}}$ for all $\pi \in \Sigma_p$ then $T^{i_1 i_2 \ldots i_p}$ are **symmetric**.
If $T^{i_1 i_2 \ldots i_p} = sgn(\pi) T_{i_{\pi(1)} i_{\pi(2)} \cdots i_{\pi(p)}}$ for all $\pi \in \Sigma_p$ then $T^{i_1 i_2 \ldots i_p}$ are **antisymmetric**.

As before if V^* has the basis $\{e^i\}$ then the tensor product induces a basis for the p-linear maps on V^* as follows, the components we just defined are simply the coordinates of the p-linear maps with respect to the induced tensor basis.

Proposition 9.3.7. *Let $T : V^* \times V^* \times \cdots \times V^* \to \mathbb{R}$ then*

$$T = T^{i_1 i_2 \ldots i_p} e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_p}.$$

Thus $\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_p}\}$ with $1 \leq i_1, i_2, \ldots, i_p \leq n$ is a basis on the vector space of all p-linear maps on V^*.

Remark 9.3.8. *I hope you can see that the tensor products of vectors enjoy all the same algebraic properties as the tensor product of dual vectors. The product is associative, distributes over addition of vectors and scalars, and so on. It would seem that almost everything is the same except that some indices are up where they were down before and vice-versa. We have spent more time on the first case because it more closely aligns with what we are ultimately interested in, differential forms. The algebraic structure of the tensor product is truly an interesting course of study in and of itself. There is much much more to say.*

9.4 tensors on a vector space V

The multilinear maps on V and V^* we have studied thus far are in fact tensors. We now give the general definition. We put all the copies of V first and then V^* but this is largely just an issue of book-keeping.

Definition 9.4.1. *A type (r, s) tensor on V is a mapping*

$$T : V^* \times V^* \times \cdots \times V^* \times V \times V \times \cdots \times V \to \mathbb{R}$$

where there are r-copies of V^* and s-copies of V that is linear in k^{th} slot,

$$T(\alpha_1, ..., \alpha_k + c\beta, ..., v_p) = T(\alpha_1, ..., \alpha_k, ..., v_p) + cT(\alpha_1, ..., \beta, ..., v_p)$$

for each $k = 1, 2, \ldots r$ for all $c \in \mathbb{R}$ and $v_i \in V$ and $\beta, \alpha_j \in V^*$ where $1 \leq i \leq r$ and $1 \leq j \leq s$. Likewise it is linear in all the V slots,

$$T(\alpha_1, ..., v_m + cy, ..., v_p) = T(\alpha_1, ..., v_m, ..., v_p) + cT(\alpha_1, ..., y, ..., v_p)$$

for each $m = 1, 2, \ldots s$ for all $c \in \mathbb{R}$ and $y, v_i \in V$ and $\alpha_j \in V^*$ where $1 \leq i \leq r$ and $1 \leq j \leq s$.

We identify that a $(p, 0)$ tensor on V is a p-linear mapping on V^* whereas a $(0, p)$ tensor on V is a p-linear mapping on V. More generally when we have a (r, s) tensor and neither r nor s are zero then we say that we have a *mixed tensor* . We follow the conventions of Frankel's *The Geometry of Physics: An Introduction*, other books might put V before V^* in the defintion above, so beware of this ambiguity. Components are defined much as before,

Definition 9.4.2. *Let T be a type (r, s) tensor on V then the* **components** *of T are defined to be*

$$T(e^{i_1}, \dots, e^{i_r}, e_{j_1}, \dots, e_{j_s}) = T^{i_1 \dots i_r}{}_{j_1 \dots j_s}.$$

we need to allow the blank space under the upper indices because we may wish to lower them in general. In case we know that indices will not be raised or lowered then we can omit the space without danger of confusion. We define the space of all type (r, s) tensors on V to be $T^r_s(V)$.

Example 9.4.3. *The metric on a vector space gives a $(0, 2)$ tensor. For example on Euclidean space,*

$$g(x, y) = x^t y$$

for all $x, y \in \mathbb{R}^n$. Or on Minkowski space,

$$g(x, y) = x^t \eta y$$

for all $x, y \in \mathbb{R}^4$.

Example 9.4.4. *The* **Riemann Tensor** *R is a $(1, 3)$ tensor with components $R^{\mu}{}_{\nu\alpha\beta}$. The different indices are really different for this tensor, they have quite distinct symmetry properties. So it would be unwise to omit the space for this mixed tensor as it would lead to much confusion. More than usual. This tensor is at the heart of General relativity which is one of the areas I hope this course helps you prepare for.*

Remark 9.4.5. *Defining symmetric and antisymmetric mixed tensors could be tricky. We cannot just haphazardly exchange any pair of indices, that would mess up the ordering of V and V^*, we could end up with something that was not a tensor according to our book-keeping. If we could make all the indices go either up or down then we could define symmetric and antisymmetric as we did before. For now let us agree to just refer to the indices of the same type(up or down) as symmetric or antisymmetric with the obvious meaning.*

We should define of the tensor product of vectors and dual vectors to be complete, we'll just exhibit the definition for the simple case of one vector and dual vector, the extension to more vectors and tensors should be obvious after the following definition.

Definition 9.4.6. *Let $v \in V$ and $\alpha \in V^*$ then $v \otimes \alpha : V^* \times V \to \mathbb{R}$ and $\alpha \otimes v : V \times V^* \to \mathbb{R}$ are defined by*

$$(v \otimes \alpha)(\beta, x) = \beta(v)\alpha(x) \qquad (\alpha \otimes v)(x, \beta) = \alpha(x)\beta(v)$$

for all $x \in V$ and $\beta \in V^*$.

As a matter of book-keeping we avoid $v \otimes \alpha$ since it has the ordering of V and V^* messed up.

Proposition 9.4.7. *Let T be a type (r, s) tensor on V then*

$$T = T^{i_1 \dots i_r}{}_{j_1 \dots j_s} e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes e^{j_1} \otimes \cdots \otimes e^{j_s}.$$

Thus $\{e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes e^{j_1} \otimes \cdots \otimes e^{j_s}\}$ with $1 \leq i_1, \dots, i_r, j_1, \dots, j_s \leq n$ is a basis for the vector space of all (r,s) tensors on on V.

9.5 raising and lowering indices on tensors

Let us suppose that the vector space V has a metric $g : V \times V \to \mathbb{R}$, for convenience we assume that $g(v, w) = g(w, v)$ for all $v, w \in V$, but we do not assume that g is positive definite. We want to allow g to include the possibilities of the Euclidean metric or the Minkowski metric. Observe that,

$$\begin{aligned} g(v, w) &= g(v^i e_i, w^j e_j) \\ &= v^i w^j g(e_i, e_j) \\ &= v^i w^j g_{ij}. \end{aligned} \tag{9.11}$$

As we discussed in the last chapter the components with indices upstairs are the *contravariant* components. The covariant components are obtained with the help of the metric,

$$\boxed{w_i \equiv w^j g_{ij}} \tag{9.12}$$

This gives us the following nice formula for $g(v, w)$

$$g(v, w) = v^i w_i \tag{9.13}$$

Notice the components of the metric are hidden in the lowered index of w. What do these equations really mean? Why should we lower the index, does that make w a covector ?

Definition 9.5.1. *Given a vector space V and a metric $g : V \times V \to \mathbb{R}$ we define a mapping $\alpha : V \to V^*$ which maps $v \mapsto \alpha_v$ as follows,*

$$\alpha_v(x) \equiv g(v, x) \tag{9.14}$$

for all $v, x \in V$. We say that α_v is the covector or dual vector that corresponds to v.

Remark 9.5.2. *The components of the dual vector corresponding to $v = v^i e_i$ are v_i. More precisely we note that*

$$\begin{aligned} \alpha_v(e_i) &= g(v, e_i) \\ &= g(v^j e_j, e_i) \\ &= v^j g(e_j, e_i) \\ &= v^j g_{ji} \\ &= v_i. \end{aligned} \tag{9.15}$$

Observe that v_i are **not** *the components of the vector v, but rather the components of the corresponding covector α_v.*

Example 9.5.3. Vectors in Minkowski Space: *Given a vector $v = v^\mu e_\mu$ in Minkowski space we call the components of the vector the* **contravariant components**. *Alternatively we can construct the corresponding dual vector $\alpha_v = v_\mu e^\mu$ where the components are the so-called* **covariant components** $v_\mu = \eta_{\mu\nu} v^\nu$. *Again it should be emphasized that without the metric there is no coordinate independent method of making such a correspondence in general. This is why we said there was no natural isomorphism between V and V^*, unless we have a metric. In the presence of a metric we can either view v as a vector or as a dual vector, both contain the same information, just packaged in a different way.*

Lets make a more concrete example,

Example 9.5.4. *In electromagnetism one considers the scalar potential V and the vector potential $\vec{A}$, remember these could be differentiated a particular way to give the electric and magnetic fields. At a particular place and time these give us a 4-vector in Minkowski space as follows,*

$$(A^\mu) \equiv (V, \vec{A})$$

The corresponding covector is obtained by lowering the index with η,

$$(A_\nu) = (\eta_{\mu\nu}A^\mu) = (-A^0, A^1, A^2, A^3) = (-V, \vec{A})$$

we observe that the time-component gains a minus sign but the spatial components stay the same. That minus sign is quite important for the equations later.

Example 9.5.5. Vectors Euclidean Space: *Assume that V is a Euclidean space with the orthonormal basis $\{e_i\}$, meaning that $g(e_i, e_j) = \delta_{ij}$. Further suppose we have a vector $v = v^i e_i$ in V. We construct the corresponding dual vector $\alpha_v = v_i e^i$ by defining $v_i = v^i$. On first glance you might say, hey where's the metric ? Didn't I just say that we needed the metric to raise and lower indices ? The metric is hidden as follows,*

$$v_j = \delta_{ij}v^j.$$

So the Euclidean metric acting on two vectors is obtained by summing against the Kronecker delta,

$$g(v, w) = v^i w_i = v^i w^j \delta_{ij}.$$

Of course if we used another weirder basis in V we would not necessarily have such a nice formula, it is important that we took the components with respect to an orthonormal basis. Anyway, we can now see clearly why it was not a problem to work with indices down in the Euclidean case, with the conventions that we have chosen in this course the Euclidean indices raise and lower without introducing any signs. Minkowski indices in contrast require more care.

Let us recall a theorem from linear algebra.

Theorem 9.5.6. *If V and W are finite dimensional vector spaces over $\mathbb{R}$ then V is isomorphic to W if and only if $dim(V) = dim(W)$.*

Notice there are $(r+s)dim(V)$ vectors in the induced tensor basis for $T^r_s(V)$ (think about the typical basis element $e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes e^{j_1} \otimes \cdots \otimes e^{j_s}$ there are $r+s$ objects tensored together and each one of those can be $dim(V)$ different things.) This means that $T^p_q(V)$ is isomorphic to $T^r_s(V)$ provided that $p+q = r+s$. In the presense of a metric we can easily construct the isomorphism in a coordinate independent fashion. We have already seen this in a special case, vectors make up $T^1_0(V)$, whereas covectors make up $T^0_1(V)$, these spaces are isomorphic as vector spaces. If you read *Gravitation* by Misner, Thorne and Wheeler you'll find they espouse the view point that the vector and covector are the same thing, but from a more pedantic perspective "same" is probably to strong a term. Mathematicians are also guilty of this abuse of language, we often say isomorphic things are the "same". Well are they really the same ? Is the set of matrices $\mathbb{R}^{2\times 2}$ the same as $\mathbb{R}^4$? I'd say no. The remedy is simple. To be careful we should say that they are the "same upto isomorphism of vector spaces".

Let us explicitly work out some higher order cases that will be of physical interest to us later,

Example 9.5.7. Field Tensor: *Let $\vec{E} = (E_1, E_2, E_3)$ and $\vec{B} = (B_1, B_2, B_3)$ be the electric and magnetic field vectors at some point then we define the* **field tensor** *to be the (0,2) tensor in Minkowski space such that $F = F_{\mu\nu} e^\mu \otimes e^\nu$ where*

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \tag{9.16}$$

Convention: *When we write the matrix version of the tensor components we take the first index to be the row index and the second index to be the column index, that means $F_{01} = -E_1$ whereas $F_{10} = E_1$. Define a type $(1,1)$ tensor by raising the first index by the inverse metric $\eta^{\alpha\mu}$ as follows,*

$$F^\alpha{}_\nu = \eta^{\alpha\mu} F_{\mu\nu}$$

The zeroth row,

$$(F^0{}_\nu) = (\eta^{0\mu} F_{\mu\nu}) = (0, E_1, E_2, E_3)$$

Then row one is unchanged since $\eta^{1\mu} = \delta^{1\mu}$,

$$(F^1{}_\nu) = (\eta^{1\mu} F_{\mu\nu}) = (E_1, 0, B_3, -B_2)$$

and likewise for rows two and three. In total the (1,1) tensor $F' = F^\alpha{}_\nu e_\alpha \otimes e^\nu$ has the components below

$$(F^\alpha{}_\nu) = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}. \tag{9.17}$$

Lets take it one step further and raise the other index to create a $(2,0)$ tensor,

$$\boxed{F^{\alpha\beta} = \eta^{\alpha\mu}\eta^{\beta\nu} F_{\mu\nu}} \tag{9.18}$$

see it takes one copy of the inverse metric to raise each index and $F^{\alpha\beta} = \eta^{\beta\nu} F^\alpha{}_\nu$ so we can pick up where we left off in the $(1,1)$ case. We could proceed case-wise like we did with the $(1,1)$ case but I think it is good to know that we can also use matrix multiplication here; $\eta^{\beta\nu} F^\alpha{}_\nu = F^\alpha{}_\nu \eta^{\nu\beta}$ and this is just the (α, β) component of the following matrix product,

$$(F^{\alpha\beta}) = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix}. \tag{9.19}$$

So we find a $(2,0)$ tensor $F'' = F^{\alpha\beta} e_\alpha \otimes e_\beta$. Other books might even use the same symbol F for F' and F'', it is in fact typically clear from the context which version of F one is thinking about. Pragmatically physicists just write the components usually so its not even an issue.

Example 9.5.8. *Lets begin with a tensor* $T = T^{ijk} e_i \otimes e_j \otimes e_k$ *then we can construct other tensors as follows,*

$$T^{ij}{}_k = g_{kn} T^{ijn}$$
$$T^i{}_{jk} = g_{jm} g_{kn} T^{imn}$$
$$T_{ijk} = g_{il} g_{jm} g_{kn} T^{lmn}$$

where T^{lmn} *are defined as usual.*

Remark 9.5.9. *Notice I've only showed how the metric converts vectors to covectors, you are not responsible for explaining in a coordinate free way how covectors can be converted to vectors by the inverse metric. You'll have to content yourself with the component version for now. Or you could bug me in office hours if you really want to know, or take a long look at Dr. Fulp's notes from Fiber bundles 2001 which I have posted on my webpage.*

Example 9.5.10. *One last example, we'll just focus on the components.*

$$\begin{aligned} S^i{}_{jk} &= g^{il} S_{ljk} \\ S^{ij}{}_k &= g^{il} g^{jm} S_{lmk} \\ S^{ijk} &= g^{il} g^{jm} g^{kn} S_{lmn} \end{aligned}$$

here we had to use the inverse metric to raise the indices.

I think that is quite enough about metric dualities for now. For the remainder of these notes we will raise and lower indices as described in this section. In summary, raise indices by using the inverse metric g^{ij}, lower indices by using the metric g_{ij}. Make sure that the free indices match up on both sides and that ought to do it.

9.6 coordinate change and tensors

In the part of the physics community it is common to define a tensor as its components. They require that the components transform in a certain manner, if it can be shown that the components transform that way then it is said to be a tensor. No mention of the tensor basis is even made sometimes, just the components are used. Let me "define" a tensor in that manner, then we'll derive that our tensors work the same,

Definition 9.6.1. $T^{i_1 \dots i_r}{}_{j_1 \dots j_s}$ *is a type* (r, s) *tensor if when the coordinates change according to* $\bar{x}^i = A^i_k x^k$ *then the tensor transforms such that*

$$\bar{T}^{i_1 \dots i_r}{}_{j_1 \dots j_s} = A^{i_1}_{k_1} \cdots A^{i_r}_{k_r} (A^{-1})^{l_1}_{j_1} \cdots (A^{-1})^{l_s}_{j_s} T^{k_1 \dots k_r}{}_{l_1 \dots l_s} \tag{9.20}$$

Lets list a few examples most relevant to us,

$$\bar{v}^j = A^j_i v^i$$

$$\bar{\alpha}_j = (A^{-1})^i_j \alpha_i$$

$$\bar{F}_{\mu\nu} = (A^{-1})^{\alpha}_{\mu}(A^{-1})^{\beta}_{\nu} F_{\alpha\beta}$$

$$\bar{F}^{\mu\nu} = A^{\mu}_{\alpha} A^{\nu}_{\beta} F^{\alpha\beta}$$

Now lets try to link this picture to the one we have developed in previous sections. We consider a tensor to be a multilinear mapping, this is a notion which is independent of basis chosen. However, we typically pick a basis, usually the standard basis and expand the tensor in components with respect to that basis. If we picked a different basis for V say $\{\bar{e}_i\}$ which has dual basis $\{\bar{e}^i\}$ for V^* then we could expand the tensor in terms of that other basis as well. So we'd have two descriptions,

$$T = T^{i_1 \dots i_r}{}_{j_1 \dots j_s} e_{i_1} \otimes \cdots \otimes e_{i_r} \otimes e^{j_1} \otimes \cdots \otimes e^{j_s}.$$

or,

$$T = \bar{T}^{i_1 \dots i_r}{}_{j_1 \dots j_s} \bar{e}_{i_1} \otimes \cdots \otimes \bar{e}_{i_r} \otimes \bar{e}^{j_1} \otimes \cdots \otimes \bar{e}^{j_s}.$$

We discovered before that the basis and dual basis transform inversely, with respect to the coordinate change $\bar{x}^i = A^i_k x^k$ we know that,

$$\bar{e}_i = (A^{-1})^k_i e_k \qquad\qquad \bar{e}^i = A^i_l e^l$$

then calculate the components in the barred coordinate system, by definition,

$$\begin{aligned}
\bar{T}^{i_1 \dots i_r}{}_{j_1 \dots j_s} &= T(\bar{e}^{i_1}, \dots, \bar{e}^{i_r}, \bar{e}_{j_1}, \dots, \bar{e}_{j_s}) \\
&= A^{i_1}_{k_1} \cdots A^{i_r}_{k_r} (A^{-1})^{l_1}_{j_1} \cdots (A^{-1})^{l_s}_{j_s} T(e^{k_1}, \dots, e^{k_s}, e_{l_1}, \dots, e_{l_s}) \\
&= A^{i_1}_{k_1} \cdots A^{i_r}_{k_r} (A^{-1})^{l_1}_{j_1} \cdots (A^{-1})^{l_s}_{j_s} T^{k_1 \dots k_s}{}_{l_1 \dots l_s}
\end{aligned} \tag{9.21}$$

thus the physicist's tensor and our tensor are really the same idea, we just write a little more. Now we can use what we've learned here in later sections when we feel inclined to check coordinate independence. We are being fairly carefree about checking that for most of our definitions, but we will on occasion check to make sure that our tensor is really a tensor. Essentially the problem is that if we give a definition in terms of one coordinate system then how do we know the definition still holds in another coordinate system ?

The quick way to verify coordinate independence is to write everything in tensor notation such that all of our indices upstairs are balanced by a partner downstairs. We'll then find that the object is invariant under coordinate change because the "covariant" and "contravariant" indices transform inversely. To be logically complete one first must show that the indices on the object really to transform "covariantly" or "contravariantly", just because we write them that way doesn't make it so. I'll come back to this point when we show that the field tensor is really a tensor.

Chapter 10

The Exterior Algebra of Forms

We continue where the last chapter left off, the next thing to discuss algebraically are special tensors which we shall call *forms*. We'll see how these forms make an interesting course of study without regard to the tensor product and as a bonus allow us to write a few nice formulas to describe the determinant. Then we will construct the wedge product from the tensor product. Finally we conclude by introducing the remarkable Hodge duality on forms.

10.1 exterior algebra from the wedge product

Let us proceed formally for a little while then we will reconnect with the tensor products.

Definition 10.1.1. *Given a vector space V we can* **define the exterior algebra** $\Lambda(V)$ *to be the span of the of the wedge products of vectors in V. Where the* **wedge product** *to be a multiplication of V that satisfies four properties (mostly following Curtis and Miller's Abstract Linear Algebra)*

(i) *the wedge product $\wedge$ is associative*
(ii) *the wedge product $\wedge$ distributes over vector addition*
(iii) *scalars pull out of the wedge products*
(iv) $e_i \wedge e_j = -e_j \wedge e_i$ *for all $i, j \in \{1, 2, \dots, dim(V)\}$*

the wedge product of p-vectors is said to have degree p. We will call the wedge product of p vectors a "p-vector".

we call $\Lambda(V)$ the *exterior algebra* because the wedge takes us outside of V. It turns out the dimension of $\Lambda(V)$ is finite. Lets see why.

Proposition 10.1.2. linear dependent vectors wedge to zero: *If $x = cy$ for some $c \in \mathbb{R}$ then $x \wedge y = 0$.*

proof: follows from (iii) and (iv), let us write out the vector's basis expansion

$$x = x^i e_i \qquad y = y^i e_i$$

clearly since $x = cy$ it follows that $x^i = cy^i$ for each i. Observe

$$\begin{aligned} x \wedge y &= (cy^i e_i) \wedge (y^j e_j) \\ &= cy^i y^j e_i \wedge e_j \qquad \text{using (ii) and (iii)} \\ &= -cy^i y^j e_j \wedge e_i \qquad \text{using (iv)} \\ &= -cy^j e_j \wedge y^i e_i \qquad \text{using (ii) and (iii)} \\ &= -x \wedge y \end{aligned} \tag{10.1}$$

The proposition is proved, $x \wedge y = 0$ if x and y are linearly dependent.

Proposition 10.1.3. *Suppose that $v_1, v_2, \dots, v_p$ are linearly dependent vectors then*

$$v_1 \wedge v_2 \wedge \cdots \wedge v_p = 0.$$

proof: by assumption of linear dependence there exist constants $c_1, c_2, \dots, c_p$ not all zero such that

$$c_1 v_1 + c_2 v_2 + \cdots c_p v_p = 0.$$

Suppose that c_k is a nonzero constant in the sum above, then we may divide by it and consequently we can write v_k in terms of all the other vectors,

$$v_k = \frac{-1}{c_k}\left(c_1 v_1 + \cdots + c_{k-1} v_{k-1} + c_{k+1} v_{k+1} + \cdots + c_p v_p\right)$$

Insert this sum into the wedge product in question,

$$\begin{aligned} v_1 \wedge v_2 \wedge \ldots \wedge v_p &= v_1 \wedge v_2 \wedge \cdots \wedge v_k \wedge \cdots \wedge v_p \\ &= (-c_1/c_k) v_1 \wedge v_2 \wedge \cdots \wedge v_1 \wedge \cdots \wedge v_p \\ &\quad + (-c_2/c_k) v_1 \wedge v_2 \wedge \cdots \wedge v_2 \wedge \cdots \wedge v_p + \cdots \\ &\quad + (-c_{k-1}/c_k) v_1 \wedge v_2 \wedge \cdots \wedge v_{k-1} \wedge \cdots \wedge v_p \\ &\quad + (-c_{k+1}/c_k) v_1 \wedge v_2 \wedge \cdots \wedge v_{k+1} \wedge \cdots \wedge v_p + \cdots \\ &\quad + (-c_p/c_k) v_1 \wedge v_2 \wedge \cdots \wedge v_p \wedge \cdots \wedge v_p \\ &= 0. \end{aligned} \tag{10.2}$$

We know all the wedge products are zero in the above because in each there is at least one vector repeated, we simply permute the wedge products till they are adjacent, then by (iv) (or the previous proposition) it is clear that $e_i \wedge e_i = 0$. The proposition is proved.

Let us pause to reflect on the meaning of the proposition above for a n-dimensional vector space V. The proposition establishes that there can be at most the wedge product n-vectors. We certainly cannot have more than n linearly independent vectors in a n-dimensional vector space, so if we did take the wedge product of $(n+1)$ vectors then by the proposition its automatically zero. Moreover, we can use the proposition to deduce the form of a basis for $\Lambda(V)$, it must consist of the wedge product of distinct linearly independent vectors. The number of ways to choose p distinct objects from a list of n distinct objects is precisely "n choose p",

$$\binom{n}{p} = \frac{n!}{(n-p)!p!} \tag{10.3}$$

for $0 \leq p \leq n$. Thus if we denote the span of all such wedges of p-linearly independent by $\Lambda^p(V)$ we can deduce that the dimension of the vector space $\Lambda^p(V)$ is precisely $\frac{n!}{(n-p)!p!}$. Naturally we induce a basis on $\Lambda(V)$ from the basis of V itself, I could attempt to write the general situation, its not that hard really, but I think it'll be more transparent if we work in three dimensions for the moment.

Example 10.1.4. exterior algebra of $\mathbb{R}^3$ *Let us begin with the standard basis $\{e_1, e_2, e_3\}$. By definition we take the $p = 0$ case to be the field itself; $\Lambda^0(V) \equiv \mathbb{R}$, it has basis 1. Next, $\Lambda^1(V) = V$. Now for something a little more interesting,*

$$\Lambda^2(V) = span(e_1 \wedge e_2, e_1 \wedge e_3, e_2 \wedge e_3)$$

and finally,

$$\Lambda^3(V) = span(e_1 \wedge e_2 \wedge e_3).$$

This makes $\Lambda(V)$ a $2^3 = 8$-dimensional vector space with basis

$$\{1, e_1, e_2, e_3, e_1 \wedge e_2, e_1 \wedge e_3, e_2 \wedge e_3, e_1 \wedge e_2 \wedge e_3\}$$

it is curious that the number of basis vectors and basis 2-vectors are equal.

10.1.1 wedge product verses cross product

Let us take a moment to contrast the wedge product and the cross product. Let $V = V^i e_i$ and $W = W^j e_j$ be vectors in $\mathbb{R}^3$, earlier we learned that

$$V \times W = \epsilon_{ijk} V^i W^j e_k$$

But what about the wedge product,

$$V \wedge W = V^i e_i \wedge W^j e_j = V^i W^j e_i \wedge ej.$$

These are very similar expressions, if we could say that $\epsilon_{ijk} e_k = e_i \wedge ej$ then we could say that $V \times W = V \wedge W$. However, we **cannot** say just that, the $V \times W$ is similar to the $V \wedge W$, but they are not equal. In particular the following seem similar,

$$\epsilon_{ijk} e_k \approx e_i \wedge ej$$

implies that,

$$\begin{aligned} e_3 &\approx e_1 \wedge e_2 \\ e_2 &\approx e_3 \wedge e_1 \\ e_1 &\approx e_2 \wedge e_3 \end{aligned} \tag{10.4}$$

the wedge product is reproducing the cross product. One big difference, the cross product takes two vectors and gives you back another vector, whereas the wedge product takes in two vectors and gives back an element of degree two which is not a vector, but rather a 2-vector. Moreover the cross product is nonassociative, but the wedge product is associative. In short the wedge product is not the cross product, rather a generalization of the cross product. It is much more general in fact, the cross product works only in three dimensions because it is the wedge product plus the tacit assumption that $\approx$ is actually equality. This is not unreasonable since degree one and two elements both have three components, but in dimensions other than three it is not the case that V and $\Lambda^2(V)$ are isomorphic. You can check this directly with equation (10.3). I'll let you argue in a homework that only in $n = 3$ do we find the situation the the dimension of V and $\Lambda^2(V)$ are equal, thus the cross product exists only in three dimensions.

10.1.2 theory of determinants via the exterior algebra

We begin by making a fundamental observation; for an n-dimensional vector space V it is clear that $dim(\Lambda^0(V)) = dim(\Lambda^n(V)) = 1$ thus if we take any set of n-vectors in V and wedge them together they must be a scalar multiple of the wedge product of the basis of V. In particular we can think about the n-columns of an $n \times n$ matrix A, these can be written as $Ae_1, Ae_2, \ldots, Ae_n$. We can then define the determinant of A as the scalar just mentioned,

Definition 10.1.5. *Let A be an $n \times n$ matrix and $e_1, e_1, \ldots, e_n$ the standard basis of $\mathbb{R}^n$ then the* **determinant of A** *is defined by the equation below*

$$\boxed{Ae_1 \wedge Ae_1 \wedge \cdots \wedge Ae_n \equiv det(A) e_1 \wedge e_1 \wedge \cdots \wedge e_n.} \tag{10.5}$$

Let us verify that this determinant is really the determinant we know and love from linear algebra. I'll work out the 2×2 case then I'll let you do the 3×3 for homework,

Example 10.1.6. *Deriving 2×2 determinant formula from the definition: the Consider the usual arbitrary 2×2 matrix,*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{10.6}$$

Consider then,

$$\begin{aligned} Ae_1 \wedge Ae_2 &= (ae_1 + ce_2) \wedge (be_1 + de_2) \\ &= abe_1 \wedge e_1 + ade_1 \wedge e_2 + cbe_2 \wedge e_1 + cde_2 \wedge e_2 \\ &= ade_1 \wedge e_2 - cbe_1 \wedge e_2 \\ &= (ad - bc) e_1 \wedge e_2 \end{aligned} \tag{10.7}$$

where all we used in the calculation above was plain old matrix multiplication plus the antisymmetry of the wedge product which tells us that $e_2 \wedge e_1 = -e_1 \wedge e_2$ and $e_1 \wedge e_1 = e_2 \wedge e_2 = 0$.

The proposition to follow is easy to prove now that we have a good definition for the determinant.

Proposition 10.1.7. *Let A be an $n \times n$ square matrix and let I be the $n \times n$ identity matrix and $r \in \mathbb{R}$ then*
(i) $det(I) = 1$
(ii) $det(A) = 0$ *if the columns of A are linearly dependent*
(iii) $det(rA) = r^n det(A)$

proof: To begin notice the k^{th} column of I is e_k thus,

$$Ie_1 \wedge Ie_2 \wedge \cdots \wedge Ie_n = e_1 \wedge e_2 \wedge \cdots \wedge e_n$$

therefore since the coefficient of the LHS of the above is defined to be the determinant we can read that $det(I) = 1$. Next to prove (ii) we appeal to proposition 10.1.3, if the columns are linearly dependent then the wedge is zero hence the $det(A) = 0$. Lastly consider (iii) we'll prove it by an indirect calculation,

$$\begin{aligned} rAe_1 \wedge rAe_2 \wedge \cdots \wedge rAe_n &= r^n Ae_1 \wedge Ae_2 \wedge \cdots \wedge Ae_n \\ &\equiv det(rA) e_1 \wedge e_2 \wedge \cdots \wedge e_n \end{aligned} \tag{10.8}$$

thus by comparing the equations we read off that $det(rA) = r^n det(A)$ just as we claimed.

Remark 10.1.8. *There is more we could do with the theory of determinants. With a little more work we could prove that $det(AB) = det(A)det(B)$. See Chapter II of Morton L. Curtis' "Abstract Linear Algebra" for a very readable treatment of these matters. Or you might also look at Chapter 5 of Hoffman and Kunze's "Linear Algebra" for a more advanced presentation of the theory of determinants.*

10.2 the wedge product constructed from the tensor product

You might notice that we dropped all mention of multilinear mappings and tensor products in the last section. The exterior algebra is interesting independent of its relation to the tensor product and I simply wanted to emphasize that. In this section we will show that there is an exterior algebra hidden inside the tensor algebra. It is simply the set of all alternating tensors. The wedge product in the last section was abstract, but now and for the rest of the course we will link the wedge product with the tensor product and as such view the wedge product of objects as a multilinear mapping on a Cartesian product of V. These special tensors are called **forms** or sometimes **alternating forms**.

Definition 10.2.1. *We define the wedge product on the dual basis of the vector space V as follows*

$$e^i \wedge e^j \equiv e^i \otimes e^j - e^j \otimes e^i$$

this is a **2-form**. *For three dual basis vectors,*

$$\begin{aligned} e^i \wedge e^j \wedge e^k \ &\equiv e^i \otimes e^j \otimes e^k + e^j \otimes e^k \otimes e^i + e^k \otimes e^i \otimes e^j \\ &-e^k \otimes e^j \otimes e^i - e^j \otimes e^i \otimes e^k - e^i \otimes e^k \otimes e^j \end{aligned} \tag{10.9}$$

this is a **3-form**. *In general we define the wedge product of p dual basis vectors,*

$$e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_p} = \sum\nolimits_{\pi \in \Sigma_p} sgn(\pi) e^{i_{\pi(1)}} \otimes e^{i_{\pi(2)}} \otimes \cdots \otimes e^{i_{\pi(p)}} \tag{10.10}$$

*this is a p-***form**. *Next define the wedge product between a p-form and a q-form,*

$$\begin{aligned} \alpha_p \wedge \beta_q \ &= (\tfrac{1}{p!}\alpha_{i_1 i_2 \dots i_p} e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_p}) \wedge (\tfrac{1}{q!}\beta_{j_1 j_2 \dots j_q} e^{j_1} \wedge e^{j_2} \wedge \cdots \wedge e^{j_q}) \\ &\equiv \tfrac{1}{p!}\tfrac{1}{!q}\alpha_{i_1 i_2 \dots i_p}\beta_{j_1 j_2 \dots j_q} e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_p} \wedge e^{j_1} \wedge e^{j_2} \wedge \cdots \wedge e^{j_q} \end{aligned} \tag{10.11}$$

*it is a $(p+q)$-***form**. *The factors of $\frac{1}{p!}$ and $\frac{1}{q!}$ are included so that the components appearing in the expressions above are the tensor components of α_p and β_q (meaning that with respect to the tensor product we could write $\alpha_p = \alpha_{i_1 i_2 \dots i_p} e^{i_1} \otimes e^{i_2} \otimes \cdots \otimes e^{i_p}$ and similarly for β_q.) Lastly we define the set of all p-forms over V to be $\Lambda^p(V)$ where we define $\Lambda^0(V) = \mathbb{R}$ and $\Lambda^1(V) = V^*$. The space of all forms over V is denoted $\Lambda(V)$.*

Proposition 10.2.2. *The wedge product defined above is a wedge product as defined in the last section. It is an associative, distributive over addition, pull scalars out, antisymmetric product.*

The proof of this proposition follows from the fact that the tensor product is an associative product with all the requisite linearity properties, and the antisymmetry follows from the fact we *antisymmetrized* the tensor product to define the wedge product, in other words the definition of $\wedge$ was chosen to select the completely antisymmetric tensor product of order p. I'll let you show it for a basic case in homework.

Remark 10.2.3. *Any completely antisymmetric tensor of type $(0,p)$ can be written in terms of a sum of wedge products. In general it works out as follows, given that*

$$T = T_{i_1 i_2 \dots i_p} e^{i_1} \otimes e^{i_2} \otimes \cdots \otimes e^{i_p}$$

is a completely antisymmetric tensor we may show that

$$T = \frac{1}{p!} T_{i_1 i_2 \dots i_p} e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_p}$$

the sum is taken over all indices which means we are not using a basis, I'll let you think about that in a homework. If we instead summed over increasing strings of indices (we'll not do that in this course, but other books do) then we would have a linearly independent set of wedges and the $p!$ would not appear in that case. Also some other books include a $\frac{1}{p!}$ in the definition of the wedge product, this will also make the $\frac{1}{p!}$ disappear. Anyway, you are not responsible for what other books say and such, I just mention it because it can be a source of great confusion if you start trying to mix and match various books dealing with wedge products.

Let us summarize what properties the wedge products of forms enjoy,

Proposition 10.2.4. *Let α, β, γ be forms on V and $c \in \mathbb{R}$ then*
(i) $(\alpha + \beta) \wedge \gamma = \alpha \wedge \gamma + \beta) \wedge \gamma$
(ii) $\alpha \wedge (c\beta) = c(\alpha \wedge \beta)$
(iii) $\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma$

We can derive another nice property of forms,

Proposition 10.2.5. *Let α_p, β_q be forms on V of degree p and q respectively then*

$$\alpha_p \wedge \beta_q = -(-1)^{pq} \beta_q \wedge \alpha_p \tag{10.12}$$

Rather than give you a formal proof of this proposition lets work out an example.

Example 10.2.6. *Let α be a 2-form defined by*

$$\alpha = ae^1 \wedge e^2 + be^2 \wedge e^3$$

And let β be a 1-form defined by

$$\beta = 3e^1$$

Consider then,

$$\begin{aligned} \alpha \wedge \beta &= (ae^1 \wedge e^2 + be^2 \wedge e^3) \wedge (3e^1) \\ &= (3ae^1 \wedge e^2 \wedge e^1 + 3be^2 \wedge e^3 \wedge e^1 \\ &= 3be^1 \wedge e^2 \wedge e^3. \end{aligned} \tag{10.13}$$

whereas,

$$\begin{aligned} \beta \wedge \alpha &= 3e^1 \wedge (ae^1 \wedge e^2 + be^2 \wedge e^3) \\ &= (3ae^1 \wedge e^1 \wedge e^2 + 3be^1 \wedge e^2 \wedge e^3 \\ &= 3be^1 \wedge e^2 \wedge e^3. \end{aligned} \tag{10.14}$$

so this agrees with the proposition, $(-1)^{pq} = (-1)^2 = 1$ so we should have found that $\alpha \wedge \beta = \beta \wedge \alpha$. This illustrates that although the wedge product is antisymmetric on the basis, it is not always antisymmetric, in particular it is commutative for even forms.

Remark 10.2.7. *The set $\Lambda(V) = \Lambda^0 \oplus \Lambda^1 \oplus \cdots \oplus \Lambda^n$ is a vector space with a wedge product. Moreover the multiplication is graded commutative as described in the proposition and exhibited in the preceding example. This makes $\Lambda(V)$ a* **Grassman Algebra**. *It is a finite dimensional algebra because V is finite dimensional, in fact it's dimension is 2^n.*

Example 10.2.8. field tensor is a 2-form: *Recall that $F = F_{\mu\nu} e^\mu \otimes e^\nu$ where the components of F were made of the components of the electric and magnetic fields. If you look at the components given in equation 9.16 it is clear that F is an antisymmetric tensor, that means $F_{\mu\nu} = -F_{\nu\mu}$. Let me demonstrate how we can rewrite it using the wedge product,*

$$\begin{aligned} F &= F_{\mu\nu} e^\mu \otimes e^\nu \\ &= \tfrac{1}{2}(F_{\mu\nu} - F_{\nu\mu}) e^\mu \otimes e^\nu \\ &= \tfrac{1}{2}(F_{\mu\nu} e^\mu \otimes e^\nu - F_{\nu\mu} e^\mu \otimes e^\nu) \\ &= \tfrac{1}{2}(F_{\mu\nu} e^\mu \otimes e^\nu - F_{\mu\nu} e^\nu \otimes e^\mu) \\ &= \tfrac{1}{2}F_{\mu\nu}(e^\mu \otimes e^\nu - e^\nu \otimes e^\mu) \\ &= \tfrac{1}{2}F_{\mu\nu} e^\mu \wedge e^\nu. \end{aligned} \tag{10.15}$$

There are two kinda tricky things I did in the calculation above. For one I used the antisymmetry of $F_{\mu\nu}$ to rewrite it in it's "antisymmetrized form" $\frac{1}{2}(F_{\mu\nu} - F_{\nu\mu})$, this is equal to $F_{\mu\nu}$ thanks to the antisymmetry. Then in the fourth line I relabeled the sums trading μ for ν and vice-versa.

10.3 Hodge duality

Hodge duality stems from the observation that $dim(\Lambda^p(V)) = dim(\Lambda^{n-p}(V))$, you'll prove this in a homework. This indicates that there is a one to one correspondence between $(n-p)$-forms and p-forms. When our vector space has a metric we can write this correspondence in a nice coordinate independent manner. Lets think about what we are trying to do here, we need to find a way to create a $(n-p)$-form from a p-form. An $(n-p)$-form has $(n-p)$ antisymmetric components, however a p-form has p-antisymmetric components. If we summed the p-components against p of the components of the n-dimensional Levi-Civita symbol then that will almost do it. We need the indices that are summed over to be half up and half down to insure coordinate independence of our correspondence, this means we should raise the p of the components. If you didn't get what I was saying in this paragraph that's ok, I was just trying to motivate the following definition.

Definition 10.3.1. *Let V be a vector space with a metric g. If $\alpha = \frac{1}{p!}\alpha_{i_1 i_2 \dots i_p} e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_p} \in \Lambda^p(V)$ then define ${}^*\alpha \in \Lambda^{n-p}(V)$ by*

$$^*\alpha \equiv \frac{1}{p!}\frac{1}{(n-p)!}\alpha^{i_1 i_2 \dots i_p}\epsilon_{i_1 i_2 \dots i_p j_1 j_2 \dots j_{n-p}} e^{j_1} \wedge e^{j_2} \wedge \cdots \wedge e^{j_{n-p}}$$

the components of ${}^*\alpha$ are

$$^*\alpha_{j_1 j_2 \dots j_{n-p}} = \frac{1}{p!}\alpha^{i_1 i_2 \dots i_p}\epsilon_{i_1 i_2 \dots i_p j_1 j_2 \dots j_{n-p}}$$

where as always we refer to the tensor components when we say components and the indices are raised with respect to the metric g as we described at length previously,

$$\alpha^{i_1 i_2 \dots i_p} = g^{i_1 j_1} g^{i_2 j_2} \cdots g^{i_p j_p} \alpha_{j_1 j_2 \dots j_p}$$

Example 10.3.2. *Consider $V = \mathbb{R}^3$ with the Euclidean metric. Let us calculate the Hodge dual of $\alpha = e^1 \wedge e^2 \wedge e^3$, to do this we'll need to figure out what the components of α are,*

$$\alpha = \frac{3!}{3!}\delta^1_i \delta^2_j \delta^3_k e^i \wedge e^j \wedge e^k$$

we need the components of the the form to be antisymmetric, so antisymmetrize,

$$\alpha = \frac{1}{3!}\delta^1_{[i} \delta^2_j \delta^3_{k]} e^i \wedge e^j \wedge e^k$$

here the [*and*] *indicate we should take all* **antisymmetric** *combinations of i, j, k in this case. But this is just the antisymmetric symbol ϵ_{ijk} and we have to divide by* 3! *to avoid double counting,*

$$\alpha = \frac{1}{3!}\epsilon_{ijk} e^i \wedge e^j \wedge e^k$$

this means that (remember that 1/3! *is used up because we are using the form expansion),*

$$\alpha_{ijk} = \epsilon_{ijk}$$

thus we find,

$$\begin{aligned} {}^*(e^1 \wedge e^2 \wedge e^3) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\epsilon_{ijk}\epsilon_{ijk} \\ &= \tfrac{1}{3!}\tfrac{1}{(0)!}6 \\ &= 1. \end{aligned} \tag{10.16}$$

the number 1 *is a* $3 - 3 = 0$ *form as we should expect. The fact that $\epsilon_{ijk}\epsilon_{ijk} = 6$ follows from summing the six nontrivial combinations of* 1, 2, 3 *where one finds*

$$\epsilon_{ijk}\epsilon_{ijk} = 1 + 1 + 1 + (-1)^2 + (-1)^2 + (-1)^2 = 6$$

Lets go the other way, lets find the Hodge dual of a number,

Example 10.3.3. *Consider $V = \mathbb{R}^3$ with the Euclidean metric. Let us calculate the Hodge dual of $\beta = 1$, the components are very easy to find, there are none. Hence,*

$$\begin{aligned} {}^*(1) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\epsilon_{ijk} e^i \wedge e^j \wedge e^k \\ &= \tfrac{1}{0!}\tfrac{1}{(3)!}6 e^1 \wedge e^2 \wedge e^3 \\ &= e^1 \wedge e^2 \wedge e^3 \end{aligned} \tag{10.17}$$

let me elaborate a little on where the 6 came from, I'll list only the nonzero terms,

$$\begin{aligned} \epsilon_{ijk} e^i \wedge e^j \wedge e^k &= \epsilon_{123} e^1 \wedge e^2 \wedge e^3 + \epsilon_{231} e^2 \wedge e^3 \wedge e^1 + \epsilon_{312} e^3 \wedge e^1 \wedge e^2 \\ &\quad + \epsilon_{321} e^3 \wedge e^2 \wedge e^1 + \epsilon_{213} e^2 \wedge e^1 \wedge e^3 + \epsilon_{132} e^1 \wedge e^3 \wedge e^2 \\ &= 6 e^1 \wedge e^3 \wedge e^2 \end{aligned} \tag{10.18}$$

since the antisymmetry of the Levi-Civita symbol and the antisymmetry of the wedge product conspire above to produce a + in each of those terms.

Remark 10.3.4. *It would seem we have come full circle,*

$${}^*\alpha = \beta \quad and \quad {}^*\beta = \alpha \implies \quad {}^{**}\alpha = \alpha.$$

Generally when you take the Hodge dual twice you'll get a minus sign that depends both on the metric used and the degree of the form in question.

That's enough of this for now. We'll do much more in the next chapter where the notation is a little friendlier for differential forms.

Chapter 11

Differential Forms

We have already done most of the foundational algebraic work in the previous two chapters. What changes here is that we now would like to consider functions from M to $\Lambda(TM)$, meaning that for each $p \in M$ we assign a form over the tangent space to p on M; $p \mapsto \alpha(p) \in \Lambda(T_pM)$. When M is a *flat space* then it can be globally identified with it's tangent space, so we would consider a form-valued function from M to $\Lambda(M)$ in that very special but important case (usually I'll assume this is the case, in fact only when we get to integration will I be forced to think otherwise, so you may assume that there are globally defined coordinates on the space M we work on in this chapter.)

Technical points aside, we simply wish to assign a form to each point in M. We could call this a form-field if we wished to make the terminology analogous to that of vectors and vector fields, however it gets tiring to always say "form field" so instead we will call such functions *differential forms*. We could also consider general tensor valued functions on V, those would be called tensor fields. We will not do that though, the reason is that differential forms are all we need for the physics in this course. Moreover, differential forms have a natural derivative which closes on forms. You can differentiate tensors in a way that gives back tensors after the derivative, that is called the *Lie derivative* and I'll let you learn about it in some other course (Riemannian geometry for example). The natural derivative on differential forms is called the *exterior derivative*, we will see how it encodes all the interesting derivatives of the usual vector calculus. In addition to defining differential forms we will delve deeper into how exactly vectors fit together with them. We will learn that differential forms provide another language for talking about the mathematics of vector calculus. It is in fact a more refined language that exposes certain facts that remain hidden in the usual calculus of vector fields. Throughout this chapter we will assume that our vector space has a metric g. This is essential because in order to use Hodge duality we need a metric.

11.1 basics and the definition of differential form

The tangent space at a point in $p \in M$ is denoted T_pM. We will actually be more interested in the dual space to T_pM, it is denoted T_pM^* and relative to the coordinate system $(x^1, x^2, \ldots, x^n)$ it has the basis

$$d_px^1,\ d_px^2,\ \ldots\ d_px^n$$

where we have placed the p to emphasize that this is the dual basis to the cotangent space at $p \in M$. We may drop the p in what follows,

$$\alpha = \alpha_i dx^i$$

this is a differential one-form on M. At a particular point $p \in M$ it will give us a 1-form,

$$(\alpha)(p) = \alpha_i(p) d_p x^i.$$

Notice that α_i is not just a number anymore, for a differential form on M the components will be functions on M. When we evaluate a differential form at a particular point then we get back to what we considered in the previous chapter. I was careful to always use the standard basis for a fixed dual vector space, we dealt with the wedge products of e^i. Well now we will deal with the wedge products of dx^i, the notation reminds us that as x varies so does the dual basis so we are not dealing with just one vector space but rather a vector space at each point of M. The algebra we did in the last chapter still holds true, we just do it one point at a time. Please don't think to hard about these matters, I admit to treat them properly we'd need much more time. The course in Manifold theory will explain what T_pM and T_pM^* are carefully, we'll really just think about the *algebraic* aspects of the differentials in this course (if you want to know more just ask me in office hours).

Definition 11.1.1. *A differential p-form on M is a "smooth" assignment of a p-form to each point in M. A differential p-form may be written in terms of a the coordinate system $\{x^1, x^2, \dots, x^n\}$ as follows,*

$$\alpha = \frac{1}{p!}\alpha_{i_1 i_2 \dots i_p} dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_p}$$

By "smooth" we simply mean the component functions $\alpha_{i_1 i_2 \dots i_p}$ are "smooth" meaning that we can take as many partial derivatives as our heart desires.

We defer the coordinate independent definition to a later course because to do things thoroughly we ought to talk about the *tangent and cotangent bundles* where we could better explain the ideas of coordinate change and smoothness.

11.2 differential forms in $\mathbb{R}^3$

As customary we begin with Cartesian coordinates $(x, y, z) = (x^1, x^2, x^3)$ on $\mathbb{R}^3$ with the standard Euclidean metric. Differential forms can be can be written in terms of dx, dy, dz as follows, In the above please note that $f, \alpha_i, \beta_{ij}, g$ are all functions. The reason I placed quotes on "basis" is that technically since the coefficients are functions not numbers its not a basis in the usual sense of linear algebra. (ask me if you wish more clarification on this idea). Also notice that these are all the nontrivial forms, the three-form is also called a **top form** because it has the highest degree possible in three dimensions.

Name	Degree	Typical Element	"Basis" for $\Lambda^p(\mathbb{R}^3)$
function	$p = 0$	f	1
one-form	$p = 1$	$\alpha = \alpha_i dx^i$	dx, dy, dz
two-form	$p = 2$	$\beta = \beta_{ij} dx^i \wedge dx^j$	$dy \wedge dz,\ dz \wedge dx,\ dx \wedge dy$
three-form	$p = 3$	$\gamma = g dx \wedge dy \wedge dz$	$dx \wedge dy \wedge dz$

Example 11.2.1. Wedge Product: *still makes sense. Let $\alpha = f dx + g dy$ and let $\beta = 3dx + dz$ where f, g are functions. Find $\alpha \wedge \beta$, write the answer in terms of the "basis" given in the table above,*

$$\begin{aligned} \alpha \wedge \beta &= (f dx + g dy) \wedge (3dx + dz) \\ &= f dx \wedge (3dx + dz) + g dy \wedge (3dx + dz) \\ &= 3f dx \wedge dx + f dx \wedge dz + 3g dy \wedge dx + g dy \wedge dz \\ &= -g dy \wedge dz - f dz \wedge dx - 3g dx \wedge dy \end{aligned} \tag{11.1}$$

Example 11.2.2. Top form: *Let $\alpha = dx \wedge dy \wedge dz$ and let β be any other form with degree $p > 0$. We argue that $\alpha \wedge \beta = 0$. Notice that if $p > 0$ then there must be at least one differential inside β so if that differential is dx^k we can rewrite $\beta = dx^k \wedge \gamma$ for some γ. Then consider,*

$$\alpha \wedge \beta = dx \wedge dy \wedge dz \wedge dx^k \wedge \gamma \tag{11.2}$$

now k has to be either $1, 2$ or 3 therefore we will have dx^k repeated, thus the wedge product will be zero. (can you prove this?).

Let us return to the issue of Hodge duality. Lets work out the action of the Hodge dual on the basis, first I'll revisit some examples in the new differential notation.

Remark 11.2.3. *the algebra has the same form, but if you think about it hard we are doing infinitely more calculations here than we did in previous chapters. Let me attempt an analogy, 1+2=3 verses say f+2f=3f for a function f. The arithmatic is like the form calculations, the function addition follows same algebra but it implicits an infinity of additions, one for each $x \in dom(f)$. Likewise equations involving differential forms implicit an infinite number of form calculations, one at each point. This is more of a conceptual hurdle than a calculational hurdle since the calculations look the same for differential forms and forms at a point.*

Example 11.2.4. *Let us calculate the Hodge dual of $\alpha = dx^1 \wedge dx^2 \wedge dx^3$, to do this we'll need to figure out what the components of α are,*

$$\alpha = \frac{3!}{3!} \delta^1_i \delta^2_j \delta^3_k dx^i \wedge dx^j \wedge dx^k$$

we need the components of the the form to be antisymmetric, so antisymmetrize,

$$\alpha = \frac{1}{3!} \delta^1_{[i} \delta^2_j \delta^3_{k]} dx^i \wedge dx^j \wedge dx^k$$

here the [and] indicate we should take all **antisymmetric** *combinations of i, j, k in this case. But this is just the antisymmetric symbol ϵ_{ijk} and we have to divide by $3!$ to avoid double counting,*

$$\alpha = \frac{1}{3!} \epsilon_{ijk} dx^i \wedge dx^j \wedge dx^k$$

this means that (remember that $1/3!$ *is used up because we are using the form expansion),*

$$\alpha_{ijk} = \epsilon_{ijk}$$

thus we find,

$$\begin{aligned} {}^*(dx^1 \wedge dx^2 \wedge dx^3) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\epsilon_{ijk}\epsilon_{ijk} \\ &= \tfrac{1}{3!}\tfrac{1}{(0)!}6 \\ &= 1. \end{aligned} \tag{11.3}$$

the number 1 *is a* $3-3=0$ *form as we should expect.*

Lets go the other way, lets find the Hodge dual of a number,

Example 11.2.5. *Let us calculate the Hodge dual of* $\beta = 1$*, the components are very easy to find, there are none. Hence,*

$$\begin{aligned} {}^*(1) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\epsilon_{ijk}dx^i \wedge dx^j \wedge dx^k \\ &= \tfrac{1}{0!}\tfrac{1}{(3)!}6dx^1 \wedge dx^2 \wedge dx^3 \\ &= dx^1 \wedge dx^2 \wedge dx^3 \end{aligned} \tag{11.4}$$

I've me elaborated where the 6 came from before (see example 10.3.3*),*

Example 11.2.6. *Let us calculate the Hodge dual of* $\gamma = dx$*, clearly it only has a* x*-component, indeed a little thought should convince you that* $\gamma = \delta^1_i dx^i$*. We expect to find a* $3-1=2$*-form. Calculate from the definition as usual,*

$$\begin{aligned} {}^*(dx) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\delta^1_i\epsilon_{ijk}dx^j \wedge dx^k \\ &= \tfrac{1}{1!}\tfrac{1}{2!}\epsilon_{1jk}dx^j \wedge dx^k \\ &= \tfrac{1}{2}(\epsilon_{123}dx^2 \wedge dx^3 + \epsilon_{132}dx^3 \wedge dx^2) \\ &= \tfrac{1}{2}(dx^2 \wedge dx^3 - (-dx^2 \wedge dx^3)) \\ &= dx^2 \wedge dx^3 \\ &= dy \wedge dz \end{aligned} \tag{11.5}$$

Example 11.2.7. *Let us calculate the Hodge dual of* $\alpha = dy \wedge dz$*, a little thought should convince you that*

$$\alpha = \delta^2_i\delta^3_j dx^i \wedge dx^j = \frac{1}{2}2\delta^2_i\delta^3_j dx^i \wedge dx^j$$

we need the components of the the form to be antisymmetric, so antisymmetrize,

$$\alpha = \delta^2_i\delta^3_j dx^i \wedge dx^j = \frac{1}{2}\delta^2_{[i}\delta^3_{j]} dx^i \wedge dx^j$$

here the [*and*] *indicate we should take all* **antisymmetric** *combinations of* i, j *in this case. Explicitly this means* $\delta^2_{[i}\delta^3_{j]} = \delta^2_i\delta^3_j - \delta^2_j\delta^3_i$ *and we had to divide by two to avoid double counting. We can antisymmetrize because we are summing against the wedge product and any symmetric combinations will vanish.*

(**this is why not antisymmetrizing worked for us in the course, we contracted against the** ϵ_{ijk} **symbol so it squashed our error of saying** $\alpha_{ij} = 2\delta^2_i\delta^3_j$**. We now see**

that instead we should have said $\alpha_{ij} = \delta^2_{[i}\delta^3_{j]}$*)*

Thus $\alpha_{ij} = \delta^2_{[i}\delta^3_{j]}$. *We expect to find an* $n-p = 3-2 = 1$*-form. Calculate from the definition as usual,*

$$\begin{aligned} {}^*(dy \wedge dz) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\delta^2_{[i}\delta^3_{j]}\epsilon_{ijk}dx^k \\ &= \tfrac{1}{2!}\tfrac{1}{1!}(\epsilon_{23k}dx^k - \epsilon_{32k}dx^k) \\ &= \tfrac{1}{2!}(\epsilon_{231}dx^1 - \epsilon_{321}dx^1) \\ &= dx^1 \\ &= dx. \end{aligned} \tag{11.6}$$

We have found the Hodge dual of a basis form of each degree. I'll collect all of the results which we found so far as well as a few more which I will let you prove in the homework,

Proposition 11.2.8. *For three dimensional Euclidean space Hodge duality gives us the following correspondences*

${}^*1 = dx \wedge dy \wedge dz$	${}^*(dx \wedge dy \wedge dz) = 1$
${}^*dx = dy \wedge dz$	${}^*(dy \wedge dz) = dx$
${}^*dy = dz \wedge dx$	${}^*(dz \wedge dx) = dy$
${}^*dz = dx \wedge dy$	${}^*(dx \wedge dy) = dz$

Observe that the wedge product plus Hodge duality is replicating the cross product, as $\hat{\imath} \times \hat{\jmath} = \hat{k}$ similarly we find that ${}^*(dx \wedge dy) = dz$. In order to better discuss the how vector fields and differential forms are related we should give a mapping from one to the other. Notice that we have two choices. Either a vector field could map to an one-form or a two-form. Both one and two forms have three components and now that we see how Hodge duality relates one and two forms it is quite evident that the following maps are natural,

Definition 11.2.9. *Let* $\vec{A} = (A^1, A^2, A^3)$ *denote a vector field in* $\mathbb{R}^3$. *Define then,*

$$\omega_A = A_i dx^i$$

the so-called "work-form" of $\vec{A}$. *Also define*

$$\Phi_A = A_i {}^*dx^i = \frac{1}{2}A_i\epsilon_{ijk}dx^i \wedge dx^j$$

the so-called "flux-form" of $\vec{A}$.

We have chosen to follow R.W.R. Darling's *Differential Forms and Connections* notation for the flux and work form mappings. These mappings are important as they provide the link between vector analysis and differential forms in $\mathbb{R}^3$.

Example 11.2.10. *Let* $\vec{A} = (a, b, c)$ *then*

$$\omega_A = adx + bdy + cdz$$

and

$$\Phi_A = ady \wedge dz + bdz \wedge dx + cdx \wedge dy$$

we have two versions of a vector field in the formalism of differential forms. In vector analysis physicists sometimes refer to certain vectors as "polar" and others as "axial". Polar vectors flip to minus themselves under a coordinate inversion whereas axial vectors are invariant under a coordinate inversion. If you analyze the differential forms here in view of that discussion you'll see that Φ_A is behaving like an axial (or pseudovector) and ω_A is behaving like a polar vector. What was hidden with the vector notation is now explicit with differential forms.

Remark 11.2.11. *Given a particular vector $\vec{A} = (a, b, c)$ we've shown that there are two possible corresponding forms, the "work-form" ω_A or the "flux-form" Φ_A. Hodge duality exchanges these two pictures, observe*

$$\begin{aligned} {}^*\omega_A &= {}^*(adx + bdy + cdz) \\ &= a^*dx + b^*dy + c^*dz \\ &= ady \wedge dz + bdz \wedge dx + cdx \wedge dy \\ &= \Phi_A \end{aligned} \tag{11.7}$$

in retrospect we can now see why we found before that $V \times W$ was similar to $V \wedge W$, we commented that $V \times W \approx V \wedge W$. We can be more precise now,

$$\omega_V \wedge \omega_W = \Phi_{\vec{V} \times \vec{W}} \tag{11.8}$$

this is the manner in which the cross-product and wedge product are related. I've left the verification of this claim for you as homework.

11.3 differential forms in Minkowski space

The logic here follows fairly close to the last section, however the wrinkle is that the metric here demands more attention. We must take care to raise the indices on the forms when we Hodge dual them. First lets list the basis differential forms, we have to add time to the mix (again $c = 1$ so $x^0 = ct = t$ if you worried about it) Remember that the Greek indices are defined to

Name	Degree	Typical Element	"Basis" for $\Lambda^p(\mathbb{R}^4)$
function	$p = 0$	f	1
one-form	$p = 1$	$\alpha = \alpha_\mu dx^\mu$	dt, dx, dy, dz
two-form	$p = 2$	$\beta = \frac{1}{2}\beta_{\mu\nu} dx^\mu \wedge dx^\nu$	$dy \wedge dz,\ dz \wedge dx,\ dx \wedge dy$
			$dt \wedge dx,\ dt \wedge dy,\ dt \wedge dz$
three-form	$p = 3$	$\gamma = \frac{1}{3!}\gamma_{\mu\nu\alpha} dx^\mu \wedge dx^\nu dx^\alpha$	$dx \wedge dy \wedge dz, dt \wedge dy \wedge dz$
			$dt \wedge dx \wedge dz, dt \wedge dx \wedge dy$
four-form	$p = 4$	$gdt \wedge dx \wedge dy \wedge dz$	$dt \wedge dx \wedge dy \wedge dz$

range over $0, 1, 2, 3$. Here the top form is degree four since in four dimensions we can have four differentials without a repeat. Wedge products work the same as they have before, just now we have dt to play with. Hodge duality may offer some surprises though.

Definition 11.3.1. *The antisymmetric symbol in* **flat** $\mathbb{R}^4$ *is denoted $\epsilon_{\mu\nu\alpha\beta}$ and it is defined by the value*

$$\epsilon_{0123} = 1$$

plus the demand that it be completely antisymmetric.

We must not assume that this symbol is invariant under a cyclic exhange of indices. Consider,

$$\begin{aligned} \epsilon_{0123} &= -\epsilon_{1023} && \text{flipped (01)} \\ &= +\epsilon_{1203} && \text{flipped (02)} \\ &= -\epsilon_{1230} && \text{flipped (03).} \end{aligned} \tag{11.9}$$

In four dimensions we'll use antisymmetry directly and forego the cyclicity shortcut. Its not a big deal if you notice it before it confuses you.

Example 11.3.2. *Find the Hodge dual of $\gamma = dx$ with respect to the Minkowski metric $\eta_{\mu\nu}$, to begin notice that dx has components $\gamma_\mu = \delta^1_\mu$ as is readily verified by the equation $dx = \delta^1_\mu dx^\mu$. Lets raise the index using η as we learned previously,*

$$\gamma^\mu = \eta^{\mu\nu}\gamma_\nu = \eta^{\mu\nu}\delta^1_\nu = \eta^{1\mu} = \delta^{1\mu}$$

Starting with the definition of Hodge duality we calculate

$$\begin{aligned} {}^*(dx) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\gamma^\mu \epsilon_{\mu\nu\alpha\beta} dx^\nu \wedge dx^\alpha \wedge dx^\beta \\ &= (1/6)\delta^{1\mu}\epsilon_{\mu\nu\alpha\beta} dx^\nu \wedge dx^\alpha \wedge dx^\beta \\ &= (1/6)\epsilon_{1\nu\alpha\beta} dx^\nu \wedge dx^\alpha \wedge dx^\beta \\ &= (1/6)[\epsilon_{1023} dt\wedge dy\wedge dz + \epsilon_{1230} dy\wedge dz\wedge dt + \epsilon_{1302} dz\wedge dt\wedge dy \\ &\qquad + \epsilon_{1320} dz\wedge dy\wedge dt + \epsilon_{1203} dy\wedge dt\wedge dz + \epsilon_{1032} dt\wedge dz\wedge dy] \\ &= (1/6)[-dt\wedge dy\wedge dz - dy\wedge dz\wedge dt - dz\wedge dt\wedge dy \\ &\qquad + dz\wedge dy\wedge dt + dy\wedge dt\wedge dz + dt\wedge dz\wedge dy] \\ &= -dy\wedge dz\wedge dt \end{aligned} \tag{11.10}$$

the difference between the three and four dimensional Hodge dual arises from two sources, for one we are using the Minkowski metric so indices up or down makes a difference, and second the antisymmetric symbol has more possibilities than before because the Greek indices take four values.

Example 11.3.3. *Find the Hodge dual of $\gamma = dt$ with respect to the Minkowski metric $\eta_{\mu\nu}$, to begin notice that dt has components $\gamma_\mu = \delta^0_\mu$ as is readily verified by the equation $dt = \delta^0_\mu dx^\mu$. Lets raise the index using η as we learned previously,*

$$\gamma^\mu = \eta^{\mu\nu}\gamma_\nu = \eta^{\mu\nu}\delta^0_\nu = -\eta^{0\mu} = -\delta^{0\mu}$$

the minus sign is due to the Minkowski metric. Starting with the definition of Hodge duality we calculate

$$\begin{aligned} {}^*(dt) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\gamma^\mu \epsilon_{\mu\nu\alpha\beta} dx^\nu \wedge dx^\alpha \wedge dx^\beta \\ &= -(1/6)\delta^{0\mu}\epsilon_{\mu\nu\alpha\beta} dx^\nu \wedge dx^\alpha \wedge dx^\beta \\ &= -(1/6)\epsilon_{0\nu\alpha\beta} dx^\nu \wedge dx^\alpha \wedge dx^\beta \\ &= -(1/6)\epsilon_{0ijk} dx^i \wedge dx^j \wedge dx^k \\ &= -(1/6)\epsilon_{ijk}\epsilon_{ijk} dx\wedge dy\wedge dz && \leftarrow \textit{sneaky step} \\ &= -dx\wedge dy\wedge dz. \end{aligned} \tag{11.11}$$

for the case here we are able to use some of our old three dimensional ideas. The Hodge dual of dt cannot have a dt in it which means our answer will only have dx, dy, dz in it and that is why we were able to shortcut some of the work, (compared to the previous example).

Example 11.3.4. *Find the Hodge dual of $\gamma = dt \wedge dx$ with respect to the Minkowski metric $\eta_{\mu\nu}$, to begin notice the following identity, it will help us find the components of γ*

$$dt \wedge dx = \frac{1}{2} 2\delta^0_\mu \delta^1_\nu dx^\mu \wedge dx^\nu$$

now we antisymmetrize to get the components of the form,

$$dt \wedge dx = \frac{1}{2} \delta^0_{[\mu} \delta^1_{\nu]} dx^\mu \wedge dx^\nu$$

where $\delta^0_{[\mu}\delta^1_{\nu]} = \delta^0_\mu \delta^1_\nu - \delta^0_\nu \delta^1_\mu$ and the factor of two is used up in the antisymmetrization. Lets raise the index using η as we learned previously,

$$\gamma^{\alpha\beta} = \eta^{\alpha\mu}\eta^{\beta\nu}\gamma_{\mu\nu} = \eta^{\alpha\mu}\eta^{\beta\nu}\delta^0_{[\mu}\delta^1_{\nu]} = -\eta^{\alpha 0}\eta^{\beta 1} + \eta^{\beta 0}\eta^{\alpha 1} = -\delta^{[\alpha 0}\delta^{\beta]1}$$

the minus sign is due to the Minkowski metric. Starting with the definition of Hodge duality we calculate

$$\begin{aligned}
{}^*(dt \wedge dx) &= \tfrac{1}{p!}\tfrac{1}{(n-p)!}\gamma^{\alpha\beta}\epsilon_{\alpha\beta\mu\nu} dx^\mu \wedge dx^\nu \\
&= (1/4)(-\delta^{[\alpha 0}\delta^{\beta]1})\epsilon_{\alpha\beta\mu\nu} dx^\mu \wedge dx^\nu \\
&= -(1/4)(\epsilon_{01\mu\nu} dx^\mu \wedge dx^\nu - \epsilon_{10\mu\nu} dx^\mu \wedge dx^\nu) \\
&= -(1/2)\epsilon_{01\mu\nu} dx^\mu \wedge dx^\nu \\
&= -(1/2)[\epsilon_{0123} dy \wedge dz + \epsilon_{0132} dz \wedge dy] \\
&= -dy \wedge dz
\end{aligned} \tag{11.12}$$

when we first did these we got lucky in getting the right answer, however once we antisymmetrize it does make it a little uglier, notice however that the end of the calculation is the same. *Since $dt \wedge dx = -dx \wedge dt$ we find $*(dx \wedge dt) = dy \wedge dz$*

the other Hodge duals of the basic two-forms calculate by almost the same calculation, I'll let you work them for homework. Let us make a table of all the basic Hodge dualities in Minkowski space, I have grouped the terms to emphasize the isomorphisms between the one-dimensional $\Lambda^0(M)$ and $\Lambda^4(M)$, the four-dimensional $\Lambda^1(M)$ and $\Lambda^3(M)$, the six-dimensional $\Lambda^2(M)$ and itself. Notice that the dimension of $\Lambda(M)$ is 16 which just happens to be 2^4. You have a homework to prove that $dim(\Lambda(V)) = 2^{dim(V)}$ in general, we've already verified the cases $n = 3$ and $n = 4$.

Now that we've established how the Hodge dual works on the differentials we can easily take the Hodge dual of arbitrary differential forms on Minkowski space. We begin with the example of the 4-current $\mathcal{J}$

${}^*1 = dt \wedge dx \wedge dy \wedge dz$	${}^*(dt \wedge dx \wedge dy \wedge dz) = -1$
${}^*(dx \wedge dy \wedge dz) = -dt$	${}^*dt = -dx \wedge dy \wedge dz$
${}^*(dt \wedge dy \wedge dz) = -dx$	${}^*dx = -dy \wedge dz \wedge dt$
${}^*(dt \wedge dz \wedge dx) = -dy$	${}^*dy = -dz \wedge dx \wedge dt$
${}^*(dt \wedge dx \wedge dy) = -dz$	${}^*dz = -dx \wedge dy \wedge dt$
${}^*(dz \wedge dt) = dx \wedge dy$	${}^*(dx \wedge dy) = -dz \wedge dt$
${}^*(dx \wedge dt) = dy \wedge dz$	${}^*(dy \wedge dz) = -dx \wedge dt$
${}^*(dy \wedge dt) = dz \wedge dx$	${}^*(dz \wedge dx) = -dy \wedge dt$

Example 11.3.5. Four Current: *often in relativistic physics we would even just call this the current, however it actually includes the charge density* ρ *and current density* $\vec{J}$. *We define,*

$$(\mathcal{J}^\mu) \equiv (\rho, \vec{J})$$

we can lower the index to obtain,

$$(\mathcal{J}_\mu) = (-\rho, \vec{J})$$

which are the components of the current one-form,

$$\mathcal{J} = \mathcal{J}_\mu dx^\mu = -\rho dt + J_x dx + J_y dy + J_z dz$$

you could also take the equation above as the definition if you wish. Now we can rewrite this using or vectors $\mapsto$ *forms mapping as,*

$$\mathcal{J} = -\rho dt + \omega_{\vec{J}}.$$

Enough notational commentary, lets take the Hodge dual,

$$\begin{aligned} {}^*\mathcal{J} &= {}^*(-\rho dt + J_x dx + J_y dy + J_z dz) \\ &= -\rho\, {}^*dt + J_x {}^*dx + J_y {}^*dy + J_z {}^*dz \\ &= \rho dx \wedge dy \wedge dz - J_x dy \wedge dz \wedge dt - J_y dz \wedge dx \wedge dt - J_z dx \wedge dy \wedge dt \\ &= \rho dx \wedge dy \wedge dz - \Phi_{\vec{J}} \wedge dt \end{aligned} \tag{11.13}$$

we'll appeal to this calculation in a later section.

Example 11.3.6. Four Potential: *often in relativistic physics we would even just call this the potential, however it actually includes the scalar potential* V *and the vector potential* $\vec{A}$. *We define,*

$$(A^\mu) \equiv (V, \vec{A})$$

we can lower the index to obtain,

$$(A_\mu) = (-V, \vec{A})$$

which are the components of the current one-form,

$$A = A_\mu dx^\mu = -V dt + A_x dx + A_y dy + A_z dz$$

you could also take the equation above as the definition if you wish. Now we can rewrite this using or vectors $\mapsto$ *forms mapping as,*

$$A = -V dt + \omega_{\vec{A}}.$$

Enough notational commentary, lets take the Hodge dual,

$$ {}^*A = V dx \wedge dy \wedge dz - \Phi_{\vec{A}} \wedge dt \tag{11.14} $$

I omitted the steps because they are identical to the last example.

Example 11.3.7. Field tensor's dual: *as we observed in example* 10.2.8 *the electromagnetic field tensor* $F = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\mu$ *is a two-form (in that example we were looking at the values of the field tensor at a point, as we have mentioned a differential two-form gives a two-form at each point moreover* $d_p x^\mu$ *was identified with* e^μ *in retrospect). Notice that we can write the field tensor compactly using the work and flux form correspondences,*

$$ F = \omega_E \wedge dt + \Phi_B $$

if this is not obvious to you then that will make your homework more interesting, otherwise I apologize. Let us calculate the Hodge dual of the field tensor,

$$
\begin{aligned}
{}^*F &= {}^*(\omega_E \wedge dt + \Phi_B) \\
&= E_x{}^*(dx \wedge dt) + E_y{}^*(dy \wedge dt) + E_z{}^*(dz \wedge dt) \\
&\quad + B_x{}^*(dy \wedge dz) + B_y{}^*(dz \wedge dx) + B_z{}^*(dx \wedge dy) \\
&= E_x dy \wedge dz + E_y dz \wedge dx + E_z dx \wedge dy \\
&\quad - B_x dx \wedge dt - B_y dy \wedge dt - B_z dz \wedge dt \\
&= \Phi_E - \omega_B \wedge dt
\end{aligned}
$$

we can present the components of ${}^*F$ *in matrix form*

$$ ({}^*F_{\mu\nu}) = \begin{pmatrix} 0 & B_1 & B_2 & B_3 \\ -B_1 & 0 & E_3 & -E_2 \\ -B_2 & -E_3 & 0 & E_1 \\ -B_3 & E_2 & -E_1 & 0 \end{pmatrix} \tag{11.15} $$

notice that the net-effect of Hodge duality on the field tensor was to make the exchanges $\vec{E} \mapsto -\vec{B}$ *and* $\vec{B} \mapsto \vec{E}$.

11.4 exterior derivative

The exterior derivative inputs a p-form and outputs a $p+1$-form. Let us begin with the definition then we will expand on its meaning through examples, eventually we will find how the exterior derivative reproduces the gradient, curl and divergence.

Definition 11.4.1. *Let* α *be a* p*-form on* M *then define*

$$ d\alpha = d(\tfrac{1}{p!}\alpha_{i_1 i_2 \dots i_p}) \wedge dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_p}) \tag{11.16} $$

where we mean the total derivative when d *acts on functions,*

$$ d(\frac{1}{p!}\alpha_{i_1 i_2 \dots i_p}) = \partial_m(\frac{1}{p!}\alpha_{i_1 i_2 \dots i_p})dx^m $$

or perhaps it would be easier to see if we just wrote the definition for f,

$$ df = \frac{\partial f}{\partial x^m} dx^m $$

all the indices are to range over the accepted range for M.

Example 11.4.2. Gradient: *Consider three-dimensional Euclidean space. Let* $f : \mathbb{R}^3 \to \mathbb{R}$ *then*

$$df = \frac{\partial f}{\partial x^i} dx^i = \omega_{\nabla f}$$

it gives the one-form corresponding to ∇f*.*

Example 11.4.3. Curl: *Consider three-dimensional Euclidean space. Let* $\vec{F}$ *be a vector field and let* $\omega_F = F_i dx^i$ *be the corresponding one-form then*

$$\begin{aligned}
d\omega_F &= dF_i \wedge dx^i \\
&= \partial_j F_i dx^j \wedge dx^i \\
&= \partial_x F_y dx \wedge dy + \partial_y F_x dy \wedge dx + \partial_z F_x dz \wedge dx + \partial_x F_z dx \wedge dz + \partial_y F_z dy \wedge dz + \partial_z F_y dz \wedge dy \\
&= (\partial_x F_y - \partial_y F_x) dx \wedge dy + (\partial_z F_x - \partial_x F_z) dz \wedge dx + (\partial_y F_z - \partial_z F_y) dy \wedge dz \\
&= \Phi_{\nabla \times \vec{F}}.
\end{aligned}$$

now we've recovered the curl.

Example 11.4.4. Divergence: *Consider three-dimensional Euclidean space. Let* $\vec{G}$ *be a vector field and let* $\Phi_G = \frac{1}{2}\epsilon_{ijk} G_i dx^j \wedge dx^k$ *be the corresponding two-form then*

$$\begin{aligned}
d\Phi_G &= d(\tfrac{1}{2}\epsilon_{ijk} G_i) \wedge dx^j \wedge dx^k \\
&= \tfrac{1}{2}\epsilon_{ijk}(\partial_m G_i) dx^m \wedge dx^j \wedge dx^k \\
&= \tfrac{1}{2}\epsilon_{ijk}(\partial_m G_i)\epsilon_{mjk} dx \wedge dy \wedge dz \\
&= \tfrac{1}{2} 2\delta_{im}(\partial_m G_i) dx \wedge dy \wedge dz \\
&= \partial_i G_i dx \wedge dy \wedge dz \\
&= (\nabla \cdot \vec{G}) dx \wedge dy \wedge dz
\end{aligned}$$

now we've recovered the divergence.

In the course of the preceding three examples we have seen that the single operation of the exterior differentiation has reproduced the gradiant, curl and divergence of vector calculus provided we make the appropriate identifications under the "work" and "flux" form mappings. We now move on to some four dimensional examples.

Example 11.4.5. Charge conservation: *Consider the 4-current we introduced in example* 11.3.5*. Take the exterior derivative of the dual to the current,*

$$\begin{aligned}
d({}^*\mathcal{J}) &= d(\rho dx \wedge dy \wedge dz - \Phi_{\vec{J}} \wedge dt) \\
&= (\partial_t \rho) dt \wedge dx \wedge dy \wedge dz - d[(J_x dy \wedge dz + J_y dz \wedge dx + J_z dx \wedge dy) \wedge dt] \\
&= d\rho \wedge dx \wedge dy \wedge dz \\
&\quad - \partial_x J_x dx \wedge dy \wedge dz \wedge dt - \partial_y J_y dy \wedge dz \wedge dx \wedge dt - \partial_z J_z dz \wedge dx \wedge dy \wedge dt \\
&= (\partial_t \rho + \nabla \cdot \vec{J}) dt \wedge dx \wedge dy \wedge dz
\end{aligned}$$

now lets do the same calculation using index techniques,

$$\begin{aligned}
d({}^*\mathcal{J}) &= d(\rho dx \wedge dy \wedge dz - \Phi_{\vec{J}} \wedge dt) \\
&= d(\rho) \wedge dx \wedge dy \wedge dz - d[\tfrac{1}{2}\epsilon_{ijk} J_i dx^j \wedge dx^k \wedge dt) \\
&= (\partial_t \rho) dt \wedge dx \wedge dy \wedge dz - \tfrac{1}{2}\epsilon_{ijk}\partial_\mu J_i dx^\mu \wedge dx^j \wedge dx^k \wedge dt \\
&= (\partial_t \rho) dt \wedge dx \wedge dy \wedge dz - \tfrac{1}{2}\epsilon_{ijk}\partial_m J_i dx^m \wedge dx^j \wedge dx^k \wedge dt \\
&= (\partial_t \rho) dt \wedge dx \wedge dy \wedge dz - \tfrac{1}{2}\epsilon_{ijk}\epsilon_{mjk}\partial_m J_i dx \wedge dy \wedge dz \wedge dt \\
&= (\partial_t \rho) dt \wedge dx \wedge dy \wedge dz - \tfrac{1}{2} 2\delta_{im}\partial_m J_i dx \wedge dy \wedge dz \wedge dt \\
&= (\partial_t \rho + \nabla \cdot \vec{J}) dt \wedge dx \wedge dy \wedge dz.
\end{aligned}$$

Observe that we can now phrase charge conservation by the following equation

$$d(^*\mathcal{J}) = 0 \qquad \iff \qquad \partial_t \rho + \nabla \cdot \vec{J} = 0.$$

In the classical scheme of things this was a derived consequence of the equations of electromagnetism, however it is possible to build the theory from this equation outward. Rindler describes that formal approach in a late chapter of "Introduction to Special Relativity".

Example 11.4.6. Field tensor from 4-potential: *Let us take the exterior derivative of the potential one-form we discussed briefly before. We anticipate that we should find the electric and magnetic fields since the derivatives of the potentials give the fields as defined in equation* 5.13.

$$\begin{aligned} dA &= d(A_\nu) \wedge dx^\nu \\ &= \partial_\mu A_\nu dx^\mu \wedge dx^\nu \\ &= \tfrac{1}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu) dx^\mu \wedge dx^\nu + \tfrac{1}{2}(\partial_\mu A_\nu + \partial_\nu A_\mu) dx^\mu \wedge dx^\nu \\ &= \tfrac{1}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu) dx^\mu \wedge dx^\nu \\ &= \tfrac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu \qquad \textit{using the lemma below} \end{aligned} \tag{11.17}$$

Lemma 11.4.7. *The field tensor's components defined by equation* 9.16 *can be calculated from the 4-potential as follows,*

$$\boxed{F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.} \tag{11.18}$$

Proof: homework.

Since the fact that $dA = F$ is quite important I'll offer another calculation in addition to the direct one you consider in the homework,

$$\begin{aligned} dA &= d(-V dt + \omega_{\vec{A}}) \\ &= -dV \wedge dt + d(\omega_{\vec{A}}) \\ &= -dV \wedge dt + (\partial_t A_i) dt \wedge dx^i + (\partial_j A_i) dx^j \wedge dx^i \\ &= \omega_{-\nabla V} + \wedge dt - \omega_{\partial_t \vec{A}} \wedge dt + \Phi_{\nabla \times \vec{A}} \\ &= (\omega_{-\nabla V} - \omega_{\partial_t \vec{A}}) \wedge dt + \Phi_{\nabla \times \vec{A}} \\ &= \omega_{-\nabla V - \partial_t \vec{A}} \wedge dt + \Phi_{\nabla \times \vec{A}} \\ &= \omega_{\vec{E}} \wedge dt + \Phi_{\vec{B}} \end{aligned}$$

where I have used many of the previous examples to aid the calculation.

Example 11.4.8. Exterior derivative of the field tensor: *We've just seen that the field tensor is the exterior derivative of the potential one-form, lets examine the next level, we should expect to find Maxwell's equations since the derivative of the fields are governed by Maxwell's equations,*

$$\begin{aligned} dF &= d(E_i dx^i \wedge dt) + d(\Phi_{\vec{B}}) \\ &= \partial_m E_i dx^m \wedge dx^i \wedge dt + (\nabla \cdot \vec{B}) dx \wedge dy \wedge dz + \tfrac{1}{2}\epsilon_{ijk}(\partial_t B_i) dt \wedge dx^j \wedge dx^k \end{aligned} \tag{11.19}$$

let me pause here to explain my logic. In the above I dropped the $\partial_t E_i dt \wedge \cdots$ term because there was another dt in the term so it vanishes. Also I broke up the exterior derivative on the flux *form of $\vec{B}$ into the space and then time derivative terms and used our work in example* 11.4.4. *Lets continue,*

$$\begin{aligned} dF &= [\partial_j E_k + \tfrac{1}{2}\epsilon_{ijk}(\partial_t B_i)]dx^j \wedge dx^k \wedge dt + (\nabla \cdot \vec{B})dx \wedge dy \wedge dz \\ &= [\partial_x E_y - \partial_y E_x + \epsilon_{i12}(\partial_t B_i)]dx \wedge dy \wedge dt \\ &\quad +[\partial_z E_x - \partial_x E_z + \epsilon_{i31}(\partial_t B_i)]dz \wedge dx \wedge dt \\ &\quad +[\partial_y E_z - \partial_z E_y + \epsilon_{i23}(\partial_t B_i)]dy \wedge dz \wedge dt \\ &\quad +(\nabla \cdot \vec{B})dx \wedge dy \wedge dz \\ &= (\nabla \times \vec{E} + \partial_t \vec{B})_i \Phi_{e_i} \wedge dt + (\nabla \cdot \vec{B})dx \wedge dy \wedge dz \\ &= \Phi_{\nabla \times \vec{E} + \partial_t \vec{B}} \wedge dt + (\nabla \cdot \vec{B})dx \wedge dy \wedge dz \end{aligned} \tag{11.20}$$

because Φ is an isomorphism of vector spaces (at a point) and $\Phi_{e_1} = dy \wedge dz$, $\Phi_{e_2} = dz \wedge dx$, and $\Phi_{e_3} = dx \wedge dy$. Behold, we can state two of Maxwell's equations as

$$\boxed{dF = 0 \quad \iff \quad \nabla \times \vec{E} + \partial_t \vec{B} = 0, \quad \nabla \cdot \vec{B} = 0} \tag{11.21}$$

Example 11.4.9. Exterior derivative of the dual to the field tensor:

$$\begin{aligned} d^*F &= d(-B_i dx^i \wedge dt) + d(\Phi_{\vec{E}}) \\ &= \Phi_{-\nabla \times \vec{B} + \partial_t \vec{E}} \wedge dt + (\nabla \cdot \vec{E})dx \wedge dy \wedge dz \end{aligned} \tag{11.22}$$

this followed directly from the last example simply replace $\vec{E} \mapsto -\vec{B}$ and also $\vec{B} \mapsto \vec{E}$. This give us the two inhomogeneous Maxwell's equations if we set it equal to the Hodge dual of the 4-current,

$$\boxed{d^*F = \mu_o {}^*\mathcal{J} \quad \iff \quad -\nabla \times \vec{B} + \partial_t \vec{E} = -\mu_o \vec{J}, \quad \nabla \cdot \vec{E} = \rho} \tag{11.23}$$

where we have looked back at example 11.3.5 *to find the RHS of the Maxwell equations.*

Now we've seen how to write Maxwell's equations via differential forms. The stage is set to prove that Maxwell's equations are Lorentz covariant, that is they have the same form in all inertial frames.

11.4.1 coderivatives and comparing to Griffith's relativitic E & M

Optional section, for those who wish to compare our tensorial E & M with that of Griffith's, you may skip ahead to the next section if not interested

I should mention that this is not the only way to phrase Maxwell's equations in terms of differential forms. If you try to see how what we have done here compares with the equations presented in Griffith's text it is not immediately obvious. He works with $F^{\mu\nu}$ and $G^{\mu\nu}$ and J^μ none of which are the components of differential forms. Nevertheless he recovers Maxwell's equations as $\partial_\mu F^{\mu\nu} = J^\nu$ and $\partial_\mu G^{\mu\nu} = 0$. If we compare equation 12.119 (the matrix form of $G^{\mu\nu}$) in Griffith's text,

$$(G^{\mu\nu}(c=1)) = \begin{pmatrix} 0 & B_1 & B_2 & B_3 \\ -B_1 & 0 & -E_3 & E_2 \\ -B_2 & -E_3 & 0 & -E_1 \\ -B_3 & E_2 & -E_1 & 0 \end{pmatrix} = -({}^*F^{\mu\nu}). \tag{11.24}$$

where to see that you just recall that raising the indices has the net-effect of multiplying the zeroth row and column by -1, so if we do that to 11.15 we'll find Griffith's "dual tensor" times negative one. The equation $\partial_\mu F^{\mu\nu} = J^\nu$ is not directly from an exterior derivative, rather it is the component form of a "coderivative". The coderivative is defined $\delta = {}^*d^*$, it takes a p-form to an $(n-p)$-form then d makes it a $(n-p+1)$-form then finally the second Hodge dual takes it to an $(n-(n-p+1))$-form. That is δ takes a p-form to a $p-1$-form. We stated Maxwell's equations as

$$dF = 0 \qquad\qquad d^*F = {}^*\mathcal{J}$$

Now we can take the Hodge dual of the inhomogeneous equation to obtain,

$${}^*d^*F = \delta F = {}^{**}\mathcal{J} = \pm\mathcal{J}$$

where I leave the sign for you to figure out. Then the other equation

$$\partial_\mu G^{\mu\nu} = 0$$

can be understood as the component form of $\delta^*F = 0$ but this is really $dF = 0$ in disguise,

$$0 = \delta^*F = {}^*d^{**}F = \pm{}^*dF \iff dF = 0$$

so even though it looks like Griffith's is using the dual field tensor for the homogeneous Maxwell's equations and the field tensor for the inhomogeneous Maxwell's equations it is in fact not the case. The key point is that there are **coderivatives** implicit within Griffith's equations, so you have to read between the lines a little to see how it matched up with what we've done here. I have not entirely proved it here, to be complete we should look at the component form of $\delta F = \mathcal{J}$ and explicitly show that this gives us $\partial_\mu F^{\mu\nu} = J^\nu$, I don't think it is terribly difficult but I'll leave it to the reader.

Comparing with Griffith's is fairly straightforward because he uses the same metric as we have. Other texts use the mostly negative metric, its just a convention. If you try to compare to such a book you'll find that our equations are almost the same upto a sign. One good careful book is Reinhold A. Bertlmann's *Anomalies in Quantum Field Theory* you will find much of what we have done here done there with respect to the other metric. Another good book which shares our conventions is Sean M. Carroll's *An Introduction to General Relativity: Spacetime and Geometry*, that text has a no-nonsense introduction to tensors forms and much more over a curved space (in contrast to our approach which has been over a vector space which is flat). By now there are probably thousands of texts on tensors, I only point out those that have benefited my understanding at various times of my mathematical youth.

11.5 Maxwell's equations are relativistically covariant

Let us begin with the definition of the field tensor once more. We define the components of the field tensor in terms of the 4-potentials as we take the view-point those are the basic objects (not the fields).

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$$

then the field tensor $F = F_{\mu\nu} dx^\mu \otimes dx^\nu$ is a tensor, or is it ? We should check that the components transform as they ought according to the discussion in section 9.6. Let $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$ then we observe,

$$\begin{array}{l} (1.)\ \bar{A}_\mu = (\Lambda^{-1})^\alpha_\mu A_\alpha \\ (2.)\ \frac{\partial}{\partial \bar{x}^\nu} = \frac{\partial x^\beta}{\partial \bar{x}^\nu}\frac{\partial}{\partial x^\beta} = (\Lambda^{-1})^\beta_\nu \frac{\partial}{\partial x^\beta} \end{array} \tag{11.25}$$

where (2.) is simply the chain rule of multivariate calculus and (1.) is not at all obvious. We'll assume that (1.) holds, that is we assume that the 4-potential transforms in the appropriate way for a one-form. In principle one could prove that from more base assumptions. Afterall electromagnetism is the study of the interaction of charged objects, we should hope that the potentials are derivable from the source charge distribution. We wrote the formulas to calculate the potentials for arbitrary sources in equation 5.15. We could take those as definitions for the potentials, then it would be possible to actually calculate if (1.) is true. We'd just change coordinates via a Lorentz transformation and verify (1.). For the sake of brevity we will just assume that (1.) holds. We should mention that alternatively one can show the electric and magnetic fields transform as to make $F_{\mu\nu}$ a tensor. Those derivations assume that charge is an invariant quantity and just apply Lorentz transformations to special physical situations to deduce the field transformation rules. See Griffith's chapter on special relativity or look in Resnick for example.

Let us find how the field tensor transforms assuming that (1.) and (2.) hold, again we consider $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$,

$$\begin{aligned} \bar{F}_{\mu\nu} &= \bar{\partial}_\mu \bar{A}_\nu - \bar{\partial}_\nu \bar{A}_\mu \\ &= (\Lambda^{-1})^\alpha_\mu \partial_\alpha((\Lambda^{-1})^\beta_\nu A_\beta) - (\Lambda^{-1})^\beta_\nu \partial_\beta((\Lambda^{-1})^\alpha_\mu A_\alpha) \\ &= (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu (\partial_\alpha A_\beta - \partial_\beta A_\alpha) \\ &= (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu F_{\alpha\beta}. \end{aligned} \tag{11.26}$$

therefore the field tensor really is a tensor over Minkowski space.

Proposition 11.5.1. *The dual to the field tensor is a tensor over Minkowski space. For the Lorentz transformation $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$ we can show*

$$^*\bar{F}_{\mu\nu} = (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu {}^*F_{\alpha\beta}$$

Proof: homework, it follows quickly from the definition and the fact we already know that the field tensor is a tensor.

Proposition 11.5.2. *The four-current is a four-vector. That is under the Lorentz transformation $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$ we can show,*

$$\bar{\mathcal{J}}_\mu = (\Lambda^{-1})^\alpha_\mu \mathcal{J}_\alpha$$

Proof: follows from arguments involving the invariance of charge, time dilation and length contraction. See Griffith's for details, sorry we have no time.

Corollary 11.5.3. *The dual to the four current transforms as a 3-form. That is under the Lorentz transformation $\bar{x}^\mu = \Lambda^\mu_\nu x^\nu$ we can show,*

$$^*\bar{\mathcal{J}}_{\mu\nu\sigma} = (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu (\Lambda^{-1})^\gamma_\sigma \mathcal{J}_{\alpha\beta\gamma}$$

Upto now the content of this section is simply an admission that we have been a little careless in defining things upto this point. The main point is that if we say that something is a tensor then we need to make sure that is in fact the case. With the knowledge that our tensors are indeed tensors the proof of the covariance of Maxwell's equations is trivial.

$$dF = 0 \qquad\qquad d^*F = {}^*\mathcal{J}$$

are coordinate invariant expressions which we have already proved give Maxwell's equations in one frame of reference, thus they must give Maxwell's equations in all frames of reference. The essential point is simply that

$$F = \frac{1}{2}F_{\mu\nu}dx^{\mu} \wedge dx^{\nu} = \frac{1}{2}\bar{F}_{\mu\nu}d\bar{x}^{\mu} \wedge d\bar{x}^{\nu}$$

Again, we have no hope for the equation above to be true unless we know that $\bar{F}_{\mu\nu} = (\Lambda^{-1})^{\alpha}_{\mu}(\Lambda^{-1})^{\beta}_{\nu}F_{\alpha\beta}$. That transformation follows from the fact that the four-potential is a four-vector. It should be mentioned that others prefer to "prove" the field tensor is a tensor by studying how the electric and magnetic fields transform under a Lorentz transformation. We in contrast have derived the field transforms based ultimately on the seemingly innocuous assumption that the four-potential transforms according to $\bar{A}_{\mu} = (\Lambda^{-1})^{\alpha}_{\mu}A_{\alpha}$. OK enough about that.

So the fact that Maxwell's equations have the same form in all **relativistically inertial** frames of reference simply stems from the fact that we found Maxwell's equation were given by an arbitrary frame, and the field tensor looks the same in the new barred frame so we can again go through all the same arguments with barred coordinates. Thus we find that Maxwell's equations are the same in all relativistic frames of reference, that is if they hold in one inertial frame then they will hold in any other frame which is related by a Lorentz transformation.

11.6 Poincaire's Lemma and $d^2 = 0$

This section is in large part inspired by M. Gockeler and T. Schucker's *Differential geometry, gauge theories, and gravity* page 20-22.

Theorem 11.6.1. *The exterior derivative of the exterior derivative is zero.* $d^2 = 0$

Proof: Let α be an arbitrary p-form then

$$d\alpha = \frac{1}{p!}(\partial_m \alpha_{i_1 i_2 \dots i_p})dx^m \wedge dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_p} \tag{11.27}$$

then differentiate again,

$$\begin{aligned} d(d\alpha) &= d\left[\tfrac{1}{p!}(\partial_m \alpha_{i_1 i_2 \dots i_p})dx^m \wedge dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_p}\right] \\ &= \tfrac{1}{p!}(\partial_k \partial_m \alpha_{i_1 i_2 \dots i_p})dx^k \wedge dx^m \wedge dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_p} \\ &= 0 \end{aligned} \tag{11.28}$$

since the partial derivatives commute whereas the wedge product anticommutes so we note that the pair of indices (k,m) is symmetric for the derivatives but antisymmetric for the wedge, as we know the sum of symmetric against antisymmetric vanishes (see equation 1.25 part *iv* if you forgot.)

Definition 11.6.2. *A differential form α is* **closed** *iff $d\alpha = 0$. A differential form β is* **exact** *iff there exists γ such that $\beta = d\gamma$.*

Proposition 11.6.3. *All exact forms are closed. However, there exist closed forms which are not exact.*

Proof: Exact implies closed is easy, let β be exact such that$\beta = d\gamma$ then

$$d\beta = d(d\gamma) = 0$$

using the theorem $d^2 = 0$. To prove that there exists a closed form which is not exact it suffices to give an example. A popular example (due to its physical significance to magnetic monopoles, Dirac Strings and the like..) is the following differential form in $\mathbb{R}^2$

$$\phi = \frac{1}{x^2 + y^2}(xdy - ydx) \tag{11.29}$$

I'll let you show that $d\phi = 0$ in homework. Observe that if ϕ were exact then there would exist f such that $\phi = df$ meaning that

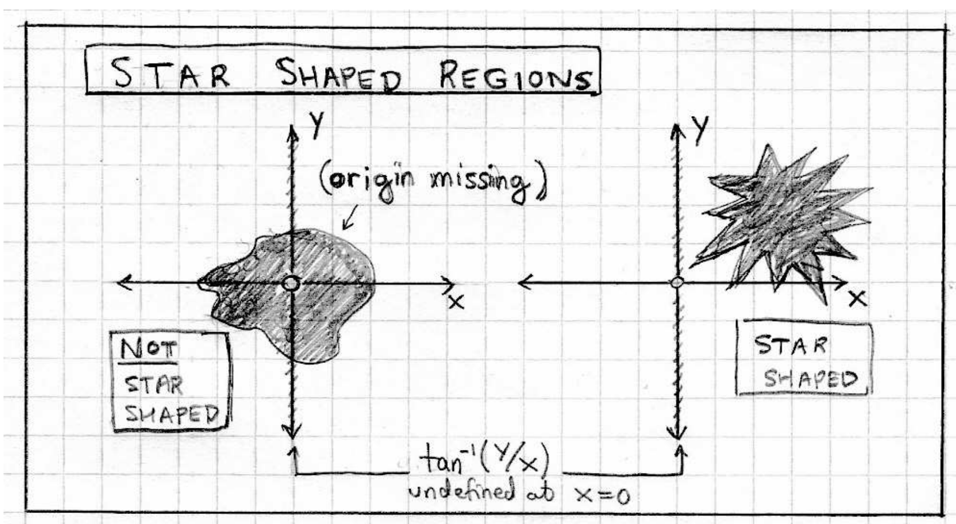
$$\frac{\partial f}{\partial x} = -\frac{y}{x^2 + y^2}, \qquad \frac{\partial f}{\partial y} = \frac{x}{x^2 + y^2}$$

which are solved by $f = arctan(y/x) + c$ where c is arbitrary. Observe that f is ill-defined along the y-axis $x = 0$ (this is the Dirac String if we put things in context), however the natural domain of ϕ is $\mathbb{R}^2 - \{(0, 0)\}$.

Poincaire suggested the following partial converse, he said closed implies exact provided we place a topological restriction on the domain of the form.

Theorem 11.6.4. *If $U \subseteq \mathbb{R}^n$ is star-shaped and $dom(\phi) = U$ then ϕ is closed iff ϕ is exact.*

I will not give proof. See the picture below to see that "star-shaped" means just that. In the diagram the region that is centered around the origin is not star-shaped because it is missing the origin, whereas the other region is such that you can reach any point by a ray from a central point. Although an example is not a proof you can see the example of ϕ satisfies the theorem, only if we restrict the natural domain of ϕ to a smaller star-shaped region can we say it is an exact form.



Definition 11.6.5. de Rahm cohomology: *We define several real vector spaces of differential forms over some subset U of $\mathbb{R}^n$,*

$$Z^p(U) \equiv \{\phi \in \Lambda^p U \mid \phi \text{ closed}\}$$

the space of closed p-forms on U. Then,

$$B^p(U) \equiv \{\phi \in \Lambda^p U \mid \phi \text{ exact}\}$$

the space of exact p-forms where by convention $B^0(U) = \{0\}$ The de Rahm cohomology groups are defined by the quotient of closed/exact,

$$H^p(U) \equiv Z^p(U)/B^p(U).$$

the $dim(H^p U) = p^{th}$ Betti number of U.

We observe that star shaped regions have all the Betti numbers zero since $Z^p(U) = B^p(U)$ implies that $H^p(U) = \{0\}$. Of course there is much more to say about Cohomology, I just wanted to give you a taste and alert you to the fact that differential forms can be used to reveal aspects of topology. Cohomology is basically a method to try to classify certain topological features of spaces. Not all algebraic topology uses differential forms though, in fact if you take the course in it here you'll spend most of your time on other calculational schemes besides the de Rahm cohomology. I digress.

Chapter 12

Integration of forms in $\mathbb{R}^n$

In this chapter we give a short introduction on how to integrate differential forms on a parametrized subset of $\mathbb{R}^n$. We demonstrate how the differential form integration recovers the usual notions of line and surface integrals in $\mathbb{R}^3$. Finally we write the Generalized Stokes's Theorem and show how it reproduces the fundamental theorem of calculus, Gauss' Theorem, and Stoke's Theorem. We will be a little sloppy throughout this chapter on the issue of convergence. It should be mentioned that the integrals cosidered will only make sense for suitably chosen regions and for reasonably behaved functions. We leave those picky details for the reader to discover. Also we mention that generally one should study how to integrate differential forms over a manifold. In a manifold we cannot generally parametrize the whole surface by just one set of parameters (in this chapter we will assume that our subsets of $\mathbb{R}^n$ have a global parametrization) so it is necessary to patch things together with something called the *partition of unity*. Just want to place what we are doing in this chapter in context, there is more to say about integrating forms. We will just do the fun part.

12.1 definitions

The definitions given here are pragmatical. There are better more abstract definitions but we'd need to know about *push-forwards* and *pull-backs* (take manifold theory or Riemannian geometry or ask me if you're interested). I also assume that you will go back and read through chapter 4 to remind yourself how line and surface integrals are formulated.

Definition 12.1.1. integral of one-form along oriented curve: *let $\alpha = \alpha_i dx^i$ be a one form and let C be an oriented curve with parametrization $X(t) : [a, b] \to C$ then we define the integral of the one-form α along the curve C as follows,*

$$\int_C \alpha \equiv \int_a^b \alpha_i(X(t)) \frac{dX^i}{dt}(t)dt$$

where $X(t) = (X^1(t), X^2(t), \ldots, X^n(t))$ so we mean X^i to be the i^{th} component of $X(t)$. Moreover, the indices are understood to range over the dimension of the ambient space, if we consider forms in $\mathbb{R}^2$ then $i = 1, 2$ if in $\mathbb{R}^3$ then $i = 1, 2, 3$ if in Minkowski $\mathbb{R}^4$ then i should be replaced with $\mu = 0, 1, 2, 3$ and so on.

Example 12.1.2. One form integrals vs. line integrals of vector fields: *We begin with a vector field $\vec{F}$ and construct the corresponding one-form $\omega_{\vec{F}} = F_i dx^i$. Next let C be an oriented curve with parametrization $X : [a,b] \subset \mathbb{R} \to C \subset \mathbb{R}^n$, observe*

$$\int_C \omega_{\vec{F}} = \int_a^b F_i(X(t)) \frac{dX^i}{dt}(t) dt = \int_C \vec{F} \cdot d\vec{l}$$

You may note that the definition of a line integral of a vector field is not special to three dimensions, we can clearly construct the line integral in n-dimensions, likewise the correspondance ω can be written between one-forms and vector fields in any dimension, provided we have a metric to lower the index of the vector field components. The same cannot be said of the flux-form correspondance, it is special to three dimensions for reasons we have explored previously.

Definition 12.1.3. integral of two-form over an oriented surface: *let $\beta = \frac{1}{2}\beta_{ij} dx^i \wedge dx^j$ be a two-form and let S be an oriented piecewise smooth surface with parametrization $X(u,v) : D_2 \subset \mathbb{R}^2 \to S \subset \mathbb{R}^n$ then we define the integral of the two-form β over the surface S as follows,*

$$\int_S \beta \equiv \int_{D_2} \beta_{ij}(X(u,v)) \frac{\partial X^i}{\partial u}(u,v) \frac{\partial X^j}{\partial v}(u,v) du dv$$

where $X(u,v) = (X^1(u,v), X^2(u,v), \ldots, X^n(u,v))$ so we mean X^i to be the i^{th} component of $X(u,v)$. Moreover, the indices are understood to range over the dimension of the ambient space, if we consider forms in $\mathbb{R}^2$ then $i,j = 1,2$ if in $\mathbb{R}^3$ then $i,j = 1,2,3$ if in Minkowski $\mathbb{R}^4$ then i,j should be replaced with $\mu,\nu = 0,1,2,3$ and so on.

Proposition 12.1.4. Two-form integrals vs. surface integrals of vector fields in $\mathbb{R}^3$: *We begin with a vector field $\vec{F}$ and construct the corresponding two-form $\Phi_{\vec{F}} = \frac{1}{2}\epsilon_{ijk} F_k dx^i \wedge dx^j$ which is to say $\Phi_{\vec{F}} = F_1 dy \wedge dz + F_2 dz \wedge dx + F_3 dx \wedge dy$. Next let S be an oriented piecewise smooth surface with parametrization $X : D \subset \mathbb{R}^2 \to S \subset \mathbb{R}^n$, then*

$$\int_S \Phi_{\vec{F}} = \int_S \vec{F} \cdot d\vec{A}$$

Proof: Recall that the normal to the surface S has the form,

$$N(u,v) = \frac{\partial X}{\partial u} \times \frac{\partial X}{\partial v} = \epsilon_{ijk} \frac{\partial X^i}{\partial u} \frac{\partial X^j}{\partial v} e_k$$

at the point $X(u,v)$. This gives us a vector which points along the outward normal to the surface and it is nonvanishing throughout the whole surface by our assumption that S is oriented. Moreover the vector surface integral of $\vec{F}$ over S was defined by the formula,

$$\int_S \vec{F} \cdot d\vec{A} \equiv \int\int_D \vec{F}(X(u,v)) \cdot \vec{N}(u,v)\ dudv.$$

now that the reader is reminded whats what, lets prove the proposition, dropping the (u,v) depence to reduce clutter we find,

$$\begin{aligned}
\int_S \vec{F} \cdot d\vec{A} &= \int\int_D \vec{F} \cdot \vec{N}\ dudv \\
&= \int\int_D F_k N_k\ dudv \\
&= \int\int_D F_k \epsilon_{ijk} \frac{\partial X^i}{\partial u}\frac{\partial X^j}{\partial v}\ dudv \\
&= \int\int_D (\Phi_{\vec{F}})_{ij} \frac{\partial X^i}{\partial u}\frac{\partial X^j}{\partial v}\ dudv \\
&= \int_S \Phi_{\vec{F}}
\end{aligned}$$

notice that we have again used our convention that $(\Phi_{\vec{F}})_{ij}$ refers to the tensor components of the 2-form $\Phi_{\vec{F}}$ meaning we have $\Phi_{\vec{F}} = (\Phi_{\vec{F}})_{ij} dx^i \otimes dx^j$ whereas with the wedge product $\Phi_{\vec{F}} = \frac{1}{2}(\Phi_{\vec{F}})_{ij} dx^i \wedge dx^j$, I mention this in case you are concerned there is a half in $\Phi_{\vec{F}}$ yet we never found a half in the integral. Well, we don't expect to because we defined the integral of the form with respect to the tensor components of the form, again they don't contain the half.

Example 12.1.5. *Consider the vector field $\vec{F} = (0,0,3)$ then the corresponding two-form is simply $\Phi_F = 3dx \wedge dy$. Lets calculate the surface integral and two-form integrals over the square $D = [0,1] \times [0,1]$ in the xy-plane, in this case the parameters can be taken to be x and y so $X(x,y) = (x,y)$ and,*

$$N(x,y) = \frac{\partial X}{\partial x} \times \frac{\partial X}{\partial y} = (1,0,0) \times (0,1,0) = (0,0,1)$$

which is nice. Now calculate,

$$\begin{aligned}
\int_S \vec{F} \cdot d\vec{A} &= \int\int_D \vec{F} \cdot \vec{N}\ dxdy \\
&= \int\int_D (0,0,3) \cdot (0,0,1) dxdy \\
&= \int_0^1\int_0^1 3dxdy \\
&= 3.
\end{aligned}$$

Consider that $\Phi_F = 3dx \wedge dy = 3dx \otimes dy - 3dy \otimes dx$ therefore we may read directly that $(\Phi_F)_{12} = -(\Phi_F)_{21} = 3$ and all other components are zero,

$$\begin{aligned}
\int_S \Phi_F &= \int\int_D (\Phi_F)_{ij} \frac{\partial X^i}{\partial x}\frac{\partial X^j}{\partial y}\ dxdy \\
&= \int\int_D \left(3\frac{\partial X^1}{\partial x}\frac{\partial X^2}{\partial y} - 3\frac{\partial X^2}{\partial x}\frac{\partial X^1}{\partial y}\right) dxdy \\
&= \int_0^1\int_0^1 \left(3\frac{\partial x}{\partial x}\frac{\partial y}{\partial y} - 3\frac{\partial y}{\partial x}\frac{\partial x}{\partial y}\right) dxdy \\
&= 3.
\end{aligned}$$

Definition 12.1.6. integral of a three-form over an oriented volume: *let* $\gamma = \frac{1}{6}\beta_{ijk}dx^i \wedge dx^j \wedge dx^k$ *be a three-form and let* V *be an oriented piecewise smooth volume with parametrization* $X(u,v,w) : D_3 \subset \mathbb{R}^3 \to V \subset \mathbb{R}^n$ *then we define the integral of the three-form* γ *in the volume* V *as follows,*

$$\int_V \gamma \equiv \int_{D_3} \gamma_{ijk}(X(u,v,w)) \frac{\partial X^i}{\partial u}\frac{\partial X^j}{\partial v}\frac{\partial X^k}{\partial w}\ dudvdw$$

where $X(u,v,w) = (X^1(u,v,w), X^2(u,v,w), \ldots, X^n(u,v,w))$ *so we mean* X^i *to be the* i^{th} *component of* $X(u,v,w)$*. Moreover, the indices are understood to range over the dimension of the ambient space, if we consider forms in* $\mathbb{R}^3$ *then* $i,j,k = 1,2,3$ *if in Minkowski* $\mathbb{R}^4$ *then* i,j,k *should be replaced with* $\mu,\nu,\sigma = 0,1,2,3$ *and so on.*

finally we define the integral of a p-form over an p-dimensional subspace of $\mathbb{R}^n$, we assume that $p \leq n$ so that it is possible to embed such a subspace in $\mathbb{R}^n$,

Definition 12.1.7. integral of a p-form over an oriented volume: *let* $\gamma = \frac{1}{p!}\beta_{i_1\ldots i_p}dx^{i_1} \wedge \cdots dx^{i_p}$ *be a p-form and let* S *be an oriented piecewise smooth subspace with parametrization* $X(u_1,\ldots,u_p) : D_p \subset \mathbb{R}^p \to S \subset \mathbb{R}^n$ *then we define the integral of the p-form* γ *in the subspace* S *as follows,*

$$\int_S \gamma \equiv \int_{D_p} \beta_{i_1\ldots i_p}(X(u_1,\ldots,u_p)) \frac{\partial X^{i_1}}{\partial u_1}\cdots\frac{\partial X^{i_p}}{\partial u_p}\ du_1\cdots du_p$$

where $X(u_1,\ldots,u_p) = (X^1(u_1,\ldots,u_p), X^2(u_1,\ldots,u_p), \ldots, X^n(u_1,\ldots,u_p))$ *so we mean* X^i *to be the* i^{th} *component of* $X(u_1,\ldots,u_p)$*. Moreover, the indices are understood to range over the dimension of the ambient space.*

12.2 Generalized Stokes Theorem

The generalized Stokes theorem contains within it most of the main theorems of integral calculus, namely the fundamental theorem of calculus, the fundamental theorem of line integrals (a.k.a the FTC in three dimensions), Greene's Theorem in the plane, Gauss' Theorem and also Stokes Theorem, not to mention a myriad of higher dimensional not so commonly named theorems. The breadth of its application is hard to overstate, yet the statement of the theorem is simple,

Theorem 12.2.1. Generalizes Stokes Theorem: *Let* S *be an oriented, piecewise smooth (p+1)-dimensional subspace of* $\mathbb{R}^n$ *where* $n \geq p+1$ *and let* ∂S *be it boundary which is consistently oriented then for a p-form* α *which behaves reasonably on* S *we have that*

$$\boxed{\int_S d\alpha = \int_{\partial S} \alpha} \tag{12.1}$$

The proof of this theorem (and a more careful statement of it) can be found in a number of places, Susan Colley's *Vector Calculus* or Steven H. Weintraub's *Differential Forms: A Complement to Vector Calculus* or Spivak's *Calculus on Manifolds* just to name a few.

Lets work out how this theorem reproduces the main integral theorems of calculus.

Example 12.2.2. Fundamental Theorem of Calculus in $\mathbb{R}$: *Let $f : \mathbb{R} \to \mathbb{R}$ be a zero-form then consider the interval $[a, b]$ in $\mathbb{R}$. If we let $S = [a, b]$ then $\partial S = \{a, b\}$. Further observe that $df = f'(x)dx$. Notice by the definition of one-form integration*

$$\int_S df = \int_a^b f'(x)dx$$

However on the other hand we find (the integral over a zero-form is taken to be the evaluation map, perhaps we should have defined this earlier, oops., but its only going to come up here so I'm leaving it.)

$$\int_{\partial S} f = f(b) - f(a)$$

Hence in view of the definition above we find that

$$\int_a^b f'(x)dx = f(b) - f(a) \quad \Longleftrightarrow \quad \int_S df = \int_{\partial S} f$$

Example 12.2.3. Fundamental Theorem of Calculus in $\mathbb{R}^3$: *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a zero-form then consider a curve C from $p \in \mathbb{R}^3$ to $q \in \mathbb{R}^3$ parametrized by $\phi : [a, b] \to \mathbb{R}^3$. Note that $\partial C = \{\phi(a) = p, \phi(b) = q\}$. Next note that*

$$df = \frac{\partial f}{\partial x^i} dx^i$$

Then consider that the exterior derivative of a function corresponds to the gradient of the function thus we are not to surprised to find that

$$\int_C df = \int_a^b \frac{\partial f}{\partial x^i} \frac{dx^i}{dt} dt = \int_C (\nabla f) \cdot d\vec{l}$$

On the other hand, we use the definition of the integral over a a two point set again to find

$$\int_{\partial C} f = f(q) - f(p)$$

Hence if the Generalized Stokes Theorem is true then so is the FTC in three dimensions,

$$\int_C (\nabla f) \cdot d\vec{l} = f(q) - f(p) \quad \Longleftrightarrow \quad \int_C df = \int_{\partial C} f$$

another popular title for this theorem is the "fundamental theorem for line integrals". As a final thought here we notice that this calculation easily generalizes to 2,4,5,6,... dimensions.

Example 12.2.4. Greene's Theorem: *Let us recall the statement of Greene's Theorem as I have not replicated it yet in the notes, let D be a region in the xy-plane and let ∂D be its consistently oriented boundary then if $\vec{F} = (M(x, y), N(x, y), 0)$ is well behaved on D*

$$\int_{\partial D} Mdx + Ndy = \int\int_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y} \right) dxdy$$

We begin by finding the one-form corresponding to $\vec{F}$ namely $\omega_F = Mdx + Ndy$ consider then that

$$d\omega_F = d(Mdx + Ndy) = dM \wedge dx + dN \wedge dy = \frac{\partial M}{\partial y} dy \wedge dx + \frac{\partial N}{\partial x} dx \wedge dy$$

which simplifies to,

$$d\omega_F = \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dx \wedge dy = \Phi_{(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y})\hat{k}}$$

Thus, using our discussion in the last section we recall

$$\int_{\partial D} \omega_F = \int_{\partial D} \vec{F} \cdot d\vec{l} = \int_{\partial D} Mdx + Ndy$$

where we have reminded the reader that the notation in the rightmost expression is just another way of denoting the line integral in question. Next observe,

$$\int_D d\omega_F = \int_D (\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y})\hat{k} \cdot d\vec{A}$$

And clearly, since $d\vec{A} = \hat{k} dxdy$ we have

$$\int_D (\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y})\hat{k} \cdot d\vec{A} = \int_D (\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}) dxdy$$

Therefore,

$$\int_{\partial D} Mdx + Ndy = \int \int_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dxdy \quad \Longleftrightarrow \quad \int_D d\omega_F = \int_{\partial D} \omega_F$$

Example 12.2.5. Gauss Theorem: *Let us recall Gauss Theorem to begin, for suitably defined $\vec{F}$ and V,*

$$\int_{\partial V} \vec{F} \cdot d\vec{A} = \int_V \nabla \cdot \vec{F} \, d\tau$$

First we recall our earlier result that

$$d(\Phi_F) = (\nabla \cdot \vec{F}) dx \wedge dy \wedge dz$$

Now note that we may integrate the three form over a volume,

$$\int_V d(\Phi_F) = \int_V (\nabla \cdot \vec{F}) dxdydz$$

whereas,

$$\int_{\partial V} \Phi_F = \int_{\partial V} \vec{F} \cdot d\vec{A}$$

so there it is,

$$\int_V (\nabla \cdot \vec{F}) d\tau = \int_{\partial V} \vec{F} \cdot d\vec{A} \quad \Longleftrightarrow \quad \int_V d(\Phi_F) = \int_{\partial V} \Phi_F$$

I have left a little detail out here, I may assign it for homework.

Example 12.2.6. Stokes Theorem: *Let us recall Stokes Theorem to begin, for suitably defined* $\vec{F}$ *and* S*,*

$$\int_S (\nabla \times \vec{F}) \cdot d\vec{A} = \int_{\partial S} \vec{F} \cdot d\vec{l}$$

Next recall we have shown in the last chapter that,

$$d(\omega_F) = \Phi_{\nabla \times \vec{F}}$$

Hence,

$$\int_S d(\omega_F) = \int_S (\nabla \times \vec{F}) \cdot d\vec{A}$$

whereas,

$$\int_{\partial S} \omega_F = \int_{\partial S} \vec{F} \cdot d\vec{l}$$

which tells us that,

$$\int_S (\nabla \times \vec{F}) \cdot d\vec{A} = \int_{\partial S} \vec{F} \cdot d\vec{l} \quad \Longleftrightarrow \quad \int_S d(\omega_F) = \int_{\partial S} \omega_F$$

The Generalized Stokes Theorem is perhaps the most persausive argument for mathematicians to be aware of differential forms, it is clear they allow for more deep and sweeping statements of the calculus. The generality of differential forms is what drives modern physicists to work with them, string theorists for example examine higher dimensional theories so they are forced to use a language more general than that of the conventional vector calculus.

12.3 Electrostatics in Five dimensions

We will endeavor to determine the electric field of a point charge in 5 dimensions where we are thinking of adding an extra spatial dimension. Lets call the fourth spatial dimension the w-direction so that a typical point in space time will be (t, x, y, z, w). First we note that the electromagnetic field tensor can still be derived from a one-form potential,

$$A = -\rho dt + A_1 dx + A_2 dy + A_3 dz + A_4 dw$$

we will find it convenient to make our convention for this section that $\mu, \nu, ... = 0, 1, 2, 3, 4$ whereas $m, n, ... = 1, 2, 3, 4$ so we can rewrite the potential one-form as,

$$A = -\rho dt + A_m dx^m$$

This is derived from the vector potential $A^\mu = (\rho, A^m)$ under the assumption we use the natural generalization of the Minkowski metric, namely the 5 by 5 matrix,

$$(\eta_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = (\eta^{\mu\nu}) \tag{12.2}$$

we could study the linear isometries of this metric, they would form the group $O(1,4)$. Now we form the field tensor by taking the exterior derivative of the one-form potential,

$$F = dA = \frac{1}{2}(\partial_\mu \partial_\nu - \partial_\nu \partial_\mu) dx^\mu \wedge dx^\nu$$

now we would like to find the electric and magnetic "fields" in 4 dimensions. Perhaps we should say 4+1 dimensions, just understand that I take there to be 4 spatial directions throughout this discussion if in doubt. Note that we are faced with a dilemma of interpretation. There are 10 independent components of a 5 by 5 antisymmetric tensor, naively we wold expect that the electric and magnetic fields each would have 4 components, but that is not possible, we'd be missing two components. The solution is this, the time components of the field tensor are understood to correspond to the electric part of the fields whereas the remaining 6 components are said to be magnetic. This aligns with what we found in 3 dimensions, its just in 3 dimensions we had the fortunate quirk that the number of linearly independent one and two forms were equal at any point. This definition means that the magnetic field will in general not be a vector field but rather a "flux" encoded by a 2-form.

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & -E_x & -E_y & -E_z & -E_w \\ E_x & 0 & B_z & -B_y & H_1 \\ E_y & -B_z & 0 & B_x & H_2 \\ E_z & B_y & -B_x & 0 & H_3 \\ E_w & -H_1 & -H_2 & -H_3 & 0 \end{pmatrix} \tag{12.3}$$

Now we can write this compactly via the following equation,

$$F = E \wedge dt + B$$

I admit there are subtle points about how exactly we should interpret the magnetic field, however I'm going to leave that to your imagination and instead focus on the electric sector. What is the generalized Maxwell's equation that E must satisfy?

$$d^*F = \mu_o {}^*\mathcal{J} \implies d^*(E \wedge dt + B) = \mu_o {}^*\mathcal{J}$$

where $\mathcal{J} = -\rho dt + J_m dx^m$ so the 5 dimensional Hodge dual will give us a $5-1=4$ form, in particular we will be interested in just the term stemming from the dual of dt,

$${}^*(-\rho dt) = \rho dx \wedge dy \wedge dz \wedge dw$$

the corresponding term in d^*F is $d^*(E \wedge dt)$ thus, using $\mu_o = \frac{1}{\epsilon_o}$,

$$d^*(E \wedge dt) = \frac{1}{\epsilon_o} \rho dx \wedge dy \wedge dz \wedge dw \tag{12.4}$$

is the 4-dimensional Gauss's equation. Now consider the case we have an isolated point charge which has somehow always existed at the origin. Moreover consider a 3-sphere that surrounds the charge. We wish to determine the generalized Coulomb field due to the point charge. First we note that the solid 3-sphere is a 4-dimensional object, it the set of all $(x,y,z,w) \in \mathbb{R}^4$ such that

$$x^2 + y^2 + z^2 + w^2 \leq r^2$$

We may parametrize a three-sphere of radius r via generalized spherical coordinates,

$$\begin{aligned} x &= rsin(\theta)cos(\phi)sin(\psi) \\ y &= rsin(\theta)sin(\phi)sin(\psi) \\ z &= rcos(\theta)sin(\psi) \\ w &= rcos(\psi) \end{aligned} \tag{12.5}$$

Now it can be shown that the volume and surface area of the radius r three-sphere are as follows,

$$vol(S^3) = \frac{\pi^2}{2}r^4 \qquad\qquad area(S^3) = 2\pi^2 r^3$$

We may write the charge density of a smeared out point charge q as,

$$\rho = \begin{cases} 2q/\pi^2 a^4, & 0 \leq r \leq a \\ 0, & r > a \end{cases}. \tag{12.6}$$

Notice that if we integrate ρ over any four-dimensional region which contains the solid three sphere of radius a will give the enclosed charge to be q. Then integrate over the Gaussian 3-sphere S^3 with radius r call it M,

$$\int_M d^*(E \wedge dt) = \frac{1}{\epsilon_o}\int_M \rho dx \wedge dy \wedge dz \wedge dw$$

now use the Generalized Stokes Theorem to deduce,

$$\int_{\partial M} {}^*(E \wedge dt) = \frac{q}{\epsilon_o}$$

but by the "spherical" symmetry of the problem we find that E must be independent of the direction it points, this means that it can only have a radial component. Thus we may calculate the integral with respect to generalized spherical coordinates and we will find that it is the product of $E_r \equiv E$ and the surface volume of the four dimensional solid three sphere. That is,

$$\int_{\partial M} {}^*(E \wedge dt) = 2\pi^2 r^3 E = \frac{q}{\epsilon_o}$$

Thus,

$$\boxed{E = \frac{q}{2\pi^2\epsilon_o r^3}}$$

the Coulomb field is weaker if it were to propogate in 4 spatial dimensions. Qualitatively what has happened is that the have taken the same net flux and spread it out over an additional dimension, this means it thins out quicker. A very similar idea is used in some *brane world* scenarios. String theorists posit that the gravitational field spreads out in more than four dimensions while in contrast the standard model fields of electromagnetism, and the strong and weak forces are confined to a four-dimensional brane. That sort of model attempts an explaination as to why gravity is so weak in comparison to the other forces. Also it gives large scale corrections to gravity that some hope will match observations which at present don't seem to fit the standard gravitational models.

This example is but a taste of the theoretical discussion that differential forms allow. As a final comment I remind the reader that we have done things for flat space in this course, when considering a curved space there are a few extra considerations that must enter. Coordinate vector fields e_i must be thought of as derivations $\partial/\partial x^\mu$ for one. Also the metric is not a constant tensor like δ_{ij} or $\eta_{\mu\nu}$ rather is depends on position, this means Hodge duality aquires a coordinate dependence as well. Doubtless I have forgotten something else in this brief warning. One more advanced treatment of many of our discussions is Dr. Fulp's Fiber Bundles 2001 notes which I have posted on my webpage. He uses the other metric but it is rather elegantly argued, all his arguments are coordinate independent. He also deals with the issue of the magnetic induction and the dielectric, issues which we have entirely ignored since we always have worked in free space.

THE END

References and Acknowledgements:

I have drawn from many sources to assemble the content of this course, the references are listed approximately in the order of their use to the course, additionally we are indebted to Dr. Fulp for his course notes from many courses (ma 430, ma 518, ma 555, ma 756, ...). Also Manuela Kulaxizi helped me towards the correct (I hope) interpretation of 5-dimensional E&M in the last example.

Vector Calculus, Susan Jane Colley

Introduction to Special Relativity, Robert Resnick

Differential Forms and Connections, R.W.R. Darling

Differential geometry, gauge theories, and gravity, M. Göckerler & T. Schücker

Anomalies in Quantum Field Theory, Reinhold A. Bertlmann

"The Differential Geometry and Physical Basis for the Applications of Feynman Diagrams", S.L. Marateck, Notices of the AMS, Vol. 53, Number 7, pp. 744-752

Abstract Linear Algebra, Morton L. Curtis

Gravitation, Misner, Thorne and Wheeler

Introduction to Special Relativity, Wolfgang Rindler

Differential Forms A Complement to Vector Calculus, Steven H. Weintraub

Differential Forms with Applications to the Physical Sciences, Harley Flanders

Introduction to Electrodynamics, (3rd ed.) David J. Griffiths

The Geometry of Physics: An Introduction, Theodore Frankel

An Introduction to General Relativity: Spacetime and Geometry, Sean M. Carroll

Gauge Theory and Variational Principles, David Bleeker

Group Theory in Physics, Wu-Ki Tung

Appendix A

Digressions and Comments

In this appendix I collect all the various corrections and additions we made throughout the semester. Some of these comments are a result of a homework problem gone astray others are simply additions or proofs that ought to have been included earlier.

I apologize for the ugly format of this chapter, if you want a clearer scan of anything in particular just email me and I'm happy to send you a pdf or something.

A.1 speed of light invariant under Lorentz transformation

I made the intial homework assignment more difficult than needed. Ignoring my hint would have been the wisest choice of action in this case. The essential point I missed in my preliminary thoughts was that light has constant speed and hence travels in a line. Lines are much easier to transform than arbitrary curves (which is what I was trying to reason through to begin).

SPEED OF LIGHT AND LORENTZ TRANSFORMATIONS

To begin we assumed that the speed of light was the same in x as it was in the x-boosted coordinate $\bar{x}$. In fact this was how we found B_x to begin with. Then we generalized the x-boost by saying that because the x-boost preserved the interval $\langle B_x X, B_x Y\rangle = \langle X, Y\rangle$ it seemed reasonable to <u>define</u> the more general notion of a Lorentz transformation by the same condition. I'll now argue that this generalization is in fact consistent with Einstein's 2^{nd} postulate, namely that the speed of light is the same in all inertial frames (inertial frames are by def$^{\underline{n}}$ related by Lorentz transformations in Special Relativity)

Before the general proof lets consider how the argument goes for an x-boost, suppose light travels in x-direction & it passes thru. origin then

$$x = ct = t$$

Now transform this eq$^{\underline{n}}$ to the x-boosted $(\bar{t}, \bar{x})$,

$$\begin{aligned} \bar{x} &= \gamma(x+\beta t) = \gamma(t+\beta t) \\ \bar{t} &= \gamma(t+\beta x) = \gamma(t+\beta t) \end{aligned} \quad \therefore \quad \frac{\bar{x}}{\bar{t}} = 1$$

Hence $\bar{x} = \bar{t}$, the light ray also travels with velocity $\dfrac{d\bar{x}}{d\bar{t}} = 1$ in the x-boosted frame

General Argument

Suppose light travels with constant velocity c in an inertial frame. Because spatial translations do not change the velocities we may without loss of generality assume that our light ray passes through the origin. Using $\vec{r}$ as the position vector we write,

$$\vec{r} = t\hat{n} \quad , \text{where } \|\hat{n}\| = 1$$

Clearly $\|\frac{d\vec{r}}{dt}\| = 1$. The "world line" of the light ray ~~is dotted~~ lies on the light cone in Minkowski space; $(t, t\hat{n})$ note

$$\langle (t, t\hat{n}), (t, t\hat{n}) \rangle = -t^2 + t^2 \hat{n}\cdot\hat{n} = 0$$

Denote the trajectory of the light in another intertial frame by $(\bar{t}, \bar{w}\hat{w})$ where $\|\hat{w}\| = 1$. we know $\exists L \in \mathcal{L}$ such that

$$L(t, t\hat{n}) = (\bar{t}, \bar{w}\hat{w}).$$

Moreover we know

$$\langle L(t, t\hat{n}), L(t, t\hat{n}) \rangle = \langle (t, t\hat{n}), (t, t\hat{n}) \rangle = 0$$

Hence,

$$\langle (\bar{t}, \bar{w}\hat{w}), (\bar{t}, \bar{w}\hat{w}) \rangle = 0$$

$$-(\bar{t})^2 + \bar{w}^2 \hat{w}\cdot\hat{w} = 0 \Rightarrow \bar{w} = \pm \bar{t}$$

Thus $\left\|\frac{d\vec{W}}{d\bar{t}}\right\| = \|\pm 1 \hat{w}\| = 1.$

This derivation holds for arbitrary Lorentz transformation

<u>Remark</u>: A similar argument shows that if $V < C$ in one frame then $\bar{V} < C$ in any other inertial frame where the velocity is $\bar{V}$. We calculated how in a very special case to find $\bar{V}$.

A.2 Campbell-Baker-Hausdorff Relation

This result is central to the theory of Lie algebras and groups. I give a proof to the third order. Iterative proofs to higher orders can be found in a number of standard texts.

ERCISE 14

JAMES COOK

laim: $e^x e^y = e^{x+y+\frac{1}{2}[x,y]+\frac{1}{12}[[x,y],y]-\frac{1}{12}[[x,y],x]+\cdots}$

Let us expand the RHS's argument,

$$C \equiv x+y+\tfrac{1}{2}(xy-yx)+\tfrac{1}{12}[xy-yx,\,y]-\tfrac{1}{12}[xy-yx,\,x]$$

$$= x+y+\tfrac{1}{2}(xy-yx)+\tfrac{1}{12}\big((xy-yx)y-y(xy-yx)\big)-\tfrac{1}{12}\big((xy-yx)x-x(xy-yx)\big)$$

$$= x+y+\tfrac{1}{2}(xy-yx)+\tfrac{1}{12}\big(xy^2-yxy-yxy+y^2x-xyx+yx^2+x^2y-xyx\big)$$

$$= x+y+\tfrac{1}{2}(xy-yx)+\tfrac{1}{12}\big(xy^2-2yxy+y^2x-2xyx+yx^2+x^2y\big)$$

Now the RHS can be expanded in power series, keep upto 3rd order,

$$e^C = 1+C+\tfrac{1}{2}C^2+\tfrac{1}{6}C^3+\cdots$$

$$= 1+C+\tfrac{1}{2}\big(x+y+\tfrac{1}{2}(xy-yx)\big)^2+\tfrac{1}{6}(x+y)^3+\cdots$$

$$= 1+C+\tfrac{1}{2}\big(x^2+xy+\tfrac{1}{2}x(xy-yx)+yx+y^2+\tfrac{1}{2}y(xy-yx)+\tfrac{1}{2}(xy-yx)(x+y)\big)$$
$$+\tfrac{1}{6}(x+y)(x^2+xy+yx+y^2)+\cdots$$

$$= 1+C+\tfrac{1}{2}(\text{—//—})+\tfrac{1}{6}\big(x^3+x^2y+xyx+xy^2+yx^2+yxy+y^2x+y^3\big)+\cdots$$

$$= 1+x+y+\tfrac{1}{2}(xy-yx)+\tfrac{1}{12}\big(xy^2-2yxy+y^2x-2xyx+yx^2+x^2y\big)$$
$$+\tfrac{1}{2}\big(x^2+xy+yx+y^2+\tfrac{1}{2}x^2y-\tfrac{1}{2}xyx+\tfrac{1}{2}yxy-\tfrac{1}{2}y^2x+\tfrac{1}{2}xyx-\tfrac{1}{2}yx^2-\tfrac{1}{2}yxy+\tfrac{1}{2}xy^2\big)$$
$$+\tfrac{1}{6}\big(x^3+x^2y+xyx+xy^2+yx^2+yxy+y^2x+y^3\big)+\cdots$$

$$= 1+x+y+\tfrac{1}{2}\big(xy-\cancel{yx}+x^2+xy+\cancel{yx}+y^2\big)$$
$$+\tfrac{1}{12}\big(2x^3+x^2y(1+3+2)+xy^2(1+3+2)+xyx(-2-3+3+2)$$
$$2y^3+y^2x(1-3+2)+yx^2(1-3+2)+yxy(-2+3-3+2)\big)+\cdots$$

$$= 1+x+y+\tfrac{1}{2}\big(x^2+2xy+y^2\big)+\tfrac{1}{6}\big(x^3+3x^2y+3xy^2+y^3\big)+\cdots$$

JAMES COOK

nsider then, again to 3rd order,

$$e^x e^y = \left(1 + x + \tfrac{1}{2}x^2 + \tfrac{1}{6}x^3\right)\left(1 + y + \tfrac{1}{2}y^2 + \tfrac{1}{6}y^3\right) + \cdots$$

$$= 1 + y + \tfrac{1}{2}y^2 + \tfrac{1}{6}y^3 + x + xy + \tfrac{1}{2}xy^2 + \tfrac{1}{6}xy^3 + \tfrac{1}{2}x^2 + \tfrac{1}{2}x^2y + \tfrac{1}{6}x^3 + \cdots$$

$$= 1 + y + \tfrac{1}{2}(x^2 + 2xy + y^2) + \tfrac{1}{6}(x^3 + 3x^2y + 3xy^2 + y^3) + \cdots$$

herefore we have shown that the BH relation holds to 3rd order.

$$\boxed{e^x e^y = e^{x + y + \frac{1}{2}[x,y] + \frac{1}{12}[[x,y],y] - \frac{1}{12}[[x,y],x] + \cdots}}$$

For the problem we had $x = A$ and $y = B$ where we were given $[[A,B],A] = 0$ & $[[A,B],B] = 0$ hence the identity,

$$\boxed{e^A e^B = e^{A + B + \frac{1}{2}[A,B]}}$$

no $+ \cdots$ req^{d} as all the other terms can be seen as nested commutators which vanish under the givens here. We leave the proof of 4^{th}, 5^{th}, ... order to someone else.

A.3 Quaternions

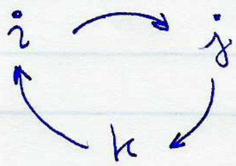
Beyond complex numbers one can consider *Quaternions* which have three imaginary units which multiply much like unit vectors in $\mathbb{R}^3$ behave under the cross-product. The next generalization are the *Octonions*, I'll let you read elsewhere about those, (I recommend the Wikipedia for such questions b.t.w.)

QUATERNIONS & VECTOR CALCULUS

t Q = quaternions then $A, B \in Q$ means

$$A = xi + yj + zk$$
$$B = ai + bj + ck$$

ere $i^2 = j^2 = k^2 = -1$ and i, j, k satisfy



$$ij = k = -ji$$
$$jk = i = -kj$$
$$ki = j = -ki$$

nsider the product of A & B given above,

$$\begin{aligned} AB &= (xi + yj + zk)(ai + bj + ck) \\ &= (xai^2 + ybj^2 + zck^2 \\ &\quad + xb\,ij + xc\,ik + ya\,ji + yc\,jk + za\,ki + zb \\ &= -xa - yb - zc \\ &\quad (xb - ay)k + (xc - za)j + (yc - zb)jk \\ &= -(\vec{A} \cdot \vec{B}) + (\vec{A} \times \vec{B})_1 i + (\vec{A} \times \vec{B})_2 j + (\vec{A} \times \vec{B}) k \end{aligned}$$

A.4 Coordinate change and Hodge Duality

Here I remedy a gaping hole in our earlier logic, I show that the Hodge dual of a differential form has the correct transformation properties under a change of coordinates. Without this calculation we could not be certain that our formulation of Maxwell's equations was truly invariant with respect to Lorentz transformations.

Hodge Duality & Coordinate Change

Consider a p-form in n-dim'l space with metric (g_{ij}), we show that the Hodge dual of $\alpha = \frac{1}{p!}\alpha_{i_1 i_2 \cdots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p}$ is in fact an $(n-p)$-form, in particular we show our def$^{\underline{n}}$ is indeed coordinate independent.

$$*\alpha_{i_1 i_2 \cdots i_{n-p}} \equiv \frac{1}{(n-p)!}\frac{1}{p!}\alpha^{j_1 j_2 \cdots j_p}\epsilon_{j_1 j_2 \cdots j_p i_1 i_2 \cdots i_{n-p}}$$

$$*\bar{\alpha}_{i_1 i_2 \cdots i_{n-p}} \equiv \frac{1}{(n-p)!}\frac{1}{p!}\bar{\alpha}^{j_1 j_2 \cdots j_p}\epsilon_{j_1 j_2 \cdots j_p i_1 i_2 \cdots i_{n-p}}$$

since $*\alpha$ is an $(n-p)$-form one ought to find that $*\bar{\alpha}_{i_1 i_2 \cdots i_{n-p}} = (\Lambda^{-1})^{k_1}_{i_1}(\Lambda^{-1})^{k_2}_{i_2}\cdots(\Lambda^{-1})^{k_{n-p}}_{i_{n-p}} *\alpha_{k_1 k_2 \cdots k_{n-p}}$. It was not obvious to me why this was previous classes. Two facts are critical to the proof,

(i.) $\bar{\alpha}^{j_1 j_2 \cdots j_p} = A^{j_1}_{i_1}A^{j_2}_{i_2}\cdots A^{j_p}_{i_p}\alpha^{i_1 i_2 \cdots i_p}$

(ii.) $\epsilon_{i_1 i_2 \cdots i_n} = B^{k_1}_{i_1}B^{k_2}_{i_2}\cdots B^{k_n}_{i_n}\epsilon_{k_1 k_2 \cdots k_n}$

Where $\det(B) = 1$ by assumption. Consider then,

$$\begin{aligned}
*\bar{\alpha}_{i_1 i_2 \cdots i_{n-p}} &= \frac{1}{(n-p)!}\frac{1}{p!}\bar{\alpha}^{j_1 j_2 \cdots j_p}\epsilon_{j_1 j_2 \cdots j_p i_1 i_2 \cdots i_{n-p}} \\
&= \frac{1}{(n-p)!}\frac{1}{p!}A^{j_1}_{l_1}A^{j_2}_{l_2}\cdots A^{j_p}_{l_p}\alpha^{l_1 l_2 \cdots l_p}(A^{-1})^{k_1}_{j_1}\cdots(A^{-1})^{k_p}_{j_p}(A^{-1})^{k_{p+1}}_{i_1}\cdots(A^{-1})^{k_n}_{i_{n-p}}\epsilon_{k_1 \cdots} \\
&= \frac{1}{(n-p)!}\frac{1}{p!}\left(A^{j_1}_{l_1}(A^{-1})^{k_1}_{j_1}\cdots A^{j_p}_{l_p}(A^{-1})^{k_p}_{j_p}\right)(A^{-1})^{k_{p+1}}_{i_1}\cdots(A^{-1})^{k_n}_{i_{n-p}}\alpha^{l_1 \cdots l_p}\epsilon_{k_1 k_2 \cdots k_n} \\
&= \frac{1}{(n-p)!}\frac{1}{p!}\delta^{k_1}_{l_1}\delta^{k_2}_{l_2}\cdots\delta^{k_p}_{l_p}(A^{-1})^{k_{p+1}}_{i_1}\cdots(A^{-1})^{k_n}_{i_{n-p}}\alpha^{l_1 l_2 \cdots l_p}\epsilon_{k_1 k_2 \cdots k_n} \\
&= (A^{-1})^{k_{p+1}}_{i_1}\cdots(A^{-1})^{k_n}_{i_{n-p}}\frac{1}{(n-p)!}\frac{1}{p!}\alpha^{k_1 k_2 \cdots k_p}\epsilon_{k_1 k_2 \cdots k_p k_{p+1} \cdots k_n} \\
&= (A^{-1})^{k_{p+1}}_{i_1}\cdots(A^{-1})^{k_n}_{i_{n-p}} *\alpha_{k_{p+1} \cdots k_n}
\end{aligned}$$

This is precisely the transformation law a $(n-p)$-form ought to follow. We elaborate on the origin of (i.) & (ii) next.

<u>More on (i) & (ii)</u>

(i) Recall $g(\Lambda v, \Lambda w) = g(v, w)$ is the condition we insist on in order that Λ be an allowed change of coordinates; $\bar{x}^\mu = \Lambda^\mu_{\ \alpha} x^\alpha$. In components,

$$(g_{ij}\, e^i \otimes e^j)(\Lambda^{-1k}_{\ \ \ \ l}\, e_k, \Lambda^{-1a}_{\ \ \ \ b}\, e_a) = g(e_l, e_b)$$

$$g_{ij}\, \Lambda^{-1i}_{\ \ \ \ l}\, \Lambda^{-1j}_{\ \ \ \ b} = g_{lb}$$

$$\bar{g}_{\mu\nu} = (\Lambda^{-1})^\alpha_{\ \mu} (\Lambda^{-1})^\beta_{\ \nu}\, g_{\alpha\beta}$$

$$\Rightarrow \bar{g}^{\mu\nu} = \Lambda^\mu_{\ \alpha} \Lambda^\nu_{\ \beta}\, g^{\alpha\beta}$$

Now then,

$$\begin{aligned}
\bar{T}^{\mu\nu} &= \bar{g}^{\mu\alpha} \bar{g}^{\nu\beta}\, \bar{T}_{\alpha\beta} \\
&= \bar{g}^{\mu\alpha} \bar{g}^{\nu\beta} (\Lambda^{-1})^\sigma_{\ \alpha} (\Lambda^{-1})^\rho_{\ \beta}\, T_{\sigma\rho} \\
&= \Lambda^\mu_{\ a} \Lambda^\alpha_{\ b}\, g^{ab}\, \Lambda^\nu_{\ c} \Lambda^\beta_{\ d}\, g^{cd} (\Lambda^{-1})^\sigma_{\ \alpha} (\Lambda^{-1})^\rho_{\ \beta}\, T_{\sigma\rho} \\
&= \Lambda^\mu_{\ a} \Lambda^\nu_{\ c}\, g^{ab} g^{cd}\, \delta^\sigma_b \delta^\rho_d\, T_{\sigma\rho} \\
&= \Lambda^\mu_{\ a} \Lambda^\nu_{\ c}\, g^{ab} g^{cd}\, T_{bd} \\
&= \Lambda^\mu_{\ a} \Lambda^\nu_{\ c}\, T^{ac}
\end{aligned}$$

which is precisely what we'd like to find.

(ii)

$$\det(\Lambda^{-1}) = 1 = (\Lambda^{-1})^{i_1}_{\ 1} (\Lambda^{-1})^{i_2}_{\ 2} (\Lambda^{-1})^{i_3}_{\ 3} (\Lambda^{-1})^{i_4}_{\ 4} \cdots (\Lambda^{-1})^{i_n}_{\ n}\, \epsilon_{i_1 i_2 \cdots i_n}$$

Since $\epsilon_{0123\cdots n}$ or $\epsilon_{123\cdots n} = 1$ we find

$$\epsilon_{12\cdots n} = (\Lambda^{-1})^{i_1}_{\ 1} (\Lambda^{-1})^{i_2}_{\ 2} \cdots (\Lambda^{-1})^{i_n}_{\ n}\, \epsilon_{i_1 i_2 \cdots i_n}$$

$$\Rightarrow \boxed{\epsilon_{j_1 j_2 \cdots j_n} = (\Lambda^{-1})^{i_1}_{\ j_1} (\Lambda^{-1})^{i_2}_{\ j_2} \cdots (\Lambda^{-1})^{i_n}_{\ j_n}\, \epsilon_{i_1 i_2 \cdots i_n}}$$

(think!)

A.5 Dimension of the exterior algebra

I assigned this as a homework, but no one quite got it correct. This is the solution which involves a slightly nontrivial induction based on Pascal's triangle.

Proof that $\dim(\Lambda(V)) = 2^{\dim(V)}$.

We argued that $\dim(\Lambda^p(V)) = \frac{n!}{(n-p)!\,p!} = \binom{n}{p}$. Moreover, we defined $\Lambda(V) = \Lambda^0(V) \oplus \Lambda^1(V) \oplus \cdots \oplus \Lambda^{n-1}(V) \oplus \Lambda^n(V)$ thus $\dim \Lambda(V) = \dim \Lambda^0(V) + \dim \Lambda^1(V) + \cdots + \dim \Lambda^n(V)$. Also Pascals triangle gives the values of $\binom{n}{p}$, you can check it,

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array} \qquad \left\{ \begin{array}{l} a_0^0 = 1 = 2^0 \\ a_0^1 + a_1^1 = 2 = 2^1 \\ a_0^2 + a_1^2 + a_2^2 = 1 + 2 + 1 = 2^2 \\ a_0^3 + a_1^3 + a_2^3 + a_3^3 = 1+3+3+1 = 2 \\ a_0^4 + a_1^4 + a_2^4 + a_3^4 + a_4^4 = 16 = 2^4. \end{array} \right.$$

Notice $1+1=2$, $1+3=4$, $3+3=6$ etc... the entries in the Δ are always the sum of the two entries above and the ones on the sides keep going. This suggests an identity for $\binom{n}{p}$, $\binom{n+1}{p} = \binom{n}{p-1} + \binom{n}{p}$ lets verify it,

$$\begin{aligned} \binom{n}{p-1} + \binom{n}{p} &= \frac{n!}{(n-p+1)!\,(p-1)!} + \frac{n!}{(n-p)!\,p!} \\ &= \frac{n!\,p}{(n+1-p)!\,p!} + \frac{n!\,(n-p+1)}{(n-p+1)!\,p!} \\ &= \frac{\cancel{n!\,p} + (n+1)n! - \cancel{n!\,p}}{(n+1-p)!\,p!} \\ &= \frac{(n+1)!}{(n+1-p)!\,p!} = \boxed{\binom{n+1}{p} = \binom{n}{p-1} + \binom{n}{p}} \end{aligned}$$

Denote the sum of $\binom{n}{p}$ for $p = 0, 1, \ldots, n$ by S_n and use notation indicated next to Δ. Assume $S_n = 2^n$ and consider

$$\begin{aligned} 2^{n+1} = 2\,2^n &= 2\left[\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k-1} + \binom{n}{k} + \cdots + \binom{n}{n}\right] \\ &= \left[\binom{n+1}{0} + \underbrace{\binom{n}{0} + \binom{n}{1}} + \underbrace{\binom{n}{1} + \binom{n}{2}} + \binom{n}{2} + \cdots + \underbrace{\binom{n}{k-1} + \binom{n}{k}} + \cdots + \binom{n+1}{n+1}\right] \\ &= \left[\binom{n+1}{0} + \binom{n+1}{1} + \binom{n+1}{2} + \cdots + \binom{n+1}{k} + \cdots + \binom{n+1}{n+1}\right] \\ &= S_{n+1} \quad \therefore \text{ By induction } \boxed{\dim \Lambda(V) = 2^{\dim V}} \end{aligned}$$