Coordinates, Vectors, Tensors

I. COORDINATES

The stage on which physics takes place is a set of points, in space or in spacetime. They are real, but have no real properties. There are many of them (and infinity of them). In order to differentiate between these points one wants to give them names. In order to manipulate them logically, one chooses those names to be numbers, such that each point has a unique name. Those numbers are called coordinates. They are arbitrary. They are names we give them so that we can talk about the points. However, clearly physics cannot depend on those arbitrary names. The world does not care if we refer to a specific point as (345), $(0,0,0)$ or $\sqrt{e}, \pi^3, 7.7$. Thus one must set up our mathematics in such a way that the results can be interpreted as being independent of those names. We will deal with two possibilities. One is that there are two types of points– spatial points, which can be labeled with three names (the so called three dimensions of space) and temporal (which is universal and can be designated by one number). In the temporal case we can assume that the only freedom we have is designating which point has the name 0, and the others fall into place. We will give an abstract class for these names by x^i where i is one of 1,2,3. Or sometimes by the names x, y, z. A change of names, from x^1, x^2, x^3 to $\tilde{x}^1, \tilde{x}^2, \tilde{x}^3$ by a differentiable function. We will assume that names are unique, so that any point has three numbers associated with that point, and that those three numbers are unique to that point (if one is given a set of three numbers, either there exists no point associated with those three numbers, or that there exists at most one point associated with those particular three numbers).

We can define a scalar function on space (something which gives a real number at each point in space) as $f(p)$, or $f(x^i)$, where the notation $f(x^i)$ is taken to mean the function of the three coordinate variables $f(x^1, x^2, x^3)$. (Theorists are lazy and do not like writing more than they have to). We will assume that these functions are differentiable, ie, that all three of the partial derivatives, $\frac{\partial f(x^1,x^2,x^3)}{\partial x^i}$ exist. Alternatively we can write the set x^1, x^2, x^3 as $\mathfrak{X}$, and $\tilde{x}^1, \tilde{x}^2, \tilde{x}^3$ as $\tilde{\mathfrak{X}}$ The components of the gradient then is

$$\frac{\partial f(\mathfrak{X})}{\partial x^i} \tag{1}$$

for i taking value of 1,2, or 3. Ie, this represents the set of devivaties for each value of i.

The problem of course is that these values depend on the coordinate. If we change coordinates then the values of these derivatives will change. Ie, these derivatives depend on arbitrary names. In particular, if we define

$$\tilde{f}(\tilde{x}^i) = f(x^i(\tilde{x}^j)) \quad \text{or} \tilde{f}(\tilde{\mathfrak{X}}) = f(\mathfrak{X}(\tilde{\mathfrak{X}}) \tag{2}$$

where we interpret say x^i not as something which has a specific value for i and for that coordinate x^i but as a symbol for the whole set of coordinates x^1, x^2, x^3. Often this is instead designated by something like $\mathfrak{X} \equiv x^i$ At present you might well say, why in the world do I not use that symbolism here. It turns out that the number of things I am going to want to have symbols for is too many to use the extreme paucity of font styles to designate.

To go back to the function f, we know that we can write the partial derivatives of $\tilde{f}$ in terms of the partial derivatives of f by the chain rule

$$\frac{\partial \tilde{f}(\tilde{x}^i)}{\partial\ \tilde{x}^i} = \sum_j \frac{\partial f(x^j(\tilde{x}^i))}{\partial x^j} \frac{\partial x^j}{\partial \tilde{x}^i} \tag{3}$$

or

$$\frac{\partial \tilde{f}(\tilde{\mathfrak{X}})}{\partial \tilde{x}^i} = \sum_j \frac{\partial f(\mathfrak{X}(\tilde{\mathfrak{X}}))}{\partial x^j} \frac{\partial^j}{\partial \tilde{x}^i} \tag{4}$$

$$\tilde{f}(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3) = f(\mathfrak{X}(\tilde{\mathfrak{X}})) \equiv f\left((x^1(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3), x^2(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3), x^3(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3)\right) \tag{5}$$

As can be seen this is starting to get messy, so in order to condense the expression while still maintaining enough symbolism to figure out what the expression means, we thus introduce another symbolism. For any function h, we define

$$\frac{\partial h}{\partial x^i} \equiv \partial_i h \tag{6}$$

$$\frac{\partial \tilde{h}}{\partial \tilde{x}^i} \equiv \partial_{\tilde{i}} \tilde{h} \tag{7}$$

Thus the above becomes

$$\partial_{\tilde{i}} \tilde{f}(\tilde{\mathfrak{X}}) = \sum_j \left(\partial_{\tilde{j}} f(\mathfrak{X}(\tilde{\mathfrak{X}}))\partial_i \tilde{x}^j\right) \tag{8}$$

To get rid of some of the tildes, regard f here as interpreted as a function of points, with the same point designated either by the $\mathfrak{X}$ or $\tilde{\mathfrak{X}}$.

This object $\partial_i f$ is what is usually called the gradient, and is a vector, if one, as mathematicians do, calls a set of objects which you can add together and get another of the same kind of object, and which you can multiply by a number and get another of the same kind of object.

And we have

$$f(p) + g(p) = h(p) \tag{9}$$

$$\partial_i f + \partial_i g = \partial_i h \tag{10}$$

$$\partial_i(\alpha f) = \alpha \partial_i f \tag{11}$$

for any functions f and g, and any constant α.

Thus the gradient is a vector, with components $\{\partial_1 f, \partial_2 f, \partial_3 f\}$. Note that if we take $\tilde{x}^1 = 7x^1, \tilde{x}^2 = x^2, \tilde{x}^3 = x^3$, then

$$\partial_{\tilde{1}} \tilde{f} = \frac{1}{7}\partial_1 f, \tag{12}$$

$$\partial_{\tilde{2}} \tilde{f} = \partial_2 f \tag{13}$$

$$\partial_{\tilde{3}} \tilde{f} = \partial_3 f \tag{14}$$

Ie, if we multiply the first coordinate by 7, then the first component of the gradient is divided by 7 and the second and third stay the same. .

The gradient, usually called the co-tangent vector, or covariant (in old literature) vector, is however only one type of vector. There is a second type called the tangent vector or the contravariant (in old literature) vector. These are the things you are used to from the prof hauling in some sticks with arrows on them, and telling you that is a vector. They are associated with curves not with functions. A function is a map from points in the space to a real numbers. A curve is the opposite, mapping from a real number (eg, the time) to a point in the spacetime. Thus $\gamma(\lambda)$ is a set of points in the space, labeled by λ. In coordinates, one can define the curve through coordinates space by

$$x^i(\gamma(\lambda)) \equiv x^i(\lambda). \tag{15}$$

Then one can define the components of the tangent vector to the curve by

$$T^i(\lambda) = \frac{dx^i(\gamma(\lambda))}{d\lambda} \tag{16}$$

If one has two curves running through the same point (which one can always assume that they are at the same point at the same value of λ just by relabelling lambda for one of the curves to make this true). Thus we can assume that $\gamma(\lambda_0)$ and $\hat{\gamma}(\lambda_0)$ are the same point in space.

Then we can add the tangent vectors of the two curves by defining

$$\gamma_{sum} = \gamma(\lambda) + \hat{\gamma}(\lambda) \equiv (p\left(x^i(\gamma(\lambda) + x^i\hat{\gamma}(\lambda) - x^i(\gamma(\lambda_0))\right)) \tag{17}$$

Then

$$T^i_{sum} = \frac{d(x^i(\lambda) + \hat{x}^i(\lambda) - x^i(\lambda_0))}{d\lambda} = T^i + \hat{T}^i \tag{18}$$

$$T^i_{prod} = \frac{dx^i(\alpha\lambda)}{d\lambda} = \alpha\frac{dx^i(\lambda)}{d\lambda} \tag{19}$$

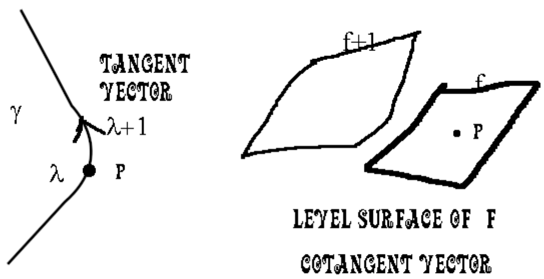

FIG. 1: Figure vector. Tangent and cotangent vectors at point p. Note that zero tangent vector is one for which the length of the curve representing the tangent vector goes to zero, while the zero cotangent vector

Ie, again the set of tangents to curves is a vector space at a point in the spacetime.

Note that although the "sum of curves" I defined above certainly depends on the coordinate system, the sum of tangent vectors at a point does not. For functions of points, the sum of functions is independent of the coordinates used, while for curves, since one cannot add points, the sum of curves depends on the coordinate system used, but in such a way that the sum of tangent vectors to the curves do not depend on the coordinate system used. The proof is a bit messy and really does not reveal any insights into tangent vectors, so I will not attempt it here.

Under a coordinate transformation we get

$$\tilde{T}^i = \sum_j T^j \frac{\partial \tilde{x}^i}{\partial x^j} = \sum_j T^j \partial_j \tilde{x}^i \tag{20}$$

Any vector which transforms in the same way as a tangent vector to a curve does is called a tangent vector, or a contravariant vector. Note that if $\tilde{x}^1 = 7x^1, \tilde{x}^2 = c^2, \tilde{x}^3 = x^3$ then $\tilde{T}^1 = 7T^1, \tilde{T}^2 = T^2, \tilde{T}^3 = T^3$. Ie, the Tangent vector grows in its first component which is the opposite of what a cotangent vector does. Tangent type vectors are designated with their index up, while cotangent vectors are designated with their index down.

No physical equation could be of the form $T^i = W_i$ since if this equation for the components was true in one coordinate system, then it would not be true in another (7 cannot equal 1/7). But it could have the form $T^i = U^i$ since both sides of this equation transform the same way under a coordinate transformation.

One can represent the tangent vector as a little piece of the curve at that point, with an arrow head representing the direction in which the parameter increases, and the piece of curve is the curve between the value of λ for which the curve goes through the point of interest, and the end of the arrowhead represents the point where the curve is when lambda increases by some fixed small amount.

For a cotangent vector, the vector represents the stacking of the level surfaces of the function at the point of interest. Eg, one could draw in the immediate vicinity of the point of interest, two level surfaces (surfaces on which the function has the same value) with one going through the point (all of the nearby points where the function has the same value as it has at the point) and a second surface on which the function has a value some fixed small amount bigger. Note that the cotangent vector is not an arrow. It is more like the stack of pages of a book.

Note that the coordinates x^i or $\tilde{x}^i$ are just functions. They are such that if $x^{j\neq i}$ constant, and $x^i(\lambda)$ has λ defined as the value of x^i at a point of the curve. The number of coordinates needed to do this is the dimension of space (in general 3 in our case, or 4 in special relativity). If you have too many coordinates then in general there are no points, or just scattered points, which obey the equations that $x^{j\neq i}$ constant. If you have too few, then you do not have curves which obey that condition, but things like surfaces or volumes.

Note that, as often written, Newton's law of gravity

$$m\frac{d^2x^i}{dt^2} = -m\partial_i\Phi_{grav} \tag{21}$$

Makes no sense when expressed in terms of general coordinates, because the two sides of the equation are different kinds of vectors. The left is a tangent type vector while the right is a co-tangent vector, and if one carries out a coordinate transformation, the two sides change differently. This has caused much special peading in the literature, and messy arguments to get around this. As we will see below, one can easily make a sensible equation using the metric tensor (see below) but because most people hide the metric, it makes things difficult. And it is not that coordinate transformations are so rare that it does not matter. For example converting to spherical polar coordinates is done "all the time".

II. RULES FOR EQUATIONS

Equations with different physical quantities must be set up so that if they are true in one coordinate system, they will also be true in another coordinate system as well.

III. DOT PRODUCT

One of the items that one often sees in physics is the dot product. This is often written as the product of the components on one vector with those of another. But we now have a problem. The dot product is supposed to represent something physical. Ie, it should be independent of which coordinates one has in the space . If one did something like $T \cdot T = \sum_i T^i T^i$ then if we did a coordinate transformation as above $(\tilde{x}^1 = 7x^i)$, then after the coordinate transformation we would have $\tilde{T} \cdot \tilde{T} = (7T^1)(7T^1) + T^2T^2 + T^3T^3 = (49T^2)^2 + (T^2)^2 + (T^3)^2 \neq (T^1)(T^1) + T^2T^2 + T^3T^3$. Ie, the value of the dot product would depend on which coordinate system one evaluated it in. This could not be a physical quantity, since the outcome depends on which coordinate system one used.

Similarly, if U is a cotangent vector, then $\mathbf{U} \cdot \mathbf{U} = U_1U_1 + U_2U_2 + U_3U_3$ would make no sense since again this would give different answers in different coordinate systems.

However, if one defined it as $\mathbf{T} \cdot \mathbf{U} = \sum_i T^i U_i$ then the 7s in T^1U_1 would cancel, and the answer would be the same in the two coordinate systems. Trying this in general

$$\sum_i \tilde{T}^i \tilde{U}_i = \sum_{ijk} T^i \frac{\partial \tilde{x}^i}{\partial x^i} \frac{\partial x^k}{\partial \tilde{x}^i} U_k = T^j U_k \sum_i \frac{\partial \tilde{x}^j}{\partial x^i} \frac{\partial x^i}{\partial x_k} \tag{22}$$

But

$$\sum_i \frac{\partial \tilde{x}^j}{\partial x^i} \frac{\partial x^i}{\partial x_k} = \frac{\partial \tilde{x}^j (x^i(\tilde{x}^k))}{\partial \tilde{x}^k} = \delta^j_k \tag{23}$$

where $\delta^j_k = 1$ if $j = k$ and zero otherwise. Thus

$$\sum_i \tilde{T}^i \tilde{U}_i = \sum_i T^i U_i \tag{24}$$

Ie, it is independent of the coordinates used to evaluate it. Thus this definition of the dot product as a product between a tangent and a cotangent vector could make sense in a physical equation.

A. Metric

One of the most important physical attributes of the world is distances and lengths. Somehow, whatever the physics is that one wants to examine, distances and lengths come into play. It is a physical attribute and thus must be independent of the coordinates one uses. What is the length of a vector? We usually say that the length is the sum of the squares of the components of a vector. But as above with the dot product, that makes no sense. It depends on what coordinates one uses. If one wants the length of a tangent vector, one must have something such that the sum of indices are over up and down, tangent and cotangent, type indices, so that the coordinate transformations can cancel. The answer is a new structure called the metric.

It is a new and entirely separate object. I want some object which will take two tangent vectors (they could be the same) such that you take some of the components of one of those vectors, and multiply it with components of the other tangent vector, and then multiply those with something else in such a way that that product is independent of coordinates.

The answer is a two component object, usually denoted by g_{ij} which has two indices, and which is such that the length of the vector is equal to the product of one tangent vector with one of the indices and the other one with the other. Ie,

$$\text{Length}^2(T^i) = \sum_{ij} g_{ij} T^i T^j \tag{25}$$

and that the components of this "metric tensor" as it is called, transform as

$$\tilde{g}_{ij} = \sum_{kl} \frac{\partial x^k}{\partial \tilde{x}^i} \frac{\partial x^l}{\partial \tilde{x}^j} g_{kl} \tag{26}$$

Now the transformation of the metric components will cancel out the transformation of the components of the tangent vectors. Furthermore this metric allows transformations from tangent vectors to cotangent vectors. Consider $\sum_j g_{ij} T^j$. Under a coordinate transformation this transforms in the same way as a cotangent vector does. It is a cotangent vector. Ie, it allows one to convert tangent vectors to cotangent vectors. Ie, for each tangent vector there is a cotangent vector associated with it via the metric. If that association is unique, ie, not two different tangent vectors can have the same cotangent vector (eg, if $g_{ij} = 0$ for all tangent vectors, then all tangent vectors would have the same associated cotangent vector) then there is an inverse metric which associates with each cotangent vector a tangent vector. Ie, there exists a tangent type tensor $\mathfrak{g}^{ij}$ which associates with each cotangent vector a tangent vector. $\sum_k \mathfrak{g}^{jk} U_k$. Now g and $\mathfrak{g}$ could be unrelated to each other. However, this would produce a physics which would be different than all classical physics. Thus the assumption is that if

$$U_i = \sum_j g_{ij} T^j; \quad \text{then} \tag{27}$$

$$T^k = \sum_j \mathfrak{g}^{ki} U_i \tag{28}$$

I.e., the tangent vector associated with the cotangent vector associated with a tangent vector is the same as that orginal tangent vector. This would give

$$\sum_i \mathfrak{g}^{ki} g_{ij} = \delta^i_k \tag{29}$$

If we regard $\mathfrak{g}$ and g as matrices, they are the inverse matrix of each other. This would also mean that the T^i and its associated cotangent vectors have the same lengths. Because of the close relationship the cotangent vector associated with T^i is usually written as T_i. They are not the same. The values of the components of T^i and T_i could be very different. depending on what the metric is. Note that there is a very special metric, the one where both the matrices g and $\mathfrak{g}$ are both the identity matrices, where the components do not change if one looks at the tangent or cotangent for a metric, and where the dot product of a vector with itself is just the sum of the squares of the components. This is called Cartesian coordinates, and was the coordinates system in which almost all physics was carried out in the 18th and 19th centuries. This meant that coordinate transformations have an almost trivial effect. It also meant that when physicists tried to go to things like spherical polar or ellipsoidal, or cylindrical coordinates they tended to make a mess of it, or at least an almost incomprehensible complication of the problems.

One way of writing the metric is to write it in infinitessimal notation. Take dx^i as small displacements of the coordinates along a curve. Define ds as the length along the curve. Then one can define the metric by the relation between ds and the dx^i.

$$ds^2 = g_{ij} dx^i dx^j \tag{30}$$

The usual Pythagorian rule is that

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2$$

which tells us that in that case the metric must be

$$g_{11} = 1; \quad g_{22} = 1; \quad g_{33} = 1$$

and all the rest of the components of the metric are zero.

The inverse metric in this case would have the same components

$$g^{11} = 1; \quad g^{22} = 1; \quad g^{33} = 1$$

. However under coordinate changes, the inverse metric changes much differently from the metric components. For the metric, the indices transform like cotangent indices, while the inverse transforms like tangent indices.

One thing to note about the metric is that it is that, because it is symmetric (like the dot product is symmetric) the off-diagonal terms in the metric ($i \neq j$) come in twice. Thus the 2-D metric

$$ds^2 = dx^2 + 6dxdy + 10dy^2 \tag{31}$$

should really be written as

$$ds^2 = dx^2 + 3dxdy + 3dydx + 10dy^2$$

which has $g_{12} = g_{21} = 3$ Not 6

IV. ANTISYMMETRIC TENSOR

There is another tensor which is useful, and that is the completely antisymmetric tensor ϵ_{ijk} (in three dimensions. In generic dimensions there will be as many indices as there are dimensions). Consider this tensor to be antisymmetric if for any pair of indexes swapped swap, the compoent changes sign. Ie,

$$\epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{ikj} = -\epsilon_{kji} \tag{32}$$

Since any swap (not necessarily adjacent) can be created by an odd number of adjacent swaps $\epsilon_{kji} = -\epsilon_{jki} = \epsilon_{jik} = -\epsilon_{ijk}$ this means that this tensor changes sign under any swap of two indices, not necessarily adjacent. If the metric matrix is the identity matrix (note that because the metric matrix is symmetric, it is always possible to choose a coordinate system such that the metric matrix is the identity), we can choose $\epsilon_{123} = 1$

Under a coordinate transformation,

$$\tilde{\epsilon}_{ijk} = \sum_{klm} \epsilon_{lmn} \frac{\partial x^l}{\partial \tilde{x}^i} \frac{\partial x^m}{\partial \tilde{x}^j} \frac{\partial x^n}{\partial \tilde{x}^k} \tag{33}$$

Since because of the antisymmetry, all non-zero terms are related to ϵ_{123} we need only look at

$$\tilde{\epsilon}_{123} = \sum_{klm} \epsilon_{lmn} \frac{\partial x^l}{\partial \tilde{x}^1} \frac{\partial x^m}{\partial \tilde{x}^2} \frac{\partial x^n}{\partial \tilde{x}^3} \tag{34}$$

Recall that we assumed that the metric elements in the original were the identity. Assume that $\epsilon_{123} = 1 = \det(g_{ij})$ ($\det(g_{ij})$ is the determinant of metric matrix with components g_{ij} which is 1 since the metric by assumption in the untilde coordinates is the identity matrix). But the expression is exactly determinant of the Jacobian matrix $J_{il} = \frac{\partial x^l}{\partial \tilde{x}^i}$. In 3 dimensions we can see this by direct calculation.

$$\det J = J_{11}(J_{22}J_{33} - J_{23}J_{32}) - J_{12}(J_{21} * J_{33} - J_{31}J_{23}) + J_{13}(J_21 * J_{32} - J_{22}J_{31}) \tag{35}$$

$$= J_{1i}J_{2j}J_{3k}\epsilon^{ijk} \tag{36}$$

$$\tilde{g}_{ij} = \sum_{kl} g_{kl} \frac{\partial x^l}{\partial \tilde{x}^i} \frac{\partial x^m}{\partial \tilde{x}^j} = (J^T \mathfrak{g} J)_{ij} \tag{37}$$

where the last term is writing the previous expression as a matrix equation and thus

$$\det(\tilde{g}_{ij}) = \det(J)^2 \det(g_{ij}) = \det(J)^2. \tag{38}$$

Thus the totally antisymmetric tensor can be written as $\epsilon_{123} = \sqrt(\det(\tilde{g}_{ij}))$. Similarly for the raised tensor ϵ^{ijk} we get

$$\tilde{\epsilon}^{123} = \frac{1}{\sqrt{\tilde{g}_{ij}}} \tag{39}$$

We note that there are actually two such tensors which differ by a minus sign, since in taking the square root, one could take either the $\pm$ of the square root (or to put it another way, the determinant of the Jacobian could have either sign, depending on whether the coordinate transformation maintained parity or reversed parity while the determinant of the metric always has the plus sign).

V. FIELD THEORY

In almost all of the physics you have learned before this, the physics is in direct contrast to the modern (meaning this last century) physics. You have learned that a good approximation to physics is particle theory. All of physics is concerned with the idea that everything is made up of tiny particles, which follow paths, and the purpose of physics is to determine how those particles move. The paradigm is the 300 year old idea encapsulated in Newton's laws, particularly the $F = ma$ equation. Particles are located at all times at definite positions in space, and the purpose of physics is to determine how those particles move. F is the influence of the outside world on those particles at the location of the particles.

Special Relativity put paid to that worldview. It is basically impossible to reconcile such a worldview with special relativity, and basically with the idea that the influence of the rest of the world on a particle is a local affair (if it was non-local, one would basically need to know everything before one could know anything).

With special relativity came field theory, with its paradigm being Faraday's ideas and Maxwell's theory about how the world behaves. Matter and gravity are not local particles with well defined locations, and the purpose of physics is not to figure out how those particles move. Of course this does not throw out all of physics, since in many situations that view is a good approximation.

Instead what one has is fields. The simplest such field is the functions mentioned above. A function is something which at each and every point in space and time, it hands you a value. Those values are the simplest sort of fields. They do not move along paths. They have no locations, but rather are everywhere. The purpose of physics is not to determine the paths that particles follow under the influence of outside forces. Instead the purpose of physics is to determine those values of the fields (how the values of those functions which are supposed to represent the physical values of those fields at various places vary from time to time).

Field theory posits that at any point in space and time, there are a whole variety of such fields, some functions, some vectors of various types, some other sorts of tensors. The dynamics of the fields are determined by local equations. Ie, the various field interact with each other (help determine and change the changes in each by their values in the immediate vicinity of the place where one is looking). There are no forces mysteriously reaching out from particles here to other particles out there. The influences spread out locally from here to there.

This also makes field theories much more difficult to handle because one has to be concerned not just about what is going on at one place (where the particle is) but with what is happening all over the place. The locality means that one can approximate the changes here by simply looking at what is going on here, at least for short periods of time (because over long times the influences here spread out to almost everywhere).

Electromagnetism has one huge advantage over most modern field theories, in that it is a linear theory in crucial aspect. What that means is that if one finds, somehow, two solutions for the equation, one can find a third, a fourth, ..., by adding together the two solutions one has, with arbitrary additional multiplying constants. As soon as one looses that advantage, which many of our modern theories do, then each of the infinitude of solutions must, in principle, be found one by one. an extremely arduous task. Even electromagnetism can lose its linearity once one takes into account the interactions with other fields. This linearity is one of the reasons why we can study electromagnetism in undergraduate education.

The complexity of having to worry about solutions everywhere at once, not just right here, also means that EM theory is a much more mathematical subject than anything you have seen before. The mathematics is going to be one of the key impediments to fully understanding what is going on in the theory and in its application to specific problems. Thus this field is going to be one that is going to take a while to get used to.

This also means that the way it is taught is going to have to occur in leaps and starts and in approximations. Thus the primary thing we will be doing this year is to impose a huge approximation on the field. That approximation will be that we are going to assume that the field does not change much. Ie, we will assume that any changes are slow so that we will be able to neglect changes. Next year you will begin to look at electrodynamics– ie situations in which the fields do change on time scales which are important to solving the problems. But that is a problem for another day.

Now to return to some of the mathematics.

VI. DERIVATIVES

.

The key derivatives that are used in Electromagnetism are the gradient, the divergence and the curl. The curl of a cotangent vector is actually a tensor. If one has a vector A_i, one can define an antisymmetric tensor $F_{ij} = \partial_i A_i - \partial_j A_i$.

It is not obvious that this is a tensor, since under a coordinate transformation, A_i transforms, and one would get

$$\partial_i A_j \tilde{-} \partial_j A_i = \sum_{kl} \left[\frac{\partial x^k}{\partial \tilde{x}_i} \frac{\partial}{\partial x_k} \left(\frac{\partial x^l}{\partial \tilde{x}^j} A_l \right) - < i \leftrightarrow j > \right] \tag{40}$$

$$= \sum_{kl} \left[\frac{\partial x^k}{\partial \tilde{x}_i} \left(\frac{\partial x^l}{\partial \tilde{x}^j} \frac{\partial}{\partial x_l} A_l \right) \right. \tag{41}$$

$$+ \frac{\partial^2 x_l}{\partial x^i \partial x^j} A_l \tag{42}$$

$$- < i \leftrightarrow j >] \tag{43}$$

The last term goes to zero when you subtract the interchanged partial derivatives, since the order of ordinary derivatives does not matter. The first terms are just the transformation one would expect for an antisymmetric tensor. Ie, the antisymmetric derivative of A_i is an antisymmetric tensor. No extra terms creep in due to the derivatives of the Jacobian matrix.

In three dimensions, one can always map a two index antisymmetric cotangent tensor onto a tangent vector (using the epsilon tensor)

$$B^i = \sum_{jk} \epsilon^{ijk} \partial_j A_k. \tag{44}$$

Note that

$$\partial_i \sqrt{g} \sum_{jk} \epsilon^{ijk} \partial_j A_k = \sum_{ijk} \partial_i e^{ijk} \partial_j A_k = \sum_{ijk} e^{ijk} \partial_i \partial_j A_k \tag{45}$$

where $e^{ijk} = \sqrt{(g)} \epsilon^{ikl}$ is a completely antisymmetric object whose value $e^{123} = 1$ so all terms in e^{ijk} are either 1 or 0– ie constants. Similarly $e_{ijk} = \frac{1}{\sqrt{g}} \epsilon_{ijk}$ which has $e_{123} = 1$. The e object is not a tensor. It does not alter properly under a change of coordinates.

But from the previous definition of $B^i = \sum_{ijk} \epsilon_{ijk} \partial_j A_k$ we get

$$\sum_i \partial_i (\sqrt{(g)} B^i = \sum_{ijk} e^{ijk} \partial_i \partial_j A_k \tag{46}$$

$$= \sum_{ijk} e^{ijk} \partial_j \partial_i A_k = - \sum_{ijk} e^{jik} \partial_j \partial_i A_k \tag{47}$$

which implies it is zero since the order of summation does not matter.

This is called the divergence of B $\nabla \cdot \mathbf{B} = \sum_i \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} B^i) = 0$. Unlike the curl it is an operation that acts on tangent vector fields, while the curl is defined for a cotangent vector field.

The relation between the curl of A_i and B^j is one of Maxwell's equations. That the divergence of the curl is 0 is an identity. It is true for any field A_i. It does not limit A_i in any way.

Similarly, if we define

$$E_i = -\partial_i \phi - \partial_t A_i \tag{48}$$

Then

$$\sum_{ijk} (\epsilon^{ijk} \partial_i E_j - \partial_j E_i) = -\epsilon^{ijk} \left(\partial_j \partial_k \phi + \partial_t \partial_j A_k - < j \leftrightarrow k > \right) \tag{49}$$

Or

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B} \tag{50}$$

This is the second of Maxwell's equations. These are identities, arising from the fact that E and B are derived from the scalar field ϕ and the vector field A_i.

The other equations are

$$\frac{1}{\sqrt{g}} \partial_j \sqrt{g} E^j = \frac{1}{\epsilon_0} \rho \tag{51}$$

$$\epsilon^{ijk} \partial_j B_k = \partial_t E^i + \mu_0 J^i \tag{52}$$

where ρ is the charge density (which can be positive or negative) and J^i (not to be confused with the Jacobian above), which is the current density. The ϵ_0 and μ_0 are constants which are steeped in history, and carry no physics with them. They are now defined (not measured) $\epsilon_0\mu_0 = \frac{1}{c^2}$ where c is the velocity of light, which is now defined. It arose because historically there was no relation between mesurements of time and of space, ie between the meter and the second. The former was defined as 1/(40000) of the circumpherence of the earth (or rather between the North pole of the earth and the equator on the line of longitude running through Paris, while the latter was was defined as 1/860000 of the the time it takes the earth to rotate from midday to the following midday. Now both are defined in terms of the oscillation time for light emitted by a Cs atom in its transition between two specific states. The meter is then obtained by **defining** the value of c times the above definition of the second.

When Maxwell wrote out his equations he used a component set of equations, in Cartesian coordinates, so made no diffeentiation between tangent and cotangent vectors. He assumed that 3-D metric was the identity. Oliver Haviside developed the "vector" notation, in which say $E_i = E^i$ (Cartesian coordinates) was designated by $\mathbf{E}$, and similarly for $\mathbf{B}$ and A. The partial derivative ∂_i was designated by ∇, the metric product by $\cdot$ (eg $\mathbf{E}\cdot\mathbf{B}$. and the cross product by $\times$ ($\mathbf{E}\times\mathbf{B}$ instead of $\epsilon^{ikl}\partial_j B_k$. The divergence was $\nabla\cdot\mathbf{B}$ rather than $\frac{1}{\sqrt{g}}\partial_i\sqrt{g}B^i$ with g=1.

This notation is more compact, but also more limited. It **only** applies if one is in Cartesian coordinates, and thus does not allow one to use use any other coordinates without a lot of work (see Griffith's appendix to see the kind of manipularions he has to go through.) Also, the cross product only works in three dimensions, not 1,2,4,5... dimensions. In particular when one goes to 4 dimensional spacetime of special relativity (never mind the general metric of General Relativity) the whole Heaviside notion falls completely apart and one must go to the tensor notation anyway (eg, the electromagnetic tensor is an antisymmetric tensor $F_{\mu\nu}$ or using the metric $F^{\mu\nu}$, where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ where the greek indices take values of (0,1,2,3) and $A_0 = -\phi$. These lead to the Lorentz transformation of Electromagnetism easily, almost trivially, instead of going through convoluted arguments as from 1905 to find the transformations. (Note that this form of the equations and the fields was already discovered by 1908, so to teach a 100 year old form now seems a bit perverse. But since most textbooks still do it, and in particular our text book still does it, much of this course will also do it.

VII. CARTESIAN COORDINATES.

If we choose special coordinates, such that g_{ij} is the identity matrix, then the tensor and cotensor components are the same. One can then get sloppy, and simply ignore factors line g_{ij} or $\sqrt{g}$, and forget that the metric plays any role in these equations. However, if one decides to change coordinates, or work in say spherical polar coordinates, because the problem is spherically symmetric, then neglecting the metric is a disaster, and one gets the mess that is the second appendix of Griffiths textbook.

Since g_{ij} in cartesian coordinates is the identity matrix, the components of tangent and cotangent vectors are the same. Thus Heaviside's notation for this case, is what is often used. A_i is $\mathbf{A}$. E_i is $\mathbf{E}$, $\frac{1}{\sqrt{g}}\partial_i\sqrt{g}B^i$ is $\nabla\cdot\mathbf{B}$, $B^i = \epsilon^{ijk}(\partial_j A_k)$ is $\mathbf{B} = \nabla\times\mathbf{A}$. Also the volume element is $dV = \sqrt{g}dx^1dx^2dx^3$ becomes simply $dV = d^3x$ and makes the proof of the Gauss and Stokes theorems more tedious– talking about dividing up the space into little cubic cells and arguing that that the integrals over the cell walls is zero. It also makes knowing what the integral over the surface more difficult, since it is almost never expressible eaasily in Cartesian coordinate.

Thus while this is certainly a compact convention, it clearly hides a lot, and produces equations which are not coordinate invariant. They are only true is a very special set of coordinates.

VIII. SUMMATION CONVENTION

In all of the above formulas, when we have summation over indices, they are always over indices where one is up (tangent vector type index) and one is down (cotangent type index). Einstein saw this and said– why do we need that summation symbol? Since whenever we have a situation where there is one index up and one down with both having the same name, why not just eliminate the summation symbol? Thus T^iW_i can be taken as shorthand for $\sum_i T^iW_i$, That way one can make the whole expression shorter and make it easier to write. Since such a summation is such that it is coordinate independent (the transformation of the up index and down index on a coordinate transformation with the summation will just cancel). Thus, if our equations are coordinate invariant equations, we can just make a rule that only an up and down index can be the same, and they must be summed over. He regarded this as one of his greatest contributions to mathematics

IX. SUMMARY:

Tangent type vector T^i.

Paradygm: Tangent vector to a curve $\frac{dx^i(\lambda)}{d\lambda}$

Coordinate transformation: $\frac{\tilde{T}^i = \sum_j T^j \partial x^i}{\partial x^j}$

Cotangent type vector U_i

Paradygm: gradient $\frac{\partial f}{\partial x^i}$

Coordinate transformation: $\frac{\tilde{\partial} \tilde{f}}{\partial x^i} = \sum_j \frac{\partial f}{\partial x^j} \frac{\partial x^j}{\partial \tilde{x}^i}$

Product of tangent and cotangent vector: $\sum_i T^i U_i = \sum_i \tilde{T}^i \tilde{U}_i$ since $\sum_j \frac{\partial x^i}{\partial \tilde{x}^j} \frac{\partial \tilde{x}^j}{\partial x^k} = \delta^i_k$

Kroneker delta δ^i_j =1 if i=j and 0 if $i \neq j$

Derivative notation $\frac{\partial_i = \partial}{\partial x^i}$ and $\frac{\tilde{\partial}_i = \partial}{\partial \tilde{x}^i}$

Metric tensor g_{ij} such that length squared of T^i is $\sum_{ij} g_{ij} T^i T^j$. Also $g_{ij} = g_{ji}$

Summation convention $\sum_i ...^i ..._i \equiv ...^i ..._i$

Map of tangent index to cotangent index $g_{ij} T^j \equiv T_i$

Inverse map $g^{ij} U_j \equiv U^i$. $T^i = g^{ij}(g_{jk} T^k)$ implies $g^{ij} g_{jk} = \delta^i_k$. g^{ik} regarded as a matrix is the inverse of the matrix g_{ij}.

Determinant of matrix g_{ij} is designated by g.

If matrix Ji^k is $\tilde{\partial}_k x^i$ and J is its determinant, then $\tilde{g} = J^2 g$.

Antisymmetric tensor in 3 dimensions $\epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{ikj}$ and $\epsilon_{123} = \sqrt(g)$. $e_{ijk} = e^{ijk}$ is antisymmetric symbol (not a tensor) with $e_{123} = 1$

Volume element: $dV = \sqrt{g} dx^1 dx^2 dx^3$.

Divergence: $\frac{1}{\sqrt{g}} \partial_i (\sqrt{g} T^i)$.

Curl: $B^i = \epsilon^{ijk} (\partial_j A_k)$

Gausses Law: $\int \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} T^i) \sqrt{g} dx^1 dx^2 dx^3 = \int \sqrt{g_{11}} T^1 \sqrt{^2 g} dx^2 x^3$ where $^2 g$ is the determinant of the two dimensional metric of the surface in coordinates where x^1 const defines the surface. Ie, it is the physical area integral over the surface of the "outward" (away from the the volume) perpendicular component of the vector field T^i.

Stokes law: $\int \epsilon^{1ij} \partial_i A_j dx^2 dx^3 = \int \epsilon^{123} A_3 dx^2$ where the surface is x^1 constant and the boundary is in the surface and has x^2 constant. Ie, the integral of the component of the curl perpendicular to the surface equals the integral of the component of the field parallel to the curve integrated using physical lengths along that curve.

Tensor equations: The "unsummed" indices must be the same names and type (super or sub-script). Additions of terms must also be of the same tensor type.

Maxwell's equations:

$$B^i = \epsilon^{ijk} \partial_j A_k \Rightarrow \frac{1}{\sqrt{g}} \partial_i \sqrt{g} B^i = 0 \tag{53}$$

$$E_i = -\partial_j \phi - \partial_t A_i \rightarrow \partial_t \epsilon^{ijk} \partial_j E_k - \partial_t B^k \tag{54}$$

$$\frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} E_i = \frac{1}{\epsilon_0} \rho = \frac{4\pi\mu_0}{c^2} \rho \tag{55}$$

$$\epsilon^{ijk} \partial_j B_k = \mu_0 \epsilon_0 \partial_t E^i + \mu_0 J^i = \frac{1}{c^2} \partial_t E^i + \mu_0 J^i \tag{56}$$

These are the 3-D coordinate invariant forms of Maxwell's equation. I.e., they allow one to write the equations in any coordinate system one wishes.

Cartesian coordinates: Coordinates in which $g_{ij} = \delta_{ij}$ everywhere. This is not a tensor. There are only a limited number of transformations one can make which go from one Cartesian coordinate system to another– namely the three rotations.

Cartesian coordinates are such that the components of the tangent and cotangent forms of a vector are numerically identical. g is always 1. They are handy, but create confusion when one tries to do calculations in other coordinates– like spherical polar coordinates for problems which are spherically symmetric.

A. Final comments

The key feature of Maxwell's equations is that they are linear equations. Ie, if one has two solutions of the equations with their charge and current densities, then the two solutions can be added to give a solution to the

Maxwell's equations with the sum of the two charge and current densities as "sources".

In this course we will be looking at the approximation where there is no time dependence. All temporal derivatives are zero. In this approximation, the electromagnetic field has dynamics of its own. It is simply tied to the charges and currents.

B. Coordinates as arguments

In the above notes I have used the notation of $f(x^i)$ to stand for $f(x^1, x^2, x^3)$. This is somewhat standard notation, but it can be confusing, as it would seem to say that the function is a function of some specific coordinated, x^1 or x^2 or x^3 rather than what it is supposed to signify x^1 and x^2 and x^3 I could try using $\mathfrak{X}$ to stand for the set $\{x^1, x^2, x^3\}$ (and $\tilde{\mathfrak{X}}$ to stand for the set $\{\tilde{x}^1, \tilde{x}^2, \tilde{x}^3\}$) when they are used as arguments for a function. I do not know if this would alleviate the confusion, of add to it.

Thus if I look at a coordinate change, I would write $\tilde{x}^i(\mathfrak{X})$ instead of $\tilde{x}^i(x^j)$, (or $x^i(\tilde{\mathfrak{X}})$ for the inverse relationship) where you need to ignore the index j of the argument since it just indicates the whole set of possible indices rather than some specific value for the index.)