Lecture I: Vectors, tensors, and forms in flat spacetime

Christopher M. Hirata
*Caltech M/C 350-17, Pasadena CA 91125, USA**
(Dated: October 1, 2012)

I. OVERVIEW

The mathematical description of curved spacetime is an involved subject and we will spend much of the first term on developing and applying the associated machinery. Key to all of this work are the concepts of vectors, forms, and tensors. I'm sure all of you have worked extensively with vectors and vector fields (the electric field in E&M, the velocity field in hydrodynamics, etc.) in previous coursework, and have studied the operations of vector algebra (norms, dot and cross products) and of vector calculus (grad, div, curl). All of these have extensions to curved spacetime, but are customarily dressed in the language of equations with tensors and funny-looking up and down indices ($A^{\alpha\beta}{}_\gamma$). Before we use this machinery to work in curved spacetime, we will repeat the description of flat spaces (Euclidean space $\mathbb{R}^n$ and Minkowski space $\mathbb{M}^4$) and special relativity to gain familiarity.

The recommended reading for this lecture is:

- MTW Ch. 1 – This is to give you a taste of where we're headed in the first term. Don't worry if you can't follow every equation in Ch. 1 yet; it will all be explained in due course! You should read this chapter as you might the first chapter of a mystery novel ...
- MTW §2.1–2.5 – This is where the mathematical material begins. It covers the same material as in lecture, and you should make sure you follow it. Since you've seen the underlying physics before, albeit in different language, this is the place for you to get a firm grasp of the notation and how to manipulate tensor expressions. So please if something in this section doesn't make sense to you – speak up! Probably someone else in class has a similar question.

II. VECTORS

[Reading: §2.3]

In freshman physics, you learned about the concept of representing a vector as an arrow (an object with a magnitude and direction – as distinct from a scalar, with only a magnitude). In special relativity, you probably learned about the notion of 4-vectors. Several types of vectors were considered, such as:

- The *displacement* of an object: if something moves from point $\mathcal{A}$ to $\mathcal{B}$, then we can write its displacement vector,

$$\boldsymbol{s}_{\mathcal{A}B} = \mathcal{B} - \mathcal{A}. \tag{1}$$

- The *velocity* of an object: if an object is moving along a curve $\mathcal{P}(\lambda)$, where λ is a parameter, then we define its velocity vector,

$$\boldsymbol{v} = \frac{d\mathcal{P}}{d\lambda}. \tag{2}$$

  To define a velocity in Newtonian mechanics, $\mathcal{P} \in \mathbb{R}^3$ and λ is the static, globally defined time. In relativity, we will take λ to be the proper time, and $\mathcal{P} \in \mathbb{M}^4$.

In this lecture, since we are working in flat spacetime, both Eqs. (1) and (2) are equally valid. When we go to curved spacetime, only Eq. (2) will make sense. Drawn on the 2-dimensional surface of the Earth, the velocity of a boat and the velocity of the ocean current are valid 2-dimensional vectors at every point, and they can be added, subtracted, etc. like any other vector. But the concept of "displacement" as a vector will no longer be meaningful

*Electronic address: `chirata@tapir.caltech.edu`

(addition doesn't commute: think about marching 20 km north and then 20 km east, versus 20 km east and then 20 km north). So for the moment you could think of a vector being tangent to the 4-dimensional surface defining spacetime, or of a vector as being defined by an expression such as Eq. (2) – e.g. a vector defined at a point $\mathcal{P}$ on Earth's surface *is* a possible velocity of an object moving on the surface of Earth that is instantaneously at $\mathcal{P}$. We will make this more mathematically precise in the context of curved spacetime later (for now we will use the traditional flat-spacetime notion of vector, and redefine it when we consider curved spacetime).

All of the material in this lecture is true for any "vector" in a finite-dimensional vector space with an inner product. That is, all of the formalism we develop here is valid so long as the operations of addition, scalar multiplication, and (for some of the material) inner product are defined and satisfy the standard rules (e.g. distributivity of scalar multiplication, etc.). The notions we will define of "up" and "down" indices, 1-forms, and tensors are also generally meaningful on any vector space, even without an inner product – although in this course we will never need to work without an inner product. This makes the formalism generally quite useful, e.g. in the analysis of data where the data can be represented as a vector (CMB maps, gravitational waveforms ...).

The fact that the vectors we are studying are tangent to a manifold will endow them with additional operations that we will need later (e.g. the covariant derivative). These additional structures make calculus on manifolds a much richer subject than simply cataloging the properties of vectors and 1-forms. But we will start with the basics.

A. Bases

For the moment, let us proceed: we work in an n-dimensional space, and the vectors $\boldsymbol{v}$ live in some n-dimensional vector space $\mathcal{V}$. Sometimes we will describe it with an arrow, but in order to do calculations you know it is usually most convenient to work with coordinates. One constructs coordinates by defining a set of basis vectors. In Euclidean geometry, these could be the basis vectors of $\mathbb{R}^3$,

$$\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}, \tag{3}$$

or in special relativity they could be those of $\mathbb{M}^4$:

$$\{\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}. \tag{4}$$

(By convention, the vector pointed in the "time" direction is usually denoted as $\boldsymbol{e}_0$. Aside from this label, we haven't done anything yet that singles out "time" as different from "space".) You are probably accustomed to working with orthonormal bases, i.e. bases where each vector has unit length and is orthogonal to the others, but we won't do this yet. We haven't even defined "orthogonal" or "unit length" yet! All that is required to be a basis is that every vector be representable in a unique way by an expression:

$$\boldsymbol{v} = v^0\boldsymbol{e}_0 + v^1\boldsymbol{e}_1 + v^2\boldsymbol{e}_2 + v^3\boldsymbol{e}_3 = \sum_{\alpha=0}^{3} v^\alpha \boldsymbol{e}_\alpha, \tag{5}$$

where v^0, v^1, v^2, and $v^3 \in \mathbb{R}$ are *components*.

Some nomenclature: if there is a way of writing every vector in the form of Eq. (5), then the set of vectors $\{\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is said to be *complete*. If no vector can be written in this form with more than one different sets of components, then $\{\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is said to be *linearly independent*. Bases are both complete and linearly independent.

B. Very important notation

You will notice that in Eq. (5), we wrote the indices on v^α in the "up" rather than "down" position. This is **very important**. The location of the index will denote how a quantity transforms under changes of basis (this could be a rotations in $\mathbb{R}^3$, but it could also be much more general – and in the more general cases these are not equivalent). In texts, the "up" indices are called *contravariant* and the "down" indices are *covariant*. I'll call them up and down.

The Greek indices are used in 4-dimensional spacetimes and can take any value. At various times we will consider cases where we want to allow an index to take only spatial values (1, 2, or 3) and deliberately exclude the time – at those times we will use Latin indices.

A final, **very important** bit of notation: the index α in Eq. (5) is summed over all of its legal values (0...3). Expressions of this form are so common in GR that we will often use the *Einstein summation convention* (EΣC) in which the sum is **implied**. That is, we will usually write Eq. (5) in the form

$$\boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha. \tag{6}$$

The EΣC applies to one up and one down index. Summed indices are dummy indices: one may consistently change their labels without consequence:

$$\boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha = v^\beta \boldsymbol{e}_\beta. \tag{7}$$

C. Linear algebra

In order for vectors to be useful, we need operations on them, and then we need to come up with laws of physics using these operations that agree with experiment. We can introduce now the standard operations of linear algebra. These are addition and scalar multiplication: in the case that we want the vector

$$\boldsymbol{w} = a\boldsymbol{u} + b\boldsymbol{v} \tag{8}$$

(with $a, b \in \mathbb{R}$ and $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{V}$) the corresponding components are

$$w^\alpha = au^\alpha + bv^\alpha. \tag{9}$$

D. Changes of basis

[Not quite covered at this point in the book, but I think it's a useful reminder.]

Sometimes one wants to convert the components of a vector from one ("unprimed") basis to another ("primed") basis. As you learned in linear algebra, the way to do this is to write the unprimed basis vectors in terms of the primed basis vectors. If the primed vectors form a basis this is always possible – one writes

$$\boldsymbol{e}_{\alpha'} = L^\beta{}_{\alpha'} \boldsymbol{e}_\beta = \sum_\beta L^\beta{}_{\alpha'} \boldsymbol{e}_\beta, \tag{10}$$

where $\mathbf{L}$ is an $n \times n$ linear transformation matrix. There must also be an inverse transformation,

$$\boldsymbol{e}_\beta = [\mathbf{L}^{-1}]^{\alpha'}{}_\beta \boldsymbol{e}_{\alpha'}, \tag{11}$$

where $\mathbf{L}^{-1}$ is the matrix inverse of $\mathbf{L}$ (a change of basis is always invertible). [Homework: Prove that the transformation matrix back to unprimed coordinates is actually $\mathbf{L}^{-1}$.]

If $\boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha$, then it follows that

$$\boldsymbol{v} = v^\beta \boldsymbol{e}_\beta = v^\beta [\mathbf{L}^{-1}]^{\alpha'}{}_\beta \boldsymbol{e}_{\alpha'} = v^{\alpha'} \boldsymbol{e}_{\alpha'}, \tag{12}$$

where we have set

$$v^{\alpha'} = [\mathbf{L}^{-1}]^{\alpha'}{}_\beta v^\beta. \tag{13}$$

You can see here that the vector components transform according to the inverse of the transformation matrix of the basis vectors. Generally the down and up indices are used to describe whether the transformation is of the type of Eq. (10) or (13).

III. METRIC TENSOR AND DOT PRODUCT

[Reading: §2.4]

In ordinary electrodynamics, we learned that a new operation was useful: the *dot product* $\cdot$. This is an operation that takes two vectors and returns a scalar. The familiar dot product had the following properties:

- Symmetric: $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$.
- Distributive: $(a\boldsymbol{u} + b\boldsymbol{v}) \cdot \boldsymbol{w} = a(\boldsymbol{u} \cdot \boldsymbol{w}) + b(\boldsymbol{v} \cdot \boldsymbol{w})$.

One can define the square-norm $|\boldsymbol{u}|^2 \equiv \boldsymbol{u}\cdot\boldsymbol{u}$, and define two vectors to be orthogonal if their dot product is zero.

You may remember an additional property from freshman physics that $|\boldsymbol{u}|^2 \geq 0$ with equality holding only if $\boldsymbol{u} = \boldsymbol{0}$. Such is **not** the case in relativity, as we will discuss in more detail below.

If we want to define a dot product satisfying the distributive property, it is sufficient to describe its behavior on the basis vectors. Let us define $g_{\alpha\beta} = \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta$; then for general vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, we find

$$\boldsymbol{u}\cdot\boldsymbol{v} = \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta u^\alpha v^\beta = g_{\alpha\beta} u^\alpha v^\beta. \tag{14}$$

The symmetry property tells us that we should have $g_{\alpha\beta} = g_{\beta\alpha}$. Thus in any basis the dot product can be represented by a symmetric matrix. This matrix changes if we do a transformation of basis – it is:

$$g_{\alpha'\beta'} = \boldsymbol{e}_{\alpha'} \cdot \boldsymbol{e}_{\beta'} = L^\alpha{}_{\alpha'} L^\beta{}_{\beta'} \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta = L^\alpha{}_{\alpha'} L^\beta{}_{\beta'} g_{\alpha\beta}. \tag{15}$$

The nature of this transformation is the reason why we write $g_{\alpha\beta}$ with down indices.

Another common notation for the dot product is to write it as a *tensor*: a "linear" (i.e. satisfying the distributive property over addition and scalar multiplication) function that takes in 2 vectors and outputs a scalar. That is, we could write

$$\mathbf{g}(\boldsymbol{u},\boldsymbol{v}) \equiv \boldsymbol{u}\cdot\boldsymbol{v} \tag{16}$$

and consider $g_{\alpha\beta}$ to be the components of the object $\mathbf{g}$.

It is a matter of taste whether we write the dot product as a function or with a $\cdot$ symbol: they are different notations for the same thing.

A. Example: $\mathbb{R}^3$

As a first example, let's consider the metric tensor in Euclidean space $\mathbb{R}^3$ with the standard orthonormal basis consisting of the unit vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$. The dot products of these form the identity matrix

$$g_{ij} = \delta_{ij} \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{17}$$

(The δ_{ij} is the Kronecker delta symbol and is defined as 1 if $i = j$ and 0 otherwise.)

In this case, the square norm of any vector is

$$|\boldsymbol{u}|^2 = g_{ij} u^i u^j = (u^1)^2 + (u^2)^2 + (u^3)^2. \tag{18}$$

So in this case the square norm is positive for all vectors except the zero vector.

B. More general case

I mentioned earlier that it is not always true that square norms have to be positive for nonzero vectors. One can see this by applying a change of basis: on the homework, you will show that (i) a change of basis can always bring the metric tensor into the form of a diagonal matrix with diagonal entries of +1, 0, or −1; and (ii) the number of +1, 0, and −1 diagonal entries cannot be changed. The number of each of these entries is called the *signature* of the metric. Metric tensors with 0s in their signatures are pathological – there are nonzero vectors that are orthogonal to all vectors, a situation that voids most of the theorems we will use – and so we will disallow them in further study. Thus the important aspect of the signature is how many +1s and how many −1s it contains. Euclidean $\mathbb{R}^3$ has signature $+++$.

C. Example: $\mathbb{M}^4$

Now we move to the spacetime of special relativity: $\mathbb{M}^4$. The physically useful dot product in this space has signature $-+++$: this means that we can choose a basis such that

$$g_{\alpha\beta} = \eta_{\alpha\beta} \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{19}$$

The matrix defines the symbol $\eta_{\alpha\beta}$:

$$\eta_{\alpha\beta} = \begin{cases} -1 & \text{if } \alpha=\beta=0 \\ 1 & \text{if } \alpha=\beta\neq 0 \\ 0 & \text{if } \alpha\neq\beta \end{cases} . \tag{20}$$

The change in sign of one number seems like it would be trivial, but it has far-reaching consequences. To see this, let's consider the square norm of a vector $\boldsymbol{u}$:

$$|\boldsymbol{u}|^2 = g_{\alpha\beta}u^\alpha u^\beta = -(u^0)^2 + (u^1)^2 + (u^2)^2 + (u^3)^2. \tag{21}$$

This is:

- Positive if the spatial components of $\boldsymbol{u}$ dominate, i.e. $\sqrt{(u^1)^2 + (u^2)^2 + (u^3)^2} > |u^0|$. Such a vector is called *spacelike*.
- Negative if $\sqrt{(u^1)^2 + (u^2)^2 + (u^3)^2} < |u^0|$. Such a vector is called *timelike*.
- Zero if $\sqrt{(u^1)^2 + (u^2)^2 + (u^3)^2} = |u^0|$. Such a vector is called *null* or *lightlike*.

One can see in special relativity that the (4-dimensional) velocity vector of an object moving slower than the speed of light is timelike, and that of an object moving faster than the speed of light is spacelike. The infamous cosmic "speed limit" is encoded into the geometry of spacetime in the form of the definition of the dot product!

[**Warning**: Some references, including most of the particle physics literature, define the metric with signature $+---$. All the physics works out the same, except that there are differing minus signs in many equations. It's like SI versus CGS units in E&M ... you'll have to get used to it. The $-+++$ signature is however in almost universal use in relativistic astrophysics.]

IV. 1-FORMS

[Reading: §2.5]

So far we've defined vectors and the dot product. We will need a few more types of objects in order to formulate the laws of special (and then general) relativity. The most elementary new construct is the *1-form*. Some of our favorite objects, such as the electromagnetic vector potential or wave vector, are naturally thought of as 1-forms. You are probably familiar with them as vectors; we will clarify the relationship in a few minutes.

A 1-form is a linear operator that takes a vector $\boldsymbol{v}$ and returns a scalar ϕ. We will sometimes denote it by a tilded letter, e.g. $\tilde{\boldsymbol{k}}$, and write the operation as:

$$\phi = \langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle. \tag{22}$$

Linearity means that the 1-form is distributive:

$$\langle \tilde{\boldsymbol{k}}, a\boldsymbol{u} + b\boldsymbol{v} \rangle = a\langle \tilde{\boldsymbol{k}}, \boldsymbol{u} \rangle + b\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle. \tag{23}$$

If the vectors live in a vector space $\mathcal{V}$, then the 1-forms are in a vector space called the *dual space* $\mathcal{V}^\star$. The operation $\langle , \rangle$ is called *contraction*.

A. Description

How are we to describe 1-forms? Again, either by geometric pictures or by components. The geometric picture presented by MTW is of a set of equally spaced parallel wave crests. Given a vector $\boldsymbol{v}$, the contraction $\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle$ is the number of times the vector pierces a crest. This number is signed (piercing a crest the "wrong way" incurs a $-$ sign) and interpolated (there are fractional crests). Physically, if $\tilde{\boldsymbol{k}}$ is a wavenumber, then $\langle \tilde{\boldsymbol{k}}, \boldsymbol{s} \rangle$ represents the number of wave cycles one traverses in a displacement $\boldsymbol{s}$.

The component description of a 1-form is straightforward: we know that a 1-form is completely described by its action on the basis vectors $\boldsymbol{e}_\alpha$. We may therefore write the components:

$$\tilde{k}_\alpha = \langle \tilde{\boldsymbol{k}}, \boldsymbol{e}_\alpha \rangle. \tag{24}$$

You should be able to prove that in a change of basis, $\tilde{k}_{\alpha'} = L^{\beta}{}_{\alpha'} k_\beta$, so it is appropriate to use the lower index. Since $\boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha$, it is easily seen that

$$\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle = \langle \tilde{\boldsymbol{k}}, v^\alpha \boldsymbol{e}_\alpha \rangle = \langle \tilde{\boldsymbol{k}}, \boldsymbol{e}_\alpha \rangle v^\alpha = \tilde{k}_\alpha v^\alpha. \tag{25}$$

Just as we had basis vectors $\boldsymbol{e}_\alpha$, so we can construct basis 1-forms $\boldsymbol{\omega}^\alpha$ with $\tilde{\boldsymbol{k}} = k_\alpha \boldsymbol{\omega}^\alpha$.

B. Relation to vectors

In a universe where we did not define the dot product (or metric tensor), vectors and 1-forms would be completely different beasts. But in a universe with a metric tensor there is a natural correspondence between vectors and 1-forms that is extremely useful in GR. Let's take any vector $\boldsymbol{k}$ (think of a wave vector) and associate to it the 1-form $\tilde{\boldsymbol{k}}$ defined by

$$\boldsymbol{k} \cdot \boldsymbol{v} = \langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle. \tag{26}$$

The components of $\tilde{\boldsymbol{k}}$ are found from Eq. (24):

$$\tilde{k}_\alpha = \langle \tilde{\boldsymbol{k}}, \boldsymbol{e}_\alpha \rangle = \boldsymbol{k} \cdot \boldsymbol{e}_\alpha = k^\beta \boldsymbol{e}_\beta \cdot \boldsymbol{e}_\alpha = g_{\alpha\beta} k^\beta. \tag{27}$$

Recall that we disallowed degenerate metrics, i.e. we required $g_{\alpha\beta}$ to form an invertible matrix (it need not be positive definite). Then the mapping from $\boldsymbol{k}$ to $\tilde{\boldsymbol{k}}$ is one-to-one and onto (establishing this was in fact the reason why we required a nondegenerate metric). The inverse mapping is

$$k^\alpha = g^{\alpha\beta} \tilde{k}_\beta, \tag{28}$$

where we have defined $g^{\alpha\beta}$ to be the matrix inverse of $g_{\alpha\beta}$: $g^{\alpha\beta} g_{\beta\gamma} = \delta^\alpha_\gamma$. On the homework, you will find the transformation properties of $g^{\alpha\beta}$ and see that it makes sense for it to have upper indices. You will also prove that:

$$g^{\alpha\beta} \tilde{k}_\alpha \tilde{l}_\alpha = \boldsymbol{k} \cdot \boldsymbol{l}. \tag{29}$$

The mapping between vectors and 1-forms is so fundamental that we will usually drop the tilde and use $\boldsymbol{k}$ to describe both. When doing actual computations, the location of the index – up or down – will tell us whether a vector or 1-form is used in evaluating an expression.

Since in GR the concept of a dot product exists, why are we even bothering with 1-forms? The reason is two-fold. First, some calculations that don't actually use the dot product are most easily carried out with 1-forms. Second, while many of the more complicated manipulations we will do in curved spacetime are formally possible with just vectors and tensors, the 1-form will dramatically simplify our lives.