Lecture II: Coordinate bases, tensor algebra in flat spacetime, and special relativity

Christopher M. Hirata
Caltech M/C 350-17, Pasadena CA 91125, USA[*]
(Dated: September 30, 2011)

I. OVERVIEW

In this lecture we will continue developing the tools of tensor algebra and calculus in flat spacetime. Some of the discussion will be longwinded, but we are working through calculus in a way that will be applicable to curved spacetime. The laws of special relativity will be used to illustrate their application.

The recommended reading for this lecture is:

- MTW §2.6–3.2. (Actually, 3.2 on tensors is covered in this lecture and 3.1 on electrodynamics is next lecture. I think the subject flows a bit more coherently this way.)

II. GRADIENTS AND THE COORDINATE BASIS

[Reading: MTW §2.6–2.7]

We will now introduce the most commonly used basis in special relativity (which generalizes readily to GR) – the coordinate basis. To do so, we have to introduce an important type of 1-form, the gradient $\boldsymbol{d}f$. We will then use this to construct the basis vectors and 1-forms, $\boldsymbol{e}_\alpha$ and $\boldsymbol{\omega}^\alpha$, associated with a coordinate system.

A. Gradient operator

Suppose that we have a scalar field f, i.e. a function that associates to each point $\mathcal{P}$ a scalar $f(\mathcal{P})$. Then you know from multivariable calculus that it was possible to define a vector field called "grad f" or "∇f." We will define the gradient here as a 1-form rather than a vector. However in the event that we have defined the metric tensor (or dot product), this 1-form can be identified with a vector by raising an index.

Let us suppose that we take a trajectory $\mathcal{P}(\lambda)$ through spacetime. Then the scalar field f defines a function of λ, $f[\mathcal{P}(\lambda)]$. This function can be differentiated and evaluated at $\lambda = 0$,

$$\frac{d}{d\lambda} f[\mathcal{P}(\lambda)]\bigg|_{\lambda=0}. \tag{1}$$

This derivative should depend only on the particular choice of $\mathcal{P}(0)$ [or the coordinates $x^\alpha(\lambda = 0)$], and their first derivatives with respect to λ. It should depend linearly on the latter. Think about the chain rule:

$$\frac{df}{d\lambda} = \frac{\partial f}{\partial x^\alpha} \frac{dx^\alpha}{d\lambda}. \tag{2}$$

Therefore we may write

$$\frac{d}{d\lambda} f[\mathcal{P}(\lambda)]\bigg|_{\lambda=0} = \left\langle \boldsymbol{d}f[\mathcal{P}(0)], \frac{d\mathcal{P}}{d\lambda}(\lambda = 0) \right\rangle. \tag{3}$$

Here $d\mathcal{P}/d\lambda$ is a vector and $\boldsymbol{d}f$ is a linear operation on this vector, i.e. $\boldsymbol{d}f$ is a 1-form. We call it the *gradient* of f.

If we set $\boldsymbol{v} = d\mathcal{P}/d\lambda$, then we may define the *directional derivative*

$$\partial_{\boldsymbol{v}} f = \langle \boldsymbol{d}f, \boldsymbol{v} \rangle = \frac{d}{d\lambda} f[\mathcal{P}(\lambda)]. \tag{4}$$

This is the unifying relation between directional derivatives, gradients, and ordinary derivatives. Note that the dot product or metric plays no role here.

---

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

B. Coordinate basis

In flat spacetime, we may choose a coordinate system in which the coordinates x^α are the components of a position vector $\boldsymbol{x} = x^\alpha \boldsymbol{e}_\alpha$. In this case, examining Eq. (2), we see that if $\boldsymbol{v} = d\mathcal{P}/d\lambda$, then

$$\langle \boldsymbol{d}(x^\alpha), \boldsymbol{v} \rangle = \frac{dx^\alpha}{d\lambda} = v^\alpha. \tag{5}$$

Since we recall that $\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle = \tilde{k}_\beta v^\beta$, it follows that the components of $\boldsymbol{d}(x^\alpha)$ are

$$[\boldsymbol{d}(x^\alpha)]_\beta = \delta^\alpha_\beta, \tag{6}$$

and thus it follows that $\boldsymbol{d}(x^\alpha)$ is a basis 1-form:

$$\boldsymbol{d}(x^\alpha) = \boldsymbol{\omega}^\alpha. \tag{7}$$

Note what is in this equation: x^α is a coordinate, i.e. it's a scalar function of position; $\boldsymbol{d}(x^\alpha)$ is its gradient, i.e. a 1-form; and $\boldsymbol{\omega}^\alpha$ is a basis 1-form (α denotes which 1-form, not which component).

So far this is all notation – nothing nontrivial has been said. But what is important is that when we go to curved spacetime (or even if we used a curved coordinate system such as spherical coordinates), the coordinates will no longer be components of a vector, although we will continue to denote them by x^α. However, the x^α are still scalar functions and their gradients $\boldsymbol{d}(x^\alpha)$ still exist. Moreover, none of the discussion here has used the metric. So when we go to curved spacetime, even though there will no longer be such a thing as a position vector $\boldsymbol{x}$, it will still be permissible to define a set of basis vectors by

$$\boldsymbol{e}_\alpha \equiv \frac{d\mathcal{P}}{d\lambda}(\lambda = 0), \qquad x^\beta[\mathcal{P}(\lambda)] = x^\beta[\mathcal{P}(0)] + \lambda \delta^\beta_\alpha \tag{8}$$

("the $\boldsymbol{e}_\alpha$ vector is the velocity of a particle whose α coordinate is increasing at unit rate, and whose other coordinates remain fixed") and the basis 1-forms by

$$\boldsymbol{\omega}^\alpha \equiv \boldsymbol{d}(x^\alpha). \tag{9}$$

This basis will be called the *coordinate basis*. For a general coordinate system it will not be orthonormal, and we may introduce other bases that are orthonormal for the purpose of displaying results; but for most calculations, computation in the coordinate basis is the most straightforward (though not necessarily efficient) approach.

[*Example of a more general coordinate system in class*: Polar coordinates in 2 dimensions. The basis $\{\boldsymbol{e}_r, \boldsymbol{e}_\theta\}$ is not orthonormal, nor is the basis $\{\boldsymbol{\omega}^r, \boldsymbol{\omega}^\theta\}$.]

In the coordinate basis, it is easily seen that the directional derivative is given by

$$\partial_{\boldsymbol{v}} = v^\alpha \frac{\partial}{\partial x^\alpha}. \tag{10}$$

III. DESCRIPTION OF SOME COMMON NOTIONS FROM SPECIAL RELATIVITY

Before we do anything complicated, we'll highlight a few simple applications of the theory. The following notions are probably familiar from special relativity, but we can give their description using the powerful new machinery we've learned. We will highlight for each concept what modifications will be required when we go to curved spacetime.

A. Proper time

Define the *world line* of a particle to be the trajectory it takes through spacetime. In general one might describe this trajectory in parametric form: $\mathcal{P}(\lambda)$, or $x^\alpha(\lambda)$ if you like coordinates. But there are an infinity of ways to choose the parameter λ. One special and often convenient choice, though, is the *proper time* τ. This is defined by the requirement

$$\frac{d\mathcal{P}}{d\tau} \cdot \frac{d\mathcal{P}}{d\tau} = -1. \tag{11}$$

One may convert any parameterization to proper time by using the conversion

$$\frac{d\tau}{d\lambda} = \sqrt{-\frac{d\mathcal{P}}{d\lambda} \cdot \frac{d\mathcal{P}}{d\lambda}}, \tag{12}$$

so long as the curve is timelike ($d\mathcal{P}/d\lambda$ has negative square norm). So proper time is an appropriate parameters for massive particles, which as we will see follow along timelike trajectories. It is not useful for massless particles such as photons, which follow trajectories where $d\mathcal{P}/d\lambda$ has zero square norm.

One notes that in special relativity, where $g_{\mu\nu} = \eta_{\mu\nu}$, we could have used the coordinate time $t = x^0$ as the initial parameter λ. Then the proper time satisfies

$$\frac{d\tau}{dt} = \sqrt{-\eta_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}} = \sqrt{1 - \left(\frac{dx^1}{dt}\right)^2 - \left(\frac{dx^2}{dt}\right)^2 - \left(\frac{dx^3}{dt}\right)^2} \equiv \gamma^{-1}. \tag{13}$$

Note that (i) proper time passes at a slower rate than coordinate time; and (ii) the timelike requirement forces the 3-dimensional spatial vector

$$^{(3)}\boldsymbol{v} = \frac{dx^i}{dt}\boldsymbol{e}_i \tag{14}$$

to have norm less than unity ($< c$). Equation (13) also defines the *Lorentz factor* γ,

$$\gamma = \frac{1}{\sqrt{1 - |^{(3)}\boldsymbol{v}|^2}}. \tag{15}$$

B. 4-velocity

In ordinary mechanics, the 3-velocity of an object is its spatial displacement per unit time – this is $^{(3)}\boldsymbol{v}$. This is not a Lorentz-invariant object. A Lorentz-invariant velocity $\boldsymbol{v}$ can be constructed – this is the 4-velocity $\boldsymbol{u}$:

$$\boldsymbol{u} \equiv \frac{d\mathcal{P}}{d\tau}. \tag{16}$$

By construction, $\boldsymbol{u} \cdot \boldsymbol{u} = -1$. The components of $\boldsymbol{u}$ can be found from

$$u^\alpha = \frac{dx^\alpha}{d\tau} = \frac{dx^\alpha}{dt}\frac{dt}{d\tau} = \frac{dx^\alpha}{dt}\gamma. \tag{17}$$

Therefore we have

$$u^0 = \gamma, \quad \text{and} \quad u^i = {}^{(3)}v^i\gamma. \tag{18}$$

C. 4-momentum

The 4-momentum vector of a massive object is defined as

$$\boldsymbol{p} = m\boldsymbol{u}, \tag{19}$$

where m is the mass. It follows that

$$\boldsymbol{p} \cdot \boldsymbol{p} = -m^2. \tag{20}$$

Thus the 4-momentum is always timelike for a massive particle.

Equation (20) suggests that the 4-momentum remains defined even for a massless particle ($m^2 = 0$), for which it has zero square norm. One might even imagine particles with negative m^2 (*tachyons*) but no such particle has been found and confirmed.

It is found experimentally that the total 4-momentum is conserved (e.g. in a particle collision). This is really a single conservation law with 4 components. But in everyday experience the time and spatial components of this law manifest themselves so differently that they have separate names. The time component of 4-momentum (p^0) is called *energy* and the spatial components p^i are the components of *3-momentum*.

In GR, the inability to construct a single global coordinate system will challenge this split notion of conservation of 4-momentum: instead we will find a local conservation law ($\nabla_\mu T^{\mu\nu} = 0$). We will have to be very careful when even defining such a thing as the total "mass" of e.g. a planet, star, or black hole. But describing the 4-momentum $\boldsymbol{p}$ of an elementary particle at a particular point on its world line will present no difficulty.

D. Force and acceleration

In either Newtonian mechanics or special relativity, a particle need not have constant velocity: forces may act on it. So we will take the derivative of the velocity of a massive particle to obtain the *4-acceleration*:

$$\boldsymbol{a} = \frac{d\boldsymbol{v}}{d\tau}, \quad \text{or (in components)} \quad a^\mu = \frac{dv^\mu}{d\tau}. \tag{21}$$

One can then define the *4-force* on the particle to be $d\boldsymbol{p}/d\tau$. This equals $m\boldsymbol{a}$ if the mass is constant.

It is always the case that 4-acceleration and 4-velocity are orthogonal:

$$\boldsymbol{a} \cdot \boldsymbol{v} = \eta_{\mu\nu} \frac{dv^\mu}{d\tau} v^\nu = \frac{1}{2} \frac{d}{d\tau} (\eta_{\mu\nu} v^\mu v^\mu) = \frac{1}{2} \frac{d}{d\tau} (-1) = 0. \tag{22}$$

So in the instantaneous rest frame of a particle, the 4-acceleration is purely spatial.

The operation of defining $\boldsymbol{a}$ – seemingly trivial and innocuous – will not make sense in curved spacetime without lots of extra work. To see why, let's write the derivative according to freshman calculus:

$$\boldsymbol{a} = \lim_{\epsilon \to 0} \frac{\boldsymbol{v}(\tau + \epsilon) - \boldsymbol{v}(\tau)}{\epsilon}. \tag{23}$$

The two vectors on the right hand side, $\boldsymbol{v}(\tau + \epsilon)$ and $\boldsymbol{v}(\tau)$, are located at different points in spacetime. In special relativity, this is no issue; we simply move the vectors to the same point and subtract. But in curved spacetime, a vector (e.g. a velocity) is defined at a point $\mathcal{P}$. The job of moving it to another point $\mathcal{Q}$ is nontrivial: imagine taking a 2-dimensional velocity vector tangent to the Earth's surface defined in Pasadena, and somehow transporting it along the globe to New York.

IV. TENSOR ALGEBRA

Now that we have built the framework of special relativity, it is time to return to our vectors and 1-forms. We have found that many useful physical concepts are vectors (velocity, momentum, acceleration, force, electric current), some are 1-forms (wave vector, electromagnetic vector potential), and there is a natural correspondence between the two. Our next task is to construct general tensors – linear functions that accept an ordered list of vectors and 1-forms and return a number. Examples of tensors will include the metric tensor, the electromagnetic field, the energy-momentum density-flux, the polarization density matrix of light, and (later) the curvature of spacetime. So here we will define tensors and then the basic operations that act on them. The algebraic operations will carry over trivially to curved spacetime, whereas the calculus operations (e.g. derivatives) will take more work.

A. Linear machines

So now we are ready to define a general tensor. Let us define a *tensor* $\mathbf{S}$ of rank $\binom{M}{N}$ as a linear operation on M 1-forms and N vectors that returns a scalar. That is, it is a function

$$\mathbf{S}(\tilde{\boldsymbol{k}}, ..., \tilde{\boldsymbol{l}}, \boldsymbol{u}, ..., \boldsymbol{v}) \tag{24}$$

that distributes over addition and scalar multiplication. We may define the components of a tensor by acting on the basis vectors,

$$S^{\alpha_1...\alpha_M}{}_{\beta_1...\beta_N} = \mathbf{S}(\boldsymbol{\omega}^{\alpha_1}, ..., \boldsymbol{\omega}^{\alpha_M}, \boldsymbol{e}_{\beta_1}, ..., \boldsymbol{e}_{\beta_N}). \tag{25}$$

Repeated use of the distributive property then tells us that

$$\mathbf{S}(\tilde{\boldsymbol{k}}, ..., \tilde{\boldsymbol{l}}, \boldsymbol{u}, ..., \boldsymbol{v}) = S^{\alpha_1...\alpha_M}{}_{\beta_1...\beta_N} \tilde{k}_{\alpha_1}...\tilde{l}_{\alpha_M} u^{\beta_1}...v^{\beta_N} \tag{26}$$

(EΣC implied as usual). The components of the tensor transform as

$$S^{\alpha'_1...\alpha'_M}{}_{\beta'_1...\beta'_N} = [\mathbf{L}^{-1}]^{\alpha'_1}{}_{\gamma_1}...[\mathbf{L}^{-1}]^{\alpha'_M}{}_{\gamma_M} L^{\delta_1}{}_{\beta'_1}...L^{\delta_N}{}_{\beta'_N} S^{\gamma_1...\gamma_M}{}_{\delta_1...\delta_N}, \tag{27}$$

i.e. one transforms each index as appropriate for a vector or 1-form.

B.   Examples

We have already seen several examples of tensors. These include:

- The metric tensor $\mathbf{g}$ is a tensor of rank $\binom{0}{2}$: it acts on 2 vectors.

- A vector $\boldsymbol{v}$ is actually a tensor of rank $\binom{1}{0}$: it takes in any 1-form $\tilde{\boldsymbol{k}}$ and returns the scalar $\langle\tilde{\boldsymbol{k}},\boldsymbol{v}\rangle$. The components of $\boldsymbol{v}$ are $\langle\boldsymbol{\omega}^\alpha,\boldsymbol{v}\rangle$ which are simply v^α.

- A 1-form $\tilde{\boldsymbol{k}}$ is actually a tensor of rank $\binom{0}{1}$: it takes in any vector $\boldsymbol{v}$ and returns the scalar $\langle\tilde{\boldsymbol{k}},\boldsymbol{v}\rangle$. Its components are $\tilde{k}_\alpha$.

C.   Raising and lowering indices

A tensor of rank $\binom{M}{N}$ can exist without reference to the metric or dot product. But if we endow the Universe with a metric, we know that there is a correspondence of 1-forms to vectors. So it follows that a tensor $\mathbf{S}$ should be able to accept either vectors or 1-forms as arguments in each slot. So really if there is a metric we need only think about the total rank $M + N$ of a tensor. There are then 2^{M+N} different forms of the tensor depending on which slots take 1-forms and which take vectors. These slots are related to each other by the operation of raising or lowering an index.

Consider for definiteness transforming from the $\binom{0}{3}$ to $\binom{1}{2}$ form of S, i.e. given $S_{\delta\epsilon\zeta}$ we wish to find $S^\alpha{}_{\beta\gamma}$. This is:

$$S^\alpha{}_{\beta\gamma} = \mathbf{S}(\boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta, \boldsymbol{e}_\gamma). \tag{28}$$

Now remember that the vector $\boldsymbol{v}$ associated with $\boldsymbol{\omega}^\alpha$ can be obtained by using the inverse-metric: its components are $v^\delta = g^{\delta\alpha}$, and the vector itself is $\boldsymbol{v} = g^{\delta\alpha}\boldsymbol{e}_\delta$. [Recall that the definition of the vector associated with a 1-form was that $\boldsymbol{v} \cdot \boldsymbol{w} = \langle\boldsymbol{\omega}^\alpha, \boldsymbol{w}\rangle$ for any vector $\boldsymbol{w} \in \mathcal{V}$.] So we will write:

$$S^\alpha{}_{\beta\gamma} = \mathbf{S}(g^{\delta\alpha}\boldsymbol{e}_\delta, \boldsymbol{e}_\beta, \boldsymbol{e}_\gamma) = g^{\delta\alpha}\mathbf{S}(\boldsymbol{e}_\delta, \boldsymbol{e}_\beta, \boldsymbol{e}_\gamma) = g^{\delta\alpha}S_{\delta\beta\gamma}. \tag{29}$$

So we may find the components of the $\binom{M+1}{N-1}$ representation of an $\binom{M}{N}$ tensor using the inverse-metric (i.e. with upper indices).

It is similarly possible to go the other way: all we need to know is that $g_{\delta\epsilon}$ is the matrix inverse of $g^{\delta\alpha}$. Therefore,

$$S_{\epsilon\beta\gamma} = g_{\delta\epsilon}g^{\delta\alpha}S_{\delta\beta\gamma} = g_{\delta\epsilon}S^\alpha{}_{\beta\gamma}. \tag{30}$$

So we have seen that just as there is a natural correspondence of vectors and 1-forms if we are given a metric, so there is also a correspondence of second rank tensors of all types, $\binom{0}{2}$, $\binom{1}{1}$, and $\binom{2}{0}$. Expressions involving the components with up and down indices allow us to indicate which version of the tensor we are talking about, but if there is a metric all these forms are equivalent. The computational rules associated with them are simple: they are generalizations of matrix multiplication by the metric tensor $\mathbf{g}$ in either direct $(g_{\mu\nu})$ or inverse $(g^{\mu\nu})$ form.