Physics in flat space-time: Special relativity

Asaf Pe'er[1]

September 18, 2017

This part of the course is based on Refs. [1] and [2].

1. Introduction

We begin by a review of special theory of relativity. The two goals are (1) to remind the basics of special relativity (SR), and (2) to introduce vectors and tensors, which are going to be crucial later on - without the complexity of curved space time. **I strongly urge** the students to read my lecture notes on special relativity for first year mechanics course (PY1052), that can be found on my webpage. This is essentially similar to what will be discussed here, but using much less confusing terms.

The principle of special relativity states that **the laws of nature are invariant in all inertial reference frames**. The experimental basis of this theory is the *Michelson-Morley experiment*, which shows that **the speed of light is constant and equals to c in every inertial frame**. This result is in contradiction to the *Galilean* transformation law of velocities, $\vec{V}' = \vec{V} + \vec{u}$, where $\vec{V}$ is the velocity of an object as seen by an observer in the frame $\mathcal{O}$, $\vec{V}'$ is its velocity in frame $\mathcal{O}'$, and $\vec{u}$ is the relative velocity between frames $\mathcal{O}$ and $\mathcal{O}'$ (you can think of one observer - in frame $\mathcal{O}$ being on a platform, while the second, in frame $\mathcal{O}'$ being in a moving train).

Consider two observers, denoted by $\mathcal{O}$ and $\mathcal{O}'$. Assume that the observer $\mathcal{O}'$ is moving (at constant velocity) with respect to the observer $\mathcal{O}$, so that both frames are inertial. We denote by x, y, z, t the (space-time) coordinate system of observer $\mathcal{O}$, and by x', y', z', t' the coordinate system of observer $\mathcal{O}'$. We further assume that the coordinate origins, $\mathcal{O}$ and $\mathcal{O}'$ coincide at $t = t' = 0$.

Consider a light source located at the origin in the frame $\mathcal{O}$ that emits light at $t = 0$. At time t, the spherical wave front of the light is at location

$$x^2 + y^2 + z^2 = c^2t^2. \tag{1}$$

A similar analysis holds for the observer $\mathcal{O}'$, thus

$$x'^2 + y'^2 + z'^2 = c^2t'^2. \tag{2}$$

---

[1]Physics Dep., University College Cork

We can assume (without loss of generality), that $\mathcal{O}'$ moves with respect to $\mathcal{O}$ in the $\hat{x}$ direction. Thus, $x' \neq x$, while $y' = y$ and $z' = z$. Equations 1, 2 thus imply that $t \neq t'$. Thus, **time separation is not the same in both frames**. In other words, *time is not an* **invariant** *quantity.*

While the time separation between two events (e.g.: emission of photon and detection of the photon) is not invariant, Equations 1 and 2 imply that one can define an **invariant** quantity, which is called the **interval** between two events. The interval is defined by

$$(interval)^2 = (\Delta s)^2 \equiv -(c\Delta t)^2 + (\Delta r)^2 = -c^2(\Delta t^2) + \left((\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\right), \quad (3)$$

where Δt is the time interval between two events, and Δr is the space interval between the same events.

Note that there is a very close analogy between the *interval* and the concept of *distance* familiar in 2 (or 3)-dimensional Euclidean space. In 2-dimensional space (see Figure 1), one can define the Cartesian coordinate system to be x, y. The distance between two points is given by $(\Delta s)^2 = (\Delta x)^2 + (\Delta y)^2$. Consider now a different (Cartesian) coordinate system, defined by x', y' axes rotated with respect to the originals. The formula for the distance in this new coordinate system is unchanged, $(\Delta s)^2 = (\Delta x')^2 + (\Delta y')^2$. We thus say that the distance is **invariant** under rotation of the coordinate system.

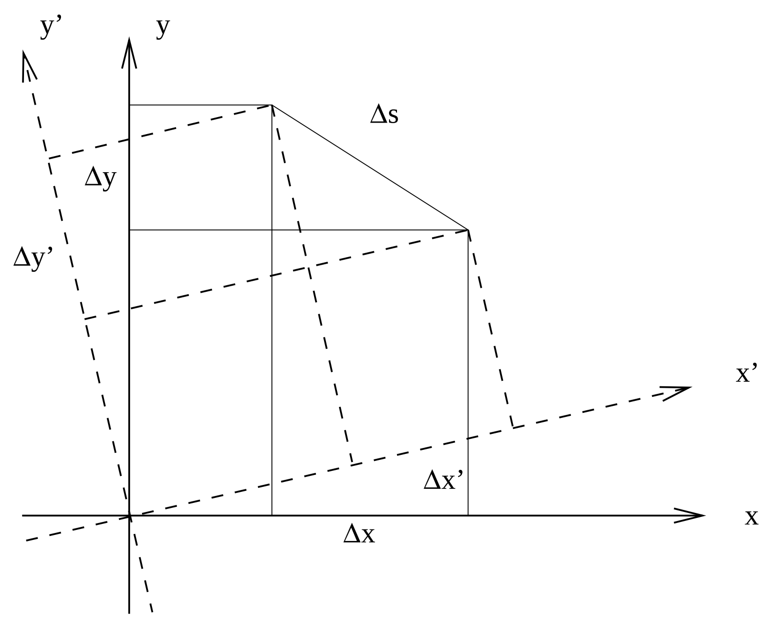

Fig. 1.— The distance Δs between two points in 2-dimensional Euclidean space is *invariant* under rotation of the coordinate system.

2. Lorentz transformations

The **interval** introduced above connects the space and time into a 4-dimensional space-time, known as **Minkowski space**. In this space, one can define an **event** by its time and space coordinates, uniquely determined by (t, x, y, z). Note that 4 quantities are required to fully determine the space-time location of every event. It is more convenient (yet, confusing !) to denote coordinates on 4-dimensional space-time by letters with Greek superscript indices running from 0 to 3, with 0 generally denoting the time coordinate. Thus any event is defined by a 4 vector of coordinates

$$x^\mu : \quad \begin{array}{l} x^0 = ct \\ x^1 = x \\ x^2 = y \\ x^3 = z \end{array} . \tag{4}$$

If we need to refer to space coordinates only, we will use Latin superscripts to stand for the space components alone:

$$x^i : \quad \begin{array}{l} x^1 = x \\ x^2 = y \\ x^3 = z \end{array} \tag{5}$$

Before we proceed, it is also convenient to write the spacetime interval in a more compact form. We introduce a 4×4 matrix, known as the **metric**, which we write using two lower indices:

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{6}$$

(Some references, especially field theory books, define the metric with the opposite sign, so be careful.) Using this new notation, we can write the interval (Equation 3) in a compact form,

$$(\Delta s)^2 = \sum_{\mu\nu=0...3} \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu. \tag{7}$$

In the last equality we used the **Einstein summation convention**, according to which indices which appear both as superscripts and subscripts are summed over. We will use this convention from here on. If you find it confusing, simply add the summation sign at the beginning of the equation. Equation 7 is thus identical to Equation 3, simply written in a more compact form.

A Lorentz transformation can now be defined as *a transformation from one coordinate system to another, that conserves the interval.* One example of such a transformation is *translation*, which is a shift of the coordinates by constants:

$$x^{\mu} \rightarrow x^{\mu'} = x^{\mu} + a^{\mu} \ , \tag{8}$$

where a^{μ} is a set of four fixed numbers. (Notice that we put the prime on the index, not on the x: because in our notation, for example, $y' \equiv (x^2)'$.) Thus, translations leave the difference Δx^{μ} unchanged, so clearly the interval is unchanged. Note that translations represent a simple shift in the coordinates origin.

The only other kind of linear transformation is to multiply the 4-vector x^{μ} by a (spacetime-independent, 4×4) matrix:

$$x^{\mu'} = \Lambda^{\mu'}{}_{\nu} x^{\nu} \ . \tag{9}$$

Equation 9 can be written in a longer form,

$$x^{\mu'} = \Lambda^{\mu'}{}_{0} x^0 + \Lambda^{\mu'}{}_{1} x^1 + \Lambda^{\mu'}{}_{2} x^2 + \Lambda^{\mu'}{}_{3} x^3 \tag{10}$$

or, in more conventional matrix notation,

$$x' = \Lambda x \ . \tag{11}$$

Let us see what conditions should the elements of the matrix Λ fulfill in order to conserve the interval:

$$\begin{aligned} (\Delta s')^2 &= \eta_{\mu'\nu'} \Delta x^{\mu'} \Delta x^{\nu'} \\ &= \eta_{\mu'\nu'} \Lambda^{\mu'}{}_{\rho} \Delta x^{\rho} \Lambda^{\nu'}{}_{\sigma} \Delta x^{\sigma} \\ &\stackrel{!}{=} \eta_{\rho\sigma} \Delta x^{\rho} \Delta x^{\sigma} = (\Delta s)^2 \end{aligned} \tag{12}$$

In deriving Equation 12, please note the following:

- We used the summation convention to sum over all indices that appear both as superscripts and subscripts. These indices are thus **dummy indices**. While we operate a multiplication between matrix and vectors, since we make the multiplication one term at a time, we are allowed to switch the order of the elements in the transition between the 2nd and 3rd line.
- The symbol $\stackrel{!}{=}$ implies "we require this equality".

The interval is thus conserved (i.e., Equation 12 is fulfilled) if the elements of the matrix Λ fulfill

$$\eta_{\rho\sigma} = \Lambda^{\mu'}{}_{\rho} \Lambda^{\nu'}{}_{\sigma} \eta_{\mu'\nu'} \ . \tag{13}$$

The matrices which satisfy Equation 13 are known as the **Lorentz transformations**; the set of them forms a group under matrix multiplication, known as the **Lorentz group**. There is a close analogy between this group and O(3), the rotation group in three-dimensional space.

Lorentz transformations fall into a number of categories. First there are the conventional **rotations**. One such example is a rotation in the x-y plane:

$$\Lambda^{\mu'}{}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{14}$$

The rotation angle θ is a periodic variable with period 2π. By direct summation, it is easy to show that Equation 13 is fulfilled. For example, if we choose $\rho = 1, \sigma = 1$, we can write the only non-zero components of the summation over μ', ν' as

$$1 = \eta_{11} = \Lambda^{\mu'}{}_{1}\Lambda^{\nu'}{}_{1}\eta_{\mu'\nu'} = \Lambda^{1}{}_{1}\Lambda^{1}{}_{1}\eta_{11} + \Lambda^{2}{}_{1}\Lambda^{2}{}_{1}\eta_{22} = \cos^2\theta + \sin^2\theta, \tag{15}$$

and similarly for every choice of ρ, σ (total 16 choices = 16 equations !). This is really identical to rotations in 3-d Euclidean space, which conserves the distance.

Another group of transformations are called **boosts**, which may be thought of as "rotations between space and time directions." As an example, consider a rotation in the t-x coordinates,

$$\Lambda^{\mu'}{}_{\nu} = \begin{pmatrix} \cosh\phi & -\sinh\phi & 0 & 0 \\ -\sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{16}$$

While the format of a boost transformation is very similar to the format of Equation 14, since the time coordinate comes with a minus (-) sign, we replace (sin) and (cos) with (sinh) and (cosh) and changed the sign of the (t, x) component. The boost parameter ϕ, unlike the rotation angle, is defined from $-\infty$ to ∞. Again, by direct substitution it is easy to show that the boost matrices fulfill the condition in Equation 13. For example, if $\rho = 1, \sigma = 1$ we find

$$1 = \eta_{11} = \Lambda^{\mu'}{}_{1}\Lambda^{\nu'}{}_{1}\eta_{\mu'\nu'} = \Lambda^{0}{}_{1}\Lambda^{0}{}_{1}\eta_{00} + \Lambda^{1}{}_{1}\Lambda^{1}{}_{1}\eta_{11} = -\sinh^2\phi + \cosh^2\phi, \tag{17}$$

which is an algebraic identity.

Additional transformations include discrete transformations which reverse the time direction or one or more of the spatial directions. A general transformation can be obtained by multiplying the individual transformations; the explicit expression for this six-parameter matrix (three boosts, three rotations) is not sufficiently pretty or useful to bother writing down. In general Lorentz transformations will not commute, so the Lorentz group is called **non-abelian** (similar to rotations !). The set of both translations and Lorentz transformations is a ten-parameter non-abelian group, known in mathematics as the **Poincaré group**.

Equation 16 can be put in a more familiar form by multiplying the Lorentz boost $\Lambda^{\mu'}{}_{\nu}$ with the four-vector x^ν and writing the result explicitly,

$$\begin{aligned} ct' &= ct\cosh\phi - x\sinh\phi \\ x' &= -ct\sinh\phi + x\cosh\phi\ . \end{aligned} \tag{18}$$

One thus find that the velocity of the point for which $x' = 0$ is

$$\beta \equiv \frac{v}{c} = \frac{x}{ct} = \frac{\sinh\phi}{\cosh\phi} = \tanh\phi\ . \tag{19}$$

Using the algebraic identities $1 - \tanh^2\phi = (\cosh^2\phi)^{-1}$ and $\sinh^2\phi = \cosh^2\phi - 1$, one can write $\cosh\phi = (1 - \tanh^2\phi)^{-1/2} = (1-\beta^2)^{-1/2} \equiv \gamma$, and $\sinh\phi = (\gamma^2 - 1)^{1/2} = \gamma\beta$. Equation 18 takes the familiar form,

$$\begin{aligned} ct' &= \gamma(ct - \beta x) \\ x' &= \gamma(x - \beta ct) \end{aligned} \tag{20}$$

This seemingly very abstract approach thus retrieves the same familiar results, although in a much more general form.

2.1. Time dilation

Consider an observer in frame $\mathcal{O}$ at rest, holding a clock. In her frame, two ticks of the clock are separated by $\Delta x = \Delta y = \Delta z = 0$, and $\Delta t = dt$, where dt is the nominal period between two ticks intended by the manufacturer. This is known as the **proper time**, often denoted by $d\tau$. In fact, according to our definition of the interval in Equation 3, we can define

$$(\Delta\tau)^2 \equiv -\frac{(\Delta s)^2}{c^2} = -\frac{1}{c^2}\eta_{\mu\nu}\Delta x^\mu \Delta x^\nu. \tag{21}$$

A second observer, O', sees the same clock moving at velocity $\vec{v}$. He will observe the two ticks separated by a time interval $\Delta t'$ and space interval $\Delta\vec{x}' = \vec{v}\Delta t'$. Thus, he will see the proper time

$$\Delta\tau' = \left(\Delta t'^2 - \frac{\Delta x'^2}{c^2}\right)^{1/2} = (1-\beta^2)^{1/2}\Delta t'. \tag{22}$$

Assuming both observers are at inertial coordinate systems, their coordinate systems are related via Lorentz transformations, and thus the interval ΔS hence the proper time $\Delta\tau$ are invariant, $\Delta\tau = \Delta\tau'$. Thus, the observer at O', that sees the clock in motion, sees it ticking with a period

$$\Delta t' = (1 - \beta^2)^{-1/2}\Delta t = \gamma dt. \tag{23}$$

3. Vectors and Tensors (part I)

Before we can proceed to discuss physics on Minkowski space-time, we need to get a more precise definition of the concept of vectors and tensors. While we do need to develop the full formalism before we can fully discuss the general theory of relativity, we will focus in this section only on the basic definitions needed for working in flat space-time. However, we will do it in a general way so that the extension to GR , which will be done in the following chapters, will be a natural one.

3.1. Four vectors

We are used to think of vectors as directed line segments in (3-d) Euclidean space. Of course, in four-dimensional space-time the vectors must be four-dimensional, and are thus referred to as **four-vectors**. We will keep this terminology at every point where there may be a confusion between four-vectors and "regular" 3-d vectors.

In §2 above, we have already used the concept of four-vector by introducing the space-time coordinates $(ct, x, y, z) \equiv x^\mu$. As we saw, the coordinate four-vector x^μ transforms under Lorentz transformation according to Equation 9. We can in fact use Equation 9 to *define* a four-vector, as any quantity **whose components** undergo the transformation

$$V^\alpha \to V^{\alpha'} = \Lambda^{\alpha'}{}_\beta V^\beta \ , \tag{24}$$

when the coordinate system is transformed by

$$x^\alpha \to x^{\alpha'} = \Lambda^{\alpha'}{}_\beta x^\beta \ . \tag{25}$$

In fact, V^α is called **contra-variant four-vector**, from reasons which will be discussed shortly.

Beyond the simple fact of dimensionality, the most important thing to emphasize is that **each vector is located at a given (single) point in spacetime**. This fact is going to turn out crucial when dealing with non-flat (curved) space-time, where the concept of "direction" is not easily defined. You may be used to thinking of vectors as stretching from one point to another in space, possibly even of "free" vectors which you can slide carelessly from point to point. These are not useful concepts in relativity.

Rather, to each point p in spacetime we associate the set of all possible vectors located at that point; this set is known as the **tangent space** at p, or T_p. Consider a point p in a curved space-time (see Figure 2). We can think of the set of all vectors attached to that point as comprising a plane which is tangent to the point (hence the name "tangent space"). The important thing here is to remember that although we draw vectors as arrows (and will continue to do so), all these vectors are attached to a single point, p.

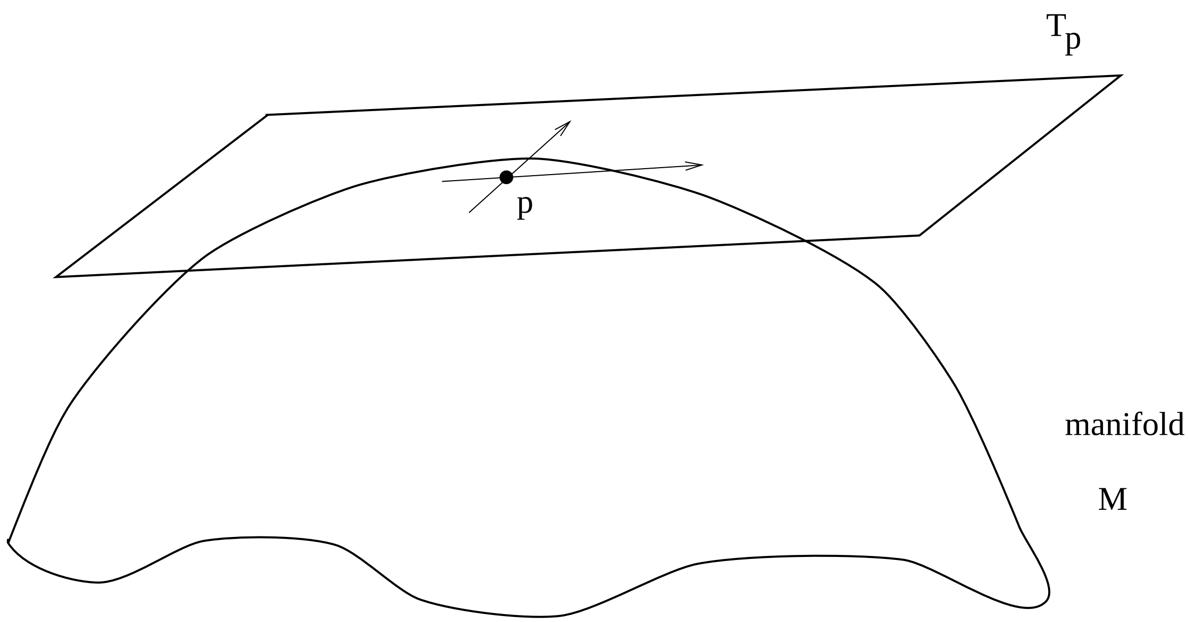

Fig. 2.— At every point p in four-dimensions space-time there is a set of (contra-variant) vectors, defined by Equation 24. The set of all these vectors form a tangent vector space, T_p.

As every point p in space-time, we can thus construct a tangent space T_p, which can be thought of as an abstract **vector space**. This concept should be familiar to you from linear algebra course. Roughly speaking, a vector space is a collection of objects (vectors) which can be added together and be multiplied by scalars (real numbers) in a linear way. Thus, for any two vectors V and W and real numbers a and b, we have

$$(a+b)(V+W) = aV + bV + aW + bW \ . \tag{26}$$

Every vector space has an origin, *i.e.* a zero vector which functions as an identity element under vector addition ($V + 0 = V$). In many vector spaces there are additional operations such as taking an inner (dot) product, but this is extra structure over and above the elementary concept of a vector space.

3.2. Basis four-vector and components

It is often useful to decompose vectors into components with respect to some set of basis vectors. A **basis** is any set of vectors which both:

- spans the vector space (any vector is a linear combination of basis vectors) and
- is linearly independent (no vector in the basis is a linear combination of other basis vectors).

For any given vector space, there will be an infinite number of legitimate bases, but each basis will consist of the same number of vectors, known as the *dimension of the space*. (For a tangent space associated with a point in Minkowski space, the dimension is of course four.) Note also that **a vector is perfectly defined by itself, regardless of its decomposition into components with respect to any arbitrary basis we may choose**; the use of basis vectors simply turns out to be very useful (you can think of the distance Δs presented in Figure 1 as an example - while its components Δx and Δy depend on the basis vectors, the distance itself is not).

Consider again a tangent space T_p at a point p. We can set up a basis of four vectors $\hat{e}_{(\mu)}$, with $\mu = \{ct, x, y, z\}$ or equivalently $\mu = \{0, 1, 2, 3\}$ (in a more precise format, we write $\mu \in \{0, 1, 2, 3\}$). For simplicity, we take the basis vectors to be aligned with the normal coordinates, x^μ. Thus, we take the basis vector $\hat{e}_{(1)}$ along the x-direction, etc. **Note the subtlety of notation**: $\hat{e}_{(\mu)}$ represent four different (basis) vecotrs, not the four components of a single vector !. Hence the parenthesis in $\hat{e}_{(\mu)}$.

By definition, any vector A can be written as a linear combination of basis vectors:

$$A = A^\mu \hat{e}_{(\mu)} \; = \left(\sum_{\mu=0}^{3} A^\mu \hat{e}_{(\mu)} \right) \tag{27}$$

the coefficients A^μ are the **components** of the vector A in the basis $\hat{e}_{(\mu)}$; clearly, in a different basis the same vector A will have different components (see again Figure 1). When the basis is changed, the transformation law of the components are given by Equation 24.

Example. A standard example of a vector in spacetime is the tangent vector to a curve (see Figure 2). Consider a general curve (or path) through space time. This curve is specified by the coordinates as a function of a free parameter, *e.g.* $x^\mu(\lambda)$. The tangent vector $V(\lambda)$ has components

$$V^\mu = \frac{dx^\mu}{d\lambda} \ . \tag{28}$$

The entire vector is thus $V = V^\mu \hat{e}_{(\mu)}$. Under a Lorentz transformation the coordinates x^μ change according to Equation 9 (or 25), while the parameterization λ is unaltered; thus the coordinates of the vector must change according to Equation 24.

Before proceeding, let us pay a careful attention to a very delicate point. In §2 above, we introduced the four-vector of coordinates label by upper indices (Equation 4), which transformed in a certain way under Lorentz transformations (Equation 9). Here, we introduced a general four vector, whose elements transform in exactly the same way under Lorentz transformation (Equations 24, 25). We thus wrote its components with upper indices as well. However, as we wanted to use the summation rule, we wrote the basis vectors $\hat{e}_{(\mu)}$, associated with a certain coordinate system that we chose with *lower* indices. This notation ensured that the invariant object constructed by summing over the components and basis vectors was left unchanged by the transformation, just as we would wish (see again Figure 1). The transformation law of the basis vectors thus **can not** be the same as those of the vector components.

Note again the subtlety: when discussing the "transformation of basis vectors" we mean transforming from one set of basis vectors to another set composed of (four) different vectors; while when we talk about "transformation law of vector component", we refer to the same vector as is being described from a different basis.

3.3. Transformation law of the basis vectors

Consider a vector V, whose components in a given basis $\hat{e}_{(\mu)}$ are V^μ, namely $V = \sum_{\mu=0}^{3} V^\mu \hat{e}_{(\mu)}$. The vector itself (as opposed to its components in some coordinate system) being an abstract geometrical entity is invariant under Lorentz transformations. We can use this fact to derive the transformation properties of the basis vectors.

Let us refer to the set of basis vectors in the transformed coordinate system as $\hat{e}_{(\nu')}$. Since the vector is invariant, we have

$$\begin{aligned} V = V^\mu \hat{e}_{(\mu)} &= V^{\nu'} \hat{e}_{(\nu')} \\ &= \Lambda^{\nu'}{}_{\mu} V^\mu \hat{e}_{(\nu')} \ . \end{aligned} \tag{29}$$

But this relation must hold no matter what the numerical values of the components V^μ are.

Therefore we can say

$$\hat{e}_{(\mu)} = \Lambda^{\nu'}{}_{\mu} \hat{e}_{(\nu')} \ . \tag{30}$$

To get the new basis $\hat{e}_{(\nu')}$ in terms of the old one $\hat{e}_{(\mu)}$ we should multiply by the inverse of the Lorentz transformation $\Lambda^{\nu'}{}_{\mu}$. But the inverse of a Lorentz transformation from the unprimed to the primed coordinates is also a Lorentz transformation, this time from the primed to the unprimed systems. We will therefore introduce a somewhat subtle notation, by writing using the same symbol for both matrices, just with primed and unprimed indices adjusted. That is,

$$(\Lambda^{-1})^{\nu'}{}_{\mu} = \Lambda_{\nu'}{}^{\mu} \ , \tag{31}$$

so that we can write the transformation rule for basis vectors:

$$\hat{e}_{(\nu')} = \Lambda_{\nu'}{}^{\mu} \hat{e}_{(\mu)} \ . \tag{32}$$

Therefore the set of basis vectors transforms via the inverse Lorentz transformation of the coordinates or vector components.

Equation 31 can also be written as

$$\Lambda_{\nu'}{}^{\mu} \Lambda^{\sigma'}{}_{\mu} = \delta^{\sigma'}_{\nu'} \ , \quad \Lambda_{\nu'}{}^{\mu} \Lambda^{\nu'}{}_{\rho} = \delta^{\mu}_{\rho} \ , \tag{33}$$

where δ^{μ}_{ρ} is the traditional Kronecker delta symbol in four dimensions. (Note that *Schutz* uses a different convention, always arranging the two indices northwest/southeast; the important thing is where the primes go.)

This explains the origin of the term "contra-variant" vector; the transformation law of the components of a contra-variant vector are opposite (contra) to the transformation law of the basis vectors.

3.4. Covariant vectors and the dual space

We have seen that the transformation law of the components of a contra-variant vector is different than the transformation law of the basis vectors associated with a given coordinate system. We thus labeled the basis vectors with lower indices, and showed that they transform via the inverse matrix. This notation ensured that the invariant object constructed by summing over the components and basis vectors was left unchanged by the transformation.

But we can think of other objects whose transformation law will be the same (rather than the opposite) to the transformation law of the basis vectors. Once we set up a vector space, we can always define another associated vector space, of equal dimension, known as the **dual vector space**. If T_p is a tangent vector space, the dual vector space to it, known as the **cotangent space** is normally denoted by T_p^*.

The dual vector space (or dual space, for short), consists of all linear maps from the original vector space to the real numbers. In math lingo, if $\omega \in T_p^*$ is a dual vector, then it acts as a map such that:

$$\omega(aV + bW) = a\omega(V) + b\omega(W) \in \mathbf{R} \ , \tag{34}$$

where V, W are vectors and a, b are real numbers. Interestingly, these maps form a vector space themselves; thus, if ω and η are dual vectors, we have

$$(a\omega + b\eta)(V) = a\omega(V) + b\eta(V) \ . \tag{35}$$

We can now introduce a set of basis dual vectors, which we will denote by $\hat{\theta}^{(\nu)}$. The requirement from these vectors to form a basis is

$$\hat{\theta}^{(\nu)}(\hat{e}_{(\mu)}) = \delta^\nu_\mu \ . \tag{36}$$

Thus, similarly to the "normal" vectors, we can write every dual vector in terms of its components, which now will be labeled with lower indices:

$$\omega = \omega_\mu \hat{\theta}^{(\mu)} \ . \tag{37}$$

Elements of the dual space T_p^* are called **covariant vectors**. (Again, the 'co-' originates from the fact that the transformation law of their components is similar to the transformation law of the basis vectors). Another name for dual vectors is **one-forms**, although we will likely not have time to discuss it any further.

The component notation leads to a simple way of writing the action of a dual vector on a vector:

$$\begin{aligned} \omega(V) &= \omega_\mu V^\nu \hat{\theta}^{(\mu)}(\hat{e}_{(\nu)}) \\ &= \omega_\mu V^\nu \delta^\mu_\nu \\ &= \omega_\mu V^\mu \in \mathbf{R} \ . \end{aligned} \tag{38}$$

This is why it is rarely necessary to write the basis vectors (and basis dual vectors) explicitly; the components do all of the work. The form of Equation 38 also suggests that we can think of vectors as linear maps on dual vectors, by defining

$$V(\omega) \equiv \omega(V) = \omega_\mu V^\mu \ . \tag{39}$$

Therefore, the dual space to the dual vector space is the original vector space itself.

For completeness, we can derive the transformation properties of dual vectors in a similar way to the derivation for vectors. The transformation law of the components are given by

$$\omega_{\mu'} = \Lambda_{\mu'}{}^{\nu}\omega_{\nu} \ , \tag{40}$$

and for basis dual vectors,

$$\hat{\theta}^{(\rho')} = \Lambda^{\rho'}{}_{\sigma}\hat{\theta}^{(\sigma)} \ . \tag{41}$$

This is just what we would expect from index placement; the components of a dual vector transform under the inverse transformation of those of a vector. Note that this ensures that the scalar (Equation 38) is invariant under Lorentz transformations, just as it should be.

Examples.

(1) Maybe the easiest way to think of dual vectors is of row vectors, as opposed to "normal" vectors which are column vectors. Consider an n-dimensional space. We can write

$$V = \begin{pmatrix} V^1 \\ V^2 \\ \cdot \\ \cdot \\ \cdot \\ V^n \end{pmatrix} \ , \quad \omega = (\omega_1 \ \omega_2 \ \cdots \ \omega_n) \ . \tag{42}$$

The action of a dual vector on a vector (Equation 38) is ordinary multiplication,

$$\omega(V) = (\omega_1 \ \omega_2 \ \cdots \ \omega_n) \begin{pmatrix} V^1 \\ V^2 \\ \cdot \\ \cdot \\ \cdot \\ V^n \end{pmatrix} = \omega_i V^i \ . \tag{43}$$

(2) Another familiar example occurs in quantum mechanics, where vectors in the Hilbert space are represented by kets, $|\psi\rangle$. In this case the dual space is the space of bras, $\langle\phi|$, and the action gives the number $\langle\phi|\psi\rangle$. (This is a complex number in quantum mechanics, but the idea is precisely the same.)

(3) The **gradient** of a scalar function ϕ, denoted by $\mathrm{d}\phi$ is a set of partial derivatives with respect to the space-time coordinates,

$$\mathrm{d}\phi = \frac{\partial\phi}{\partial x^{\mu}}\hat{\theta}^{(\mu)} \ . \tag{44}$$

We can use the conventional chain rule often used to transform partial derivatives, to determine the transformation rule of dual vector components:

$$\begin{aligned} \frac{\partial\phi}{\partial x^{\mu'}} &= \frac{\partial x^{\mu}}{\partial x^{\mu'}}\frac{\partial\phi}{\partial x^{\mu}} \\ &= \Lambda_{\mu'}{}^{\mu}\frac{\partial\phi}{\partial x^{\mu}} \ , \end{aligned} \tag{45}$$

where we have used Equations 9 and 31 to relate the Lorentz transformation to the coordinates. The fact that the gradient is a dual vector leads to the following shorthand notations for partial derivatives:

$$\frac{\partial\phi}{\partial x^{\mu}} \equiv \partial_{\mu}\phi \equiv \phi,_{\mu} \ . \tag{46}$$

(Very roughly speaking, "while x^{μ} has an upper index, when it is in the denominator of a derivative it implies a lower index on the resulting object.")

Note that when we operate the dual vector of a gradient $\mathrm{d}\phi$ on a vector $V = V^{\mu}\hat{e}_{(\mu)}$ defined along a curve, $V^{\mu} = dx^{\mu}/d\lambda$, we obtain the natural result of an ordinary derivative of the function ϕ along the curve,

$$\partial_{\mu}\phi\frac{\partial x^{\mu}}{\partial\lambda} = \frac{d\phi}{d\lambda} \ . \tag{47}$$

3.5. Tensors

A straightforward generalization of vectors and dual vectors is the notion of a **tensor**. Formally, just as a dual vector is a linear map from vectors to $\mathbf{R}$, a tensor T of type (or rank) (k, l) is a multilinear map from a collection of (k) dual vectors and (l) vectors to $\mathbf{R}$. Multilinearity means that the tensor acts linearly in each of its arguments. From this point of view, a scalar is a type $(0, 0)$ tensor, a vector is a type $(1, 0)$ tensor, and a dual vector is a type $(0, 1)$ tensor.

Transformation law of tensors. Basically, a tensor of rank (k, l) have k contravariant indices and l covariant indices, with the corresponding Lorentz transformation properties. Thus, the transformation law of tensors is

$$T^{\mu'_1\cdots\mu'_k}{}_{\nu'_1\cdots\nu'_l} = \Lambda^{\mu'_1}{}_{\mu_1}\cdots\Lambda^{\mu'_k}{}_{\mu_k}\Lambda_{\nu'_1}{}^{\nu_1}\cdots\Lambda_{\nu'_l}{}^{\nu_l}T^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} \ . \tag{48}$$

In other words, each upper index gets transformed like a vector, and each lower index gets transformed like a dual vector.

Examples. An extremely important $(0, 2)$ tensor which we have already encountered - although without calling it in this name - is the **metric tensor**, $\eta_{\mu\nu}$ (see Equation 6). The action of the metric on two vectors is so useful that it gets its own name, the **inner product** (or scalar product):

$$\begin{aligned} \eta(V,W) &= \eta_{\mu\nu}V^\mu W^\nu = V\cdot W = W\cdot V \\ &= -V^0W^0 + V^xW^x + V^yW^y + V^zW^z\ , \end{aligned} \tag{49}$$

where the last equality in the first line originates from the symmetry properties of $\eta_{\mu\nu}$.

Just as with the conventional Euclidean dot product, we will refer to two vectors whose dot product vanishes as **orthogonal**. Since the dot product is a scalar, it is left invariant under Lorentz transformations; therefore the basis vectors of any Cartesian inertial frame, which are chosen to be orthogonal by definition, are still orthogonal after a Lorentz transformation.

The **norm** of a vector is defined to be inner product of the vector with itself:

$$V^2 \equiv \eta_{\mu\nu}V^\mu V^\nu = V_\nu V^\nu \tag{50}$$

Unlike in Euclidean space, this number is not positive definite:

$$\text{if}\quad \eta_{\mu\nu}V^\mu V^\nu \quad \text{is} \begin{cases} <0\ , & V^\mu \text{ is timelike} \\ =0\ , & V^\mu \text{ is lightlike or null} \\ >0\ , & V^\mu \text{ is spacelike}\ . \end{cases}$$

(A vector in Minkowski space can have zero norm without being the zero vector.) You will notice that the terminology is the same as that which we earlier used to classify the relationship between two points in spacetime; it's no accident, of course, and we will go into more detail later.

Clearly, the norm of a vector, being a scalar is Lorentz invariant. We can check this directly:

$$V'^2 \equiv \eta_{\mu'\nu'}V^{\mu'}V^{\nu'} = \eta_{\mu'\nu'}\Lambda^{\mu'}{}_\rho\Lambda^{\nu'}{}_\sigma V^\rho V^\sigma = \eta_{\rho\sigma}V^\rho V^\sigma = V^2 \tag{51}$$

In an identical way, the scalar product between two vectors is also conserved.

(2) Another tensor is the Kronecker delta δ^μ_ν, of type $(1,1)$, whose components are simply

$$\delta^\mu_\nu \equiv \begin{cases} 1\ , & \mu=\nu \\ 0\ , & \mu\neq\nu\ . \end{cases}$$

(3) The **inverse metric** tensor $\eta^{\mu\nu}$, is a type $(2,0)$ tensor defined as the inverse of the metric:

$$\eta^{\mu\nu}\eta_{\nu\rho} = \eta_{\rho\nu}\eta^{\nu\mu} = \delta^\mu_\rho\ . \tag{52}$$

In fact, as you can check, the inverse metric has exactly the same components as the metric itself. (This is only true in flat space-time in Cartesian coordinates, and will fail to hold in more general situations.)

(4) The **Levi-Civita tensor** is a $(0,4)$ tensor defined by

$$\epsilon_{\mu\nu\rho\sigma} = \begin{cases} +1 \text{ if } \mu\nu\rho\sigma \text{ is an even permutation of } 0123 \\ -1 \text{ if } \mu\nu\rho\sigma \text{ is an odd permutation of } 0123 \\ 0 \text{ otherwise .} \end{cases} \tag{53}$$

Here, a "permutation of 0123" is an ordering of the numbers 0, 1, 2, 3 which can be obtained by starting with 0123 and exchanging two of the digits; an even permutation is obtained by an even number of such exchanges, and an odd permutation is obtained by an odd number. Thus, for example, $\epsilon_{0321} = -1$.

3.5.1. Manipulating tensors

Raising and lowering indices. The example of the norm (Equation 50) presented above, is an example of the use of the metric tensor to lower indices. Indeed, both the metric and inverse metric can be used to **raise and lower indices** on tensors. That is, given a tensor $T^{\alpha\beta}{}_{\gamma\delta}$, we can use the metric to define new tensors which we choose to denote by the same letter T:

$$\begin{aligned} T^{\alpha\beta\mu}{}_{\delta} &= \eta^{\mu\gamma} T^{\alpha\beta}{}_{\gamma\delta} \ , \\ T_{\mu}{}^{\beta}{}_{\gamma\delta} &= \eta_{\mu\alpha} T^{\alpha\beta}{}_{\gamma\delta} \ , \\ T_{\mu\nu}{}^{\rho\sigma} &= \eta_{\mu\alpha}\eta_{\nu\beta}\eta^{\rho\gamma}\eta^{\sigma\delta} T^{\alpha\beta}{}_{\gamma\delta} \ , \end{aligned} \tag{54}$$

and so forth. Notice that raising and lowering does not change the position of an index relative to other indices, and also that "free" indices (which are not summed over) must be the same on both sides of an equation, while "dummy" indices (which are summed over) only appear on one side. As an example, we can turn vectors and dual vectors into each other by raising and lowering indices:

$$\begin{aligned} V_{\mu} &= \eta_{\mu\nu} V^{\nu} \\ \omega^{\mu} &= \eta^{\mu\nu} \omega_{\nu} \ . \end{aligned} \tag{55}$$

This explains why the gradient in three-dimensional flat Euclidean space is usually thought of as an ordinary vector, even though we have seen that it arises as a dual vector; in Euclidean space (where the metric is diagonal with all entries $+1$) a dual vector is turned into a vector with precisely the same components when we raise its index. You may then wonder why we have belabored the distinction at all. One simple reason, of course, is that in a Lorentzian spacetime the components are not equal:

$$\omega^\mu = (-\omega_0, \omega_1, \omega_2, \omega_3) \ . \tag{56}$$

In a curved spacetime, where the form of the metric is generally more complicated, the difference is rather more dramatic. But there is a deeper reason, namely that tensors generally have a "natural" definition which is independent of the metric. Even though we will always have a metric available, it is helpful to be aware of the logical status of each mathematical object we introduce. The gradient, and its action on vectors, is perfectly well defined regardless of any metric, whereas the "gradient with upper indices" is not. (As an example, we will eventually want to take variations of functionals with respect to the metric, and will therefore have to know exactly how the functional depends on the metric, something that is easily obscured by the index notation.)

Contraction. The operation of contraction turns a (k, l) tensor into a $(k-1, l-1)$ tensor, by summing over one upper and one lower index:

$$S^{\mu\rho}{}_{\sigma} = T^{\mu\nu\rho}{}_{\sigma\nu}. \tag{57}$$

The result is a well-defined tensor. Note that *it is only permissible to contract one upper and one lower indices* (as opposed to two upper or two lower indices). Further, note that the order of indices does matter - in general,

$$T^{\mu\nu\rho}{}_{\sigma\nu} \neq T^{\mu\rho\nu}{}_{\sigma\nu}. \tag{58}$$

While this is a very (perhaps too) brief introduction to tensors, we will stop here and may return to this later, if needed.

Having completed this abstract mathematical part, let us go now to kinematic in a flat space-time.

4. Special relativistic kinematics

4.1. Four velocity, acceleration and force

Let us now discuss the motion of a particle in space-time. In three-dimensions we are used to think of the position of a particle x^i as a function of time t in a particular inertial frame. However, in four-dimensions it is easier to think at the space-time location of the moving particle as describing a curve in space-time. This curve is called **world line**. (Figure 3).

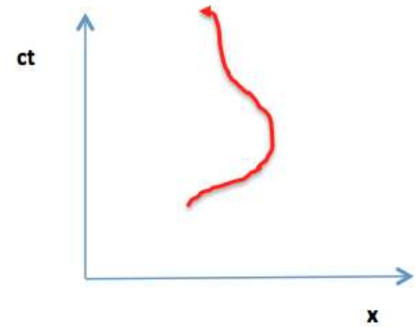

Fig. 3.— The space-time location of an object is represented by a dot in 4-dimensions space-time. As the object changes its position, its trajectory forms a line, known as **world line** of the object.

The description of the motion is done by giving all four-coordinates of the particle, x^μ as a function of a single parameter, σ, which varies along the world line. For each value of σ, $x^\mu = x^\mu(\sigma)$ determine a point along the curve. While we can have many choices for σ, a natural choice is the proper time, τ. Thus, the world line is described by $x^\mu = x^\mu(\tau)$.

We can now define the **four-velocity**, U^μ by:

$$U^\mu \equiv \frac{dx^\mu}{d\tau} \ . \tag{59}$$

Note that as all vectors, U^μ is defined at a single point along the world line. The proper time $d\tau$ was defined in Equation 21 ($(\Delta\tau)^2 = -\frac{1}{c^2}\eta_{\mu\nu}\Delta x^\mu \Delta x^\nu$),

$$d\tau = \sqrt{dt^2 - (dx)^2/c^2} = dt\sqrt{1 - \frac{(dx/dt)^2}{c^2}} = dt\sqrt{1-\beta^2} \tag{60}$$

We can write Equation 59 in components,

$$\begin{aligned} U^0 &= \frac{d(ct)}{\sqrt{1-\beta^2}dt} = \frac{c}{\sqrt{1-\beta^2}} = \gamma c \\ U^i &= \frac{dx^i}{\sqrt{1-\beta^2}dt} = \frac{1}{\sqrt{1-\beta^2}}\frac{dx^i}{dt} = \gamma v^i = \gamma \beta^i c \end{aligned} \tag{61}$$

(Note that in many textbooks, at this point and even earlier, c is taken to be 1, and is omitted from the equations. I keep it here for clarity).

The norm of the four velocity is

$$U^2 \equiv U_\mu U^\mu = \eta_{\mu\nu} U^\nu U^\mu = \eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = \frac{\eta_{\mu\nu} dx^\mu dx^\nu}{d\tau^2} = -c^2 \tag{62}$$

where we used Equation 21.

We can further define the four acceleration by

$$a^\mu \equiv \frac{dU^\mu}{d\tau} \tag{63}$$

and the relativistic four-force by

$$f^\mu = ma^\mu = m\frac{dU^\mu}{d\tau}. \tag{64}$$

This definition replaces Newton's second law of motion. It has all the necessary features: it reduces to Newton's familiar law at any inertial frame where the speed of the particle is much smaller than c; it satisfies the principle of relativity; and in the absence of force, $dU^\mu/d\tau = 0$, and thus Newton's first law holds. The constant m is called the particle's **rest mass**.

Interestingly, the four-acceleration is perpendicular to the four velocity:

$$a^\mu U_\mu = \eta_{\mu\nu}\frac{dU^\mu}{d\tau}U^\nu = \frac{1}{2}\frac{d(\eta_{\mu\nu}U^\mu U^\nu)}{d\tau} = \frac{1}{2}\frac{d(-c^2)}{d\tau} = 0. \tag{65}$$

Equation 65 simply implies that there are only three independent equations of motion, as in Newtonian mechanics; the fourth is being determined.

4.2. Energy and momentum (massive particles)

A natural extension is to define the four-momentum of a massive particle by

$$p^\mu \equiv m\frac{dx^\mu}{d\tau} = mU^\mu \tag{66}$$

Using Equations 61 and 62, we find immediately that

$$\begin{aligned} p^2 \equiv p_\mu p^\mu &= -m^2c^2; \\ p^0 &= \gamma mc; \\ p^i &= \gamma\beta^i mc = \gamma mv^i. \end{aligned} \tag{67}$$

In the limit of small velocities, $v \ll c$ or $\beta^i \ll 1$, $\gamma \simeq 1 + \beta^2/2$, and one gets

$$\begin{aligned} &p^0 \simeq mc + \tfrac{1}{2}\tfrac{mv^2}{c} \\ &p^i \simeq mv^i \end{aligned} \tag{68}$$

The spatial term in Equation 68 is the Newtonian momentum. If we multiply the time component of p by c, we get

$$cp^0 \simeq mc^2 + \frac{1}{2}mv^2. \qquad (69)$$

The second term is immediately recognized as the kinetic energy of a particle, while the first term is a quantity of units of energy, which Einstein referred to as the "rest mass energy" of a massive particle. The term cp^0 is thus simply the **energy** of a particle. This is why the four-momentum is also called **energy-momentum four vector**, whose components in any inertial frame are given by Equation 67.

Using Equation 67, we can write

$$p^0 \equiv \frac{E}{c} = \left((p^i)^2 + m^2c^2\right)^{1/2}. \qquad (70)$$

4.3. Massless particles

So far we considered massive particles, that move at speed $v < c$. Consider now massless particles (e.g., photons), that travel at speed c. Thus, if the photon travels in the x direction, $x = ct$. We cannot use the proper time $d\tau$ as a parameter along a world line of a light ray, since $d\tau$ between any points in it is always zero ! (see Equation 60).

In this case, we have to choose a different, arbitrary parameter (let us denote it by λ) instead of τ as a parameter along the world line of the photon. The four-velocity U is thus defined by $U^\mu = dx^\mu/d\lambda$.

The four velocity is a **null vector**, namely

$$U^2 = 0 \qquad (71)$$

In the absence of forces acting on a photon, the equation of motion is

$$\frac{dU}{d\lambda} = 0 \qquad (72)$$

(compare to the equation of motion of a massive particle, Equation 63).

As was shown by Einstein himself, the energy of a photon E is connected to its angular frequency ω by

$$E = \hbar\omega, \qquad (73)$$

where $\hbar$ is Planck's constant.

We can now use Equation 70 with $m = 0$, to write the momentum of a photon, $p^i = E/c = \hbar\vec{k}$, where $\vec{k}$ is the **wave vector** of a photon.

Thus, the photon four-momentum is

$$p^\mu = (\hbar\omega/c, \hbar\vec{k}). \tag{74}$$

Clearly, $p^2 = p_\mu p^\mu = 0$.

4.4. Many particle system: the energy momentum tensor

So far we considered a single particle moving is space-time. For a single particle, p^μ provides a complete description of its energy and momentum. However, for extended systems it is necessary to go further and define the **energy-momentum tensor** (sometimes called the stress-energy tensor), $T^{\mu\nu}$. (Hopefully, you have already encountered this tensor in your studies of electromagnetism, solid state physics or fluid dynamics !). This is a symmetric $(2, 0)$ tensor which tells us all we need to know about the energy-like aspects of a system: energy density, pressure, stress, and so forth.

Roughly speaking, one can think of $T^{\mu\nu}$ as "the flux of four-momentum p^μ across a surface of constant x^ν". To make this more concrete, let's consider the very general category of matter which may be characterized as a **fluid** — a continuum of matter described by macroscopic quantities such as temperature, pressure, entropy, viscosity, etc. In fact this definition is so general that it is of little use. In general relativity essentially all interesting types of matter can be thought of as **perfect fluids**, from stars to electromagnetic fields to the entire universe. *Schutz* defines a perfect fluid to be one with no heat conduction and no viscosity, while *Weinberg* defines it as a fluid which looks isotropic in its rest frame; these two viewpoints turn out to be equivalent. Operationally, you should think of a perfect fluid as one which may be completely characterized by its pressure and density.

4.4.1. Dust

To understand perfect fluids, let's start with the even simpler example of **dust**. Dust is defined as a collection of particles at rest with respect to each other, or alternatively as a perfect fluid with zero pressure. Since the particles all have an equal velocity in any fixed inertial frame, we can imagine a *"four-velocity field"* $U^\mu(x)$ defined all over spacetime. (Indeed, its components are the same at each point.) Define the **number-flux four-vector** to be

$$N^\mu = nU^\mu \ , \tag{75}$$

where n is the number density of the particles as measured in their rest frame.

When viewed from an arbitrary frame (not necessarily the one in which the dust is at rest), N^0 is the number density of particles, while N^i is the flux of particles in the x^i direction. In the frame in which the dust is at rest, the number flux four vector is $(n, 0, 0, 0)$.

Let's now imagine that each of the particles have the same mass m. Then in the rest frame the energy density of the dust is given by

$$\rho = nmc^2 \ . \tag{76}$$

By definition, the energy density completely specifies the dust. But ρ only measures the energy density in the rest frame; what about other frames? We notice that both n and mc^2 are 0-components of four-vectors in their rest frame; specifically, $N^\mu = (n, 0, 0, 0)$ and $p^\mu = (mc, 0, 0, 0)$. We can thus think of ρ as the $(0, 0)$ component of a more general tensor, as measured in the dust rest frame. This leads us to define the energy-momentum tensor for dust:

$$T^{\mu\nu}_{\rm dust} = p^\mu N^\nu = nmU^\mu U^\nu = \rho U^\mu U^\nu \ , \tag{77}$$

where ρ is defined as the energy density in the rest frame.

The conservation of particle number can be written in a very elegant way using the gradient operator (see Equation 46),

$$\partial_\mu N^\mu \equiv \frac{\partial N^\mu}{\partial x^\mu} = 0. \tag{78}$$

Using Equation 61, we can write Equation 78 explicitly,

$$\begin{aligned} \frac{\partial(\gamma n)}{\partial t} + \frac{\partial(\gamma n\vec{v})}{\partial \vec{x}} &= 0, \\ \frac{\partial(\gamma n)}{\partial t} + \nabla(\gamma n\vec{v}) &= 0. \end{aligned} \tag{79}$$

This equation reduces to the classical continuity equation in the limit $\gamma \to 1$, and is the relativistic generalization of it. When multiplied by the charge q, Equation 78 represents **charge conservation**.

4.4.2. Fluids

Having mastered dust, more general perfect fluids are not much more complicated. Remember that "perfect" can be taken to mean "isotropic in its rest frame." This in turn means that $T^{\mu\nu}$ is diagonal — there is no net flux of any component of momentum in an orthogonal direction. Furthermore, the nonzero spacelike components must all be equal, $T^{11} = T^{22} = T^{33}$. The only two independent numbers are therefore T^{00} and one of the T^{ii}; we can choose to call the first of these the energy density ρ, and the second the pressure p. (Sorry that it's the same letter as the momentum.) The energy-momentum tensor of a perfect fluid therefore takes the following form in its rest frame:

$$T^{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} . \tag{80}$$

We now need to generalize this formula to any frame, in which the fluid is not necessarily at rest. For dust we had $T^{\mu\nu} = \rho U^\mu U^\nu$, so we might begin by guessing $(\rho + p)U^\mu U^\nu$, which gives

$$\begin{pmatrix} \rho + p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \tag{81}$$

To get the answer we want we must therefore add

$$\begin{pmatrix} -p & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} . \tag{82}$$

Fortunately, this has an obvious covariant generalization, namely $p\eta^{\mu\nu}$. Thus, the general form of the energy-momentum tensor for a perfect fluid is

$$T^{\mu\nu} = (\rho + p)U^\mu U^\nu + p\eta^{\mu\nu} . \tag{83}$$

This is an important formula for applications such as stellar structure and cosmology.

Besides being symmetric, $T^{\mu\nu}$ has the even more important property of being **conserved**. In this context, conservation is expressed as the vanishing of the "divergence":

$$\partial_\mu T^{\mu\nu} = 0 . \tag{84}$$

This is a set of four equations, one for each value of ν. The $\nu = 0$ equation corresponds to conservation of energy, while $\partial_\mu T^{\mu k} = 0$ expresses conservation of the $k^{\rm th}$ component of the momentum. We are not going to prove this in general; the proof follows for any individual source of matter from the equations of motion obeyed by that kind of matter. In fact, one way to define $T^{\mu\nu}$ would be "a $(2, 0)$ tensor with units of energy per volume, which is conserved."

4.5. Maxwell's equations

(This section is given for completeness. Unfortunately, I don't think that we will have time to discuss it in class.)

Let us have a new look at **Maxwell's equations** of electrodynamics. In 19$^{\rm th}$-century notation, these are

$$\begin{aligned}\nabla\times\vec{B}-\partial_t\vec{E} &= 4\pi\vec{J}\\ \nabla\cdot\vec{E} &= 4\pi\rho\\ \nabla\times\vec{E}+\partial_t\vec{B} &= 0\\ \nabla\cdot\vec{B} &= 0\ .\end{aligned}\tag{85}$$

Here, $\vec{E}$ and $\vec{B}$ are the electric and magnetic field 3-vectors, $\vec{J}$ is the current, ρ is the charge density (don't confuse with the energy density introduced in Equation 75), and $\nabla\times$ and $\nabla\cdot$ are the conventional curl and divergence. These equations are invariant under Lorentz transformations, of course; that's how the whole business got started. But they don't look obviously invariant; our tensor notation can fix that. Let's begin by writing these equations in just a slightly different notation,

$$\begin{aligned}\epsilon^{ijk}\partial_j B_k-\partial_0 E^i &= 4\pi J^i\\ \partial_i E^i &= 4\pi J^0\\ \epsilon^{ijk}\partial_j E_k+\partial_0 B^i &= 0\\ \partial_i B^i &= 0\ .\end{aligned}\tag{86}$$

In these expressions, spatial indices have been raised and lowered with abandon, without any attempt to keep straight where the metric appears. This is because δ_{ij} is the metric on flat 3-space, with δ^{ij} its inverse (they are equal as matrices). We can therefore raise and lower indices at will, since the components don't change. Meanwhile, the three-dimensional Levi-Civita tensor ϵ^{ijk} is defined just as the four-dimensional one, although with one fewer index. We have replaced the charge density by J^0; this is legitimate because the density and current together form the **current 4-vector**, $J^\mu=(\rho,J^1,J^2,J^3)$.

We can now define the **electromagnetic field strength tensor**, $F_{\mu\nu}$, by

$$F_{\mu\nu}=\begin{pmatrix}0 & -E_1 & -E_2 & -E_3\\ E_1 & 0 & B_3 & -B_2\\ E_2 & -B_3 & 0 & B_1\\ E_3 & B_2 & -B_1 & 0\end{pmatrix}=-F_{\nu\mu}\ .\tag{87}$$

Using this definition, we can now get a completely tensorial 20$^{\rm th}$-century version of Maxwell's equations. Begin by noting that we can express the field strength with upper indices as

$$\begin{aligned}F^{0i} &= E^i\\ F^{ij} &= \epsilon^{ijk}B_k\ .\end{aligned}\tag{88}$$

(To check this, note for example that $F^{01} = \eta^{00}\eta^{11}F_{01}$ and $F^{12} = \epsilon^{123}B_3$.) Then the first two equations in (86) become

$$\begin{aligned} \partial_j F^{ij} - \partial_0 F^{0i} &= 4\pi J^i \\ \partial_i F^{0i} &= 4\pi J^0 \ . \end{aligned} \tag{89}$$

Using the antisymmetry of $F^{\mu\nu}$, we see that these may be combined into the single tensor equation

$$\partial_\mu F^{\nu\mu} = 4\pi J^\nu \ . \tag{90}$$

A similar line of reasoning, which is left as an exercise to you, reveals that the third and fourth equations in (86) can be written

$$\partial_{[\mu} F_{\nu\lambda]} = 0 \ , \tag{91}$$

where the square bracket symbol [] denotes antisymmetrization of the indices, namely summing over all index permutations when reversing the sign for an odd permutation. For example,

$$T_{[\mu\nu\rho]\sigma} = \frac{1}{6}\left(T_{\mu\nu\rho\sigma} - T_{\mu\rho\nu\sigma} + T_{\rho\mu\nu\sigma} - T_{\nu\mu\rho\sigma} + T_{\nu\rho\mu\sigma} - T_{\rho\nu\mu\sigma}\right) \ . \tag{92}$$

The four traditional Maxwell equations are thus replaced by two, thus demonstrating the economy of tensor notation. More importantly, however, both sides of Equations 90 and 91 manifestly transform as tensors; therefore, if they are true in one inertial frame, they must be true in any Lorentz-transformed frame. This is why tensors are so useful in relativity — we often want to express relationships without recourse to any reference frame, and it is necessary that the quantities on each side of an equation transform in the same way under change of coordinates. As a matter of jargon, we will sometimes refer to quantities which are written in terms of tensors as **covariant** (which has nothing to do with "covariant" as opposed to "contravariant"). Thus, we say that Equations 90 and 91 together serve as the covariant form of Maxwell's equations, while Equations 85 or 86 are non-covariant.

A final aside: we have already mentioned that in general relativity gravitation does not count as a "force." As a related point, the gravitational field also does not have an energy-momentum tensor. In fact it is very hard to come up with a sensible local expression for the energy of a gravitational field; a number of suggestions have been made, but they all have their drawbacks. Although there is no "correct" answer, it is an important issue from the point of view of asking seemingly reasonable questions such as "What is the energy emitted per second from a binary pulsar as the result of gravitational radiation?"

REFERENCES

[1] S. Carroll, *Lecture Notes on General Relativity*, part 1. Special Relativity and Flat Spacetime (http://preposterousuniverse.com/grnotes/).

[2] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (John Wiley & Sons), chapter 2.

[3] B. Schutz, *A First Course in General Relativity* (Cambridge), chapters 2, 3, 4.