Chapter 4

General relativity

4.1. The framework of special relativity

4.1.1. Introduction. The goal of this lecture is to motivate and present a geometric treatment of the theory of special relativity. Even readers who have had some experience with special relativity are encouraged to work through this lecture and the corresponding exercises. The terminology and the notation used here is integrated with the terminology and the notation developed thus far; they set the stage for the remainder of the course.

4.1.2. Principles of relativity. From a physics standpoint, there are two fundamental principles of special relativity: *the Principle of Relativity* (which dates back to Galileo) and *the Universality of the Speed of Light* (which is what makes special relativity "special"). The Principle of Relativity encompasses the idea that *no one is at absolute rest* and that all unaccelerated (also known as inertial) observers are "created equal". A typical statement of the Principle of Relativity goes along the lines of "all inertial observers observe the same laws of physics", but a more careful investigation reveals that such a statement relies on some further (often implicit) assumptions. For example, it is assumed that each inertial observer has their own way of assigning spatial (x, y, z) and temporal t-coordinates to an event, that the spatial geometry is always Euclidean and independent of t, etc.

A careful treatment of such implicit assumptions can, for example, be found in [**10**].

The principle of the Universality of the Speed of Light was postulated by Albert Einstein based on earlier experimental evidence. It states that all inertial observers measure the same speed of light[1]. Experimental evidence suggests that this speed, denoted c, is about $299, 792, 458$ meters per second, although in mathematical relativity it is common to choose the units of space measurement based on the units of time measurement so that the speed of light is always equal to 1. (Note that in particular this makes speed dimensionless.)

4.1.3. Inertial observers. For an inertial observer the set of all events makes up a *space-time*, that is, the set

$$\mathbb{R}^{1+3} = \{(t, x, y, z) \,|\, t, x, y, z \in \mathbb{R}\}.$$

The t-axis

$$\{(t, 0, 0, 0) \,|\, t \in \mathbb{R}\}$$

represents the observer's own trajectory. The space-time $\mathbb{R}^{1+3}$ under component-wise operations, much like $\mathbb{R}^4$, has the structure of a vector space. However, it pays huge dividends to turn this perspective upside-down and think of space-time as an abstract vector space and *inertial observer* as a manifestation of choosing a basis $(\partial_t, \partial_x, \partial_y, \partial_z)$ and coordinates (t, x, y, z).

Next, we make an additional assumption that the spatial geometry at each $t = \text{const}$ moment is Euclidean with (x, y, z) serving as Cartesian coordinates. Under such an assumption the Universality of the Speed of Light states that

$$\frac{\sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}}{\Delta t} = c = 1$$

because the units (calibration of the axes) are chosen so that the speed of light is equal to 1. Consequently, the trajectory of light is located along *light-cones*

$$-(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = 0.$$

[1]Built into this is the assumption that light moves along straight paths with time-independent speed which is equal in each direction.

Now suppose a different inertial observer: this means a different choice of basis $(\partial_{\tilde{t}}, \partial_{\tilde{x}}, \partial_{\tilde{y}}, \partial_{\tilde{z}})$ and coordinates $(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z})$. Since we are in a linear-algebraic framework, the connection between the two coordinatizations is linear. The Galilean Principle of Relativity and the Universality of the Speed of Light imply that

$$-(\Delta\tilde{t})^2 + (\Delta\tilde{x})^2 + (\Delta\tilde{y})^2 + (\Delta\tilde{z})^2 = 0$$

along the trajectories of light, i.e., along

$$-(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = 0.$$

In other words, for two different inertial observers the nullsets of the expressions

$$\begin{cases} I = & -(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2, \\ \tilde{I} = & -(\Delta\tilde{t})^2 + (\Delta\tilde{x})^2 + (\Delta\tilde{y})^2 + (\Delta\tilde{z})^2, \end{cases}$$

are identical. A mathematical argument[2] can now be made here showing that I and $\tilde{I}$ are scalar multiples of one another. The most natural postulate one can make at this point is that the calibration (e.g., choice of units) of the $(\partial_t, \partial_x, \partial_y, \partial_z)$ and $(\partial_{\tilde{t}}, \partial_{\tilde{x}}, \partial_{\tilde{y}}, \partial_{\tilde{z}})$ coordinate systems is such that the two expressions, I and $\tilde{I}$, are identically equal to one another. In other words we assume

$$-(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = -(\Delta\tilde{t})^2 + (\Delta\tilde{x})^2 + (\Delta\tilde{y})^2 + (\Delta\tilde{z})^2$$

for every pair of points in the space-time. This invariant of special relativity is called the *interval*. In general, when the units of measurement are not chosen so that the speed of light is 1, the expression for the interval is given by

$$-c^2(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2.$$

4.1.4. Minkowski space-time. The concept of interval motivates the use of an inner-product

$$\mathfrak{m}(U_1, U_2) = -t_1t_2 + x_1x_2 + y_1y_2 + z_1z_2 \tag{4.1}$$

[2] The linear relationship between $(\Delta t, \Delta x, \Delta y, \Delta z)$ and $(\Delta\tilde{t}, \Delta\tilde{x}, \Delta\tilde{y}, \Delta\tilde{z})$ converts the expression for $\tilde{I}$ into a homogeneous quadratic polynomial in $(\Delta t, \Delta x, \Delta y, \Delta z)$. As a polynomial in $(\Delta t, \Delta x, \Delta y, \Delta z)$ the expression I is irreducible. By Hilbert's Nullstellensatz, for example, it follows that I divides $\tilde{I}$. Both polynomials are quadratic and thus the polynomials are scalar multiples of one another.

on the vector space $\mathbb{R}^{1+3} = \{(t, x, y, z)^T\}$, and an articulation of laws of special relativity in terms of geometry of this inner-product space. To a large extent this idea goes back to the German mathematician Hermann Minkowski who in 1907/8 reformulated Albert Einstein's 1905 work on special relativity (combined with an even earlier work of the Dutch physicist Hendrik Antoon Lorentz) in this new geometric language. In his honor the inner-product space defined by (4.1) is called the *Minkowski space-time*.

As different inertial observers "measure" the same value of interval, different observers "measure" the same value of $\mathfrak{m}(U, U)$. In other words, while the individual values of t, x, y, z change with the change of observers the values of $\mathfrak{m}(U, U)$ have to remain the same. It then follows that the values of

$$\mathfrak{m}(U_1, U_2) = \tfrac{1}{2}\left(\mathfrak{m}(U_1 + U_2, U_1 + U_2) - \mathfrak{m}(U_1, U_1) - \mathfrak{m}(U_2, U_2)\right)$$

when $U_1 \neq U_2$ also do not change with the change of observers. In particular, one gets to think about the Minkowski inner-product as providing an "objective", observer-independent framework into which the basic relativity postulates are already built in. The laws of special relativity are thus barely more then consequences of the Minkowski geometry, i.e., geometry of $\mathbb{R}^{1+3}$ equipped with the inner-product $\mathfrak{m}$.

A careful reader has certainly noticed that the Minkowski inner-product $\mathfrak{m}$ is not an inner-product in a classical, positive-definite sense of the word. It is an example of a *non-degenerate inner-product*, i.e., $\mathbb{R}$-valued symmetric bilinear map $\langle ., . \rangle$ for which the zero-vector is the only vector U such that $\langle U, V \rangle = 0$ for all V. The concepts of magnitude, angle and distance are at least somewhat problematic in this more general framework. This is due to the fact that non-degenerate inner-products permit (non-zero) vectors $U \neq 0$ with $\langle U, U \rangle < 0$ or even $\langle U, U \rangle = 0$. For example, the vector ∂_t in Minkowski space-time is such that $\mathfrak{m}(\partial_t, \partial_t) = -1$ while the vector $\partial_t + \partial_x$ lies on the light cone and satisfies $\mathfrak{m}(\partial_t + \partial_x, \partial_t + \partial_x) = 0$. In analogy with these special-relativistic interpretations vectors U such that $\langle U, U \rangle < 0$ are called *time-like* while those for which $\langle U, U \rangle = 0$ are called light-like (or *null*) vectors. The phrase *space-like* vectors is reserved for vectors U with $\langle U, U \rangle \geq 0$.

An argument involving diagonalization of symmetric matrices shows that each non-degenerate inner-product permits what is sometimes called a *pseudo-orthonormal basis*: a basis $(..., e_i, ...)$ of vectors for which

$$\langle e_i, e_j \rangle = \begin{cases} \pm 1 & \text{if } \; i = j, \\ 0 & \text{if } \; i \neq j. \end{cases}$$

For example, each inertial observer gives rise to a pseudo-orthonormal basis $(\partial_t, \partial_x, \partial_y, \partial_z)$ of Minkowski space-time. In general, one shows that the number r of time-like vectors of a pseudo-orthonormal basis does not depend on the choice of the pseudo-orthonormal basis. Assuming the underlying vector space is n-dimensional, the ordered pair $(r, n-r)$ is called *the signature* of the inner-product. The signature of the Minkowski inner-product $\mathfrak{m}$ is $(1, 3)$. In the literature signature $(1, 3)$ is commonly referred to as the *Lorentz signature*.

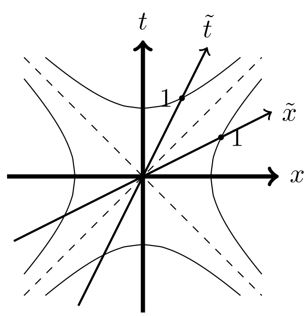

Figure 4.1. Space-time diagram for Minkowski geometry

Minkowski geometry is often depicted using so-called space-time diagrams, i.e., diagrams such as the one in Figure 4.1. Note that the dashed lines in the diagram depict the light cone. The interior of the light cone corresponds to the set of all time-like vectors and is called a *time-cone*. The time-cone has two connected components: one future- and one past-pointing. The set of space-like vectors corresponds to the exterior of the time- and light-cones. The hyperbolas in the diagram indicate the calibration of the axes. Specifically, the intersection of the $\tilde{t}$-axis with $\mathfrak{m}(U, U) = -1$, also called the *unit* *time-like sphere*[3], determines the unit of time measurement for the observer $\partial_{\tilde{t}}$. The situation is completely analogous with the *unit space-like sphere* $\mathfrak{m}(U,U)=1$.

To summarize: the main mathematical idea of special relativity is that inertial observers correspond to different choices of forward-pointing unit time-like vectors ∂_t. They give rise to different choices of pseudo-orthonormal bases $(\partial_t, \partial_x, \partial_y, \partial_z)$ of the Minkowski space-time $\mathbb{R}^{1+3}$. There is no absolute, preferred observer-independent choice of time coordinate t just as there is no preferred set of spatial coordinates (x, y, z). This is a familiar situation: as geometers we are taught to be careful and not confuse coordinate-based identities for actual geometric phenomena. Likewise, a relativist is careful not to confuse coordinate-based perspectives with objective physical phenomena. Special relativity is a study of consequences of the Minkowski geometry, i.e., geometry of $\mathbb{R}^{1+3}$ equipped with the inner-product $\mathfrak{m}$, and not a study of coordinate expressions.

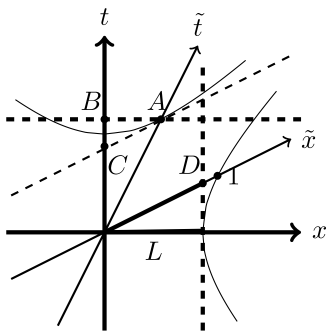

Figure 4.2. According to the observer ∂_t the event A appears to be simultaneous to the event B. However, according to the observer $\partial_{\tilde{t}}$ the events A and C appear to be simultaneous.

4.1.5. Lorentz transformations, time dilation and length contraction. For a given observer ∂_t events whose time coordinates t are the same are seen as *simultaneous* with respect to ∂_t. Events which are simultaneous with respect to ∂_t are located within the

[3]This is not literally a sphere; it is a hyperbola/hyperboloid depending on the dimension.

$\mathfrak{m}$-orthogonal complement to ∂_t within $\mathbb{R}^{1+3}$. These 3-dimensional subspaces serve as the rest spaces, i.e., simultaneity spaces for ∂_t. Do note that the concept of rest space is highly observer-dependent. In the diagram in Figure 4.2 lines parallel to x- and $\tilde{x}$-axes are simultaneity lines according their respective observers. The geometry in each rest space is Euclidean.

Relative to the observer ∂_t, the observer $\partial_{\tilde{t}}$ defines a unique vector $\mathbf{v}$ in the rest space of ∂_t such that $\partial_{\tilde{t}}$ is collinear with $\partial_t + \mathbf{v}$. Since $\mathfrak{m}(\partial_{\tilde{t}}, \partial_{\tilde{t}}) = -1$ it follows that

$$\partial_{\tilde{t}} = \tfrac{1}{\sqrt{1-v^2}} \left(\partial_t + \mathbf{v}\right)$$

where $v = |\mathbf{v}|$. The vector $\mathbf{v}$ is often called the *spatial velocity* of $\partial_{\tilde{t}}$ with respect to the observer ∂_t; meanwhile, $v = |\mathbf{v}|$ is called the *spatial speed*[4]. Contrary to the corresponding classical Gallilean concepts the spatial velocity $\widetilde{\mathbf{v}}$ of ∂_t with respect to $\partial_{\tilde{t}}$ is <u>*not*</u> $-\mathbf{v}$. The reason for this is very simple: $-\mathbf{v}$ is not in the rest space of $\partial_{\tilde{t}}$. Instead, the relative velocity $\widetilde{\mathbf{v}}$ is the (appropriately scaled) orthogonal projection of ∂_t onto the rest space of $\partial_{\tilde{t}}$. However, it is true that

$$|\mathbf{v}| = |\widetilde{\mathbf{v}}| = v = \sqrt{1 - \frac{1}{\mathfrak{m}(\partial_t, \partial_{\tilde{t}})}}.$$

The situation with relativistic addition of velocities and the like is addressed in the exercises following this lecture. The reader should note that the concept of 4-velocity, which is discussed below, is much more suitable for relativistic purposes than the concept of spatial velocity.

To compare observations made by ∂_t and $\partial_{\tilde{t}}$ it is customary to *choose* the coordinate axes ∂_x and $\partial_{\tilde{x}}$ so that they are collinear with $\mathbf{v}$ and $\widetilde{\mathbf{v}}$,

$$\mathbf{v} = v\partial_x \quad \text{and} \quad \widetilde{\mathbf{v}} = -v\partial_{\tilde{x}}.$$

Under such choices we have

$$\partial_{\tilde{t}} = \tfrac{1}{\sqrt{1-v^2}} \left(\partial_t + v\partial_x\right), \quad \partial_{\tilde{x}} = \tfrac{1}{\sqrt{1-v^2}} \left(v\partial_t + \partial_x\right).$$

[4]Hidden in here is the fact that in the present framework spatial speed is always less than 1, i.e., always less than the speed of light. In our narrative this fact is a consequence of being able to "carry a clock" i.e., have a coordinate trajectory of the form $(\tilde{t}, 0, 0, 0)$.

In matrix notation this change of basis/inertial observers can be recorded as

$$(4.2)\qquad \begin{aligned} (\partial_{\tilde{t}}\ \partial_{\tilde{x}}) &= (\partial_t\ \partial_x)\begin{pmatrix} \frac{1}{\sqrt{1-v^2}} & \frac{v}{\sqrt{1-v^2}} \\ \frac{v}{\sqrt{1-v^2}} & \frac{1}{\sqrt{1-v^2}} \end{pmatrix}, \\ \begin{pmatrix} t \\ x \end{pmatrix} &= \begin{pmatrix} \frac{1}{\sqrt{1-v^2}} & \frac{v}{\sqrt{1-v^2}} \\ \frac{v}{\sqrt{1-v^2}} & \frac{1}{\sqrt{1-v^2}} \end{pmatrix}\begin{pmatrix} \tilde{t} \\ \tilde{x} \end{pmatrix}. \end{aligned}$$

Just as the phrase *orthogonal* is used for transformations/matrices which map orthonormal bases to orthonormal bases, the phrase *Lorentz transformations* is used for transformations/matrices which map pseudo-orthonormal bases to pseudo-orthonormal bases. Specifically, changes of inertial observers correspond to changes of pseudo-orthonormal bases and amount to Lorentz transformations. The matrix in (4.2) is oftentimes called a *boost*; alternatively we say that it corresponds to a *boosted observer*.

In the diagram in Figure 4.2 the t-coordinate of A is strictly bigger than its $\tilde{t}$-coordinate, as evidenced by the fact that the unit time-like sphere intersects the t-axis "below" the event B. This phenomenon is known as *time dilation*. The exact relationship follows from (4.2) and is often represented in the form of

$$\Delta t = \frac{\Delta \tilde{t}}{\sqrt{1-v^2}}.$$

The situation is similar with taking length measurements. The endpoints of a line segment L in Figure 4.2 can be interpreted as the endpoints of a meter stick in the rest space of the observer ∂_t. However, the observer $\partial_{\tilde{t}}$ does *not* perceive the endpoints of the line segment L as the endpoints of a "meter stick" if for no other reason than because the far end of the L is not simultaneous with the close end. The event D when the far end of the meter stick finally "arrives" is marked by an $\tilde{x}$-coordinate which is most certainly less than 1. This phenomenon is known as *length contraction*. It follows from (4.2), applied to the event D, i.e.

$$\begin{pmatrix} t \\ 1 \end{pmatrix} = \frac{1}{\sqrt{1-v^2}}\begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}\begin{pmatrix} 0 \\ \tilde{x} \end{pmatrix},$$

that the exact value here is $\tilde{x} = \sqrt{1-v^2}$.

Time dilation and length contraction under boost are relativistic effects which, due to $\frac{1}{\sqrt{1-v^2}} \approx 1 + \frac{1}{2}v^2$, are not as observable when $v \ll 1$. Another way to see this is by taking what is called the *Newtonian limit*. A quick way to take a Newtonian limit of a relativistic expression is to restore the value of c in the expression and then schematically take $c \to \infty$. For example, after restoring c the expression for boost takes the form of

$$\begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1-(v/c)^2}} & \frac{v/c}{\sqrt{1-(v/c)^2}} \\ \frac{v}{\sqrt{1-(v/c)^2}} & \frac{1}{\sqrt{1-(v/c)^2}} \end{pmatrix} \begin{pmatrix} \tilde{t} \\ \tilde{x} \end{pmatrix}.$$

The Newtonian limit here yields the Gallilean transformations

$$t = \tilde{t}, \quad x = \tilde{x} + v\tilde{t}$$

in which neither time dilations nor length contractions exist.

4.1.6. On causality and proper time. Observers and objects (inertial or not) trace out curves γ through space-time. A curve can be parametrized in many different ways but, in Riemannian geometry at least, it is often natural to employ a unit speed parametrization; up to an additive constant the parameter involved is the same as the arc-length. The arc-length approach does not easily extend to Minkowski space-time because of the difficulty with the concept of (vector) magnitude. We now discuss certain genres of curves in Minkowski space-time for which the broad idea of arc-length does carry through.

Reparametrizations lead to tangent vectors $\dot{\gamma}$ which are scalar multiples of one another and thus the time-like, space-like or null nature of $\dot{\gamma}$ is independent of the choice of the parametrization of γ. A curve γ is called *time-like* if $\dot{\gamma}$ is time-like. For example, trajectories of (inertial) observers (capable of measuring time) are by definition time-like. The concept of *space-like curves* is defined analogously.

It can be shown (see Exercises 2 and 9 following Lecture 1.2) that the value of

$$\tau = \int_{u_{\text{Initial}}}^{u_{\text{Final}}} \sqrt{-\mathfrak{m}(\dot{\gamma}, \dot{\gamma})}\, du$$

is independent of the choice of the parametrization u. (In the case of space-like curves the minus sign is omitted.) For inertial observers where $\dot{\gamma} = \partial_t$ is unit time-like, the value of τ is simply the amount of elapsed time, as measured by ∂_t. It is a geometrically/physically meaningful quantity associated with an observer and not an artifact of a choice of time coordinate. This particular observation leads to the following interpretation of τ associated to time-like γ: it is the amount of time measured by γ's clock. In literature τ is often called *proper time*. The concept of proper time for a time-like curve γ is analogous to the concept of arc-length for a space-like or Riemannian curve. To find proper time, one, just as in Riemannian geometry, integrates $d\tau$ where

$$d\tau^2 = -dt^2 + dx^2 + dy^2 + dz^2.$$

In Riemannian geometry we frequently employ unit-speed, i.e., arc-length parametrizations. Unless specifically stated otherwise, every time-like curve γ throughout these lecture notes should be assumed to be parametrized by its proper time τ; we then have

$$\mathfrak{m}(\dot\gamma, \dot\gamma) = \mathfrak{m}(\partial_\tau, \partial_\tau) = -1.$$

So far the concepts of tangential and velocity vector fields have been used interchangeably. Henceforth all velocities should be understood as 4-*velocities* ∂_τ.

In addition to time-like curves we also often restrict our attention to *causal curves*, that is, curves whose tangent vector field is time-like or null. The idea here is that a physical influence can only "travel" along causal curves. For an event A the set of events B for which there exists a (future-pointing) causal curve from A to B is called a *(future) causal cone* at A. Rephrased in plain words, the future causal cone at A is the set of events which are in theory influenceable by A. It can be shown that it is precisely the union of (future-pointing) time- and light-cones with vertex at A. The future time-cone at A can also be seen as a set of endpoints of future-pointing time-like curves starting at A.

4.1.7. Matter in special relativity.

4.1.7.1. *An example from classical mechanics.* Classical mechanics is built around concepts such as mass, energy and momentum. (Linear) *momentum* of a point object of mass m moving with velocity $\mathbf{v}$ is defined as

$$\mathbf{p} = m\mathbf{v}.$$

It is a subject of a conservation law, which in turn is commonly seen in the form of Newton's second law "$F = m\mathbf{a}$". Many an introductory physics exercise is a variation on the theme of conservation of mass and conservation momentum upon collision of point bodies[5]:

$$\sum_{i=1}^{n_{\rm in}} m_{i,\rm in} = \sum_{j=1}^{n_{\rm out}} m_{j,\rm out}, \quad \sum_{i=1}^{n_{\rm in}} \mathbf{p}_{i,\rm in} = \sum_{j=1}^{n_{\rm out}} \mathbf{p}_{j,\rm out}. \tag{4.3}$$

Point objects are just a convenient idealization. A more "realistic" object such as a cloud of dust is perceived as being made out of individual small objects. Typically, the number of such objects is too numerous to make keeping track of each individual object useful or even possible. Instead, we keep track of average quantities, attach them to a concept of a point-object and then compute with point-objects using ideas of calculus. For example, in place of mass m of a point object we frequently make use of the concept of *mass density* $\rho(\mathbf{x})$ of a dust cloud. The idea here is that the infinitesimal volume element at $\mathbf{x}$ has mass $\rho(\mathbf{x})\,\mathrm{dvol}_{\mathbf{x}}$, a quantity which can then be integrated as needed. The stated description of a dust cloud applies to the classical, Euclidean situation in $\mathbb{R}^3$; we use $\mathbf{x}$ as a shorthand for spatial variables (x, y, z) and $\rho(\mathbf{x})$ is assumed to be a non-negative function on $\mathbb{R}^3$.

To be fair, the cloud of dust is likely to evolve with time and the function $\rho(\mathbf{x})$ should be viewed not only as a function of spatial position but also time:

$$\rho(t, \mathbf{x}).$$

The idea here is that point objects within the dust cloud are likely to have non-zero relative momentum, making the individual objects cluster towards or move away from certain locations. This makes it clear that it is of interest not only to keep track of energy/mass density but also of *momentum density*

$$\rho(t, \mathbf{x})\mathbf{u}(t, \mathbf{x});$$

[5] Here $(m_{i,\rm in}, \mathbf{p}_{i,\rm in})$ are mass and momentum of $n_{\rm in}$ particles going into the collision, and $(m_{i,\rm in}, \mathbf{p}_{i,\rm in})$ are mass and momentum of $n_{\rm out}$ particles leaving the collision.

here $\mathbf{u}(t, \mathbf{x})$ denotes the classical spatial velocity at time t of the point object located at $\mathbf{x}$.

There is a conservation law known as *the continuity equation*. It is derived from *conservation of mass* and it informs us about the dynamics of the function $\rho(t, \mathbf{x})$. It relies on the observation that the quantity $-\mathrm{div}(\rho\mathbf{u})$ computes the amount (in terms of mass density) of objects which cluster at a location in a unit of time. The continuity equation states that

$$\frac{\partial \rho}{\partial t} + \mathrm{div}(\rho\mathbf{u}) = 0.$$

Another conservation law addresses the dynamics of the vector field $\mathbf{u}(t, \mathbf{x})$. It comes from the *conservation of momentum*, i.e., Newton's second law. This is neither the time nor place to go into details of fluid mechanics, but the broad idea is that the motion of dust particles is due to variation in pressure and that particles move towards zones where the pressure is lower. In other words, it is assumed that there is a potential function $\mathscr{P}$ called the *pressure* whose negative gradient $-\mathrm{grad}\,\mathscr{P}$ provides the force under which dust particles move. By comparing the velocity $\mathbf{u}(t, \mathbf{x})$ a point object has at a moment t to the velocity $\mathbf{u}(t+\delta t, \mathbf{x}+\mathbf{u}\cdot\delta t)$ it has δt later we see that the acceleration $\mathbf{a}(t, \mathbf{x})$ of the point object located at $\mathbf{x}$ at time t is

$$\mathbf{a} = \tfrac{\partial}{\partial t}\mathbf{u} + \nabla_{\mathbf{u}}\mathbf{u}.$$

From here we obtain what is known as *Euler's equation*:

$$\rho \cdot \big(\tfrac{\partial}{\partial t}\mathbf{u} + \nabla_{\mathbf{u}}\mathbf{u}\big) = -\mathrm{grad}\,\mathscr{P}.$$

In many situations of interest $\mathscr{P}$ is presumed to be a function of ρ and the exact relation between them is called *the equation of state.*

To summarize: The partnership of the continuity and the Euler's equation (along with the equation of state) is what governs the evolution of a dust cloud. The two equations stem from the conservation laws for mass and for momentum, and can be seen as continuous versions of (4.3).

4.1.7.2. *Relativistic version.* We now examine the kind of adjustments needed to be made to the above in order to account for (special) relativistic effects.

Point objects in special relativity are assumed to have a conserved quantity associated to them, the 4-*momentum* P. For a point-object with 4-velocity U the 4-momentum P is assumed to be a constant scalar multiple of U. Its *mass* m is then *defined* by

$$-m^2 = \mathfrak{m}(P, P), \quad \text{i.e.,} \quad P = mU.$$

The 4-momentum P described above is often referred to as *energy-momentum*. The reason for this nomenclature is as follows: Suppose a particle is moving with spatial velocity $\mathbf{v}$ relative to the observer $(\partial_t, \partial_x, \partial_y, \partial_z)$; then

$$U = \tfrac{1}{\sqrt{1-v^2}}(\partial_t + \mathbf{v}) \quad \text{and} \quad P = \tfrac{m}{\sqrt{1-v^2}}\partial_t + \tfrac{m}{\sqrt{1-v^2}}\mathbf{v}.$$

To second order in v the latter expands as

$$P = (m + \tfrac{1}{2}mv^2)\partial_t + m\mathbf{v};$$

the latter suggests that the time component of the 4-momentum P can be thought of as energy while the spatial components constitute spatial (linear) momentum. The said formula for energy is more recognizable in the form in which the value of c is appropriately restored:

$$mc^2 + \tfrac{1}{2}mv^2.$$

What we are seeing here is the much-celebrated equivalence of mass and energy. For this reason the ∂_t-component of P is called *mass-energy*; we interpret it as mass-energy measured by the observer ∂_t.

There are good reasons, relativistic and non-relativistic alike, to view *momentum as a covector*. For example, in our context here each given inertial observer ∂_t arrives at their own measurement of mass-energy:

$$\partial_t \;\mapsto\; \tfrac{m}{\sqrt{1-v^2}} = \mathfrak{m}(P, \partial_t).$$

Likewise, for each given unit direction $\mathbf{w}$ in their rest space the observer ∂_t could measure the component of the momentum in the direction $\mathbf{w}$:

$$\mathbf{u} \;\mapsto\; \tfrac{m}{\sqrt{1-v^2}}\langle \mathbf{v}, \mathbf{w}\rangle = \mathfrak{m}(P, \mathbf{w}).$$

For us here, therefore, it makes a lot of sense to consider the covector field $P_\flat$ instead of P with the operation of lowering indices defined using the Minkowski metric $\mathfrak{m}$:

$$P_\flat = m\,U_\flat.$$

We adopt this perspective henceforth.

Any attempt at introducing mass-energy-momentum *density* for a relativistic dust cloud is immediately faced with the following issue: density expresses the amount per unit of *spatial volume* and from a relativistic standpoint there is no absolute concept of space! In other words, different inertial observers would not only come up with different measurements for energy-momentum due to the reasons discussed in the previous paragraph but also because of phenomena such as length contraction. Different observers observing the same collection of dust particles would agree on their number but not on the amount of space they occupy and not on their energy-momenta. The observations of the sheer number *density* would change from observer to observer. An observer ∂_t moving at relative speed v relative to the dust cloud would observe the same cloud as occupying only $\sqrt{1-v^2}$ the volume. This, for example, means that mass-energy-density as measured by ∂_t satisfies

$$\frac{\rho}{\sqrt{1-v^2}} \cdot \frac{1}{\sqrt{1-v^2}} = \frac{\rho}{1-v^2}.$$

When evaluated at the observer ∂_t the straightforward generalization $\rho\,\mathfrak{m}(U,\cdot)$ of the "Euclidean" $\rho\mathbf{u}$ leads to $\frac{\rho}{\sqrt{1-v^2}}$ instead of the correct mass-energy density $\frac{\rho}{1-v^2}$. In particular, $\rho\,\mathfrak{m}(U,\cdot)$ cannot possibly be the expression for relativistic energy-momentum density.

However, the evaluation $T(\partial_t, \partial_t)$ of the covariant 2-tensor

$$T = \rho\,\mathfrak{m}(U,\cdot)\,\mathfrak{m}(U,\cdot) \tag{4.4}$$

does lead to the correct mass-energy density:

$$T(\partial_t, \partial_t) = \rho\left(\mathfrak{m}\left(\tfrac{1}{\sqrt{1-v^2}}(\partial_t + \mathbf{v}), \partial_t\right)\right)^2 = \frac{\rho}{1-v^2}.$$

In fact, one can show (see Exercise 1 following Lecture 4.2) that the tensor T in (4.4) is the <u>only</u> symmetric 2-tensor for which

$$T(\partial_t, \partial_t) = \frac{\rho}{1-v^2} \quad \text{for all observers} \quad \partial_t.$$

Overall, we are brought to the idea that a continuum of relativistic matter is best expressed in the language of symmetric 2*-tensors.* Symmetric 2-tensors T such as the one above are called *stress-energy tensors*. The terminology here is partly borrowed from classical fluid dynamics where one makes use of symmetric 2-tensors called stress tensors. The expressions for the stress-energy tensor T vary depending on the physical theory under consideration. The expression for T in (4.4) is the simplest form of the relativistic perfect fluid. Discussion of stress-energy tensors in general or of other physical theories (e.g. electromagnetism) is beyond the scope of these notes.

Conservation laws for physical quantities, such as those for mass and momentum, are expected to have a relativistic counterpart in terms of the stress-energy tensor. To gain an insight into how this might work recall the continuity equation $\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{u}) = 0$ discussed earlier in this lecture. It schematically resembles a statement that the space-time 4-vector field

$$\rho \partial_t + \rho \mathbf{u}$$

is divergence-free in some Minkowski space-time sense of the word. This resemblance might inspire one to study *divergence of stress-energy tensor*, i.e., *contracted* ∇T:

$$(\operatorname{div} T)_i = \sum_{j,k=0}^{3} \frac{\partial}{\partial x^j} \left(\mathfrak{m}^{jk} T_{ik} \right), \tag{4.5}$$

where (x^0, x^1, x^2, x^3) denote the coordinates $(t, \mathbf{x})$ of some inertial observer. Note that $\operatorname{div} T$ in this situation is a covector field. An ambitious reader could verify that for T as in (4.4) the requirement that $\operatorname{div} T = 0$ is equivalent to a pair of a scalar and a vector equation:

$$\nabla_U \rho + \rho \operatorname{div} U = 0 \quad \text{and} \quad \rho \nabla_U U = 0.$$

Furthermore, the Newtonian limits (taken as in Section 4.1.5 above) of the obtained equations correspond to the continuity and the Euler equation with $\mathscr{P} = 0$. It is for these reasons that we see (4.4) as the stress-energy tensor of the pressure-less *relativistic perfect fluid*.

The stress-energy tensor for a relativistic perfect fluid with non-zero pressure is found to have the form

$$T = \mathscr{P} \mathfrak{m} + (\rho + \tfrac{\mathscr{P}}{c^2})\, \mathfrak{m}(U, \cdot)\, \mathfrak{m}(U, \cdot). \tag{4.6}$$

Recovering the continuity and the Euler equations is a doable but a substantially more involved exercise. We shall continue our discussion of stress-energy tensors when we discuss general-relativistic matter in Lecture 4.2.

4.1.8. Exercises for Lecture 4.1.

Non-degenerate inner-product spaces.

(1) Let $\langle ., . \rangle$ be a non-degenerate inner-product on a vector space $\mathscr{V}$ of dimension n. Let U be a fixed (non-zero) space-like or time-like vector. Investigate the orthogonal complement of U following the guidelines given below.
   (a) Show that $V \mapsto \langle U, V \rangle$ is a linear map of rank 1 and nullity $n - 1$.
   (b) Show that $U^\perp = \{V \,|\, \langle V, U \rangle = 0\}$ is a subspace of dimension $n - 1$.
   (c) Show that each vector V in $\mathscr{V}$ can uniquely be decomposed as
   $$V = U_1 + U_2,$$
   with U_1 in $\mathrm{Span}(U)$ and U_2 in $U^\perp$.
   (d) Show that the restriction of the inner-product $\langle ., . \rangle$ to $U^\perp$ is still non-degenerate.
   (e) How many of the above claims still hold if U is non-zero null?

(2) Let $\langle ., . \rangle$ be a non-degenerate inner-product on a vector space $\mathscr{V}$ of dimension n.
   (a) Iterate the arguments of the previous problem to show the existence of a pseudo-orthonormal basis
   $$(e_1^-, \ldots, e_r^-, e_1^+, \ldots, e_s^+)$$
   with $\langle e_i^-, e_i^- \rangle = -1$ and $\langle e_i^+, e_i^+ \rangle = 1$ for all i.
   (b) Let $(f_1^-, \ldots, f_p^-, f_1^+, \ldots, f_q^+)$, with $\langle f_i^-, f_i^- \rangle = -1$ and $\langle f_i^+, f_i^+ \rangle = 1$ for all i, be another pseudo-orthonormal basis for V. Show that $(r, s) = (p, q)$.

Geometry in Minkowski space-time.

(3) (a) Show that the future (top) sheet of the unit time-like sphere

$$-t^2 + x^2 + y^2 + z^2 = -1$$

in $\mathbb{R}^{1+3}$ can be parametrized by

$$\begin{cases} t = \cosh\sigma, \\ x = \sinh\sigma\cos\theta\sin\phi, \\ y = \sinh\sigma\sin\theta\sin\phi, \\ z = \sinh\sigma\cos\phi \end{cases}$$

with $\sigma \in \mathbb{R}$ and the usual domains for θ and ϕ.

(b) In analogy with spherical coordinates for Euclidean $\mathbb{R}^n$, future time-cones can be parametrized by $(r, \mathbf{u})$ where $\mathbf{u}$ is future pointing and unit time-like. Show that within the time-cone the Minkowski metric can be expressed as

$$d\tau^2 = -dr^2 + r^2\left(d\sigma^2 + \sinh^2\sigma\, ds^2_{S^2}\right),$$

where $ds^2_{S^2}$ denotes the metric on the standard unit sphere. (Gauss's Lemma for Minkowski space-time).

(4) Let $\gamma_0(r) = r\mathbf{u}$ for unit time-like $\mathbf{u}$ join event A to event B. Let γ be any time-like curve joining A to B. Show that the proper time of γ is no bigger than the proper time of γ_0:

$$\tau(\gamma) \leq \tau(\gamma_0).$$

In other words, show that inertial trajectories maximize proper time. (This is analogous to geodesics minimizing arc-length, at least within the domain of geodesic normal coordinates.)

(5) Let M denote the future (top) sheet of the unit time-like sphere in Minkowski space-time $\mathbb{R}^{1+n}$.

(a) Analyze the geometry of M by following the prompts below:

- Suppose $\gamma(u)$ is a curve on M. Show that $\dot{\gamma}$ is orthogonal, in the Minkowski sense of the word, to M. Use this observation to describe the tangent spaces to M.
- Is the Minkowski dot-product, when restricted to the tangent spaces to M, non-degenerate? If so, what is the inherited signature?
- Use a symmetry argument similar to the one we used on S^n in Lecture 1.3 to identify geodesics on M.
- Find the expression for the metric on M in coordinates from Exercise 3 above. (Is it a familiar expression? What can you say about the curvature of M?)

(b) Discuss what would change if the radius of the time-like sphere were changed from 1 to some ϱ.

(6) Let M, as above, denote the future (top) sheet of the unit time-like sphere in $\mathbb{R}^{1+n}$ and let $L = (-1, 0, \ldots, 0)$. For a point P on M consider the point $(0, u^1, \ldots, u^n)$ at which the line LP intersects the $x^0 = 0$-coordinate plane. (This is analogous to the stereographic projection for spheres seen in Lecture 1.1.)

(a) Convince yourself that the $(u^1, \ldots, u^n)$ parametrize M. What region of $\mathbb{R}^n$ is occupied by $(u^1, \ldots, u^n)$?

(b) Find the rules which relate $(x^0, \ldots, x^n)$ to $(u^1, \ldots, u^n)$ and vice versa.

(c) Compute an explicit formula for the metric on M in $(u^1, \ldots, u^n)$ coordinates. Once again, you should get a familiar result.

Special relativity. [6]

[6] Exercises and examples from relativity textbooks are heavily coated in the language of clocks, trains, rockets, spaceships etc., in addition to being decorated with some explicit (yet questionably realistic) numerical values. While some readers might find such rhetorical tools a motivation for learning, they are by no means necessary nor enlightening for everyone. While we have chosen to avoid such presentation styles thus

(7) This exercise is on the so-called Twin Paradox. While Albert Einstein in 1905 only described it as "a peculiar consequence" regarding clocks, the formulations of the "paradox" now commonly found in textbooks are believed to go back to the 1911 lecture "L'evolution de l'espace et du temps" ("The Evolution of Space and Time") by the French physicist Paul Langevin. We quote his formulation here[7].

> ... our traveler would need only to agree to being shut up inside a projectile that the Earth would launch at a velocity sufficiently close to that of light, but still less than it, which is physically possible, arranging for an encounter with, say, a star to take place at the end of one year in the lifetime of the traveler and to send him back toward the Earth at the same velocity. Having returned to Earth two years older, he will emerge from his ark to find that our globe has aged two hundred years....

(a) Find the velocity of the traveler from Langevin's formulation.

(b) What is the time on Earth just before the traveler turns around? How much time does the traveler, just before he turns around, think passed on Earth since his departure?

(c) What is the time on Earth just after the traveler turns around? How much time does the traveler, just after he turns around, think passed on Earth since his departure?

(d) From the traveler's standpoint, how much did Earth age in the moments of his U-turn?

(e) Is Earth an inertial observer in this situation? Is the traveller an inertial observer?

---

far, it would be misleading to completely deprive the readers of exposure to traditional relativity folklore. It is for that reason, and that reason alone, that the formulation styles of the remaining exercises are so varied.

[7]The translation used is by J. Sykes; see AMS Historica 108:285-300, 1973.

(f) The Earth could equally "say" that it was moving away from the traveler for 200 years, and then turned around to reunite with him. If such a perspective was valid the traveler would have aged 20, 000 years during his travels. Why is it that the traveler has not aged more than the Earth?

(g) The phenomenon known as the Twin Paradox is tapping into the subject addressed in Exercise 4 above. Discuss this relationship.

(8) Leaving his twin Peter at rest in their freely falling spaceship, Paul departs at constant relative speed 0.8 on a trip around a circle of radius of three light years in the rest space of their spaceship. What is the age difference between the brothers when they reunite? (Be sure to use correct units.)

(9) This is a problem on addition of spatial velocities or, rather, on relating relative speeds between three inertial observers. Suppose ∂_{t_1} and ∂_{t_2} are two inertial observers whose spatial velocities with respect to a third inertial observer ∂_t are $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively.

(a) Compute $\mathfrak{m}(\partial_{t_1}, \partial_{t_2})$ and relate the obtained expression with the relative spatial speed v between the observers ∂_{t_1} and ∂_{t_2}.

(b) Show that the relative spatial speed v between the observers ∂_{t_1} and ∂_{t_2} is

$$v = \frac{1}{1 - \langle \mathbf{v}_1, \mathbf{v}_2 \rangle} \sqrt{|\mathbf{v}_2 - \mathbf{v}_1|^2 - |\mathbf{v}_1 \times \mathbf{v}_2|^2}.$$

(c) Suppose the vectors ∂_{t_1}, ∂_{t_2} and ∂_t are coplanar, i.e., suppose that the spatial velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ are collinear. Show that[8]

$$v = \frac{|v_2 - v_1|}{1 - v_1 v_2}.$$

[8] The identity is often seen in the form of addition of spatial velocities $v_2 = \frac{v+v_1}{1+vv_1}$. Do note that, similar to the situation in classical Newtonian mechanics, this particular expression only applies when the motions of the three observers are aligned.

Is this speed more or less than the speed we would obtain in classical Gallilean mechanics?

(d) This time suppose that the spatial velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. Show that

$$v = \sqrt{v_1^2 + v_2^2 - v_1^2 v_2^2}.$$

Is this speed more or less than the speed we would obtain in classical Gallilean mechanics?

(10) Suppose 1-dimensional motion of a relativistic train leaving a train station after stopping there. The conductor of the train has the train accelerate with constant acceleration a. Let $v(\tau)$ denote the speed of the train relative to the stationmaster when the conductor's clock reads τ.

(a) Show that

$$\frac{1}{1-v^2} \cdot \frac{dv}{d\tau} = a.$$

Hint: consult Exercise 9.

(b) Find $v(\tau)$, under the assumption that the train left the station at time $\tau = 0$.

(c) Suppose t denotes the time on the stationmaster's clock. Find $t(\tau)$ assuming that the train left the station at time $t = 0$.

(d) Find the expression for the relative velocity v as a function of the station master's clock. What is the acceleration of the train as measured by the stationmaster?

(e) According to the stationmaster, how far away is the conductor from the stationmaster at time t?

(f) According to the conductor, how far away is the stationmaster from the conductor at time τ?