A BRIEF INTRODUCTION TO MATHEMATICAL RELATIVITY PART 1: SPECIAL RELATIVITY

ARICK SHAO

These notes are the first of a pair of brief articles containing an informal introduction to the mathematics behind the theory of relativity. Here, we survey *special relativity*, which represents the case in which gravitational effects are negligible, and which served as a precursor to general relativity.

In order to keep these notes appropriately short, most of the details—e.g., technical definitions, proofs, and computations—are omitted. Background knowledge in differential geometry would be helpful here for understanding various points, but will not be strictly required due to the informal nature of the discussions.

An important disclaimer is that these notes focus primarily on the mathematical, rather than the physical, aspects of the theory. This is mostly by necessity, since I am a mathematician (and not a physicist), with background in partial differential equations and differential geometry (and not in theoretical physics). Consequently, this article will approach the subject from a mathematical viewpoint, in particular in terms of geometry of Minkowski spacetime. Physicists will rightfully have a different perspective on many of these points.

For a detailed mathematical reference, see [3, Ch. 6] for a formal development of Minkowski geometry and special relativity. For a (physics) text containing both mathematical and physical elements, see [4, Ch. 4]. Furthermore, on the physics side, many elementary references to the topics touched upon here are widely available on the Internet.

1. Minkowski Geometry

The postulates for special relativity were first formulated by Albert Einstein in 1905.[1] Quoting *Nobelprize.org* [2], Einstein's postulates state that:

(1) **The principle of relativity**: *The laws of physics are the same in all inertial frames of reference.*
(2) **The constancy of speed of light in vacuum**: *The speed of light in vacuum has the same value c in all inertial frames of references.*

From these postulates and various physical considerations, Einstein was able to derive many mind-blowing consequences. For instance, two observers moving at different velocities will

[1]Unfortunately, here we are ignoring vast amounts of physics history which eventually led to this point; see, for instance, [1, 5] for some basic summaries.

perceive lengths of objects differently. These results had a profound effect on modern physics and on the way physicists viewed the world.

On the mathematical side, a major contribution came from Hermann Minkowski, who in 1908 gave a mathematical formulation of special relativity in terms of differential geometry. Aside from providing a rigorous mathematical model for the universe under special relativity, Minkowski's viewpoint was also important because it would later be extended by Einstein into the theory of general relativity.

1.1. **Minkowski Spacetime.** Minkowski's mathematical model melded "time" and "space" into a single entity "spacetime". More specifically, this spacetime is modeled by[2]

$$\mathbb{R}^4 := \{(t, x, y, z) | t, x, y, z \in \mathbb{R}\},$$

for which we can naively think of the first component as the time component and the remaining three components as the spatial components.

Now, the description of spacetime as the set of 4-tuples (t, x, y, z) is philosophically not so accurate. In light of the relativity principle, all inertial reference frames, or coordinate systems in mathematical lingo, should be treated the same. In other words, since there should not be a single preferred system of coordinates, we should not privilege the Cartesian coordinates t, x, y, z, as in the above. Thus, the mathematical object for describing spacetime should not just be $\mathbb{R}^4$, but instead this $\mathbb{R}^4$ "modulo coordinate systems".

The notion for capturing this coordinate-independent object is precisely that of a (differential) *manifold*. As a result, we can formally define the Minkowski universe as $\mathbb{R}^4$, but viewed as a 4-dimensional manifold.[3]

Remark. The above should be contrasted with Newtonian theory, for which space itself can be modelled by the manifold $\mathbb{R}^3$ (modulo coordinate systems), and for which time is similarly described by $\mathbb{R}$ modulo coordinates. The difference, however, is that the Newtonian theory posits a clear split between space and time, whereas in special relativity, space and time are combined into a single aggregate object $\mathbb{R}^4$.

Next, recall that in the classical Newtonian setting, the squared distance between two points in space is the Euclidean distance:

$$d_E^2((x_1, y_1, z_1), (x_2, y_2, z_2)) := (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2.$$

[2]There are some caveats; see further below.

[3]A manifold is, roughly, an object that near each point "looks like" a copy of $\mathbb{R}^n$. While $\mathbb{R}^4$ trivially satisfies this characterization, the pedagogical point here is that this manifold description of $\mathbb{R}^4$ does not make any particular coordinate system more special than any other.

In Minkowski spacetime, an analogous "squared distance" is constructed on R^4, *except that we flip the sign of the time component*:[4]

$$d_M^2((t_1, x_1, y_1, z_1), (t_2, x_2, y_2, z_2)) \\ := -(t_2 - t_1)^2 + (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2.$$

Remark. We note that although d_E^2 and d_M^2 are defined above using special fixed coordinates, both can in fact also be expressed in a coordinate-independent manner.

Since we would like to take derivatives, we wish to state the above in a differential form. In the Newtonian setting, this is captured by the *Euclidean metric* on $\mathbb{R}^3$:

$$g_E := dx^2 + dy^2 + dz^2.$$

More explicitly, given two tangent vectors about some point in space,

$$v = (v^x, v^y, v^z), \qquad w = (w^x, w^y, w^z),$$

their inner product with respect to g_E is

$$g_E(v, w) = v^x w^x + v^y w^y + v^z w^z,$$

that is, the standard dot product in 3 dimensions.[5]

On the other hand, the metric on $\mathbb{R}^4$ corresponding to d_M^2 is

$$g_M := -dt^2 + dx^2 + dy^2 + dz^2.$$

In other words, given two tangent vectors about a point in spacetime,

$$v = (v^t, v^x, v^y, v^z), \qquad w = (w^t, w^x, w^y, w^z),$$

we measure their "inner product" with respect to this Minkowski metric as

$$g_M(v, w) = -v^t w^t + v^x w^x + v^y w^y + v^z w^z.$$

Thus, we formally define *Minkowski spacetime* to be the manifold-and-metric pair $(\mathbb{R}^4, g_M)$. This geometric object describes the universe as modeled by special relativity.[6] Consequently, the mathematics behind special relativity is precisely the study of this Minkowski geometry. Because of this change in sign in the "time" component, Minkowski geometry has vastly different properties compared to the more familiar Euclidean geometry that is associated with classical Newtonian physics.

[4]Since we are more concerned with the mathematical aspects here, we simplify notations by assuming units such that the speed of light is 1.

[5]We remark that d_E can be recovered from g_E by integrating along curves.

[6]This is not entirely accurate, as in addition to $(\mathbb{R}^4, g_M)$, one also has a *time orientation*, that is, a way of distinguishing between "future" and "past" directions.

1.2. **Causal Character.** In contrast to the Euclidean metric and distance, which are always positive-definite, the Minkowski product $g_M(v, v)$ no longer has a definite sign. This lends itself to the notion of the *causal character* of a direction:

- A tangent vector v is *timelike* iff $g_M(v, v) < 0$.
- v is *spacelike* iff either $g_M(v, v) > 0$ or $v = 0$.
- v is *null*, or *lightlike*, iff $g_M(v, v) = 0$ and $v \neq 0$.

Examples of each type of vector can be found in Figure 1. In particular, the set of all null directions from the origin form a double cone about the origin (represented by the dotted lines), called the *light*, or *null*, *cone*. Directions lying within the light cone are timelike, while directions lying in the exterior are spacelike.

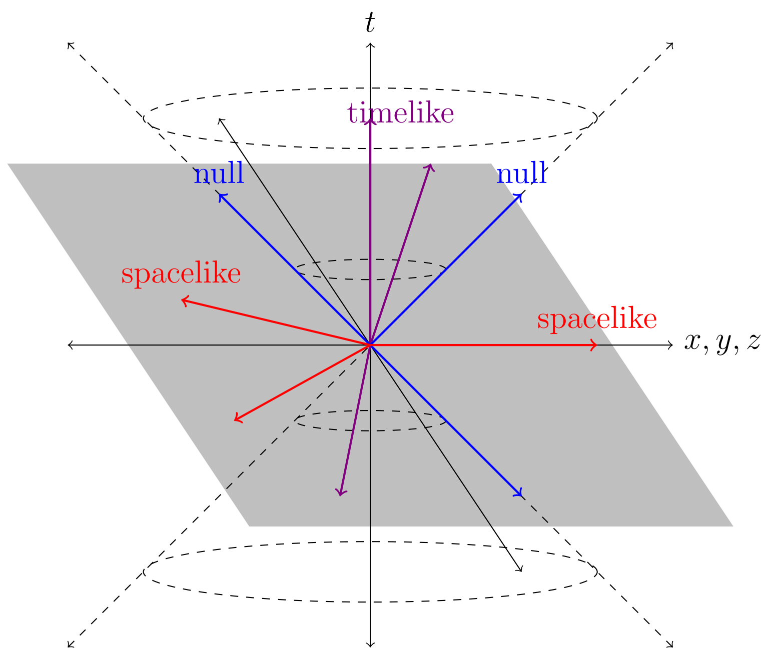

Figure 1. Timelike, spacelike, and null directions about the origin. For visual clarity, the three spatial dimensions are compacted into two dimensions.

This partition into causal characters has physical interpretations as well. A point in Minkowski spacetime, called an *event*, represents a single particle at a particular time, while a timelike curve in $\mathbb{R}^4$ represents an *observer*, that is, a single particle existing throughout time. Null lines, on the other hand, represent *light rays*.

Note these prescriptions of observers and of timelike directions being within the light cone automatically imply that objects cannot travel faster than the speed of light. Furthermore, this leads to the basic tenets of *causality*: an event A at some time $t > 0$ can be affected by an event O at the origin if and only if A lies in the (future) null cone centered at O, or equivalently, O and A are connected by a timelike curve.

1.3. **Inertial Coordinates.** Recall the principle of relativity states that the laws of physics should not change in any inertial, or nonaccelerating, frame of reference. Since the physical content is given through the Minkowski metric g_M, the relativity principle can thus be formulated as follows: if one were to change from the usual Cartesian coordinates to another "inertial" coordinate system $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$, then the metric should look the same with respect to these new coordinates,

$$g_M = -d\bar{t}^2 + d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2.$$

Another way to view this mathematically is to think of this change of (global) coordinates as a mapping from $\mathbb{R}^4$ to itself,

$$(t, x, y, z) \mapsto \Phi(t, x, y, z) := (\bar{t}, \bar{x}, \bar{y}, \bar{z}).$$

From this perspective, the preceding statement about the metric being preserved can be precisely expressed as this map Φ being an *isometry*, that is, a map that does not change the geometry of the spacetime (represented by the metric g_M).[7]

So, then, what are some of these isometries, or equivalently, changes of inertial coordinate systems? The simplest ones are the *translations*,

$$(t, x, y, z) \mapsto (t + t_0, x + x_0, y + y_0, z + z_0),$$

which shift the time and spatial coordinates by a fixed amount. Another family of isometries are the *spatial rotations*, which leave t constant but rotate the (x, y, z)-coordinates. Note that these translations and rotations are also isometries of Euclidean geometry.

What is interesting, and new to Minkowski geometry, are the isometries known as *Lorentz boosts*. These can be thought of as spacetime "rotations" involving also the time coordinate. However, because of this reversal in sign for the time coordinate in g_M, the Lorentz boosts are quite different in nature from the usual rotations.

For simplicity let us consider a boost involving only t and x. This is given by

$$(t, x, y, z) \mapsto (\bar{t}, \bar{x}, \bar{y}, \bar{z}) = \left(\frac{1}{\sqrt{1 - v^2}}(t - vx), \frac{1}{\sqrt{1 - v^2}}(x - vt), y, z \right),$$

where $v \in \mathbb{R}$ satisfies $|v| < 1$.[8] This is the *Lorentz transformation* that is usually found in elementary physics texts, and it represents the change of coordinates from a constant shift in velocity in the x-direction. More specifically, if an observer is moving with velocity $(v, 0, 0)$ with respect to (t, x, y, z)-coordinates, then the same observer is at rest with respect to the $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$-coordinates. Analogous boosts can, of course, be defined involving (t, y) or (t, z), or for any composition of the the preceding three boosts. Finally, any combination of translations, rotations, and boosts produces yet another isometry.

---

[7] In differential geometric language, this means that the pullback $\Phi^* g_M$ is g_M itself.

[8] Recall that the speed of light is 1 here.

A natural consequence of this discussion is the fact that, similar to Newtonian mechanics, there is no absolute notion of velocity in special relativity. Consider, for example, an observer who is at rest with respect to one inertial coordinate system (t, x, y, z). Then, by a Lorentz transformation, we can produce another inertial coordinate system in which this observer is moving at a constant velocity. Since neither coordinate system should be favored over the other due to the relativity principle, one has no way of determining what is the "absolute velocity" of the observer, or whether this observer is "at rest". Indeed, these notions of "velocity" and "at rest" only exist relative to an observer or a coordinate system.

However, one drastic departure from Newtonian theory is that *in special relativity, there is also no canonical notion of "time elapsed" between two events*. Indeed, given two events $P, Q \in \mathbb{R}^4$, the elapsed times $t(Q) - t(P)$ and $\bar{t}(Q) - \bar{t}(P)$ with respect to two inertial coordinate systems (t, x, y, z) and $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$ need not be the same.[9] We will demonstrate this in more detail in the upcoming discussions on physical consequences.

Remark. Given an observer A, one does have the notion of *proper time*, i.e., the time lapse as measured by A itself. If P and Q lie on the curve A, then this proper time is the "length" of the curve segment between P and Q, *measured with respect to the "distance"* d_M.

1.4. **Energy-Momentum.** Consider a timelike curve α in $\mathbb{R}^4$ satisfying

$$g_M(\alpha'(\tau), \alpha'(\tau)) = -1.$$

Here, $\alpha'(\tau)$, which we note is a vector with four components, denotes the tangent vector of α at the point $\alpha(\tau)$. One can view this as the Minkowski analogue of arc length parametrization (or in other words, α is parametrized by proper time).

Suppose now that α represents a particle with mass m. We then define the *energy-momentum* vector field for this particle α by

$$P(\tau) := m \cdot \alpha'(\tau).$$

The idea is that both the energy and the momentum contents of the particle are contained within this single 4-vector field. Roughly speaking, the energy is expressed in the timelike part of P, while the momentum is expressed in the spacelike part. However, like for the notions of "velocity" at "at rest" above, this "timelike part" and "spacelike part" can only be defined relative to an observer.

To be more specific, consider another observer A, which intersects α at $\alpha(\tau)$, at which A measures the energy-momentum of α (see Figure 2). At $\alpha(\tau)$, we can decompose $P(\tau)$ as

$$P(\tau) = ET + X,$$

where:

[9] For example, this can be the case if the coordinate systems are related via a nontrivial Lorentz boost.

- $E \geq 0$ is a number.
- The vector T is unit (with respect to g_M), timelike, and points in the direction of A.
- The vector X is (Minkowski) orthogonal to T (i.e., $g_M(T, X) = 0$).[10]

The number E and the vector X represent the *energy* and the *momentum* of α at $\alpha(\tau)$, as measured by A. A quick computation yields that

$$-m^2 = g_M(P(\tau), P(\tau)) = -E^2 + g_M(X, X) \geq -E^2.$$

In particular, this implies that $m \leq E$, and that the difference $E - m$ arises from the measured momentum of the particle.

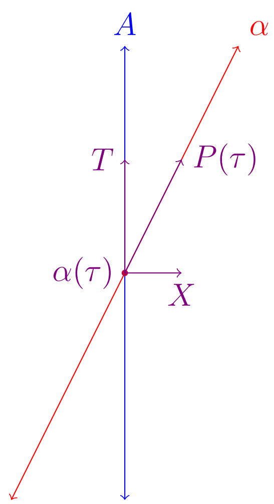

FIGURE 2. An observer A measuring the energy-momentum of α at the event $\alpha(\tau)$.

Again, the above is only defined relative to an observer A. Measuring with respect to a different observer would result in a different decomposition

$$P(\tau) = E'T' + X',$$

which would yield a different measurement of energy and momentum.

It is natural, though, to measure with respect to the particle α itself. From this frame of reference, the particle is of course at rest, resulting in the decomposition

$$P(\tau) = m \cdot \alpha'(\tau) + 0 = E_r T_r + X_r.$$

In particular, we have that

$$X_r = 0, \qquad T_r = \alpha'(\tau), \qquad E_r = m,$$

i.e., the momentum vanishes, and the *rest energy*, E_r, is just the mass m itself.

[10]In particular, X is necessarily spacelike.

Note that throughout this discussion, we had, for mathematical convenience, assumed that the speed of light is 1. If we were to have worked with a different value c for the speed of light throughout all the preceding discussions, then the revised expression for the rest energy would now be given by the infamous formula

$$E_r = mc^2.$$

In other words, from the mass, one can measure the (rest) energy of the particle, and vice versa. In physics, this is known as *mass-energy equivalence*.

2. Some Physical Consequences

To conclude our discussions and to further demonstrate the mathematical theory, we show that many of the seemingly strange physical consequences of Einstein's postulates can be quite straightforwardly derived from understanding Minkowski geometry.[11]

2.1. **Simultaneity.** The first seemingly counterintuitive consequence is that there is no absolute notion of two events happening "at the same time". This is closely tied to the fact that there is no absolute measure of time or elapsed time.

To demonstrate, we consider two observers, A and B, in Minkowski spacetime. Suppose A is at rest with respect to some inertial coordinates (t, x, y, z), that is, the curve represented by A is precisely given by $x = y = z = 0$. Furthermore, suppose B is moving at a constant velocity away from A, as depicted in Figure 3.

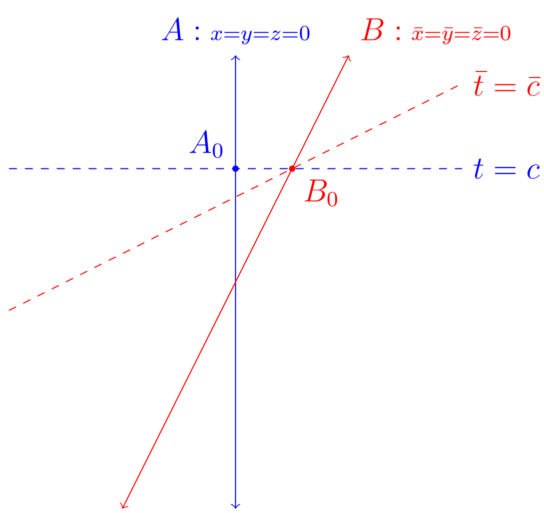

Figure 3. Two observers A and B (figure centered about A).

[11]Of course, there is a certain amount of "cheating" here. The significant achievement is Einstein's discovery of these physical consequences and, on the mathematical side, Minkowski's discovery of the rigorous geometric model for capturing these observations. However, it is still quite instructive to show just how these physical observations are naturally manifested in the mathematical theory.

Consider the event A_0—observer A at a given time. What observer A perceives as simultaneous to A_0 is the hypersurface $t = t(A_0) := c$, or, in coordinate-independent terms, the hypersurface normal to A at A_0 *with respect to the Minkowski metric*. In particular, in Figure 3, according to observer A, the event B_0 is occurring at the same time as A_0.

Next, let us consider the perspective of observer B. For this, we examine new inertial coordinates $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$, with respect to which B is at rest and A is moving at a constant nonzero velocity. At the event B_0, what B perceives as simultaneous is the hypersurface $\bar{t} = \bar{t}(B_0) := \bar{c}$, i.e., the hypersurface normal to B at B_0, *again with respect to Minkowski metric*. Note that $\bar{t} = \bar{c}$ intersects observer A at a different point than A_0. In summary, these two observers, moving at different velocities, see different events as occurring simultaneously.

Figure 4 below displays the same information as Figure 3, but with respect to the inertial coordinates centered about observer B.

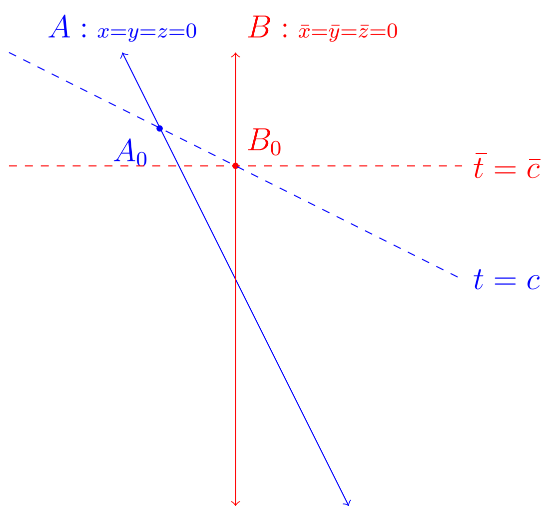

FIGURE 4. Two observers A and B (figure centered about B).

2.2. **Length Contraction.** The next consequence, often referred to as *Lorentz-Fitzgerald contraction*, is the property that observers moving at different velocities will perceive the length of an object differently. Consider an observer A and a rod, which are at rest with respect to each other. Moreover, consider another observer B, which is moving at a constant velocity away from A and the rod.

This basic setup is sketched in Figure 5. Observers A and B are represented as before, while the rod is represented by the gray shaded region. The length of the rod as measured by observer A is represented by the "length" of the bolded blue line segment ($t = c$) within the gray area. On the other hand, the length of the rod according to B is represented by the "length" of the bolded red line segment ($\bar{t} = \bar{c}$) within the gray area.

Now, Euclidean intuitions would suggest that the red line would be longer. However, recall that in Minkowski metric, the time direction has the opposite sign as opposed to the spatial

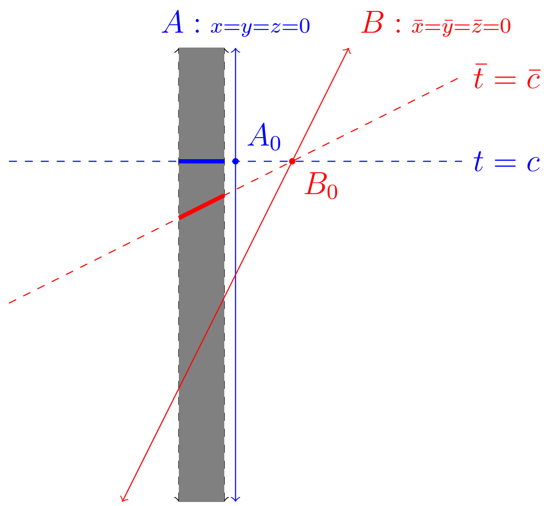

FIGURE 5. Length contraction: two observers A and B and a rod.

directions. As a result of this, the "tilt" in the dotted red line results in this red segment having a *shorter* length than the blue segment. Consequently, observer B's measurement of the rod will yield a shorter length than observer A's measurement.

In summary, this principle of length contraction states that *an observer in motion with respect to an object will perceive that object to have shorter length than an observer at rest with respect to the object.*

2.3. **Time Dilation.** Another similar phenomenon is that of *time dilation*: two observers moving at different velocities will measure different time lapses between two events. To demonstrate, we consider the situation depicted in Figure 6.

In this figure, we have two (inertial) observers A and B, both carrying clocks that are synchronized at the event O. First, A measures from its own clock the time elapsed between events O and A_0 on A. This is equal to the proper time elapsed for A, i.e., the Minkowski length of the bolded blue line segment in Figure 6 connecting O to A_0.

Suppose that, at the same time, A also measures the time elapsed on B's clock. This is the time elapsed between O and the event B_0 on B occuring at the same time (according to A) as A_0. Note this is the proper time elapsed for B, that is, the g_M-length of the bolded red segment between O and B_0. Again, Euclidean intuitions would have one believe that the red line segment would be longer than the blue. However, since both curves are timelike, and the red line picks up an extra *spacelike* component, which in the metric has the opposite sign, the actual effect is that the red line segment is *shorter* than the blue.

Thus, *A will measure less time elapsed on B's clock, which is in motion relative to A, than on its own clock, which is relatively at rest.*

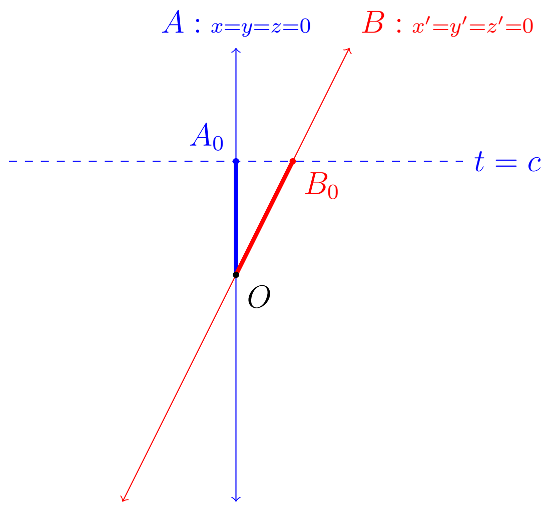

FIGURE 6. Time dilation: two observers A and B.

2.4. **The Twin Paradox.** One of the most widely discussed points of confusion when first studying special relativity is the *twin paradox* (though not so appropriately named, since there is in fact no paradox here). Mathematically speaking, this phenomenon is captured by the following rather unprovocative observation: *two different timelike curves connecting two events can have different (Minkowski) lengths.*[12]

To explain more concretely, consider the scenario depicted in Figure 7. Suppose one person, A, stays in bed at home, while a second person, B, gets in a spaceship and blasts away on a long journey. Eventually, B gets tired of traveling and flies back on the spaceship to meet A, who is still in bed and has not yet bothered to move. When A and B meet, they will observe that A *has aged more than* B.

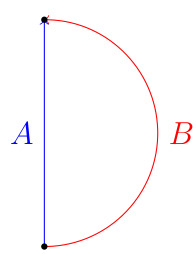

FIGURE 7. Twin paradox: A at rest, while B flies off on a spaceship.

Here, the time elapsed for A is given by its proper time, that is, the Minkowski length of the blue curve segment. Similarly, the time elapsed for B is the length of its red curve segment. When comparing the two lengths, the idea is similar to the preceding time dilation setting: the direction of B has an extra spacelike component compared to that of A. Since

[12]The analogous principle also holds, of course, in Euclidean geometry.

both curves are timelike, this extra spacelike component yields the opposite sign and makes the length of the red segment for B smaller than that of the blue segment for A.

So what exactly is the supposed paradox here? Consider now a different coordinate system $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$, in which B is at rest and A now travels in an arc; see Figure 8. If we rerun the above argument, then it would seem that B, who is now at rest, should age slower than A!

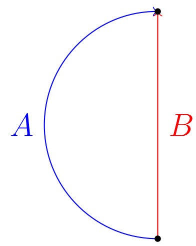

FIGURE 8. Twin paradox: coordinates adapted to B.

The error hidden in the above argument is that B is not engaged in inertial (i.e., nonaccelerated) motion, so the $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$-coordinates fail to be inertial, that is,

$$g_M \neq -d\bar{t}^2 + d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2.$$

In other words, the map from the original coordinates to the barred coordinates fails to preserve g_M. Thus, the preceding (correct) argument fails to hold in the setting of Figure 8, where the roles of A and B are interchanged.

REFERENCES

1. *Nobelprize.org: History of special relativity*, http://www.nobelprize.org/educational/physics/relativity/history-1.html.
2. *Nobelprize.org: The postulates of special relativity*, http://www.nobelprize.org/educational/physics/relativity/postulates-1.html.
3. B. O'Neill, *Semi-Riemannian geometry with applications to relativity*, Academic Press, 1983.
4. R. Wald, *General relativity*, The University of Chicago Press, 1984.
5. *Wikipedia: History of special relativity*, http://en.wikipedia.org/wiki/History_of_special_relativity.