The Geometry of Spacetime

Randy S

Abstract Article 21808 explained how to define the geometry of space by assigning a length to every finite path. This article introduces the geometry of spacetime, with emphasis on flat spacetime (special relativity). The geometry of spacetime includes a distinction between spacelike paths and timelike paths. A spacelike path has an intrinsic length, but a timelike path has an intrinsic *duration* instead. The journey of a physical object is represented by a path with an intrinsic duration.

Contents

1 The geometry of flat space: a quick review

Here's a quick review of article 21808. In three-dimensional space, a coordinate system labels each point with a unique triple of numbers x, y, z. Given a coordinate system, we can specify an arbitrary path using three functions

$$x(\lambda),\ y(\lambda),\ z(\lambda) \tag{1}$$

whose derivatives are not all zero for any λ. Different values of the real variable λ specify different points along the path, and the functions (1) give the coordinates of each of those points.

The geometry of space can be defined by assigning a length to every finite path. The familiar geometry of flat space is defined by taking

$$\text{length} = \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} d\lambda\ |\dot{s}| \tag{2}$$

to be the length of the part of the path that goes from λ_{min} to λ_{max}, where $s(\lambda)$ is the function defined by

$$\dot{s}^{\,2} = \dot{x}^2 + \dot{y}^2 + \dot{z}^2, \tag{3}$$

using an overhead dot to denote a derivative with respect to λ. Equation (3) is more commonly written as the line element

$$ds^2 = dx^2 + dy^2 + dz^2. \tag{4}$$

Intuitively, if dx, dy, dz are the changes in the coordinates x, y, z for some infinitesimal segment of a path, then ds is the corresponding length of that segment. We recognize (4) as the "Pythagorean theorem," but for infinitesimal segments so that we can use it to define the length of arbitrary curved paths. Equation (3) or (4) implicitly defines the **metric** of flat three-dimensional space.

2 Different kinds of worldlines in flat spacetime

The previous section reviewed the geometry of flat space. The geometry of flat spacetime is defined similarly, with two differences. First, we need four coordinates w, x, y, z instead of only three. We can specify an arbitrary path in spacetime (called a **worldline**) using four functions $w(\lambda), x(\lambda), y(\lambda), z(\lambda)$. Second, the quantity $\dot{x}^2 + \dot{y}^2 + \dot{z}^2$ on the right-hand side of equation (3) is replaced by

$$-\dot{w}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2. \tag{5}$$

This can be positive, negative, or zero, so clearly the left-hand side of (3) also needs to be generalized. This will be done in section 4. Each case has a different name:[1]

- A worldline is called **spacelike** wherever $\dot{w}^2 < \dot{x}^2 + \dot{y}^2 + \dot{z}^2$.
- A worldline is called **timelike** wherever $\dot{w}^2 > \dot{x}^2 + \dot{y}^2 + \dot{z}^2$.
- A worldline is called **lightlike** (or **null**) wherever $\dot{w}^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2$.

A generic worldline can have segments that are spacelike, segments that are timelike, and segments that are lightlike. If a worldline is spacelike everywhere, then we simply call it spacelike, and likewise for a worldline that is timelike or lightlike everywhere. More vocabulary:

- A worldline is called **causal**[2] if it is not spacelike anywhere.
- A point in spacetime is often called an **event**, whether or not anything actually happens there.

These definitions refer to the **causal structure** of flat spacetime. The *geometry* of flat spacetime will be defined in section 4.

[1] The definitions listed here are valid for flat spacetime in this special coordinate system. Section 19 gives a more general version of these definitions, one that works in any spacetime (flat or curved) and any coordinate system.

[2] This is prounounced "cause-uhl." The root word is *cause*, as in cause-and-effect (section 5).

3 The light cone

Each event p in spacetime has an associated **light cone**, which is is the boundary between those events that can be connected to p by a timelike worldline and those that cannot. Those that can are said to be **inside** p's light cone, and those that cannot are **outside** p's light cone.

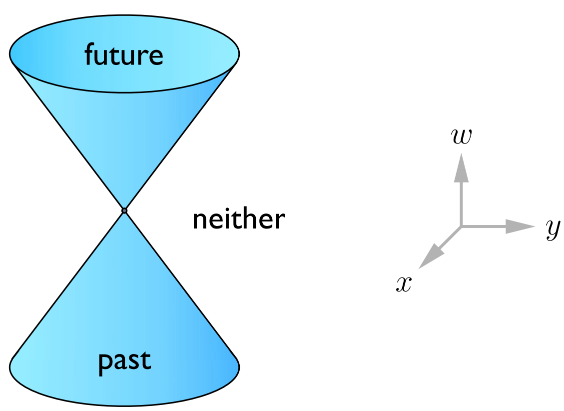
As an example, consider the event p with coordinates $(0, 0, 0, 0)$. Then an event with coordinates (w, x, y, z) can be connected to p by a timelike worldline if and only if[3]

$$w^2 > x^2 + y^2 + z^2. \tag{6}$$

This light cone is the boundary of this, which is the set of events with $w^2 = x^2 + y^2 + z^2$. The light cone separates spacetime into three regions:[4]

- Events inside p's light cone are in the **causal future** of p if $w > 0$.
- Events inside p's light cone are in the **causal past** of p if $w < 0$.
- Events outside p's light cone are not in the causal future or causal past of p, not even if $w \neq 0$. This replaces the naïve everyday concept of an event being "simultaneous" with p.

The shape of the light cone is illustrated here, without the z-dimension:

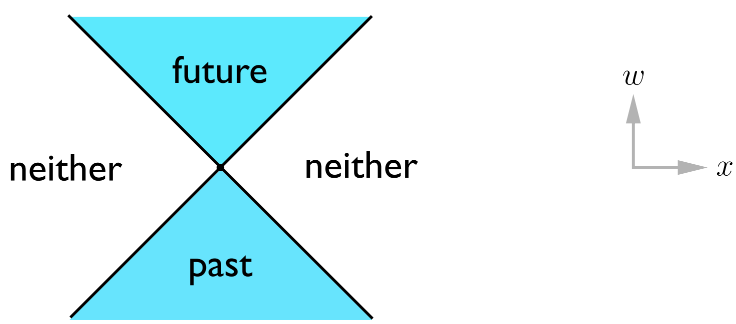


[3]This assumes that we're using the coordinate system that is used in section 2.

[4]These descriptions use a particular choice of time orientation (section 5).

Depicting the light cone in all four dimensions is difficult to do on paper, but we can easily picture it using a movie in which the w-coordinate is movie-time. As before, consider the light cone of the event with coordinates $(0,0,0,0)$. At any given movie-time w, the 3d picture shows the surface of the sphere with radius $\sqrt{x^2+y^2+z^2} = |w|$ centered on the origin $(x,y,z) = (0,0,0)$. At movie-times $w < 0$, the radius of the sphere is shrinking. At movie-time $w = 0$, the sphere is reduced to a single point $(x,y,z) = (0,0,0)$, which represents the event p itself. At movie times $w > 0$, the radius of the sphere is growing. Events inside the $w > 0$ spheres are in the causal future of p, and events inside the $w < 0$ spheres are in the causal past of p. Events outside the spheres are neither in the causal future nor past of p, which replaces the naïve everyday concept of events being "simultaneous" with p. (The following sections implicitly explain why the concept of "simultaneous" is naïve, and why the concept of events being outside each other's light cones is an appropriate replacement.)

Depicting only two dimensions, say w and x, is often more convenient. Then the light cone of a given event p consists of a pair of diagonal lines, as shown here:



In this picture,

- A worldline is lightlike wherever its slope has magnitude 45°.
- A worldline is timelike wherever its slope is closer to vertical.
- A worldline is spacelike wherever its slope is closer to horizontal.

This type of picture is used in a few of the following sections.

4 The geometry of flat spacetime

A spacelike worldline has an intrinsic length Δs, called its **proper length**, with $s(\lambda)$ defined by

$$\begin{aligned} \dot{s}^2 &= -\dot{w}^2 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2 && \text{(derivative notation)} \\ ds^2 &= -dw^2 + dx^2 + dy^2 + dz^2 && \text{(differential notation).} \end{aligned} \tag{7}$$

This replaces equations (3)-(4). A timelike worldline has an intrinsic duration $\Delta\tau$, called its **proper duration**, with $\tau(\lambda)$ defined by

$$\begin{aligned} c^2\,\dot{\tau}^2 &= \dot{w}^2 - \big(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\big) && \text{(derivative notation)} \\ c^2\,d\tau^2 &= dw^2 - \big(dx^2 + dy^2 + dz^2\big) && \text{(differential notation),} \end{aligned} \tag{8}$$

where c is a units-conversion factor that allows s and τ to be expressed in different units, such as meters and seconds.[5] The rest of this article uses **natural units** (article 37431) in which $c = 1$.

Equations (7) and (8) make sense only when their right-hand sides are non-negative, so a timelike worldline doesn't have a proper length, and a spacelike worldline doesn't have a proper duration. For a lightlike worldline, the proper length and duration are both defined, and they're both zero.

The (proper) length of a spacelike worldline is given by equation (2), but now with $\dot{s}$ defined by (7) instead of by (3). Similarly, the (proper) duration of a timelike worldline is given by

$$\text{duration} = \int_{\lambda_{\min}}^{\lambda_{\max}} d\lambda\; |\dot{\tau}| \tag{9}$$

with $\dot{\tau}$ defined by (8). Equation (7) or (8) implicitly defines the **metric** of flat spacetime.

[5] c turns out to be the local speed of light in a vacuum. often (dangerously) abbreviated “the speed of light.”

5 Physical principles

The previous section defined the geometry of flat spacetime, but physics needs more than just mathematical definitions. Physics also needs principles that relate those mathematical definitions to the real world. Here a few simple principles that we can use for classical objects with negligible size and no structure (pointlike objects):[6]

- The **principle of causality**[7] says that something which happens at one event cannot affect (cause) anything at another event unless the two events can be connected to each other by a causal (timelike or lightlike) worldline. This principle implies that the journey of a physical object can only be represented by a causal worldline, so every journey has a well-defined duration.

- The mathematical definition of duration introduced in the previous section is consistent with the physical concept of duration measured by the object's internal clock.

- If two objects meet twice, then they agree about which meeting occurred first. The equations in the previous section don't specify the sign of $\dot{\tau}$, so they don't specify which of the two directions along a causal worldline is *future* and which is *past*. However, we can choose a **time orientation**[8] that specifies *future* and *past* along all causal worldlines so that they always agree about the sequence – even though they don't always agree about the duration, as demonstrated in the next section.

[6] This is clearly an idealization, but it's good enough for many applications.

[7] This is pronounced "cause-ality."

[8] The future/past language in section 3 assumed a particular choice of time orientation.

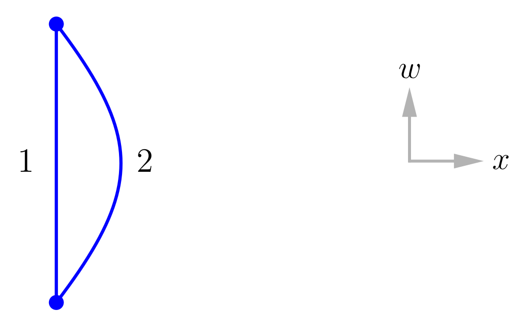
6 Comparing durations: a simple example

This section illustrates the use of equation (8) by showing that two different timelike worldlines connecting the same pair of events can have different durations. This is analogous to the familiar fact that two paths connecting the same pair of points in three-dimensional space can have different lengths (section 1).

Consider these two causal worldlines:[9]

$$\begin{aligned}&\text{Worldline 1: } (w,x,y,z) = (\lambda_1, 0, 0, 0)\\&\text{Worldline 2: } (w,x,y,z) = (A\sinh\lambda_2,\ B - A\cosh\lambda_2, 0, 0)\end{aligned}$$

where A, B are constants with $B > A > 0$. The two worldlines are parameterized by λ_1 and λ_2, respectively. These wordlines intersect each other at two events,[10] namely $(w,x,y,z) = (\pm(B^2 - A^2)^{1/2},\ 0,\ 0,\ 0)$, as illustrated here:



At the intersections, the w-coordinates of the two worldlines are equal to each other, so the changes in their parameters are related to each other by

$$(\lambda_1)_{\text{max}} - (\lambda_1)_{\text{min}} = (A\sinh\lambda_2)_{\text{max}} - (A\sinh\lambda_2)_{\text{min}}. \tag{10}$$

We want to compare the duration between the two intersections. We can calculate the duration along each worldline using (9) with $|\dot{\tau}|$ given by equation (8). For the

[9] The functions sinh and cosh are reviewed in article 77597.

[10] These are the two events at which worldline 2 has $x = 0$.

worldlines shown above, the derivatives are

$$\begin{aligned}&\text{Worldline 1: } (\dot{w},\dot{x},\dot{y},\dot{z}) = (1,0,0,0) && \Rightarrow |\dot{\tau}| = 1\\ &\text{Worldline 2: } (\dot{w},\dot{x},\dot{y},\dot{z}) = (A\cosh\lambda_2,\ -A\sinh\lambda_2,\ 0,\ 0) && \Rightarrow |\dot{\tau}| = A\end{aligned}$$

Using these results for $\dot{\tau}$ in (9) gives

$$\begin{aligned}\text{duration}_1 &= (\lambda_1)_{\text{max}} - (\lambda_1)_{\text{min}}\\ \text{duration}_2 &= (A\lambda_2)_{\text{max}} - (A\lambda_2)_{\text{min}}.\end{aligned}$$

Equation (10) implies

$$\text{duration}_1 > \text{duration}_2. \tag{11}$$

This illustrates the fact that the duration depends on the worldline, not just on the pair of events. That geometrically-obvious statement is sometimes called the **twin paradox**, because the two worldlines could represent the journeys of two identical twins – and we just demonstrated that they age different amounts between consecutive meetings.[11] It's called a paradox because we don't normally notice it,[12] so the intuition we developed as children doesn't account for it, but now we can understand it just as clearly as we understand the familiar fact that two different paths connecting the same pair of points in three-dimensional space can have different lengths.

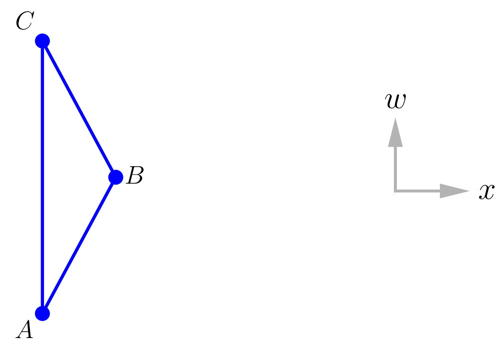
Worldline 2 looks longer than worldline 1 as drawn on the page (contrary to the inequality (11)), but that's misleading. The picture only conveys the *coordinates* of the events, not the geometry. The geometry is defined by equations (7)-(8), not by the picture. The result (11) comes from using those equations. Spacetime geometry is different than the purely spatial geometry of a page.

[11] The "aging" here is not a matter of physiology. It's a matter of spacetime *geometry*. We're not comparing how old the twins look, we're comparing how old they actually are.

[12] We don't normally notice it because, between meetings, the macroscopic objects of everyday experience all have nearly-identical worldlines: their relative speeds are all much less than the speed of light.

7 Comparing durations when one worldline is kinked

The previous example compared the durations of two smooth worldlines that start at the same event and end at the same event. We get a similar result if worldline 2 is replaced by a piecewise-smooth worldline, as illustrated here:



In this case, we want to compare the duration of segment AC (previously called worldline 1) to the duration of the kinked worldline ABC. We can handle the kink in worldline ABC just like we would when calculating the length of a path in 3d space: we subdivide the worldline into smooth segments, AB and BC, and we add their durations to get the total duration of ABC. The result is that the duration of segment AC (previously called worldline 1) is greater than the total duration of the kinked worldline ABC, just like we would have anticipated from the previous result (11).

8 Maximizing the duration

Given two events A and B that can be connected by a timelike worldline, which timelike worldline from A to B has the longest (proper) duration?

We can answer this using the calculus of variations (article 46044). The details are worked out in Martin (1988), starting in section 3.6. Here, I'll just quote the result, specialized to flat spacetime in a coordinate system where the proper time is given by equation (8). In this special case, a necessary condition for a timelike worldline to maximize the duration between a given pair of events is[13]

$$\ddot{w} = \ddot{x} = \ddot{y} = \ddot{z} = 0. \tag{12}$$

Notice that the worldline that was called "worldline 1" in section 6 satisfies this condition.

In contrast, if A and B are two events that can be connected by a spacelike worldline, then no spacelike worldline with those endpoints maximizes the proper length: the proper length can be made arbitrarily large.[14] This is possible because the right-hand side of equation (7) has three positive terms, so a spacelike worldline can "turn around and go back to where it was" without becoming non-spacelike anywhere (without making the right-hand side of (7) non-positive anywhere). A timelike worldline cannot "turn around and go back to where it was" without becoming non-timelike somewhere, because the right-hand side of equation (8) has only one positive term. That's why the duration of a timelike worldline with given endpoints has an upper limit, even though the length of a spacelike worldline with given endpoints does not.

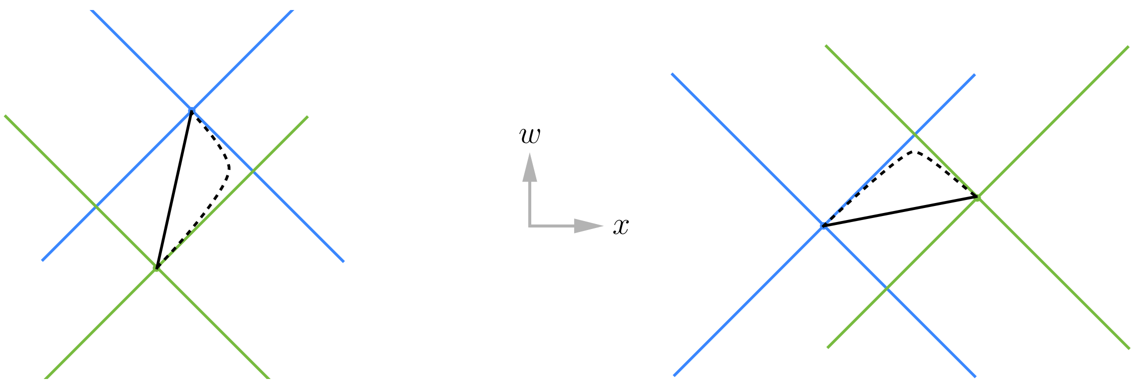
[13] In an arbitrary coordinate system, or when spacetime is not flat, the condition (12) is replaced by the general equation for a **geodesic** (section 11).

[14] This statement assumes that space is at least two-dimensional. If space were only one-dimensional (so that spacetime is two-dimensional), then the length of a spacelike worldline with given endpoints would have a finite upper limit, just like the duration of a timelike worldline with given endpoints has an upper limit.

9 Minimizing duration or length: intuition

The previous section considered the *maximum* duration (which is finite) among all timelike worldlines having the same endpoints, and the *maximum* length (which is infinite) among all spacelike worldlines having the same endpoints.

What about the *minimum* duration or length? In both cases, the minimum – actually the infimum[15] – is always *zero*. For any smooth timelike (respectively spcelike) worldline with given endpoints, we can always construct other smooth timelike (respectively spcelike) worldlines with the same endpoints and whose durations (respectively lengths) come arbitrarily close to zero. Intuitively, this is because any pair of events can always be connected by a piecewise-lightlike worldline, whose proper duration and length are both zero. To make that intuition precise, consider these pictures:



The light cones of two events are shown in blue and green, respectively. In the picture on the left (respectively right), the solid line shows a timelike (respectively spacelike) worldline connecting the two events. The dashed line shows how we can distort the original worldline, keeping it timelike (respectively spacelike) but pushing it closer and closer to the light cones in order to make its duration (respectively length) arbitrarily close to zero.

[15]The **infimum** of a set S of real numbers is the largest real number that is $\leq$ every number in S. If the infimum itself happens to be in S, then it's called the minimum.

10 Minimizing duration or length: calculation

This section confirms the intuition in the previous section by constructing an explicit family of timelike (respectively spacelike) worldlines, all with the same endpoints, whose durations (respectively lengths) come arbitrarily close to zero.

First consider the spacelike case. Suppose that the coordinates of events A and B are

$$(w,x,y,z)_A = (-a,-b,0,0) \qquad\qquad (w,x,y,z)_B = (a,b,0,0) \tag{13}$$

with $a^2 < b^2$ so that they can be connected by a smooth spacelike worldline. The general case can be reduced to (13) by translations and rotations of the x, y, z coordinates. For any $\Lambda > 0$, they are connected by the worldline

$$\begin{aligned} w(\lambda) &= S\epsilon\sinh\lambda - C\epsilon\cosh\lambda + C\epsilon\cosh\Lambda & \qquad y(\lambda) &= 0 \\ x(\lambda) &= C\epsilon\sinh\lambda - S\epsilon\cosh\lambda + S\epsilon\cosh\Lambda & \qquad z(\lambda) &= 0 \end{aligned}$$

with $-\Lambda < \lambda < \Lambda$ and $C \equiv \sqrt{S^2+1}$, where ϵ and S are chosen to satisfy the conditions $w(\Lambda) = a$ and $x(\Lambda) = b$, which imply

$$\epsilon^2\sinh^2\Lambda = b^2 - a^2 \qquad\qquad S/C = a/b. \tag{14}$$

This worldline is spacelike because $\dot{x}^2 - \dot{w}^2 = \epsilon^2 > 0$, and equation (7) says that its proper length is

$$\int_{-\Lambda}^{\Lambda} d\lambda\,|-\dot{w}^2 + \dot{x}^2| = 2\Lambda|\epsilon| = \frac{2\Lambda}{\sinh\Lambda}\sqrt{b^2-a^2}.$$

The right-hand side approaches zero as $\Lambda \to \infty$, so this completes the proof.

The proof in the timelike case is similar: simply exchange the roles of w and z.

11 Straightest and shortest: two distinct concepts

The preceding sections showed that if two events can be connected by a smooth spacelike worldline, then they can also be connected by other smooth spacelike worldlines whose lengths are arbitrarily close to zero. How do we reconcile this with the everyday idea that the distance between two points in space has a nonzero minimum, namely the length of the straightest path? To resolve this, we need to recognize that *straightest* and *shortest* are two distinct concepts. The distinction doesn't matter in space, but it does matter in spacetime.

In spacetime, the everyday concept of the *straightest* path is replaced by the concept of a **geodesic** (article 33547).[16] Section 8 mentioned that a timelike geodesic has the maximum possible duration among all timelike worldlines with the same endpoints. A spacelike geodesic has neither the maximum nor minimum possible length among all spacelike worldlines with the same endpoints (sections 8 and 9), but the length of the spacelike geodesic does correspond to the everyday concept of the length of the straightest path. As in everyday life, we can use this as the definition of *the* distance between two events.[17,18] Of course, this only makes sense if the two events can be connected by a spacelike worldline.

We should also distinguish between the concept of the distance between two *events* (previous paragraph) and the concept of the distance between two *objects*, because each object is represented by a whole timelike wordline, not by a single event. The next section addresses concept of the distance between two objects.

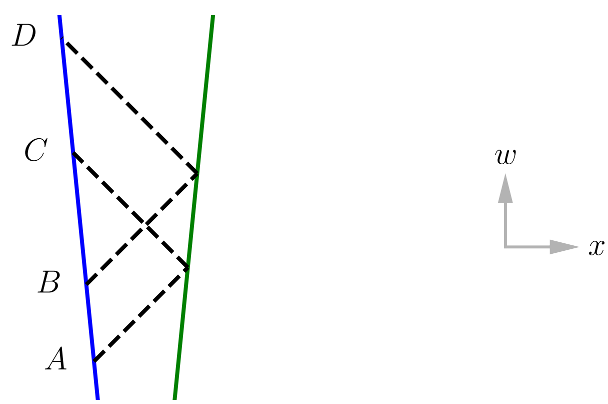
[16] In flat spacetime, in a coordinate system where geometry is defined by (7)-(8), a geodesic is a worldline satisfying (12). For other coordinate systems or in curved spacetime, the condition (12) is modified.

[17] Similarly, the everyday concept of *the* duration between two events corresponds to the length of the timelike *geodesic* connecting those two events – because the objects everyday experience all have worldlines that are close enough to being geodesics that we don't notice the difference in their durations (section 6).

[18] Any two events can be connected by a geodesic, and that geodesic is unique if spacetime is flat and topologically trivial, as we are assuming here, and in more general spacetimes if one event is within a sufficiently small neighborhood of the other (theorem 8.1.2 in Wald (1984), and page 8 in section 2 in Witten (2019)). Otherwise, two or more geodesics connecting the same pair of events may exist (section 21).

12 The distance between two objects

The previous section mentioned how a standard distance between two *events* can be defined. The concept of the distance between two *objects* is different, because each object is represented by a whole timelike wordline, not by a single event. To define the everyday concept of the distance between two objects (two timelike worldlines), we can use the time required for a lightlike signal to leave one object, reach the other object, and return to the original object. This time interval is defined along the first object's worldline, as the duration between a transmission event and the corresponding reception event. Of course, if the two objects are moving relative to each other, then this distance depends on when the signal is transmitted, as illustrated here:



The solid blue and green lines are the worldlines of the two objects. The dashed lines represent lightlike signals. The duration between events A and C along the first object's worldline (blue line) provides one measure of the distance to the other object, and the duration between events B and D along the first object's worldline provides another measure of the distance. Both measurements take time, because the transmission and receptions events are separated by a finite duration. This highlights the fact that the everyday concept of the *instantaneous* distance between two objects is only approximately meaningful: it is meaningful only if the distance is not changing too quickly. This isn't just a limitation of the measurement technology. It's a limitation of the concept itself.

13 Linear coordinate transformations

The coordinate system w, x, y, z used in the previous sections is called **Minkowski coordinates**. Now consider a new coordinate system u, v, y, z that is related to the original one by

$$w = (u + v)/2 \qquad\qquad x = (u - v)/2. \tag{15}$$

Substitute this into equation (7) to get

$$\dot{s}^2 = -\dot{u}\,\dot{v} + \dot{y}^2 + \dot{z}^2. \tag{16}$$

This defines the same spacetime geometry as equation (7), but expressed in a different coordinate system. This illustrates the fact that the same line element can look different in different coordinate systems.

More generally, if M is any invertible 4×4 matrix with constant (coordinate-independent) components, then[19]

$$(w\ x\ y\ z) = (w'\ x'\ y'\ z')M \tag{17}$$

defines a new coordinate system (w', x', y', z'). Substitute (17) into (7) to get

$$\dot{s}^2 = -(\dot{w}'\ \dot{x}'\ \dot{y}'\ \dot{z}')M\eta M^T(\dot{w}'\ \dot{x}'\ \dot{y}'\ \dot{z}')^T$$

with $\eta = \text{diag}(1, -1, -1, -1)$, and the superscript T means transpose. The previous example (15)-(16) is a special case of this.

Another important family of special cases are the **Lorentz transformations**, defined by the condition $M\eta M^T = \eta$. After a Lorentz transformation, the equation for proper length (or proper time) looks just like (7) (respectively (8)) but with the new letters in place of the old ones, so Lorentz transformations describe symmetries of flat spacetime, just like rotations describe symmetries of flat space. This symmetry of flat spacetime is called **Lorentz symmetry**.

[19] I'm using matrix notation: $(w'\ x'\ y'\ z')$ is a matrix with one row, which when multiplied into the square matrix M gives another matrix $(w\ x\ y\ z)$ with one row.

14 Comparing durations again

A coordinate transformation re-labels things, but it doesn't change the geometry. As an example, this section outlines how the analysis in section 6 can be expressed in the u, v, y, z coordinate system defined by (15).

Equations (15) imply $u = w + x$ and $v = w - x$, so the two worldlines in section 6 are described in the new coordinate system by

$$\begin{aligned}&\text{Worldline 1: } (u, v, y, z) = (\lambda_1, \lambda_1, 0, 0)\\&\text{Worldline 2: } (u, v, y, z) = \big(B - A\exp(-\lambda_2),\ -B + A\exp(\lambda_2), 0, 0\big).\end{aligned}$$

These wordlines intersect each other at two events, namely those with $u = v = \pm(B^2 - A^2)^{1/2}$ and $y = z = 0$. Between these two events, the changes in the two worldlines' parameters are related to each other by (10), as before. We want to compare the duration between the two intersections. We can calculate the duration along each worldline using (9), as before, but now $|\dot{\tau}|$ given by

$$\dot{\tau}^2 = \dot{u}\,\dot{v} - \dot{y}^2 - \dot{z}^2, \tag{18}$$

which is the proper-time equation (in units for which $c = 1$) corresponding to the proper-length equation (16). For the worldlines shown above, the derivatives are

$$\begin{aligned}&\text{Worldline 1: } (\dot{u}, \dot{v}, \dot{y}, \dot{z}) = (1, 1, 0, 0) && \Rightarrow |\dot{\tau}| = 1\\&\text{Worldline 2: } (\dot{u}, \dot{v}, \dot{y}, \dot{z}) = \big(A\exp(-\lambda_2),\ A\exp(\lambda_2),\ 0,\ 0\big) && \Rightarrow |\dot{\tau}| = A\end{aligned}$$

Using these in equation (9) gives the same durations as in section 6.

Again: coordinates are just labels. The duration of a given timelike worldline with given endpoints is the same no matter what coordinate system we use to calculate it. On the other hand, the *picture* shown in section 6 is specific to the original w, x, y, z coordinate system. If the picture were re-drawn for the u, v, y, z coordinates, then vertical and horizontal lines would be lightlike, worldline 1 would be diagonal, and worldline 2 would be tilted similarly. The picture only conveys the coordinates of the events, not the geometry. The geometry is defined by (18).

15 Other examples of coordinate transformations

Sometimes, using a coordinate system that is defined for only part of the spacetime can be useful. A familiar example is the coordinate system w, r, ϕ, z related to the original one by

$$x = r\cos\phi \qquad\qquad y = r\sin\phi \tag{19}$$

with $0 \leq \phi < 2\pi$. This is an example of **polar coordinates**. They are defined for only part of the spacetime: the coordinate ϕ is undefined at $x = y = 0$, and it's not smooth across $y = 0$ when $x > 0$. We can still use them for worldlines that avoid those parts of spacetime, and then substituting (19) into (7) gives

$$\dot{s}^2 = -\dot{w}^2 + \dot{r}^2 + r^2\dot{\phi}^2 + \dot{z}^2. \tag{20}$$

as shown in article 21808.

Another important example is the coordinate system r, ϕ, y, z related to the original one by

$$w = r\sinh\phi \qquad\qquad x = r\cosh\phi \tag{21}$$

with $-\infty < r < \infty$ and $-\infty < \phi < \infty$. This is an example of **Rindler coordinates**. They are not defined where $w^2 \geq x^2$, but we can still use them for worldlines that avoid those parts of spacetime. Use the product (Leibniz) rule to get the identities

$$\dot{w} = (\sinh\phi)\,\dot{r} + (r\cosh\phi)\,\dot{\phi} \qquad\qquad \dot{x} = (\cosh\phi)\,\dot{r} + (r\sinh\phi)\,\dot{\phi},$$

and then use these in (7) to get

$$\dot{s}^2 = -r^2\dot{\phi}^2 + \dot{r}^2 + \dot{y}^2 + \dot{z}^2. \tag{22}$$

This coordinate system is called **Rindler coordinates**.

Equations (20) and (22) both still define the same geometry as (7), at least for the part of the spacetime where the new coordinates are defined. A coordinate transformation merely re-labels things. It doesn't change the geometry.

16 Generic coordinate system

Instead of using different letters for the different coordinates, we can use an index, like this (in the four-dimensional case):[20]

$$(w, x, y, z) \to (x^0, x^1, x^2, x^3).$$

The index notation is useful when we want the equations to be valid in arbitrary coordinate systems instead of being specialized to one coordinate system. It also facilicates the generalization to N-dimensional spacetime, simply by allowing the index to take values in $\{0, 1, ..., N-1\}$. A generic coordinate transformation can be described by expressing the old coordinates x^a as functions of new coordinates X^a:

$$x^a(X^0,\, X^1,\, ...,\, X^{N-1}).$$

A worldline can be described by specifying either the old or new coordinates as functions of another parameter λ that increases monotonically along the worldline. Then the product (Leibniz) rule implies the useful identity

$$\dot{x}^a = \frac{\partial x^a}{\partial X^b}\dot{X}^b. \tag{23}$$

I'm using the **(Einstein) summation convention**: a sum is implied over any index that occurs both as a superscript and as a subscript in the same term, with the understanding that a superscript (respectively subscript) in the denominator is treated as though it were a subscript (respectively superscript) in the numerator. In particular, a sum over b is implied in equation (23).

[20] Writing the index as a superscript is conventional, as explained at the end of article 21808.

17 Flat spacetime in a generic coordinate system

Substitute (23) into equations (7)-(8) to get[21]

$$\begin{aligned} \dot{s}^2 &= -g_{ab}(X)\,\dot{X}^a\,\dot{X}^b && \text{(derivative notation)} \\ ds^2 &= -g_{ab}(X)\,dX^a\,dX^b && \text{(differential notation)} \end{aligned} \tag{24}$$

and

$$\begin{aligned} c^2\,\dot{\tau}^2 &= g_{ab}(X)\,\dot{X}^a\,\dot{X}^b && \text{(derivative notation)} \\ c^2\,d\tau^2 &= g_{ab}(X)\,dX^a\,dX^b && \text{(differential notation)} \end{aligned} \tag{25}$$

with

$$g_{ab}(X) \equiv \eta_{cd}\frac{\partial x^c}{\partial X^a}\frac{\partial x^d}{\partial X^b} \tag{26}$$

and

$$\eta_{ab} \equiv \begin{cases} 1 & \text{if } a = b = 0, \\ -1 & \text{if } a = b \in \{1, 2, ..., N-1\} \\ 0 & \text{if } a \neq b. \end{cases} \tag{27}$$

I'm using the **mostly-minus** convention, which is more convenient in contexts where causal worldlines get more attention than spacelike worldlines do, because it avoids an explicit minus sign in the proper-time equation (25). The price for this is the explicit minus sign in the proper-length equation (24), which is awkward in contexts where spacelike worldlines get more attention. The **mostly-plus** convention pushes the explicit minus sign to the proper-time equation instead, by flipping the signs in the definition of η, equation (27).

Equations (24)-(27) define the same geometry as (7)-(8), at least for the part of the spacetime where the new coordinates are defined. A coordinate transformation merely re-labels things. It doesn't change the geometry.

[21]Here, I'm using the letter X to represent the whole list of coordinates: $X = (X^0, X^1, ..., X^{N-1})$.

18 Generalization to curved spacetime

Suppose we are given the coefficients $g_{ab}(x)$ in the equations

$$\dot{s}^2 = -g_{ab}(x)\,\dot{x}^a\,\dot{x}^b \qquad \text{(derivative notation)} \tag{28}$$
$$ds^2 = -g_{ab}(x)\,dx^a\,dx^b \qquad \text{(differential notation)}$$

and

$$c^2\,\dot{\tau}^2 = g_{ab}(x)\,\dot{x}^a\,\dot{x}^b \qquad \text{(derivative notation)} \tag{29}$$
$$c^2\,d\tau^2 = g_{ab}(x)\,dx^a\,dx^b \qquad \text{(differential notation).}$$

How do we know whether these equations define the geometry of flat spacetime? Determining this can be tedious in practice, but the principle is simple: they define the geometry of flat spacetime if and only if they can be obtained from equations (7)-(8) by a coordinate transformation, as explained in the previous section.[22]

If no such coordinate transformation exists, then equations (28)-(29) may still define a valid spacetime geometry, but one that is *different* than the geometry of flat spacetime: the spacetime is curved instead of flat. To define a valid spacetime geometry (curved or flat), the only requirement is that the **metric**[23] with components $g_{ab}(x)$ has **lorentzian signature**. Section 20 explains what this means.

Article 24902 introduces one important example of a curved spacetime.

[22]The previous section used lowercase and uppercase letters to distinguish the two coordinate systems. We don't need to do that here, because here we'll only be considering one (arbitrary) coordinate system.

[23]The concept of a metric does not rely on coordinates (article 09894), but a metric can be *represented* by its components $g_{ab}(x)$ in a particular coordinate system.

19 Types of worldlines in curved spacetime

In a generic spacetime (including flat spacetime in a generic coordinate system), a worldline is called[24]

- **spacelike** if the right-hand side of (28) is positive,
- **timelike** if the right-hand side of (29) is positive,
- **lightlike** (or **null**) if the right-hand sides are zero.

A spacelike worldline (with endpoints) has a proper length, and a timelike worldline (with endpoints) has a proper duration. A lightlike worldline has both, and they're both zero. A worldline is called **causal** if it is not spacelike anywhere.

With these definitions, the principle of causality in section 5 still applies. And just like in section 3, each event p has an associated light cone, defined as the boundary between those events that can be connected to p by a timelike worldline and those that cannot. The equation for the light cone depends on g_{ab}, but the concept is the same for any g_{ab}.

[24] Recall (section 17) that I'm using the mostly-minus convention for the metric.

20 The signature of the metric

The coefficients $g_{ab}(x)$ in equations (28)-(29) might as well be **symmetric**, which means

$$g_{ab}(x) = g_{ba}(x),$$

because only the symmetric part contributes to those equations. To define a valid geometry, the matrix with components $g_{ab}(x)$ should also be **invertible** (also called **nondegenerate**) at every point x in spacetime.

If those requirements are both satisfied, then the metric has a **signature**, and its signature is the same everywhere. To define the signature, first observe that for any given point x in spacetime, the fact that the matrix with components $g_{ab}(x)$ is symmetric implies that we can choose a coordinate system in which it is diagonal at the given point x. The invertibility requirement (the second requirement highlighted above) implies that the number of positive and negative components of this diagonal matrix must add up to N, the number of spacetime dimensions. The **signature** is the pair (N_+, N_-), where N_+ and N_- are the numbers of positive and negative components, respectively, in this diagonal matrix. The signature is the same at all points in spacetime, because the invertibility requirement does not allow it to vary. These special cases have names:

- The signature $(N, 0)$ is **euclidean**. In this case, the metric defines a geometry of **space**.[25] The signature is called euclidean even if the space is curved, even though the words "euclidean geometry" typically refer to flat space.

- The signatures $(N-1, 1)$ and $(1, N-1)$ are called **lorentzian**, in the **mostly-plus** and **mostly-minus** conventions, respectively. A metric with lorentzian signature defines a geometry of **spacetime**. The signature is called lorentzian even if the spacetime is curved, even though the words "Lorentz symmetry" refer to a property of flat spacetime.

[25]The word **space** is often used more generally, not necessarily implying euclidean signature. Sometimes a spacetime (with lorentzian signature) is called a "space." As always in physics, we must rely on the context to clarify the meanings of the words.

21 Topologically nontrivial spacetimes

Most of this article assumed that spacetime has trivial topology. For a simple example of a spacetime with nontrivial topology, start with flat spacetime with the geometry defined by equations (7)-(8), and **compactify** the x-dimension by imposing the equivalence relation[26]

$$(w,\, x+\kappa,\, y,\, z) \sim (w,\, x,\, y,\, z) \tag{30}$$

for some constant κ. This means that for all w, x, y, z, the point with coordinates $(w, x+\kappa, y, z)$ is considered to be the *same point* as the one with coordinates (w, x, y, z). If we start with equations (7)-(8) and then impose the equivalence relation (30), the resulting spacetime is still geometrically flat, but it is also topologically nontrivial.

In this spacetime, consider two events A and B that can be connected by a timelike worldline. These two events can be connected by a timelike worldline that satisfies (12) and does not wrap around the x-dimension, but they can also be connected by many other worldlines that satisfy (12) and *do* wrap around the x-dimension. Most of these worldlines are spacelike, and some of them (the ones that don't wrap too many times) may be timelike if the maximum duration between A and B is large enough compared to κ. This illustrates the reason for footnote 18 in section 11.

Mathematically, we could also consider imposing an equivalence relation on the w-coordinate, but then the spacetime would have causal worldlines that close back on themselves, so two events could both be in each other's causal futures. Even in areas of physics where topologically nontrivial spacetimes are routinely considered, they are normally required to be **globally hyperbolic** so that such pathologies do not occur. Section 3 in Witten (2019) gives a good introduction to this subject.

[26] In case this sounds mysterious, here's a familiar analogy. Start with a flat two-dimensional piece of paper with a coordinate system x, y, and curl it up into a cylinder of radius R so that points with coordinates $(x + 2\pi R, y)$ and (x, y) coincide with each other. Mathematically, this doesn't require any third dimension: the equivalence relation $(x + 2\pi R, y) \sim (x, y)$ can be imposed directly, without curling anything into a third dimension.

22 The local flatness theorem

Most spacetimes are curved, not flat,[27] but learning about the geometry of flat spacetime is still an important part of understanding curved spacetime. This is partly because flat spacetime is a relatively easy special case, but it's also because of the **local flatness theorem**:[28] any spacetime is approximately flat in a sufficiently small neighborhood of any given point.[29] More precisely, for any given metric and any given point p with coordinates x_p, we can choose a coordinate system in which the Taylor expansion[30] of the components $g_{ab}(x)$ of the metric has the form

$$g_{ab}(x) = \eta_{ab} + O\big((x - x_p)^2\big), \tag{31}$$

without any term linear in $x - x_p$, where η_{ab} is defined by (27).

The significance of not having a linear term becomes more clear after learning how all of this mathematical formalism relates to gravity. That subject is beyond the scope of this article, but I'll mention the idea:[31] the worldline of an object in free-fall is a geodesic, and equation (31) implies that objects in free-fall are not being "accelerated" in the relative sense[32] at x_p. This is the essence of the **equivalence principle**, which is just the local flatness theorem applied to the physics of gravity.

[27] Among the $g_{ab}(x)$s that define valid spacetimes, most of them don't have the form (26). The euclidean-signature version of this assertion is derived in article 21808, and the derivation for lorentzian signature is similar.

[28] This theorem is well-known, but the name doesn't appear to be standard. I borrowed this name from Schutz (1985), which states the result in equation (6.3) and proves it at the end of section 6.2. You can also learn about it by searching online for **Riemann normal coordinates**. Beware that Schutz (1985) uses the name **Riemannian manifold** for what most mathematicians would call a **pseudo-Riemannian manifold**: "Riemannian" usually implies euclidean signature (except in Schutz's book), whereas "pseudo-Riemannian" encompasses all signatures.

[29] Sometimes we consider spacetimes that have **singularities**, where the geometry is not well-defined. The equivalence principle applies only in neighborhoods where the geometry is well-defined.

[30] The idea of a Taylor expansion is introduced in article 93169.

[31] Article 33547 gives a little more insight.

[32] Like many words in physics, the word "acceleration" is overloaded. An object that is being accelerated in the **absolute** (or **intrinsic**) sense has a weight. Conversely, weightlessness is the absense of absolute acceleration. The magnitude of an object's absolute acceleration is the same no matter what coordinate system we use. In contrast, acceleration in the **relative** (or **extrinsic**) sense refers to the second-derivative of the object's coordinates, so it clearly does depend on which coordinate system we use.

23 References

Martin, 1988. *General Relativity: A guide to its Consequences for Gravity and Cosmology.* John Wiley & Sons

Schutz, 1985. *A First Course in General Relativity.* Cambridge University Press

Wald, 1984. *General Relativity.* University of Chicago Press

Witten, 2019. "Light Rays, Singularities, and All That" *Rev. Mod. Phys.* **92**: 45004, https://arxiv.org/abs/1901.03928

24 References in this series

Article **09894** (https://cphysics.org/article/09894):
"Tensor Fields on Smooth Manifolds" (version 2023-11-12)

Article **21808** (https://cphysics.org/article/21808):
"Flat Space and Curved Space" (version 2023-11-12)

Article **24902** (https://cphysics.org/article/24902):
"The Ideal Non-Rotating Black Hole" (version 2023-12-09)

Article **33547** (https://cphysics.org/article/33547):
"Free-Fall, Weightlessness, and Geodesics" (version 2024-02-25)

Article **37431** (https://cphysics.org/article/37431):
"How to Think About Units" (version 2022-02-05)

Article **46044** (https://cphysics.org/article/46044):
"The Action Principle in Newtonian Physics" (version 2022-02-05)

Article **77597** (https://cphysics.org/article/77597):
"Energy and Momentum at All Speeds" (version 2022-02-18)

Article **93169** (https://cphysics.org/article/93169):
"Derivatives and Differentials" (version 2023-05-14)