8. Invariance of the Interval between Events

We have identified a variety of things that people using different inertial frames of reference disagree about: the rate of a clock, the length of a stick, whether two events are simultaneous, whether two clocks are synchronized. There are also some things people using different frames of reference do agree on: people in all frames of reference agree about space-time coincidences — whether or not two events occur both at the same time *and* at the same place; and people in all frames of reference agree about whether or not something moves at the speed of light c.

There are other quantities, that people using different frames of reference agree about. The constancy of the speed of light is, in fact, only a special case of this broader group of so-called invariant quantities. Things that everybody agrees on, regardless of the frame of reference they use, play a more important role in our understanding of the world than things whose values vary from one frame of reference to another. The theory of relativity is concerned with identifying such invariant quantities, and in this sense "Theory of Relativity" is a terrible name. "Theory of Invariance" would have been better, since the most interesting content of the theory is its identification of quantities that do not change when you change your frame of reference.

We can get a hint at what these invariants might be, by first giving a somewhat more abstract statement of the constancy of the speed of light:

Consider two distinct events[1] E_1 and E_2. Let D and T be the distance and time between the events in a particular frame of reference. If the two events happen to be events in the history of a single photon moving uniformly at speed c (for example the photon leaving a slide projector and arriving at a screen) then $D/T = c$.

Now since the speed of the photon is the same in all frames of reference, in any other frame of reference the distance D' and the time T' between E_1 and E_2 are also related by $D'/T' = c$, even though D' need not be equal to D nor need T' be equal to T. We can turn this into an alternative statement of the constancy of the velocity of light:

[1] I remind you that as used in relativity, the term "event" means something that happens at a definite place and a definite time. People using different frames of reference may, of course, use different numbers to identify the place and the time of the event, but everybody will agree that the event was not something spread out over a region of space and a period of time, but something that occurred at a specific position and at a specific moment.

If the time T and distance D between two events are related by $D = cT$ in one frame of reference, then they will be related in the same way in any other frame of reference. Putting it another way, if the time between two events in nanoseconds is equal to the distance between them in feet in any one frame of reference, then the time between them in nanoseconds in any other frame of reference will be equal to the distance between them in feet in that other frame.

We can express this relation between the time and distance between the two events in the form[2] $(cT)^2 = D^2$ or, equivalently,

$$c^2T^2 - D^2 = 0. \tag{8.1}$$

Two events which are separated by a time and a distance satisfying (8.1) are said to be "light-like separated" or to have a "light-like separation". The term is intended to remind you that a single photon can be present at both events—i.e. a photon can be produced at the earlier event that arrives at the later event just as the later event is taking place. Two such events can be bridged by a light signal. Using this terminology we can give an alternative statement of the constancy of the velocity of light: if two events are light-like separated in one frame of reference, they will be light-like separated in all frames of reference.

When stated in this way, the principle of the constancy of the velocity of light is a special case of a much more general principle. We show below that if T is the time and D is the distance between *any* events E_1 and E_2 in one frame of reference, then even when $c^2T^2 - D^2$ is not zero, its value is still the same in all frames of reference, even though T and D separately vary from one frame to another. This is called the (principle of the) invariance of the interval:[3]

For any pair of events a time T and a distance D apart, the value of $c^2T^2 - D^2$ does not depend on the frame of reference in which T and D are specified.

To see why this is so, we consider separately the two different ways in which $c^2T^2 - D^2$ can be non-zero: either cT is bigger than D or cT is less than D.

[2] A good reason for putting this in terms of the squares is that it might sometimes be useful to define T or D to be positive or negative depending on conventions about the time order of the events or the direction from one event to the other. Since two quantities that differ only in their signs have the same squares, we can include all these alternatives by writing the relation in terms of the squares.

[3] I comment below on the significance of the term "interval".

Suppose first that cT is bigger than D. Then D/T must be less than c, so it is possible for an object travelling at a speed

$$v = D/T \tag{8.2}$$

less than the speed of light, to be present at both events. The proper frame of such an object is the (unique) frame in which those two events happen in the same place. Let T_0 be the time between the two events in this special frame in which they happen in the same place. One can think of T_0 as the time between the events according to a clock that is present at both of them.

If we are given the time T and distance D between the events in any frame at all, we can figure out what T_0 must be in terms of T and D from the form $s = \sqrt{1 - v^2/c^2}$ of the slowing-down factor. For in the frame in which the events are separated in space and time by D and T, a clock that is present at both events moves with the speed $v = D/T$. Therefore in the time T between the events the clock only advances by sT. So the amount T_0 the clock has advanced between the events must be related to T by

$$T_0 = sT = T\sqrt{1 - v^2/c^2}. \tag{8.3}$$

Since $v = D/T$, it follows from (8.3) that

$$T_0^2 = T^2 - D^2/c^2. \tag{8.4}$$

So when the time T and distance D between two events are related by $T > D/c$, then no matter what frame of reference you calculate it in, $T^2 - D^2/c^2$ has the same value: it is the square of the time T_0 between the two events in that special frame in which the events occur at the same place.

There is also the case in which $cT < D$. Now D/T exceeds c so it is impossible for anything moving at less than the speed of light to be present at both events. There is no frame of reference in which the two events happen at the same place. Now, however, there is a frame of reference in which the two events happen at the same time!

To see why, consider two clocks that are stationary and synchronized in a frame in which the events are separated in space and time by D and T, with one clock present at each event. Since the time between the events is T in that frame and the clocks are stationary and synchronized in that frame, if the clock at the earlier event reads 0 then the clock at the later one must read T. Since the distance between the stationary clocks in that frame is D, we can arange for the clocks to be attached to the two ends of a stick of proper length D that is also stationary in that frame. But in a new frame, moving with speed v along the stick in the direction from the later event to the earlier one, the clock at the later event must be behind the clock at the earlier one by Dv/c^2. If we could pick v so that Dv/c^2 were equal to T, then the events would be simultaneous in the new frame. This requires that

$$v = c^2T/D = \frac{c}{D/(cT)}. \tag{8.5}$$

Since D is larger than cT the required speed v is less than c, and there is indeed a frame in which the two events are simultaneous.

Since the two events occur at opposite ends of a stick of proper length D that is moving with speed $v = c^2T/D$ in the new frame, and since the events are simultaneous in the new frame, the distance D_0 between the two events in the new frame is just the shrunken length of the moving stick. It is therefore given by

$$D_0 = sD = D\sqrt{1 - v^2/c^2}. \tag{8.6}$$

Since the speed v of the new frame is given by (8.5) we deduce from (8.6) that

$$D_0^2 = D^2 - c^2T^2. \tag{8.7}$$

So when the time T and distance D between two events are related by $D/c > T$, then no matter what frame of reference you calculate it in, $D^2 - c^2T^2$ has the same value: it is the square of the distance D_0 between the two events in a special frame in which they occur at the same time.[4]

To summarize, if D is the distance and T the time between two events then the quantity $c^2T^2 - D^2$ is independent of the frame of reference in which D and T are measured, and it is useful to distinguish between three cases:

(a) $c^2T^2 - D^2 > 0$. The events are said to be *time–like separated*, because there is a frame of reference in which they happen at the same place. In that frame they are separated *only* in time, and the time T_0 between them is given by[5] $c^2T_0^2 = c^2T^2 - D^2$.

[4] Note the pleasing resemblance between this italicized conclusion and the italicized conclusion immediately following (8.4). When distances and times are measured in feet and nanoseconds (so that $c = 1$) the two statements differ only by the interchange of space and time.

[5] Note that once you *know* that $c^2T^2 - D^2$ is independent of the frame in which D and T are measured, then it is obvious that $c^2T^2 - D^2$ is given by $c^2T_0^2$ since T_0 is the time between the events in the frame in which the distance D_0 between them is 0. It is also clear that in this case there can be no frame in which the events happen at the same time, since in such a frame T would be zero and $c^2T^2 - D^2$ could not be positive.

(b) $c^2T^2 - D^2 < 0$. The events are said to be *space–like separated*, because there is a frame of reference in which they happen at the same time. In that frame they are separated *only* in position, and the distance D_0 between them is given by[6] $D_0^2 = D^2 - c^2T^2$.

(c) $c^2T^2 - D^2 = 0$. The events are said to be *light-like separated*, because a single photon can be present at both events.

The quantity

$$\sqrt{|c^2T^2 - D^2|}$$

is called the *interval* between the two events. "Interval" is a word carefully selected to be neutral as to whether the separation it suggests is in space or in time. When $c^2T^2 - D^2$ is positive, the interval between the events (divided by c)[7] is just the time between them in the frame of reference in which they happen at the same place. When $c^2T^2 - D^2$ is negative the interval between the events is just the distance between them in the frame of reference in which they happen at the same time.

There is an intriguing analogy between this state of affairs and the purely spatial description of points in a plane. Suppose we have two points P_1 and P_2 and suppose that P_1 is a distance x to the *east* of P_2, and a distance y to the *north*. Then by the Pythagorean theorem, the direct distance d between the points is given by

$$d^2 = x^2 + y^2. \tag{8.8}$$

If, on the other hand, P_1 is a distance x' to the *northeast* of P_2 and a distance y' to the *northwest*, then again by the Pythagorean theorem, the direct distance d between the points satisfies

$$d^2 = x'^2 + y'^2. \tag{8.9}$$

Since the direct distance between the points has nothing to do whether you calculate it out of eastern and northern separations or north-eastern and north-western separations, we conclude that the value of $x^2 + y^2$ doesn't depend on the purely spatial frame of reference (known as a coordinate system) used to measure x and y.

[6] Given that $c^2T^2 - D^2$ is indeed invariant, the value of $D^2 - c^2T^2$ is obviously D_0^2 in the frame in which the events happen at the same time, since in that frame the time T_0 between them is zero. In this case there can be no frame in which the events happen at the same place, since in such a frame D would be zero and $c^2T^2 - D^2$ could not be negative.

[7] Yet another advantage of using feet and nanoseconds as the units of space and time is that this parenthetical remark would then have been unnecessary.

The remarkable discovery of relativity is that a similar relation holds for combined spatial and temporal separations. The only difference is that one subtracts rather than adds the squares to get the invariant quantity. The fact that an additional factor of c appears in the invariant quantity $c^2T^2 - D^2$ is not a significant difference, for if we had chosen to measure eastern separation x in one set of units (say yards) and northern separation y in another (say feet), a similar conversion factor between the units would have had to appear in the purely geometrical relation (8.8).[8] The factor c, which disappears if we use "natural units" of space and time like feet and nanoseconds, is just a conversion factor that comes in when we choose to use inappropriate units like feet and seconds (c=1,000,000,000 f/s) instead of feet and nanoseconds ($c = 1$ f/ns).

The reason nobody noticed the invariance of the interval for so long is again a consequence of the enormous size of the speed of light c on the kinds of scales we are used to using. For the kinds of temporal and spatial separations we are used to under everyday terrestrial conditions, T is simply not small enough, nor D large enough, for cT not to be enormously larger than D, so that $c^2T^2 - D^2$ is hardly distinguishable from c^2T^2. Under these circumstances the invariance of the interval reduces to the assertion that the time between any pair of events is the same in all frames of reference, which is exactly what people used to believe. Only when D becomes so large and/or T so small that D/T is no longer tiny compared with c does the invariance of the interval have the richer implications we now know it to have.

Here is an entertaining consequence of the invariance of the interval:

Consider two events in the history of a uniformly moving clock, a time T and a distance D apart. Since the distance between the two events is $D_0 = 0$ in the proper frame of the clock, the time T_0 ticked off by the clock between the events satisfies $T_0^2 = T^2 - D^2/c^2$, as we have already noted in (8.4). We can rewrite this relation in the form $T_0^2 + D^2/c^2 = T^2$ or, dividing both sides by T^2, as

$$T_0^2/T^2 + D^2/c^2T^2 = 1. \tag{8.10}$$

Since the clock is present at both events, D/T is just the speed v of the clock in the frame in which it moves; it tells us how many feet the position of the clock changes per nanosecond of time. On the other hand T_0/T tells us how many nanoseconds the clock ticks off per nanosecond of time in the frame in which it moves. So (continuing to use feet and nanoseconds) we have

$$(T_0/T)^2 + v^2 = 1. \tag{8.11}$$

[8] If we continued to express d in feet, the relation would become $d^2 = 9x^2 + y^2$.

The relation (8.11) tells us that the sum of the square of the speed at which a uniformly moving clock runs (in nanoseconds of clock reading per nanosecond of time) plus the square of the speed at which the clock moves through space (in feet of space per nanosecond of time) is one.[9]

Now a stationary clock moves through time at one nanosecond per nanosecond and does not move through space at all. But if the clock moves, there is a tradeoff: the faster it moves through space — i.e. the larger v is — the slower it moves through time — i.e. the smaller T_0/T is — in such a way as to maintain the sum of the squares of the two at 1. It is as if the clock is always moving through a union of space and time — spacetime — at the speed of light. If the clock is stationary then the motion is entirely through time (at a speed of one nanosecond per nanosecond). But in order to move through space as well, the clock must sacrifice some of its speed through time, in order to keep the total speed through spacetime equal to 1, as required by (8.11).

The analogy with ordinary speed along a highway is striking: a car moving east with its cruise control set to a fixed speed of 55 mph must sacrifice part of its easterly speed v_e to acquire some northerly speed v_n, because the cruise control keeps the speed of the car fixed at 55 mph, while the Pythagorean theorem requires the easterly and northerly speeds to be related by $55^2 = v_e^2 + v_n^2$.

[9] You can check for yourself that this fact can also be deduced directly from the form of the slowing down factor $s = \sqrt{1 - v^2/c^2}$ for moving clocks, without exploiting the concept of interval. There is a sense in which the invariance of the interval is the deeper of the two concepts.

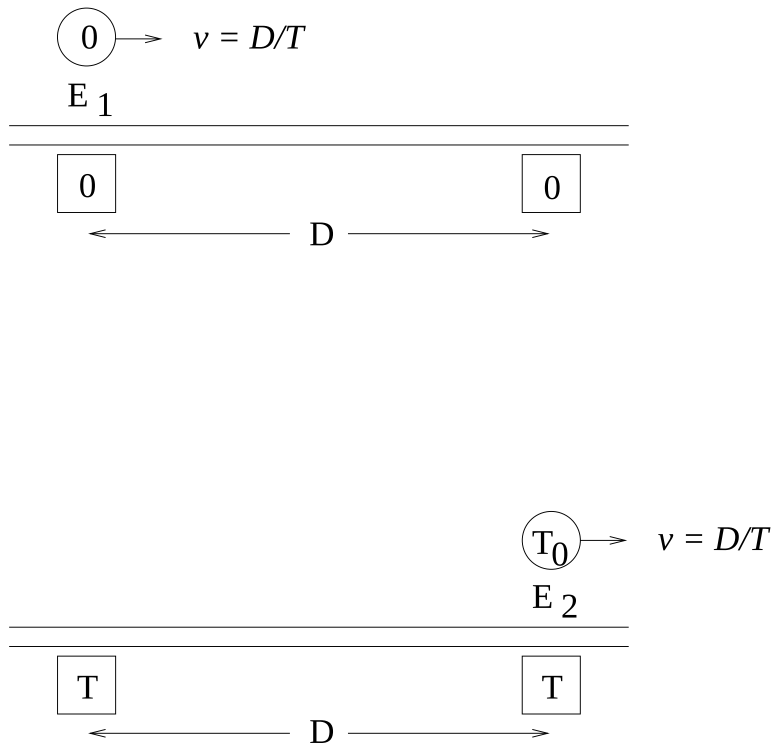

Figure 1. The invariance of the interval between time-like separated events.

The figure shows two events E_1 and E_2 that occur a time T apart and a distance D apart in the track frame. The upper part of the figure shows the track-frame situation when event E_1 occurs at track-frame time 0, as indicated by the two clocks (square boxes pictured just below the tracks) synchronized and stationary in the track frame. The lower part of the figure shows the track-frame situation when event E_2 occurs, a track-frame time T later. Both clocks have advanced by T. The track-frame distance D between the events is indicated in both parts of the figure. Because the tracks are stationary in the track frame, D is just the proper length of the portion of track stretching between the two clocks.

A third clock (the round object above the tracks) is shown, moving with speed $v = D/T$ (which is less than the speed of light c, when cT is greater than D). In the time T between the two pictures the moving clock goes a distance $vT = D$, so since it is at event E_1 in the upper picture, it has gone just the distance necessary for it to be at E_2 in the lower picture. Because the clock is moving with speed v, it runs slowly in the track frame and in a time T it only advances by $T_0 = T\sqrt{1 - v^2/c^2}$. As a result T_0 is related to T by $T_0^2 = T^2 - v^2T^2/c^2 = T^2 - D^2/c^2$.

This establishes that $T^2 - D^2/c^2$ is the time that has elapsed between the events in the special frame in which they happen at the same place.

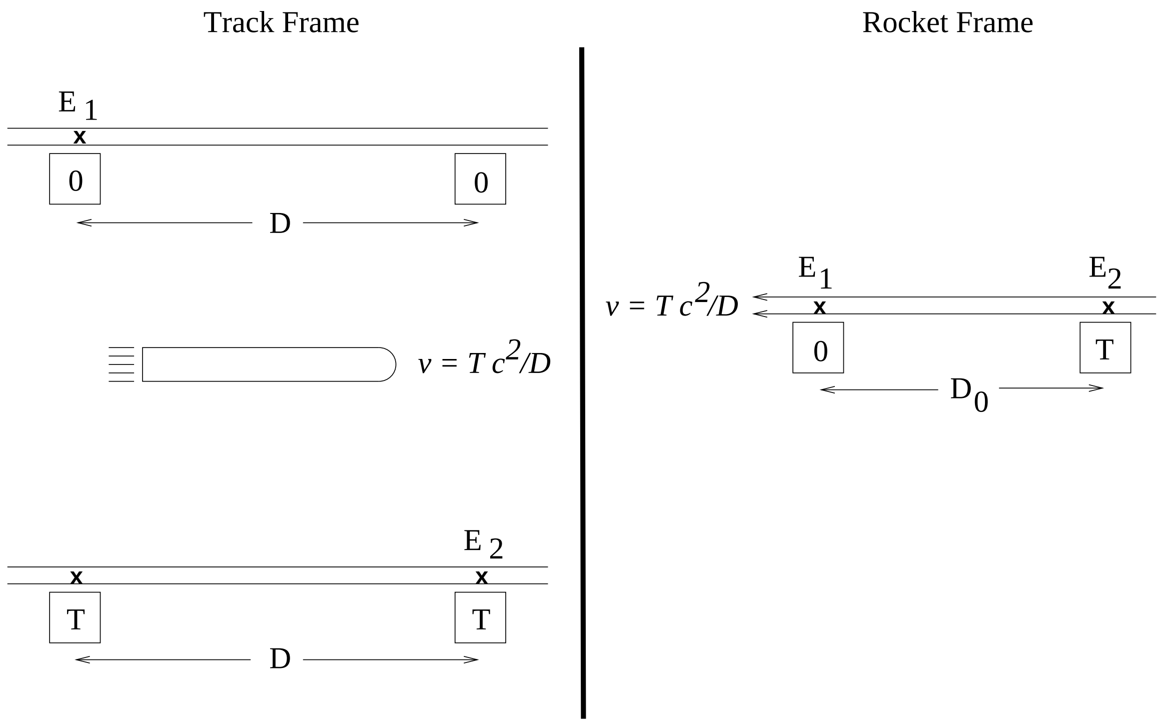

Figure 2. The invariance of the interval between space-like separated events.

The part of the figure to the left of the heavy line shows events E_1 and E_2 that occur a time T and distance D apart in the track frame. The upper part on the left shows the track-frame situation when E_1 occurs at track-frame time 0, as indicated by two clocks (square boxes pictured just below the tracks) synchronized and stationary in the track frame. The lower part on the left shows the track-frame situation when event E_2 occurs, a track-frame time T later. The track-frame distance between the events is D. Because the tracks are stationary in the track frame, D is just the proper length of the track that stretches between the two clocks. As each event occurs it makes a mark (x) on the tracks.

A rocket (the long object in the middle of the left half of the figure) is shown, moving to the right with speed $v = c(cT/D) = Tc^2/D$ (which is less than the speed of light c, when D is greater than cT). In the rocket frame the track-frame clocks move to the left with speed v and the clock on the left (the clock in front) is behind the clock on the right (the clock in the rear) by Dv/c^2. Because $v = Tc^2/D$, the clock on the left is behind the clock on the right by exactly T in the rocket frame, so in the rocket frame when the clock on the right reads T the clock on the left only reads 0. This means that the events are simultaneous in the rocket frame. Because the events are simultaneous in the rocket frame, the moving track has no time to change its position between the events, so the distance D_0 between the events is given by the distance between the two marks (x) on the moving track. This distance is the length D of track between the two marks in the track-frame, reduced by the shrinking factor $\sqrt{1 - v^2/c^2}$. As a result D_0 is related to D by $D_0^2 = D^2 - v^2D^2/c^2 = D^2 - c^2T^2$.

This establishes that $D^2 - c^2T^2$ is the distance between the events in a frame in which they happen at the same time.