II. Review of Special Relativity

The universe is described by a 4-dimensional space-time continuum. A proper description of standard cosmological models requires the use of relativistic mechanics. Although a full treatment of cosmological models requires the use of general relativity, it turns out that most essential features can be derived by slight generalizations of special relativity. This chapter reviews the essential concept of special relativity and then focuses on the use of metrics in describing the space-time continuum.

Principles of Special Relativity

An important precept of special relativity is that one can establish global inertial frames of reference. Although such frames have 4 dimensions (three space plus one time), it is sufficient to consider just one spatial dimension (called it the x direction) and time t. Within such a frame (call it frame 1), one can distribute observers who are at rest with respect to one another, all of whom have the same coordinated time t and who are separated by well-defined distances x. x and t are each arbitrary in two ways: they can have an arbitrary origin and arbitrary dimensions. An event is defined to be a particular point in the space-time continuum that is specified by a unique x and t. From an operational point of view, the coordinates of any event are measured by the observer who happens to be local to that event (such that light travel time delays can be ignored). One can then imagine that that observer then transmits the information to more distant observers in its frame of reference if necessary. A moving particle follows a path $x = x(t)$. If the particle is in free motion and passes throught the point $x = 0$ at $t = 0$, then that path is $x = vt$, where v is a constant velocity. There is nothing unconventional about these concepts, and they are common to Newtonian dynamics.

Consider now a second inertial frame (frame 2) moving at velocity v with respect to the first in the $+x$ direction which has its own measures of time t' and distances x'. Let the origin of the space and time dimensions be defined so that the event $x = 0$, $t = 0$ coincides with the event $x' = 0$, $t' = 0$. The unconventional aspect of special relativity is how the coordinates of other events in the two inertial frames are related. Let $\beta = v/c$ (where c is the speed of light). Then

$$x' = \frac{x - vt}{\sqrt{1-\beta^2}}, \qquad (2.1a) \qquad\qquad x = \frac{x' + vt'}{\sqrt{1-\beta^2}}, \qquad (2.1c)$$

$$t' = \frac{t - vx/c^2}{\sqrt{1-\beta^2}}, \qquad (2.1b) \qquad\qquad t = \frac{t' + vx'/c^2}{\sqrt{1-\beta^2}}. \qquad (2.1d)$$

These equations abandon the concepts of the existence of a universal time and of universal lengths. It is important to recognize that not only are length and time scales altered, but also the concept of simultaneity of events between different frames of reference is destroyed (equations 2 and 4). For example, this means that if all the observers in frame 1 compare their clocks at $t = 0$ with their neighboring counterparts in frame 2, then the times showing on the clocks of the frame 2 observers will not all be the same but rather will vary with position according to $t' = -vx/c^2\sqrt{1-\beta^2}$. This destruction of simultaneity leads to virtually all of the apparent paradoxes in special relativity. It means that one must be particularly careful about what is meant by measuring lengths and times of observers and objects in moving frames.

The origin of frame 2 follows the path $x' = 0$. In frame 1, this path is given by $x = vt$. The origin of frame 1 follows the path $x = 0$. In frame 2, this path is given by $x' = -vt'$. Hence aside from the sign of v, the two frames observe each other in a symmetric fashion.

Consider the path of a ray of light. In frame 1, the ray follows the path $x = ct$. In frame 2, the above equations (either 2.1*a* and 2.1*b* together or 2.1*c* and 2.1*d* together) give $x' = ct'$. Hence the speed of a light ray is independent of the velocity of the frame of reference. In special relativity, this property is assumed to be fundamental, and one can then derive equations 2.1 accordingly.

To observers in frame 1, a standard yard stick moving by in frame 2 appears to be shorter than 1 yard. This result appears paradoxical since an observer in frame 2 ought to see a frame 1 yardstick to be shorter as well, yet both ought not to be true at the same time. The paradox is resolved by considering in detail just how length measurements are made. Consider two observers in frame 1 who measure the length of a yardstick moving by in frame 2. Since the yardstick is moving, it is necessary to sample the beginning and end positions of the yardstick at the same time t in order to make a meaningful measurement. ΘBy equation 2.1*a* above, if the sampling occurs at $t = 0$ as measured in frame 1 and one end of the yardstick lies at $x = x' = 0$ at that time, then the other end of the yardstick has a coordinate $x = x'\sqrt{1-\beta^2}$. with $x' = 1$ yard. Imagine that the two observers (at 0 and x in frame 1) slap paint brushes on the yardstick at that time. The observers in frame 2 will not see the paint slaps occurring simultaneously, but the one at the front of the yardstick will see a slap occuring earlier, at time $t' = -vx'/c^2$. Hence the frame 2 observers will think that the frame 1 observers have not made a proper measurement.

Another paradox of special relativity is the effect of "moving clocks run slow". The easiest way to resolve this paradox is to consider two ways that observers in frame 1 can compare their clocks with observers in frame 2. The first way is to consider a clock in frame 2 that sits at $x' = 0$ and is compared with the nearest clock in frame 1 as time passes and the clock in frame 2 moves relative to frame 1. By equation 2.1*d* above, $t' = t\sqrt{1-\beta^2}$, *i.e.*, the moving clock ticks at a slower rate than the clock in frame 1. However, another way to compare the clocks is for an observer in frame 1 to sit at rest at $x = 0$ and look at the nearest clock in frame 2 as it passes by. Then by equation 2 above, one finds $t' = t/\sqrt{1-\beta^2}$, and so clocks in the moving frame appear to tick faster. The differences arise because the details of the measurement process are different.

The above examples show that the concepts of measuring distance and time do not have the same meaning in special relativity as they do in Newtonian physics, where time and distances can be measured independent of the motion of the frame of reference. However, the concepts of absolute time and distances are so useful that we might ask if they can be recovered even within special relativity. The answer is yes. Consider once again the process by which we measure "lengths" of moving yardsticks: because the yardstick is moving, what we do is specify two "events" which happen to be the locations of each end of the yardstick at a specific time as measured in frame 1 and then measure the distance between the two events. Now we have not specified yet how to measure the distance between two events that do not occur at the same time but we shall do so imminently. In any case, if the two events do occur at the same time (as measured in our rest frame), then it makes sense to define the distance between them to be the conventional linear distance as measured with a yardstick. This reasoning applies to conventional Newtonian physics as well. The major difference arises when we consider how to measure the distances between two events that do not occur simultaneously. One measures both a difference in position (Δx) and a difference in time (Δt). In Newtonian physics, both invariant with respect to the observing frame. However, in special relativity, that is not the case. Nevertheless, one can construct a single invariant quantity by the following logic. Consider two spatially separated events that occur nearly simulaneously in some reference frame. One can find a moving reference where the two events are in fact precisely simultaneous and then the distance between them is the conventional distance measured with a yardstick. We define this distance to be the proper distance s between the two events. Let us say that that occurs in frame 1 and that both events occurs at $t = 0$ with one being $x = 0$ and the other at $x = s$. The question then arises, if we measure coordinates for the the other event x', t', in some other frame of reference, can we combine them in some fashion so as to recover the proper distance? The answer is yes, and to do so one can take equations 2.1a and 2.1b above, set $t = 0$, combine the two to eliminate v, and then solve for x. The solution is $s^2 = x^2 = x'^2 - c^2 t'^2$.

If the two events occurs close spatially but well separated in time, then it may not be possible to find a frame where both occur simultaneously. In this case, however, it is possible to find a frame where both occur at the same spatial position but separated in time. The time difference between the two events as measured by the single observer in this frame is called the proper time τ between the two events. By the same reasoning as before, we find that $\tau^2 = t'^2 - x'^2/c^2$. The choice between s or τ between two events depends on whether the interval between them is "spacelike" or "timelike".

To recapitulate, the way to measure the lengths of things is to go to a frame where those things are at rest and then specify two events which are at either end of the "thing" and which occur simultaneously in that rest frame. The proper distance between those events is the convential yardstick measurement of distance. Likewise, the way to measure elapsed time intervals is to go to a frame where the two events that one wants to time occur at the same position. Then the proper time is the ordinary elapsed clock time measured by an observer at that position.

Motion in Special Relativity

The concept of proper time can be generalized to the case of motion by an accelerated particle. Along any small segment of the path followed by the particle, the change in proper time $\Delta\tau$ is given by $\Delta\tau^2 = \Delta t^2 - \Delta x^2/c^2$ where x and t are coordinates of any inertial frame. Suppose the path of the particle is parameterized by $x = x(t)$. Then the elapsed proper time (= elapsed time of a clock carried along with the particle) is

$$\tau = \int \sqrt{1 - (dx/dt)^2}dt. \tag{2.2}$$

Consider once again two events separated by a time-like interval. One can draw many paths that connect those two events, corresponding to different accelerated observers. Each path measures its own elapsed time τ. One of those paths has a maximum; it is called a geodesic. The equation describing a geodesic can be found as follows. Equation 2.2 gives the elapsed proper time τ along some path $x = x(t)$. Pick a new path $x'(t) = x(t) + \delta x(t)$, where δx is a small deviation that is 0 at either endpoint of the path. The extra proper time along the new path is (after substitution and expansion),

$$\delta\tau = -\frac{1}{c^2}\int \frac{(dx/dt)(d\delta/dt)}{\sqrt{1-(dx/dt)^2/c^2}}dt. \tag{2.3}$$

Integrating by parts, and remembering that δx is 0 at either endpoint of the integral, one obtains

$$\delta\tau = +\int \delta x(t)\ddot{x}\frac{d}{dt}\left[\frac{\dot{x}}{\sqrt{1-\dot{x}^2}}\right]dt. \tag{2.4}$$

If this is to be true for any arbitrary path deviation $\delta x(t)$, then the quantity in brackets must be independent of time. We recognize this quantity as the relativistic expression for linear momentum per unit rest mass. Thus, if the elapsed proper time is an extremum, then the linear momentum is conserved, which is equivalent to saying that the velocity is constant.