Invariant Distance in Euclidean Geometry

According to Euclidean geometry, it is possible to label all space with coordinates $x, y, z$ such that the square of the distance between a point labeled by $x_1, y_1, z_1$ and a point labeled by $x_2, y_2, z_2$ is given by $(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$. If points 1 and 2 are only infinitesimally separated, and we call the square of the distance between them $(d\ell^2)$, then we could write this rule, that gives the square of the distance as

\[d\ell^2 = dx^2 + dy^2 + dz^2 \tag{1}\]

This rule has physical significance. The physical content is that if you place a ruler between these two points, and it is a good ruler, it will show a length of $\sqrt{d\ell^2}$. Since it is difficult to find rulers good at measuring infinitesimal lengths, we can turn this into a macroscopic rule. Imagine a string following a path parameterized by $\lambda$, from $\lambda = 0$ to $\lambda = 1$, then the length of the string is $\int_0^1 d\lambda (d\ell/d\lambda)$. That is, every infinitesimal increment $(d\lambda)$ corresponds to some length $(d\ell)$. If we add them all up, that's the length of the string.

There are many ways to label the same set of points in space. For example, we could rotate our coordinate system about the $z$ axis by angle $(\theta)$ to form a primed coordinate system with this transformation rule

\[z' = z\]

\[y' = x \sin\theta + y \cos\theta\]

\[x' = x \cos\theta - y \sin\theta\]

Under such a relabeling, the distance between points 1 and 2 is unchanged. Physically, this has to be the case. All we've done is used a different labeling system. That can't affect what a ruler would tell us about the distance between any pair of points.
Further, for this particular transformation, the equation that gives us the distance between infinitesimally separated points has the same form. Show that this distance rule:

\[(d\ell')^2 = (dx')^2 + (dy')^2 + (dz')^2 \tag{2}\]

applied to the prime coordinates, gives the same distance; i.e, show that \((d\ell' = d\ell)\). Because this distance is invariant under rotations of the coordinate system, we call it the invariant distance.

We want to emphasize that the labels themselves, \(x, y, z\) or \(x', y', z'\) have no physical meaning. All physical meaning associated with the coordinates comes from an equation that tells us how to calculate distances along paths. To drive this point home, note that we could also label space with a value of \(x, y, z\) at every point, but do it in such a way that we would have the distance between \(x, y, z\) and \((x+dx, y+dy, z+dz)\) have a square given by

\[(d\ell^2 = dx^2 + x^2(dy^2 + \sin^2y dz^2) \tag{3}\]

For many readers, this result would look more familiar if we renamed the coordinates \((r=x), (\theta = y),\) and \((\phi = z)\) so that we get another expression for the invariant distance,

\[(d\ell^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{4}\]

This is the usual spherical coordinate labeling of a 3-dimensional Euclidean space by distance from origin, \(r\), a latitude-like angle, \(\theta\), and a longitudinal angle, \(\phi\). Show that the invariant distance given by Eq. TBD and the invariant distance given by Eq. TBD are consistent if the coordinates are related via

\[z = r\cos\theta\]
\[y = r\sin\theta \cos\phi\]
\[x = r \sin \theta \sin\phi\].

We now want to use this, perhaps, unfamiliar formulation, to derive some well-known results. We are going to calculate the ratio of circumference to radius for a circle. Imagine a path parameterized by \(\lambda\) such that \((z=0), (x=r_1 \cos\lambda)\) and \((y = r_1\sin\lambda)\). This path is clearly periodic in \((\lambda)\). What is the length of the path over one period? Use our rule for how to calculate the length of a path. You already have in your mind that this path is that of a circle. Let's prove it. To do so, we need to know that every point on the path is the same distance from the origin. This is more easily done in spherical coordinates. Consider this path:

\[\phi = \lambda\]
\[\theta = \pi/2\]
\[r = r_1\]

You can easily check, using the Spherical to Cartesian coordinate transformation, that this is the same path, just expressed in
different coordinates. You can also use equation 4 to show that the length of the path from $\lambda = 0$ to $2\pi$ is $2\pi r_1$ which should be a confirmation of the result you got in Cartesian coordinates.

In these new coordinates we want to show that this is a circle. So consider the path parameterized by $\mu$ that runs from the origin of coordinates $(r=0)$ out to $(r=r_1)$ at fixed $(\phi = \lambda)$ and $(\theta = \pi/2)$. The path is simply given by

\[
\begin{align*}
\phi &= \lambda \\
\theta &= \pi/2 \\
r &= \mu
\end{align*}
\]

for $\mu$ going from 0 to $\mu = r_1$. Show that $\int_0^{r_1} (d\ell/d\mu) d\mu = r_1$, a result that is independent of $\lambda$ and thus all points in the path parameterized by $\lambda$ are the same distance, $r_1$, from the origin.

You have now demonstrated that the path parameterized by $\lambda$ is a circle, that it has radius $r_1$ and it has circumference $2\pi r_1$. This may seem tedious. All we have managed to achieve is to derive a simple geometrical fact that you already know. The value will come soon, when you will use the same techniques to calculate the relationship between radius and circumference in geometries for which this well-known result no longer holds.

Before going on, we should take a little more care. We have shown that a particular path that takes us from the origin to $(r=r_1)$, $(\theta = \pi/2)$ and $(\phi = \lambda)$ has distance $r_1$ independent of $\lambda$ and thus all points in the path parameterized by $\lambda$ are the same distance, $r_1$, from the origin.

For $J = \int_1^2 d\mu f(q_i, \dot q_i, \mu)\ d\mu$ where $(\dot q_i \equiv dq_i/d\mu)$, the path from point 1 to 2 that extremizes $\int J$ satisfies these equations

\[
\left[\frac{d}{d\mu}\left(\frac{\partial f}{\partial \dot q_i}\right)\right] = \frac{\partial f}{\partial q_i}
\]

This is a mathematical result with more than one application. In mechanics, the action is given as an integral over the Lagrangian so that

\[
S = \int dt L(q_i, \dot q_i, t)
\]

and because a system passes from point 1 to point 2 by minimizing the action, the equations of motion will satisfy

\[
\left[\frac{d}{dt}\left(\frac{\partial f}{\partial \dot q_i}\right)\right] = \frac{\partial f}{\partial q_i}
\]

which you know as the Euler-Lagrange equations.

In the case at hand we have length = $\int d\mu \left[\sqrt{\left(\dot r^2 + r^2(\dot \theta^2 + \sin^2 \theta \dot \phi^2)\right)}\right]$ where

\[
f = \left[\sqrt{\left(\dot r^2 + r^2(\dot \theta^2 + \sin^2 \theta \dot \phi^2)\right)}\right]
\]
so the path with shortest length path between any two points should satisfy

\[
\frac{d}{d\mu} \left( \frac{\partial f}{\partial \dot{r}} \right) = \frac{\partial f}{\partial r}
\]

\[
\frac{d}{d\mu} \left( \frac{\partial f}{\partial \dot{\theta}} \right) = \frac{\partial f}{\partial \theta}
\]

\[
\frac{d}{d\mu} \left( \frac{\partial f}{\partial \dot{\phi}} \right) = \frac{\partial f}{\partial \phi}
\]

These equations are kind of hairy, if you work them out in generality. However, we are testing to see if a particular path satisfies them, the path from the origin to \(r=r_1\), \(\theta = \pi/2\) and \(\phi = \lambda\) that proceeds at fixed \(\theta\) and \(\phi\). For this reason we have \(\dot{\theta} = 0\) and \(\dot{\phi} = 0\) which really simplifies the evaluation of the above equations. We will just do one term out of the first equation as an example, and leave evaluation of the rest of the terms as an exercise. In particular, we evaluate \(\frac{\partial f}{\partial r} = \frac{r}{f} (\dot{\theta}^2 + \sin^2 \theta \dot{\phi}^2) = 0\), where the last equality follows since \(\dot{\theta}\) and \(\dot{\phi}\) are both zero. The other terms rapidly simplify for the same reason and one can verify that the given path does indeed satisfy the equations for paths that connect two points with the shortest possible distance.

The notion of relativity of motion is not something new with Einstein. It's built into Newtonian mechanics as well. The idea is that the results of experiments done in any inertial frame should be the same. An inertial frame is one in which Newton's laws of motion are satisfied. These laws prescribe accelerations, not velocities, so a frame that is moving at constant relative velocity with respect to an inertial frame must itself be an inertial frame. If the unprimed coordinates are Euclidean coordinates then the primed ones given by

\[
[t' = t]
\]

\[
[x' = x - vt]
\]

\[
[y' = y]
\]

\[
[z' = z]
\]

presumably are as well. Because the relationship is time dependent (and because we are anticipating the generalization to Lorentz transformations), we have explicitly included time in the transformation, even though time transforms trivially.

We will call such a transformation a "Galilean Boost."

One can quickly verify that, if evaluated at the same time \(t\), the distance between two infinitesimally separated points is unchanged by this transformation. Further, the transformation is symmetric. A particle that is fixed in the unprimed coordinate system, has speed \(\sqrt{(dx'/dt)^2} = v\) in the primed coordinate system, and a particle that is fixed in the primed coordinate system has speed \(\sqrt{(dx/dt)^2} = v\) in the unprimed system.

With the Galilean boost transformation, velocities add in a simple manner. If \(u' = dx'/dt\) where \(x'(t)\) is the \(x'\) location of some object over time, then \(u = dx/dt = u' + v\) follows straight from the transformation equations.
The Galilean Boost equations are not wrong. They simply define a new way of labeling space and time. However, somebody using these coordinates would find some odd things. For example, a clock staying fixed to the origin of the primed system, in going from \(t' = t'_1\) to \(t' = t'_2\), would indicate the amount of time elapsed was not \(t_2' - t_1'\) but instead \(\sqrt{1-v^2/c^2}(t_2' - t_1')\) where \(c\) is the speed of light.

This looks like a failure of relativity. Has the clock started behaving weirdly because it's moving? Are the laws of physics not obeyed in all inertial frames? As you know, because you have studied special relativity, this is not the case. It turns out that the "Galilean Boost" can be generalized to a "Lorentz Boost",

\[
\begin{align*}
  t' &= \gamma (t-vx/c^2) \\
  x' &= \gamma (x - vt) \\
  y' &= y \\
  z' &= z
\end{align*}
\]

where \(\gamma \equiv 1/\sqrt{1-v^2/c^2}\). I refer to this as a generalization because in the limit that \(c \rightarrow \infty\) this reduces to the Galilean boost. These primed coordinates, as opposed to the primed coordinates of the Galilean boost, give sensible results. For example, a clock staying fixed to the origin of the primed system, in going from \(t' = t'_1\) to \(t' = t'_2\), would indicate the amount of time elapsed was indeed equal to \(t_2' - t_1'\).

Unlike rotational coordinate transformations, that preserve the rule governing spatial distances between pairs of points, a Lorentz transformation does not. The spatial separation between \((x,t)\) and \((x+dx,t)\) is \(dx\). The spatial separation between these points in the prime frame is \(\gamma dx\), as one can see from the transformation rule. How can length depend on reference frame? Key to resolving this apparent paradox is the fact that in the primed frame the two events are not simultaneous. We won't go through sorting out these apparent paradoxes here.

We will, however, introduce a quantity that, unlike length, has a rule that is invariant under Lorentz transformations. For Cartesian coordinates, the square of the invariant distance between event \((t,x,y,z)\) and event \((t+dt,x+dx, y+dy, z+dz)\) is given by

\[
\begin{align*}
  ds^2 &= -c^2 dt^2 + dx^2 + dy^2 + dz^2
\end{align*}
\]

This quantity has the following two-part physical interpretation:

1. For \(ds^2 > 0\), \(\sqrt{ds^2}\) is the length of a ruler that connects the two events and is at rest in the frame in which the two events are simultaneous.
2. For \(ds^2 < 0\), \(\sqrt{-ds^2}\) is the time elapsed on a clock that moves between the two events with no acceleration.

Why is this quantity invariant under boosts? That's a deep question, and I'm not sure anyone knows. Essentially, we use this rule because it is consistent with what we have noticed from experiments. Maxwell's equations are a synthesis from experiments. Their form is invariant under a Lorentz transformation, not a Galilean transformation. The Lorentz
transformation preserves the invariant distance (as you should verify).

Exercise 1) Bob and Alice are trying to figure out how to put their 5 meter ladder into the shed, but the shed is only 4 meters long. Bob proposes that he takes the ladder and runs as fast as he can, 0.80 times the speed of light, so that the ladder is shrunk in the shed's reference frame. The shed has doors on either end, so when Bob runs into the shed Alice will close the door behind him. What does Alice see? Does Bob get the whole ladder inside the shed before he bursts through the back door? What does Bob see?

Exercise 2) Events A and B occur 10 meters and 100 ns apart in time in frame 1. If they occur 25 ns apart in frame 2, what must their spatial separation be?

Exercise 3) Events A and B occur at the same time at a different space in reference frame 1. Is there another reference frame in which events A and B occur at the same space, but not at the same time?

Before continuing with spacetime, we temporarily retreat to the study of space alone. We do so to introduce the notion of "curvature." Previously we asserted that one could label space with coordinates \( r \), \( \theta \), and \( \phi \) such that points separated by \( dr \), \( d\theta \), and \( d\phi \) would be separated by a distance (as one would measure with a ruler) with square given by

\[
s^2 = dr^2 + r^2(\,d\theta^2 + \sin^2 \theta \,d\phi^2)\]

A space that can be labeled in this way is homogeneous (invariant under translations) and isotropic (invariant under rotations). The easiest way to see this is to remember that there's a coordinate transformation to Cartesian coordinates for which

\[
s^2 = dx^2 + dy^2 + dz^2\]

Now the homogeneity is more evident, since transforming \( x \) to \( x' = x + L \) would clearly leave the distance rule unchanged. We've also already seen that rotations leave the distance rule unchanged. So, the space is homogeneous and isotropic. If we choose to label it with spherical coordinates about a particular origin, our labeling is inhomogeneous and anisotropic, but not the space itself.

It turns out that whether one can label space in this way or not is a matter to be settled by experiment. It's not necessarily true. Even if we restrict ourselves to completely homogeneous and isotropic geometries, we can mathematically describe spaces that cannot be labeled in this way.

What is generally true is that all three-dimensional homogeneous and isotropic spaces can be labeled with coordinates \( r \), \( \theta \), and \( \phi \) such that

\[
s^2 = \frac{dr^2}{1-kr^2} + r^2(\,d\theta^2 + \sin^2 \theta \,d\phi^2)\]

for \( k \) a constant that can be positive, negative or zero. Euclidean space is a special case with \( k=0 \).

We will get around to deriving this invariant distance expression, in a way that demonstrates that it is for homogeneous and isotropic spaces, but for now we are going to apply it to learn what it would be like to live in a space with \( k \neq 0 \).
Exercise 4: Let \(k<0\). Consider the length of the line from \(r=0\) to \(r=a\). Call this R. What is the circumference of a circle of radius R? Remember, this isn't our familiar Euclidean space, so work out the integral with \(r=a\). Compare this to the circumference of a circle with the same radius R in Euclidean space (\(k=0\)). Repeat the exercise, except with \(k>0\).