Chapter 2 – Rethinking Newton's Laws of Motion

2–1. Finding momentum.

Although Einstein used the Two Postulates of Relativity as his basis for deriving the details of special relativity, this approach is not necessary. We will derive Einstein's results from a different starting point, one that does not require us to accept the constant velocity of light. But first, let us look again at Newton's Laws of Motion. We can state these as follows:

1) *The natural motion of an object is straight-line motion at constant speed.* The quantity of motion is a vector constant called momentum, ***p***.

2) *A force is necessary to change the momentum of an object. We define force as*

$$\boldsymbol{F} = \frac{d\boldsymbol{p}}{dt}.$$

Furthermore, the total force on an object is the vector sum over all the forces between the object and every other object with which it interacts.

3) *If two objects experience a short-range interaction, the total momentum following the interaction is the same as the total momentum prior to the interaction.* This is known as "asymptotic momentum conservation." Note that this is much more restrictive than Newton's statement of the Third Law, which is that the total momentum must be conserved at every moment throughout the entire interaction.

These laws require us to 1) find a quantity that is conserved in interactions, the momentum, and 2) define force as the time rate of change of this quantity.

René Descartes first proposed that the quantity of motion in the universe should be conserved because God made the right amount of motion in the universe and, hence, this amount of motion could not be changed. However, he did not suggest what should be interpreted as this conserved quantity of motion. Based on experimental observation, Christiaan Huygens proposed that the conserved quantity should be *mv*. Newton adopted this interpretation of momentum. But we must remember that this is just an experimental observation.

Let us propose a thought experiment as a means of defining the momentum of an object. This is not quite as easy as it may seem at first sight, because if momentum is a complicated function of velocity, it may hard to deduce the functional form. We could, however, make an educated guess that we should be able to make a series expansion for momentum. That is,

$$p = mv + Amv^2 + Bmv^3 + \ldots$$

If an object has a very small velocity, then the first term can adequately represent its momentum. One way to insure that a colliding object has a very small velocity is for it to have a very large mass. So now, we can set up our thought experiment: Let a small test object with momentum p_0 collide elastically with a very large object. The small object will reflect off the large object with essentially the same speed that it had before the collision. In the limit that the mass of the second object gets very large, we define the momentum of the test object to be:

$$2p_0 = mv.$$

It turns out that $p = mv$ does not work when objects have speeds over about 10% of the speed of light. We are forced to multiply *mv* by a function which we call $\gamma(v)$.

$$\boldsymbol{p} = m\gamma(v)\,\boldsymbol{v} = m\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\,\boldsymbol{v} \equiv m\frac{1}{\sqrt{1 - \beta^2}}\,\boldsymbol{v}.$$

In this equation *c* represents the speed of light and v / c is written as **β** for convenience.

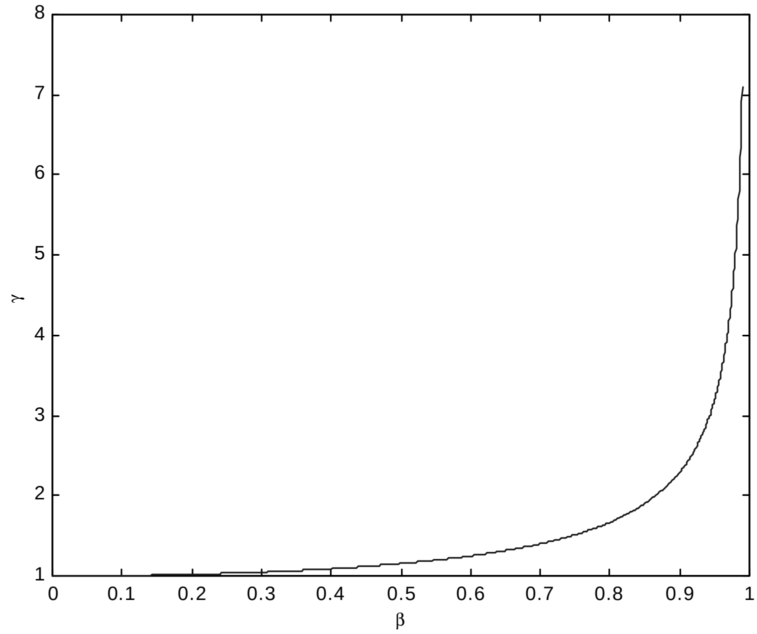

Figure 2–1 γ as a function of β.

We can say that as the speed of an object increases, its mass increases. When we think of mass, we think of two rather different concepts. The first concept is gravitational mass. This is measured by the amount of gravitational force, or weight, that is associated with an object. The second concept is inertial mass. Inertial mass is measured by collisions between objects. There is no reason to believe, *a priori*, that gravitational and inertial mass should always be the same. Einstein, however, proposed in his General Theory of Relativity, a geometric theory of gravitation, that they should be identical.

2–2. The Correspondence Principle

Of course, Newtonian mechanics makes no such allowance for the momentum of an object to be anything other than *mv*. However we can think of Newtonian physics as the regime in which $\gamma(v) \approx 1$. This brings up one of the fundamental guideposts of 20th Century physics, the "Correspondence Principle." This principle states that since we know Newtonian mechanics works for objects with small velocities, any new theory must algebraically reduce to the Newtonian result in the limit of small velocities. This is just another way of saying that since Newtonian mechanics works experimentally for slowly moving objects, our theory must agree with experiment in this range. Of course, we recognize that our theory must also agree with measurements made at high velocities to be an acceptable theory.

2-3. The World Line

To describe the motion of an object, we can specify its location in space by using standard Cartesian coordinates. The position of a particle is then described by a three-dimensional vector

$$\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z}.$$

This could also be written as a column vector

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

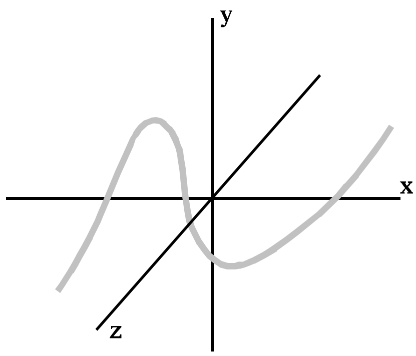

Figure 2–2 Particle trajectory in three spatial dimensions.

The meaning of both of these vectors is the same. We use the column vector because this notation is frequently encountered in discussions of relativity.

Let us think of the motion of a simple point particle. The path the particle traces out is called its trajectory. A trajectory only tells us, however, the points occupied by a particle, and says nothing about when the particle is at a given point.

We can generalize the idea of Cartesian spatial coordinates to include a coordinate for time as well. The particle would then be described by a four-dimensional vector. For convenience, we choose to create this new variable so that it will have dimensions of length like the three spatial dimensions. We call this variable $w = ct$, where c is the velocity of light so that w has units of length. (At this point, any value of c, such as 34.7 m/s, would be just as good as the velocity of light, but we'll see later why the velocity of light is the natural choice.) Thus the space-time vector, **x**, for the particle is written:

$$\mathbf{x} = w\hat{t} + x\hat{x} + y\hat{y} + z\hat{z} \qquad \text{or} \qquad \mathbf{x} = \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}$$

A trajectory in both space and time is called the "world line" of the particle. To plot the world line would require us to draw a four-dimensional graph. Since that's rather hard, we'll just graph the world lines for motion in two spatial dimensions.

Let us try to use these ideas to not only describe the motion of the particle, but to predict it as well. In order to determine the world line of a particle, we need four things:

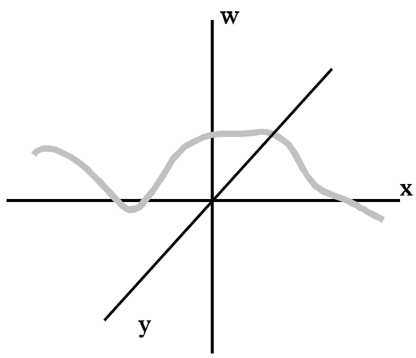

Figure 2–3 Particle trajectory in two spatial dimensions and one time dimension, the "world line."

1) the initial coordinates of the particle (its position in space-time).
2) the initial velocity of the particle
3) the mass of the particle.
4) a means of determining the force on the particle at every point along the world line.

The basic algorithm for mapping out the world line is simple:
1) Start with the particle at a given point in space.
2) Create a "propagator," a four-dimensional vector that points in the direction of the world line.
3) Pick a new point a very small distance (really an infinitesimal distance in the limit) from the first point in the direction of the propagator.
4) Repeat steps 2 and 3.

For details of how the propagator can be used in this way to construct the world line of an object, please see Appendix A.

2–4. Finding the Propagator

The next step is to determine the propagator. In general we know that it must be a vector with four dimensions. We can write this as:

$$\mathbf{P} = \begin{pmatrix} P_t \\ P_x \\ P_y \\ P_z \end{pmatrix}$$

The effect of the propagator is to determine the new space-time coordinate of the particle given the old space-time coordinate. If we let α be a small constant that determines the step length (so that in the limit $\alpha \to 0$),

$$\Delta\mathbf{x} = \begin{pmatrix} \Delta w \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \alpha \begin{pmatrix} P_t \\ P_x \\ P_y \\ P_z \end{pmatrix} \tag{2–1}$$

It is easy to see that the direction the particle moves in space is given by the velocity vector. Let's then write:

$$\mathbf{P} = \begin{pmatrix} P_t \\ v_x \\ v_y \\ v_z \end{pmatrix} \tag{2–2}$$

Now we must find P_t. To do this, we need to make use of (2–1) and (2–2):

$$\Delta w = \alpha P_t \qquad \Delta x = \alpha v_x$$

$$v_x = \frac{\Delta x}{\Delta t} = \frac{c\Delta x}{\Delta w} = \frac{c\alpha v_x}{\alpha P_t}$$

$$P_t = c$$

This now gives us the complete propagator.

$$\mathrm{P} = \begin{pmatrix} c \\ v_x \\ v_y \\ v_z \end{pmatrix} \tag{2–3}$$

2–5. The Energy-Momentum Vector

The propagator we have computed is a vector that points in the direction of the particle's motion in space-time. It is clear that we could multiply this vector by any constant and still have a perfectly good propagator, as multiplication by a constant will not change its direction. One particularly useful choice is to multiply the propagator by *mc* where *m* is the mass of the particle. Using the name **E** for this four-component vector, we have what is called the "energy-momentum" vector for the particle:

$$\mathbf{E} = \begin{pmatrix} m\gamma c^2 \\ m\gamma v_x c \\ m\gamma v_y c \\ m\gamma v_z c \end{pmatrix} \equiv \begin{pmatrix} E \\ p_x c \\ p_y c \\ p_z c \end{pmatrix} \tag{2–4}$$

The spatial parts of this vector (the three lower components) are just the components of momentum multiplied by c. The time (upper) component is something with units of energy that satisfies the relation $E = m\gamma c^2$. Notice that in the Newtonian limit ($\gamma = 1$), this reduces to $E = mc^2$. Although this looks like the familiar equation we associate with relativity, the meaning is different. The constant c could have been any number that has units of velocity; c is just a scale factor for the time axis of the four-dimensional coordinate system. What we do know, though, is that $p_x c$ tells us the size of space steps needed to generate the world line in the x-direction, and E tells us the size of time steps necessary to generate the world line in the w-direction.

Using the energy-momentum vector as the new propagator, $\mathbf{P}'$, we can revise (2–1) and (2–3) to read as:

$$\begin{pmatrix} \Delta w \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \varepsilon \begin{pmatrix} P_0' \\ P_1' \\ P_2' \\ P_3' \end{pmatrix} = \varepsilon \begin{pmatrix} m\gamma c^2 \\ p_x c \\ p_y c \\ p_z c \end{pmatrix} \tag{2–5}$$

where ε is a small constant, different in magnitude but identical in purpose to α.

2–6. The Energy-Momentum Relationship

The last thing we need in order to calculate the world line of the particle is to find what effect force has on its motion.

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d\mathbf{p}c}{dw}$$

With this, we can write the rate of change in the energy-momentum vector as

$$\frac{d\mathbf{E}}{dw} = \begin{pmatrix} F_t \\ F_x \\ F_y \\ F_z \end{pmatrix}$$

We know what the spatial components of this four-vector mean – they are usual components of force in three dimensions. However, we need to find F_t. Straightforward, but tedious differentiation gives:

$$F_t = \frac{dE}{dw} = \frac{1}{c}\frac{d}{dt} m\gamma c^2 = \vec{\beta} \cdot \mathbf{F}$$

However, let us assume that we do not know the functional form of γ for now, and see if we can use physical arguments to suggest that this expression is valid. In Newtonian kinematics $\gamma = 1$, so $F_t = 0$. This means that the direction of the world line in the time coordinate can never change for a particle of fixed mass. On the other hand, if γ is not constant, E must change with velocity. We postulate that the change in E must be given by the change in classical kinetic energy K. That is:

$$K = \tfrac{1}{2} m v^2 = \frac{m^2 v^2}{2m} = \frac{\mathbf{p} \cdot \mathbf{p}}{2m}$$

$$\frac{dK}{dt} = \frac{1}{2m}\left(2p_x \frac{dp_x}{dt} + 2p_y \frac{dp_y}{dt} + 2p_z \frac{dp_z}{dt} \right) = \frac{1}{m}\mathbf{p} \cdot \frac{d\mathbf{p}}{dt} = \mathbf{v} \cdot \mathbf{F}$$

$$\frac{dK}{dw} = \frac{1}{c}\frac{dK}{dt} = \vec{\beta} \cdot \mathbf{F}$$

Note that there is no way that we can prove that F_t must be $\vec{\beta} \cdot \mathbf{F}$, but we must accept this either as a postulate based on experimental evidence or as a postulate based on the reasonable assumption that the change in E must be equal to the change in K in the Newtonian regime.

Now let us see what the consequences of this result must be. We have:

$$\frac{dE}{dw} = \vec{\beta} \cdot \mathbf{F} = \vec{\beta} \cdot \frac{d\mathbf{p}c}{dw}$$

$$dE = \frac{mc^2 \vec{\beta}\gamma}{mc^2 \gamma} \cdot d\mathbf{p}c$$

$$EdE = \mathbf{p}c \cdot d\mathbf{p}c \qquad (2\text{-}6)$$

Equation (2–6) is a differential equation that relates changes in energy to changes in momentum. Notice how energy, the component of the propagator in the time direction, is symmetrically related to momentum, the components of the propagator in the spatial directions. The solution of this equation can be shown to be:

$$E^2 = \mathbf{p}c \cdot \mathbf{p}c + C = p^2 c^2 + C$$

where C is a constant. To find C, we make use of the Correspondence Principle which suggests that when the momentum is zero, the energy should have the non-relativistic value of $E_0 = mc^2$. Thus:

$$E^2 = p^2 c^2 + m^2 c^4 = p^2 c^2 + E_0^2 \qquad (2\text{–}7)$$

This is called the "Relativistic Energy-Momentum Relationship" and is at the heart of relativistic kinematics.

2.7 Finding γ

Now we are in a position to find γ. We already know E and p as functions of γ, so we can substitute these into the energy-momentum relationship above.

$$
\begin{aligned}
E^2 &= p^2c^2 + m^2c^4 \\
m^2\gamma^2c^4 &= m^2\gamma^2\beta^2c^4 + m^2c^4 \\
\gamma^2 &= \gamma^2\beta^2 + 1 \\
\gamma^2(1-\beta^2) &= 1 \\
\gamma &= \frac{1}{\sqrt{1-\beta^2}} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}
\end{aligned}
$$

From this equation we see that γ has a value very nearly 1 when v is small but becomes infinite as v approaches c. In fact, we can see that v can never get larger than c, as that would force γ to become imaginary. This means that c is not just some arbitrary constant as it was non-relativistically, but it is a limit to the speed of any particle. We can determine what this limit is by considering the following:

$$\beta = \frac{v}{c} = \frac{pc}{E}, \quad v = c\frac{pc}{E} = c\frac{\sqrt{E^2 - m^2c^4}}{E}$$

The photon is a "particle of light" with energy but no mass. From the above expression, we see that the velocity of a photon must be c. Hence, c is the velocity of light (in vacuum), 2.997×10^8 m / s.

2.8 Relativistic Energies

Now we have a relationship between energy and momentum that implies that total energy is something more than a factor needed for the abstract propagation of a world line. When the velocity of a particle is zero, the particle still has energy $E_0 = mc^2$. This is called the "rest energy" of an object. Although we saw this formula before, we must now interpret this as a real energy residing in the mass of the particle, and if we convert mass to energy, it tells us how much energy will be produced. The quantity E is called "total energy" of an object. This energy includes the kinetic and rest energy of a particle. We can then define the kinetic energy as the difference between the rest energy and the total energy. The following table summarizes these quantities.

Total Energy	$E = mc^2\gamma$
Rest Energy	$E_0 = mc^2$
Kinetic Energy	$E = mc^2(\gamma-1)$
Momentum	$pc = mc^2\beta\gamma$

We suggested that all relativistic quantities should, because of the Correspondence Principle, reduce to Newtonian quantities when the velocity is small. Let's see if this is true for the kinetic energy.

$$
\begin{aligned}
K &= mc^2(\gamma - 1) \\
&= mc^2\left(\frac{1}{\sqrt{1-\beta^2}} - 1\right) = mc^2\left[(1-\beta^2)^{-1/2} - 1\right] \\
&\approx mc^2\left[\left(1+\tfrac{1}{2}\beta^2\right) - 1\right] \quad \text{if } \beta << 1 \\
&= \tfrac{1}{2}m\beta^2 c^2 \\
&= \tfrac{1}{2}mv^2
\end{aligned}
$$

Note that we made use of the Taylor Series expansion, $(1+\delta)^n \approx 1 + n\delta, \quad \delta << 1$.

2.9 What is Relativity Then?

The difference between relativistic kinematics and Newtonian kinematics all boils down to the fact that in Newtonian mechanics F_t is taken to be zero, so that $\gamma = 1$, but in relativistic kinematics $F_t = \vec{\beta} \cdot \mathbf{F}$ so that γ becomes $1/\sqrt{1-\beta^2}$. This small difference, however, leads to a number of important consequences. This is because, although the mathematical differences are small, the philosophical differences are immense. In Newtonian mechanics, space and time are in two different worlds, wholly unconnected. In Einsteinian mechanics, space can depend on time and time on space. Hence the propagator of time, energy, can be related to the propagator of space, momentum. By simply allowing such interdependence, we were able to find a relationship between energy and momentum, another relationship between energy and mass, and a limiting speed for all matter.

But it does not stop here. We now need to go back and see what relationships between space and time are implied by our newly found relationships between kinematic variables.