General Relativity and Differential Geometry

By

Harry Hausner

* * * * * * * * *

Submitted in partial fulfillment

of the requirements for

Honors in the Departments of Math & Physics and Astronomy

UNION COLLEGE

June, 2016

Contents

Chapter 1

Introduction

Our purpose here is to give a basic introduction to Einstein's theory of relativity, both in the special and general cases. The basis of the theory is that the laws of physics should be independent of reference frame: physics should, in essence, look the same regardless of where a person is, how fast they are moving, or how much they are accelerating. This leads to a theory of physics which has pronounced effects on what one might have assumed we well understood classical problems.

We first start by talking about the action principle, the classical principle of motion which can be used to find the laws of motion given the potential energy of the system. This is the main principle which will give us our laws of motion, however, we will increasingly redefine the action to make the principle more "geometric". We will make it of a form such that the laws of motion come from the very shape of the universe.

This is first done in special relativity which introduces the idea of spacetime. This lets us treat both time and space as geometric quantities and make the fact that the speed of light is the same for all observers a simple geometric result. General relativity then makes the force of gravity a mere mater of the curvature of spacetime.

Underlying all of this is the mathematics of differential geometry. Differential geometry is the study of generalized spaces upon which one can perform calculus. These spaces are of most interest to physics since it is hard to imagine how one could perform physics without calculus, and allowing are spaces to be more general than the usual Euclidean 3-space we are use to allows us to encode physical laws as more fundamental properties of geometry the universe.

While continuing to improve the theory of relativity, we shall examine the effects on the test case of a satellite in orbit around a star. In particular we will look at how increasingly relativistic theories effect the shape of bound orbits. This is a system which is well understood, and most readers should be familiar with Kepler's laws which govern the behavior classically. The effects of relativity will at first distort the orbits from their classical closed ellipses, and then make them only calculable numerically.

However, before we can get into all that, we must first start with the the principle of least action.

Chapter 2

The Principle of Least Action

In order to develop the theory of relativity we will find it useful to forgo Newton's laws in favor of the principle of least action. Not only is this a far more elegant tool, it makes for a straightforward transition to the relativistic regime and allows for a greater intuition.

The Principle of Least Action simply states that the path an object takes is that which minimize the action. However, the action does not have a simple formula, and it changes depending on the physical system. One of the goals of relativity is to generate a more natural explanation for the action principle.

2.1 Fermat's Principle

The simplest application of the action principle is *Fermat's principle*, which states that the path that light takes between two points is that which minimizes the time it takes the light to make the journey. This is an incredibly powerful assumption to make; we can derive many of our basic optical laws, such as the laws of reflection and refraction from this one idea.

Newtonian mechanics conceives of physics as actions and reactions, or a objects acting on each other. Fermat's principle introduces a new way of thinking about physics: things occur they way they do because it is the optimal way of doing them for some idea of "optimal". This shifts the picture away from causes and effects, and forces us to think of physics being linked to the away things are arranged. In other word, Fermat's principle derives physical law from the geometry system instead of the interactions of its components. This is also the way we will think of physics in relativity.

The idea of minimizing the time the light travels will come back when we explore relativity, where we will minimize the *proper* time an object experiences. This is not quite the case for Classical Mechanics, though a similar principle applies which is of less natural significance.

2.2 The Action in Classical Mechanics

In classical mechanics, we define the Lagrangian, L, to be the difference between the kinetic and potential energy of a system. We will start in one dimension, so the Lagrangian is a function of position q, velocity $\dot{q}$, and time t. The action is then defined as:

$$S[q] = \int_{t_i}^{t_f} L(q, \dot{q}, t)dt, \tag{2.1}$$

where the integral is taken over the time the object is traveling. From here it is just a mater of finding which paths extremize the action.

It is worth noting that that action is not an ordinary function, but is a mathematical object called a *functional*. Where a function maps one value to another, a functional maps a function to a value. In this case the action is mapping the path, which is a function of time, to a number. We indicate that the action is a functional by using square brackets around its argument, as opposed to the parentheses used for ordinary functions.

The fact that the action is a functional means that we cannot use our standard methods for extremiziation from ordinary calculus; instead we must use a method called the calculus of variations.

Regardless, the action is an odd quantity to consider, and it is not immediately obvious that extremizing the action will yield the physical laws we expect. It does turn out to work, but moreover, it is a limit case of extremizing the proper time, as we shall see later.

The extremize the action integral, we will consider varying the path the integral is taken over while fixing the end points. The idea is that at an extremum, the value of the intergral will not change for a small variation. To this end, we can imagine an arbitrary, small path difference δq which perturbs the path the object takes, but leaves the end points the same, i.e., $\delta q(t_i) = \delta q(t_f) = 0$. The action taken along this new path $q + \delta q$ is:

$$S[q+\delta q] = \int_{t_i}^{t_f} L(q+\delta q, \dot{q}+\delta\dot{q}, t)dt = \int_{t_i}^{t_f} L(q,\dot{q},t) + \delta q\frac{\partial L}{\partial q} + \delta\dot{q}\frac{\partial L}{\partial \dot{q}} + \mathcal{O}(\delta q)^2 dt, \tag{2.2}$$

where in the last step we expand for small δq and assume $\delta\dot{q}$ is on the order of δq. We may note that, by integration be parts:

$$\int_{t_i}^{t_f} \delta\dot{q}\frac{\partial L}{\partial\dot{q}}dt = \left[\delta q\frac{\partial L}{\partial\dot{q}}\right]_{t_i}^{t_f} - \int_{t_i}^{t_f}\delta q\frac{d}{dt}\frac{\partial L}{\partial\dot{q}}dt = -\int_{t_i}^{t_f}\delta q\frac{d}{dt}\frac{\partial L}{\partial\dot{q}}dt, \tag{2.3}$$

where the first term is a zero because we required $\delta q(t_i) = \delta q(t_f) = 0$. Thus the action from caused by a small path change is:

$$\begin{aligned} S[q+\delta q] &= \int_{t_i}^{t_f} L(q,\dot{q},t) + \delta q\left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial\dot{q}}\right) + \mathcal{O}(\delta q)^2 dt \\ &= S[q] + \int_{t_i}^{t_f}\delta q\left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial\dot{q}}\right) + \mathcal{O}(\delta q)^2 dt. \end{aligned} \tag{2.4}$$

For $S(q)$ to be an extremum, we must require $S[q+\delta q] - S[q] = \mathcal{O}(\delta q)^2$ for an arbitrarily small path differences δq. This analogous to requiring the derivative of a function to be zero; it means that any path "near" the path q must not effect the value of the action to first order. By equation (2.4) this means:

$$\int_{t_i}^{t_f}\delta q\left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial\dot{q}}\right)dt = 0 \tag{2.5}$$

Where we can neglect the order δq^2 terms since δq is sufficiently small. For equation (2.5) to be true regardless of the choice of δq, it must be that:

$$\frac{\partial L}{\partial q} = \frac{d}{dt}\frac{\partial L}{\partial\dot{q}}$$

This can easily be extended to any number of coordinates. If we had coordinates q_i, for $i = 1, 2, \ldots, N$, we could repeat the process above for a path difference δq_i which effects only the q_i position. Since S must be at a minimum for each such path difference we have:

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i}, \tag{2.6}$$

for each $i = 1, 2, \ldots, N$. These equations are known as the Euler-Lagrange equations.

For each coordinate q_i, we call the factor:

$$p_i = \frac{\partial L}{\partial \dot{q}},$$

the conjugate, (or canonical) momentum of q_i. We call it such because for a free particle moving in the x direction, $L = \frac{1}{2}m\dot{x}^2$, and so $p_x = m\dot{x}$, which is exactly the momentum term we would expect. However, the conjugate momentum does not *need* to be ordinary momentum.

We call the quantity

$$H = \sum_{i=1}^{N} \dot{q}_i p_i - L \tag{2.7}$$

the *Hamiltonian*. In most cases[1] this works out to be the energy of the system. In both special and general relativity, the Lagrangian will not be the difference in kinetic and potential energy, and so it will become more convenient to arrive at the Lagrangian from the Hamiltonian.

2.3 Lagrangian Densities

The above discussion allows us to find the laws of motion for particles which follow a path through space. This is only one class of objects of interest in physics. If we wish for the Action Principle to give us a full description of physics we need it to generate laws for objects which exist simultaneously everywhere in space: we need a principle of least action for fields.

As before we will use the one dimensional case for our derivation with the multidimensional case being nearly identical. For a field, $\phi(x, t)$, defined at every point x in space at all times t we will define the action as

$$S[\phi] = \int_{t_i}^{t_f} \int_{-\infty}^{\infty} \mathcal{L}(x, t, \phi, \phi', \dot{\phi}) dx dt, \tag{2.8}$$

where the prime indicates a derivative with respect to x, the dot represents a derivative with respect to t, and $\mathcal{L}$ is called the *Lagrangian density*. The Lagrangian density, like the Lagrangian from before, is some function which will describe the physics of the field. It is related to the Lagrangian by

$$L = \int_{-\infty}^{\infty} \mathcal{L} dx$$

To extremize the action we will consider some deviation of the field $\delta\phi$ which does not change the boundary conditions of the field, much as we considered a path variation previously. The process here is essentially the same as it was previously, except that deviation is not in a path,

[1] Essentially boiling down to the coordinates being *scleromic*, meaning that they are not dependent on time. For every situation we will consider this will be the case.

but a field. This means out deviation will be a function of both time and space, just like the field is, and must also fix the boundary conditions of the field. This requires:

$$\lim_{x\to\infty}\delta\phi=\lim_{x\to-\infty}\delta\phi=\lim_{t\to t_i}\delta\phi=\lim_{t\to t_f}\delta\phi=0 \tag{2.9}$$

Which is analogous to to having the path deviation, δq, be zero at the end points of the path.

We then consider the action for the new field $\phi+\delta\phi$ and Taylor expand about $\delta\phi=\delta\phi'=\delta\dot{\phi}=0$

$$\begin{aligned}S[\phi+\delta\phi]&=\int_{t_i}^{t_f}\int_{-\infty}^{\infty}\mathcal{L}(x,t,\phi+\delta\phi,\phi'+\delta\phi',\dot{\phi}+\delta\dot{\phi})dxdt\\&=\int_{t_i}^{t_f}\int_{-\infty}^{\infty}\mathcal{L}(x,t,\phi,\phi',\dot{\phi})+\delta\phi\frac{\partial L}{\partial\phi}+\delta\phi'\frac{\partial L}{\partial\phi'}+\delta\dot{\phi}\frac{\partial L}{\partial\dot{\phi}}+\mathcal{O}(\delta\phi)^2dxdt\end{aligned} \tag{2.10}$$

We then use integration by parts on the $\delta\phi'$ and $\delta\dot{\phi}$ term similar to our previous method in equation (2.3), except that for the first we instead integrate the $\delta\phi'$ terms over x. This yields:

$$\int_{-\infty}^{\infty}\delta\phi'\frac{\partial\mathcal{L}}{\partial\phi'}dx=-\int_{-\infty}^{\infty}\delta\phi\frac{d}{dx}\frac{\partial\mathcal{L}}{\partial\phi'}dx, \tag{2.11a}$$

and

$$\int_{t_i}^{t_f}\delta\dot{\phi}\frac{\partial\mathcal{L}}{\partial\dot{\phi}}dt=-\int_{t_i}^{t_f}\delta\phi\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\phi}}dt, \tag{2.11b}$$

since $\delta\phi$ vanishes at the boundaries.

Substituting equations (2.8), (2.11a), and (2.11b), into equation (2.10) gives us:

$$S[\phi+\delta\phi]=S[\phi]+\int_{t_i}^{t_f}\int_{-\infty}^{\infty}\delta\phi\left(\frac{\partial\mathcal{L}}{\partial\phi}-\frac{d}{dx}\frac{\partial\mathcal{L}}{\partial\phi'}-\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\phi}}\right)dxdt+\mathcal{O}(\delta\phi)^2 \tag{2.12}$$

The conversation from here is nearly identical to that in section 2.2. The change in the action due to $\delta\phi$ must vanish, and so the integral must as well. Since the function $\delta\phi$ is arbitrary, this must mean that:

$$\frac{\partial\mathcal{L}}{\partial\phi}-\frac{d}{dx}\frac{\partial\mathcal{L}}{\partial\phi'}-\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{\phi}}=0 \tag{2.13}$$

This process can be easily extended for any number of fields, ϕ_i for $i=1,2,\ldots,N$, dependent on any number of independent variables x_j for $j=1,2,\ldots,M$. The more general form of equation (2.13) is:

$$\frac{\partial\mathcal{L}}{\partial\phi_i}-\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\left(\frac{\partial\phi_i}{\partial t}\right)}-\sum_{j=1}^{M}\frac{d}{dx_j}\frac{\partial\mathcal{L}}{\partial\left(\frac{\partial\phi_i}{\partial x_j}\right)}=0, \tag{2.14}$$

which holds for every field ϕ_i, $i=1,2,\ldots,N$. In most cases we will have $n=3$, where $x_1=x, x_2=y$, and $x_3=z$.

Just as we defined a Lagrangian density, we would also like to have a *Hamiltonian density*, $\mathcal{H}$, analogously to equation (2.7). Since ϕ_i and $\mathcal{L}$ are analogous to q_i and L respectively, we need only come up with a substitute for the conjugate momentum p_i. We defined the conjugate momentum

for a coordinate, we define the *conjugate momentum field*, π_i, for a field ϕ_i as:

$$\pi_i = \frac{\partial \mathcal{L}}{\partial \dot{\phi}_i}. \tag{2.15}$$

This allows us to define the Hamiltonian density as:

$$\mathcal{H} = \sum_{i=1}^{N} \dot{\phi}_i \pi_i - \mathcal{L}, \tag{2.16}$$

and for most cases, this quantity is equivalent to the energy density of the system.

2.4 Symmetries and Conservation Laws

A quantity X is a conserved quantity if

$$\frac{dX}{dt} = 0 \tag{2.17}$$

If the Lagrangian is not explicitly dependent on a coordinate q_i, then Euler-Lagrange equation, (2.6), for this coordinate becomes:

$$\frac{dp_i}{dt} = 0, \tag{2.18}$$

and hence the conjugate momentum p_i is conserved.

We can also get a similar conserved quantity associated with time dependence. Consider the time derivative of the Hamiltonian from equation (2.7).

$$\begin{aligned}
\frac{dH}{dt} &= \sum \left(\ddot{q}_i p_i + \dot{q}_i \dot{p}_i\right) - \frac{dL}{dt} \\
&= \sum \left(\ddot{q}_i \frac{\partial L}{\partial \dot{q}_i} + \dot{q}_i \frac{\partial L}{\partial q_i}\right) - \left(\sum \left(\dot{q}_i \frac{\partial L}{\partial q_i} + \ddot{q}_i \frac{\partial L}{\partial \dot{q}_i}\right) + \frac{\partial L}{\partial t}\right) \\
&= -\frac{\partial L}{\partial t}
\end{aligned} \tag{2.19}$$

Which means that the Hamiltonian is conserved if the Lagrangian has no explicit time dependence. Looking in terms of the Euler-Lagrange equation. This makes the Hamiltonian the time conjugate momentum.

We will say we have a *translation symmetry* with respect to some coordinate if a translation that coordinate preserves the Lagrangian, i.e., $L(q) = L(q + \Delta q)$. In most cases this will mean that the coordinate in question does not appear explicitly in the Lagrangian, and in particular, the Hamiltonian is conserved if the Lagrangian does not explicitly depend on time. There are other types of symmetry, but we will not consider them here. It turns out that every continuous symmetry of the Lagrangian generates a conservation law, and that conservation laws come only from such symmetries. This is called *Nöther's theorem*, and is an extremely important tool in physics which we will use to our advantage.

Chapter 3

The Classical Treatment of the Two Body Problem

To illustrate how to use the principle of least action in a physical example we will look at the two body problem, where two objects interact gravitationally with each other. This example will also allow us to compare the results of relativity to the classical case. We will use this example because it is easy to look at in both classical and relativistic way, and shows some clear differences in behavior. The two body problem can always be thought of as a single body acting in a central potential. It is not useful for our purposes to go into why here and an explanation can be found in any classical mechanics textbook[1].

We will be particularly interested in the case where the object is bound in an orbit: unable to escape or crash into the larger object. This partially to limit the number of comparisons we will have to do once we move into relativity, but it is also the most physically interesting case. These bound orbits are what governs the motion of the Moon around the Earth, and the planets around the Sun. In fact, as we shall see later, it was in the motion of Mercury that we found some of the earliest evidence for the validity of general relativity.

3.1 Set Up

. We will start by considering an object of mass m and a central potential induced by a mass M of the form:

$$U = -\frac{GMm}{r}, \tag{3.1}$$

where G is Newton's gravitational constant and r is the radial position of our object of interest.

It is worth noting that since we are dealing with a central potential, which cannot cause a torque about the center, and we have no external forces to cause torques; the angular momentum of the system must be conserved. This can also be though of as a consequence of Nöher's theorem with the rotational symmetry of the system. The angular momentum is, as always:

$$\mathbf{\Lambda} = \mathbf{r} \times \mathbf{p} \tag{3.2}$$

Where $\mathbf{r}$ and $\mathbf{p}$ are the position and momentum vectors respectively, and we use $\mathbf{\Lambda}$ for angular momentum to distinguish it from the Lagrangian. This conservation means that the plane spanned

[1] I recommend Taylor's *Classical Mechanics*.
Taylor, John R. *Classical Mechanics*, Sausalito, Calif.: University Science Books, 2005. Print.

by **r** and **p** is constant. We will call this plane the orbital plane. The facts that the motion of the system is confined to this plane and that the potential depends only on r makes it natural for us to work in polar coordinates.

This allows us to write our Lagrangian as:

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + \frac{GMm}{r}. \tag{3.3}$$

We note from the Lagrangian that there are two conserved quantities; one associated with time translational symmetry and one with rotational symmetry. These quantities are:

$$E = H = \sum \dot{q}_i p_i - L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - \frac{GMm}{r}, \tag{3.4}$$

$$\Lambda = \frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta}, \tag{3.5}$$

where the first, which is the Hamiltonian, is the same as the total energy, and that the conjugate momentum of θ is the angular momentum.

This gives us two partially coupled, first order differential equations with two unknowns: $r(t)$ and $\theta(t)$. Thus the system has a unique solution for r and θ so long as we specify the initial conditions. We can set the initial $t = 0$ position of the orbiting body at whatever angle in the orbital plane we desire by simply choosing a value for $\theta(0)$. This will have no impact (up to rotation) on the overall shape of the orbit, because as we noted before, we have a symmetry with respect to θ.

In Newtonian mechanics, one may be used to having to specify *four* initial conditions to solve differential equations of this form. The reason we can get away with not specifying $\dot{r}(0)$ and $\dot{\theta}(0)$ is because we instead substitute in E and Λ as specified in (3.4) and (3.5) for $\dot{r}$ and $\dot{\theta}$ respectively. This allows us to solve the system with these constant quantities being arbitrary, and we can then find a specific solution by choosing values for E and Λ.

3.2 Non-Dimensionalization

With all of the G, M, and m's in these equations, it is convenient at this point to non-dimensionalize. We will take our units to be:

$$E_0 = mc^2 \qquad e = \frac{E}{E_0} \tag{3.6a}$$

$$\rho_0 = \frac{GM}{c^2} \qquad \tilde{r} = \frac{r}{\rho_0} \tag{3.6b}$$

$$t_0 = \frac{GM}{c^3} \qquad \tilde{t} = \frac{t}{t_0} \tag{3.6c}$$

$$\Lambda_0 = \frac{GMm}{c} \qquad l = \frac{\Lambda}{\Lambda_0} \tag{3.6d}$$

where the second column lists our dimensionless quantities for energy, the Lagrangian, distance, time, and angular momentum respectively. We choose these to be our units mostly because they are suggestive of what is to come. Readers already familiar with relativity may recognize E_0 as the rest-energy of our object and ρ_0 as the Schwarzschild radius of the mass causing the potential.

This is unimportant for the classical material, but will make the transition to relativistic mechanics smoother.

Factoring out the units, equations (3.4) and (3.5) become:

$$e = \frac{1}{2}\left(\dot{\tilde{r}}^2 + \tilde{r}^2\dot{\tilde{\theta}}^2\right) - \frac{1}{\tilde{r}}, \tag{3.7a}$$

$$l = \tilde{r}^2\dot{\tilde{\theta}}. \tag{3.7b}$$

For notational convenience we will now drop the tildes from our dimensionless quantities. All quantities for the rest of this section may be assumed to be dimensionless.

3.3 The Effective Potential

Something to note is that we can find information about the orbits using an effective potential, allowing us to think of this system as a single object moving in a one dimension with position $r(t)$. We will rewrite e as $e = \mathrm{KE}(\dot{r}) + U_{\text{eff}}(r)$. The first step is to solve for $\dot{\theta}$ in equation (3.5), which give us:

$$\dot{\theta} = \frac{l}{r^2}$$

Using the total energy from equation (3.7a) and removing the one dimensional kinetic portion by setting $\dot{r} = 0$ we can then find an effective potential:

$$U_{\text{eff}}(r) = \frac{1}{2}\frac{l^2}{r^2} - \frac{1}{r}. \tag{3.8}$$

A plot of U_{eff} can be seen in figure 3.1 for various values of l. By choosing values for e and l we can see qualitatively what possible trajectories should look like, as in figure 3.2. Note that since the kinetic energy must always be greater than zero, the total energy of the system must greater than U_{eff}. When the total energy is less than zero, the object will be bound between two positions $r_{\min}$ and $r_{\max}$, neither falling in nor escaping but bouncing back-and-forth between its maximum and minimum distances from the center. For $e \geq 0$, the object will be unbound to the star, but will instead ricochet off the potential and escape to $r \to \infty$, having no maximum distance from the center. In fact, for any value of $l \neq 0$, the object will never fall in to the center due to the "barrier" in front of $r = 0$.

We can use the effective potential to find the minimum acceptable value of e, whoch is the minimum value of U_{eff}. By taking the derivative of equation (3.8), we find that this corresponds to:

$$r_c = l^2, \tag{3.9}$$

And so:

$$e_{\min} = -\frac{1}{2l^2} \tag{3.10}$$

We can also learn the maximum possible energy for a bound orbit from the effective potential. Since the effective goes to infinity at $r = 0$ and zero for $r \to \infty$ for any value of $l > 0$, the maximum value for e must be 0, since this otherwise there is no barrier to stop the satellite from going out infinitely.

We can solve for acceptable values of l which will produce bound orbits. We require that l

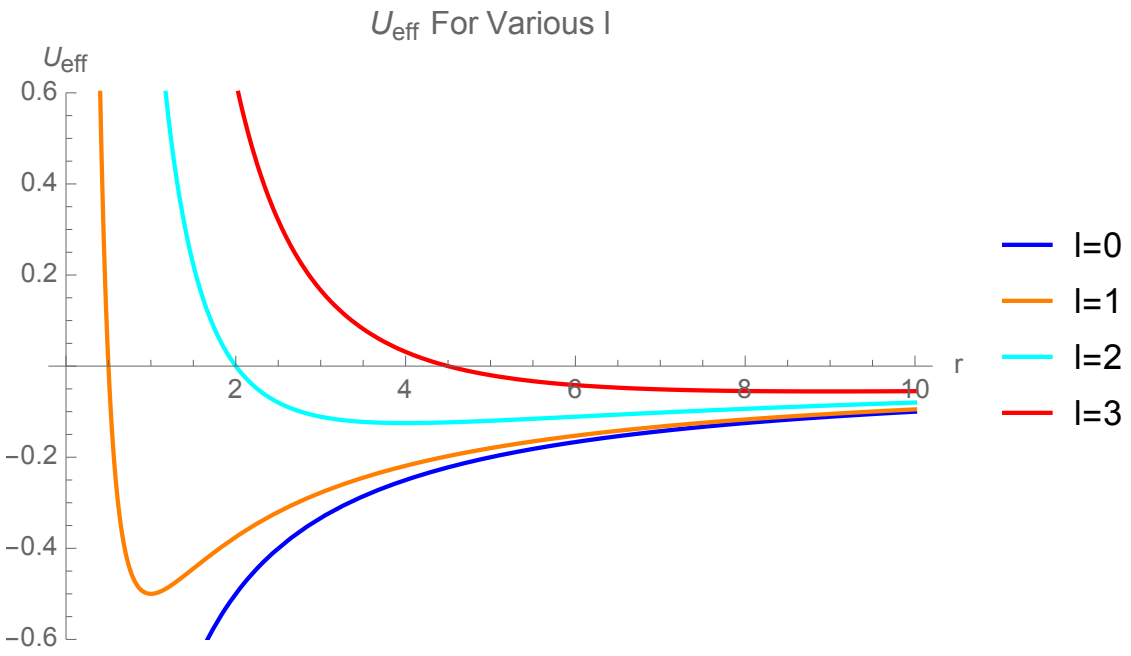


Figure 3.1: The effective potential for various values of l. When $l = 0$, the potential goes to $-\infty$ as r becomes small. For all other values of l there is a barrier before $r = 0$, and so we have bound orbits for $e < 0$. The minima of these plots correspond to circular orbits.

must be greater than zero[2], and the eccentricity must have a real value, so $l < \dfrac{1}{\sqrt{-2e}}$.

The effective potential gives us intuition for what should happen, but it is not the same as actually solving the system. We will find that in the relativistic problem we will be able to construct an object similar to the effective potential. However, it will not have the same meaning. Nonetheless it will again provide helpful intuition.

3.4 Solving the System

The effective potential allows us to qualitatively describe the orbits, but how can we get a quantitative treatment? For this we must solve our differential equations (3.7a) and (3.7b). However, there is a problem: radius has no analytic solution[3] as function of time. The solution to this is to think of the radius as a function of angle. This allows to find the shape, if not the time evolution, of the orbits. The main obstacle to this is that we have time derivatives and so to resolve this we write $u(\theta) = \dfrac{1}{r(t(\theta))}$. (Working with $\dfrac{1}{r}$ actually simplifies our calculations). This allows us to express $\dot{r}$ as:

$$\dot{r} = \frac{d\theta}{dt}\frac{dr}{d\theta} = \frac{l}{r^2}\frac{d}{d\theta}\left(\frac{1}{u}\right) = lu^2\left(-\frac{u'}{u^2}\right) = -lu', \tag{3.11}$$

[2]Technically, we could allow l to be negative, but physically this would be identical to the case where l were positive accept the object would move in the opposite direction. Since the direction of the orbit is not terribly important, we simplify by only considering the positive l cases.

[3]Here we us "analytic" in the sense of an exact solution being derived through analysis rather as opposed to being solved numerically. This is not to be confused with the mathematical notion of an analytic, or holomorphic, function.

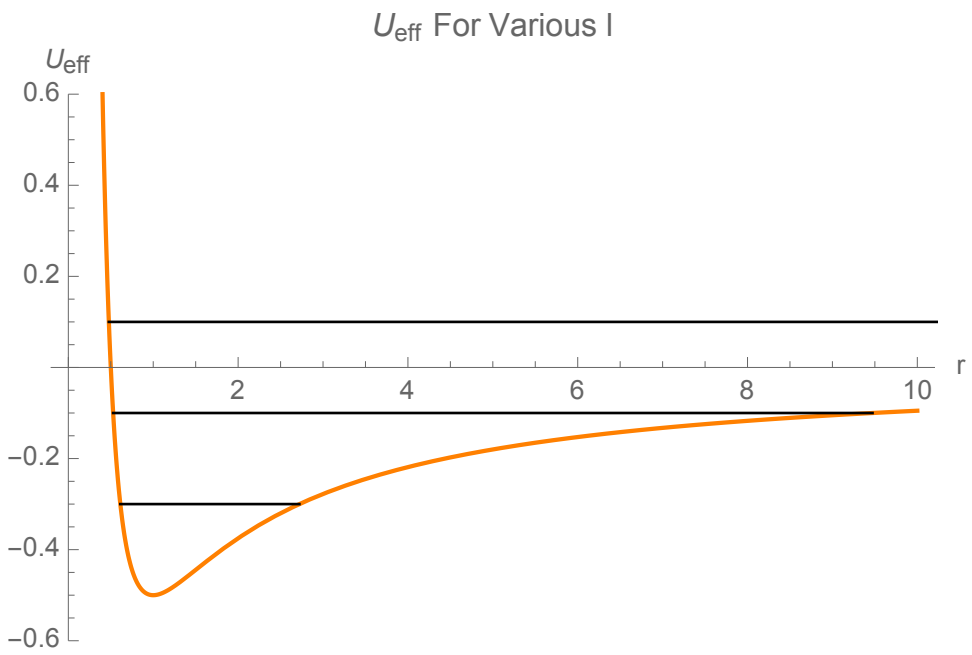


Figure 3.2: Choosing $l = 1$, we consider objects with total energy $e = -0.3, -0.1,$ and 0.1. We can note that the orbits with $e < 0$ are bound by the potential, being confined to a smaller region as e decreases. Orbits at $e = 0$ will not be bound, but will have an interesting shape, since U_{eff} approaches 0 as r approaches infinity. For values of $e > 0$, there is no outer boundary, but there is a barrier before $r = 0$. The orbits for $e \geq 0$ will "bounce" back from the source of the potential before striking.

where primes indicate a derivative with respect to θ. We may now rewrite our energy as:

$$e = \frac{1}{2}\left(l^2 u'^2 + l^2 u^2\right) - u, \tag{3.12}$$

and solving for u' yields:

$$u' = \frac{du}{d\theta} = \pm\frac{\sqrt{2e - l^2u^2 + 2u}}{l} \tag{3.13}$$

This is a separable differential equation, which can be solved by integration. We will pick our constant of integration to be zero as this simply forces the orbit to take place in the plane and does not effect the orbit's shape. Thus:

$$\pm\,\theta = \pm\int d\theta = \int \frac{l\,du}{\sqrt{2e - l^2u^2 + 2u}}. \tag{3.14}$$

We can compute this integral by completing the square in the radical and substituting in:

$$\varepsilon = \sqrt{1 + 2el^2} \tag{3.15}$$

We call this quantity the eccentricity of the orbit. This gives us:

$$\cos\theta = \frac{l^2u - 1}{\varepsilon} \tag{3.16}$$

and so

$$r(\theta) = \frac{1}{u} = \frac{r_0}{1 + \varepsilon \cos\theta}, \tag{3.17}$$

where

$$r_0 = l^2 \tag{3.18}$$

is a factor relating to the scale of the orbit.

3.5 Analysis

This equation gives us the general formula for conic sections with eccentricity ε for our orbits. Geometrically speaking, the eccentricity represents how "circular" a conic section is. For real eccentricities $0 < \varepsilon < 1$ we get ellipses, for $\varepsilon = 1$ we get a parabola, for $\varepsilon > 1$ we get hyperbolas, and for $\varepsilon = 0$ we get circular orbits where $r(\theta) = r_0 = r_c$. Parabolas and hyperbolas are unbounded, and so for a bound orbit we must have $0 \leq \varepsilon < 1$. The eccentricity we found will always be a real number when $e > -\frac{1}{2l^2}$, will always be non-negative when real, and will be less than one for $e < 0$. This imposes a limit on e in agreement with our limits from equation (**??**).

Examples of bound orbits can be seen in figure 3.3. We can see that all of the orbits are closed, meaning they are periodic and return to some initial position. The fact that we get closed orbits is important, because we will see that once we consider special relativity we will no longer have this behavior. Instead, the orbit will precess and not close back on itself.

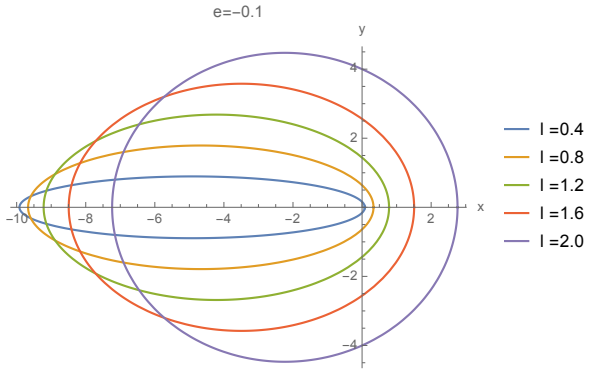


Figure 3.3: This is the orbit, $r(\theta) = \frac{r_0}{1 + \varepsilon \cos\theta}$, for an energy of -0.1 and various angular momenta. The orbit is a closed ellipse. Here we can see that the larger the angular momentum, the less eccentric the orbit is.

Chapter 4

Special Relativity

The main idea behind Einstein's special relativity is that the laws of physics should work in *all inertial* reference frames. Most important is that requiring that Maxwell's equations hold for all frames means requiring the speed of light to be the same in all reference frames. Moreover, the invariance of the speed of light is an experimental fact, and as such we need an adequate theory to explain why this must be the case. The principal idea behind special relativity is the idea of space-time.

4.1 Space-Time

In order to keep the speed of light invariant, it is necessary to think of space and time as a unified entity called space-time. To make our discussion of space-time simpler we will use a unit convention called geometric units. In this system distance is measured in light seconds. This allows us to write the speed of light as one light-second per second and greatly simplify our equations and can be thought of as a non-dimensionalization of sorts, in agreement with our previous non-dimensionalization in section 3.2. We will also treat light-seconds and ordinary seconds as the same, omitting the one light-second per second conversion factor. The main benefit of this system is that it allows us to treat space and time on a more equal footing, and helps us realize Einstein's idea of treating time as simply another dimension. The conversion back to conventional units is a simple matter of dividing by a factor of c in the time coordinate.

To illustrate how to preserve the speed of light, consider the following example. Consider two inertial reference frames, each of one spatial and one temporal dimension, one moving with a speed v relative to the other as in figure 4.1. The higher dimensional system is completely analogous as the only interesting geometry happens on the axis parallel to the relative motion of the frames. We will collectively call the rays emitted from the origin in the $\pm x$ directions at the speed of light the *light cone*. For the speed of light to be constant in any other frame the transformation from one to the other must preserve the light cone. This transformation must also be linear, i.e., map straight lines to straight lines, because while the two frames may disagree about the speed and location of an object, they must agree that inertial objects have constant velocity.

This can then be broken down into an eigenvalue problem. Say we have two frames moving with constant velocity v relative to each other. We must find a linear transform which has the vectors $(1, 1)$ and $(1, -1)$ as eigenvectors, i.e., returns the vectors back multiplies by some scalar. We also know that, by the definition of inertial reference frames, the map must send the $x = 0$

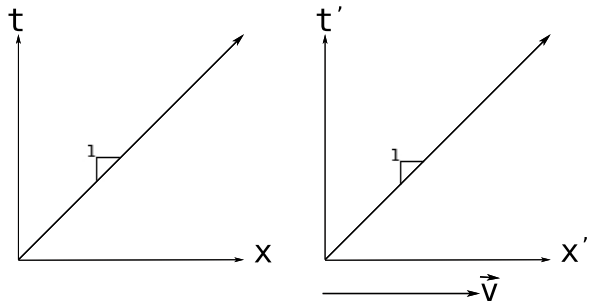


Figure 4.1: Two t-x space-time reference frames where the primed frame is moving with a speed of v relative to the other. In geometric units the speed of light is 1, so we must conceive of a transformation which converts from unprimed to primed coordinates yet also preserves the line of slope 1.

line to a line with slope v. This is enough to ensure that this transformation must be:

$$B_v = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}. \tag{4.1}$$

We will call this matrix the Boost or Lorentz matrix. Since it can be written as a matrix the transformation is linear. To see that this matrix satisfies our problem, first apply it to the vector $(t, 0)$:

$$B_v(t,0) = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix} \begin{pmatrix} t \\ 0 \end{pmatrix} = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} t \\ vt \end{pmatrix} = \frac{t}{\sqrt{1-v^2}} \begin{pmatrix} 1 \\ v \end{pmatrix}, \tag{4.2}$$

which is a point along the line of slope v as desired. We can also show:

$$B_v(1,1) = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1+v \\ v+1 \end{pmatrix} = \frac{1+v}{\sqrt{1-v^2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{4.3}$$

so the vector $(1, 1)$ is an eigenvector of our boost matrix. It is easy to show that the same is true of $(1, -1)$.

By applying this linear transformation to an arbitrary (t, x) space-time vector, this leads to the standard special relativity coordinate transformation of:

$$x' = \frac{x}{\sqrt{1-v^2}} + \frac{vt}{\sqrt{1-v^2}}, \tag{4.4}$$

$$t' = \frac{t}{\sqrt{1-v^2}} + \frac{vx}{\sqrt{1-v^2}}, \tag{4.5}$$

which should look familiar to readers with exposure to relativity.

We should note here that these equations do not make sense for $v > 1$, i.e., velocities faster than light. This is not the easiest idea to accept, but we will give a physical justification later. For now, the best way of understanding this is that there is no linear transformation which preserves the light cone from a rest frame to one moving superluminally. Nothing can go faster than the speed of light.

4.2 Minkowski Space

What are the implications of this? For one thing, the spatial distance between two points is no longer the same between two frames. It would appear that by fixing the speed of light in all frames we have destroyed any meaningful way of measuring spatial separation. However, there is a quantity that is preserved. By design this system preserves the light cone, the set of x and t such that $t^2 - x^2 = 0$. It turns out that this holds for $t^2 - x^2 = \tau^2$ for any constant τ, and furthermore for $t^2 - x^2 - y^2 - z^2 = \tau^2$ in a full four dimensional space-time[1]. This quantity can be thought of as the distance of (t, x, y, z) from the origin, and so we could easily extend this to see that the distance between any two points in space-time is preserved. The quantity τ is of great physical importance. We will call τ the *proper time* because it is the invariant quantity of the time as perceived by object. To see this, imagine a reference frame which kept the object at the origin. This would cause $t^2 - 0^2 - 0^2 - 0^2 = t^2 = \tau^2$. As all objects see themselves at rest, τ in this "special" frame is the time as felt by the object in question[2].

It will be useful for later to define a differential element of proper time: the amount of proper time which infinitesimally infinitesimal two events. This can be done as:

$$d\tau^2 = dt^2 - dx^2 - dy^2 - dz^2 = dt^2 \left(1 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2\right) = dt^2 \left(1 - v^2\right), \qquad (4.6)$$

and so $\dfrac{dt}{d\tau} = \gamma(v)$, where

$$\gamma(v) = \frac{1}{\sqrt{1 - v^2}}$$

is call the *Lorentz Factor*. This will allow us to take derivatives with respect to proper time.

We can think of the property of the Lorentz matrix preserving proper time as the result of a modified dot product of a space-time vector with itself. We define the *Minkowski Inner Product* for two space-time vectors $u = (t_1, x_1, y_1, z_1)$ and $w = (t_2, x_2, y_2, z_2)$:

$$\eta(u, w) = \begin{pmatrix} t_1 & x_1 & y_1 & z_1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} t_2 \\ x_2 \\ y_2 \\ z_2 \end{pmatrix} = t_1 t_2 - x_1 x_2 - y_1 y_2 - z_1 z_2 \qquad (4.7)$$

We will also use η to represent the matrix used in this definition, and will call our choice of η the *metric*[3]. One may note that we could construct the classical (or Euclidean) dot product by changing η to the identity matrix and ignoring the terms associated with time.

Similar to how we can define the length of a vector to be the square root of the vector dotted into itself, we can use the Minkowski inner product to define a new kind of "length" for a space-time vector. An issue arises though, as the Minkowski inner product is not positive-definite (always

[1] We could also talk about conserving $-t^2 + x^2 + y^2 + z^2$, which would lead to the same geometry, but with some sign changes. The literature on relativity uses both with equal frequency, and so it is important to know which is being discussed. We will say the system we are using has signs $(+ - - -)$, and the other $(- - - +)$.

[2] The importance of the proper time cannot be overstated. Philosophically it is nice because it means that while two observers may not agree on how old an object is, they will always agree on how old the object believes itself to be. More importantly for physics, is that an unstable particle's half life is measured with respect to its proper time, not the time of an observer.

[3] Were we to use the $(- - - +)$ convention, we would choose η to have opposite sign.

greater than zero unless involving a zero vector), so we cannot simply take the square root and get a real value. To rectify this we will consider three cases.

We will call a spacetime vector, u, *time-like* if $\eta(u,u) > 0$, *space-like* if $\eta(u,u) < 0$, and *light-like* or *null* if $\eta(u,u) = 0$. We then define the *Minkowski norm*, which will act differently on time, space, and light-like vectors. The Minkowski norm of a spacetime vector, u, takes the form of:

$$\|u\|_M = \begin{cases} \sqrt{\eta(u,u)}, \text{ if } u \text{ is time-like} \\ \sqrt{-\eta(u,u)}, \text{ if } u \text{ is space-like} \\ 0 \text{ if } u \text{ is light-like} \end{cases} \tag{4.8}$$

Where we use the subscript M to distinguish the Minkowski norm from the Euclidean norm. This gives us a concept of length for space-time vectors which will be equal in all inertial frames. It is also simple to define something analogous to distance, which we will call the *space-time interval* between u and w, as $\|u - w\|_M$.

Now, none of the results of the Minkowski norm would have changed had we simply defined the matrix η differently. Depending on our choice of η, we can define a different geometry, meaning different notions of lengths and distances. We call the geometry generated by choosing the identity matrix *Euclidean Geometry*, and we call the geometry we generated above *Minkowski geometry* or *Minkowski space*. We will consider more general spaces and geometries when we talk about relativity, the main difference being that we will define a different matrix for our inner product.

4.2.1 Index Notation

It is useful here to define some new notation. We shall relabel our coordinates t, x, y, and z as x^0, x^1, x^2, and x^3 respectively. This is helpful for thinking about the space and time coordinates in the same way, but is also more powerful than that. We will also define $\eta_{\mu\nu}$ as the element in the μth row and νth column of η and introduce implicit summation whenever we have matching upper and lower indices. This allows us to rewrite the Minkowski norm as $\|u\|_M^2 = x^\mu \eta_{\mu\nu} x^\nu$. This notation allows for much more elegant equations.

We will also define $x_\mu = \eta_{\mu\nu} x^\nu$, which we will call the *contravariant* coordinates. We call them this to distinguish them from the x^μ coordinates, which we will call the *covariant* coordinates. For our case here $x_0 = x^0$ and $x_\mu = -x^\mu$.

4.3 Mechanics in Minkowski Space

Now that we have a geometry which explains the invariance of the speed of light, the question becomes: what does this mean for physics?

We shall call the location of an event in space-time the *4-position*:

$$\mathbb{r} = (x^0, x^1, x^2, x^3) = (t, \mathbf{r}), \tag{4.9}$$

where $\mathbf{r}$ is the position in 3-space. It is natural to consider how velocity looks in space-time. We want some facsimile of the velocity which behaves nicely under Lorentz transformations, i.e., has invariant magnitude. We shall define the *4-velocity*:

$$\mathbb{V} = \frac{d}{d\tau}(x^0, x^1, x^2, x^3) = \frac{dt}{d\tau}(1, v_x, v_y, v_z) = \gamma(v)(1, \mathbf{v}), \tag{4.10}$$

where $\mathbf{v}$ denotes the ordinary velocity vector we are used to in 3-space. The differential position transforms as expected, and the differential element of proper time is invariant. This means the ratio of the two will transform in the same way that the 4-position does when we switch reference frames, that is to say, it can be transformed by applying the Lorentz transformation. We call all vectors which transform in the same way as the 4-position *4-vectors*. Also, we can easily note that in a frame where $v \ll 1$, $\gamma(v) \approx 1$, and so the 4-velocity becomes $(1, \mathbf{v})$, where the spatial component is the classical velocity.

To do mechanics in Euclidean space, we usually think in terms of the momentum of an object, and so we would like to define a 4-momentum so that we can do physics in Minkowski space. The obvious idea is to multiply our 4-velocity by the mass of the object. This will clearly create another 4-vector since we are simply multiplying a 4-vector by a scalar. Thus we define the 4-momentum as:

$$\mathbb{P} = m\mathbb{V} = \gamma m(1, \mathbf{v}). \tag{4.11}$$

We will call the spatial component of the 4-momentum $\mathbf{p} = \gamma m\mathbf{v}$ the *relativistic momentum*.

We would like to preserve the form of Newton's Laws for our space-time physics, and so we will define the *4-force* as:

$$\mathbb{F} = \frac{d\mathbb{P}}{d\tau}, \tag{4.12}$$

Which is a 4-vector for the same reasons the 4-velocity was. We will call the spatial component of the 4-force $\mathbf{F} = \dfrac{d\mathbf{p}}{d\tau}$ the *relativistic force*.

From this we can justify why objects cannot travel faster than light. Consider the work it would take to accelerate an object to light speed:

$$W = \int_{\text{path}} \mathbf{F} \cdot d\mathbf{x} = \int_{\text{path}} \frac{d\mathbf{p}}{d\tau} \cdot d\mathbf{x} = \int_{\text{path}} \mathbf{v} \cdot d\mathbf{p} = \int_{\text{path}} v\, dp \tag{4.13}$$

We know the expression for the relativistic momentum $p = \dfrac{mv}{\sqrt{1-v^2}}$, and can find $\dfrac{dp}{dv} = \dfrac{m}{(1-v^2)^{3/2}}$. Substituting into equation (4.13) yields:

$$W = \int_0^1 \frac{mv}{(1-v^2)^{3/2}} dv = \left.\frac{m}{\sqrt{1-v^2}}\right]_0^1, \tag{4.14}$$

which is clearly divergent, meaning it would take infinite energy to accelerate a massive object to light speed.

This also gives us a way to find the energy of a moving object. By changing the limits on the integral in (4.14), we can find the kinetic energy as:

$$KE = \frac{m}{\sqrt{1-v^2}} - m, \tag{4.15}$$

which we can see approaches the classical expression for $v \ll 1$ by Taylor expanding the first term:

$$KE = \left(m + \frac{1}{2}mv^2 + \mathcal{O}\left(v^4\right)\right) - m \approx \frac{1}{2}mv^2 \tag{4.16}$$

Note that the first term in equation (4.15) is the first component of the 4-momentum $\mathbb{P}^0 = \gamma m$.

Since this quantity seems to be associated with energy, we shall call it relativistic energy E. This approaches the classical energy in the non-relativistic limit plus a mass term, but this extra term is unimportant since it is just a constant additive factor. We will call this the *rest energy*.

We should note that in classical situations the magnitude squared of the 4-momentum is

$$\|\mathbb{P}\|^2 = \gamma^2 m^2(1 - v^2) = m^2 = E^2 - p^2 \tag{4.17}$$

This quantity is invariant by construction since $\mathbb{P}$ is a 4-vector. If we convert back to conventional units, this gives us Einstein's famous energy equation:

$$E^2 = p^2c^2 + m^2c^4, \tag{4.18}$$

or, for an object at rest:

$$E = mc^2. \tag{4.19}$$

One of the things we saw in special relativity was that nothing can go faster than the speed of light. However, we can use the same basic principal as before, the minimization of action, to find a solution which takes special relativity into account. For this section we will revert back to conventional units to better elucidate the physics.

To do this we must find the expression for the special relativistic Lagrangian, which we will derive from the Hamiltonian, assumed to be the total energy. We note that energy of a relativistic free particle is given by $E^2 = p^2c^2 + m^2c^4$, where p is the magnitude of the momentum, and so the Hamiltonian must be:

$$H = \sqrt{p^2c^2 + m^2c^4} = \gamma mc^2 \tag{4.20}$$

Where γ is the Lorentz factor.

The relationship between the Hamiltonian and the Lagrangian is $H = \sum \dot{q}_i p_i - L$, where the factors of $\dot{q}_i p_i$ here become $\gamma m v_i^2$, and so the Lagrangian for a free particle in special relativity is:

$$L_{\text{free}} = \gamma m v^2 - \gamma mc^2 = -mc^2\sqrt{1 - \frac{v^2}{c^2}} \tag{4.21}$$

Reconfiguring terms this becomes $L = -\dfrac{mc}{dt}\sqrt{dt^2 - dx^2 - dy^2 - dz^2}$. This last factor should familiar. It is the differential element of proper time, $d\tau$. This makes the action:

$$S = \int L dt = -mc \int d\tau = -mc(\tau_2 - \tau_1) \tag{4.22}$$

This means that the action principle for a free particle can really be thought of as the extremization of change in proper time! The importance of this is not explored here, but will become important in later sections where we will see that objects in a curved spacetime follow geodesics, paths which extremized proper time. This can be thought of as an extension of Fermat's Principle, in which light is thought to travel along the path which minimizes the time traveled. To account for interactions, we will write:

$$L = -mc^2\sqrt{1 - \frac{v^2}{c^2}} - U, \tag{4.23}$$

Where U is the potential energy.

The action principal itself is unchanged in Special Relativity, so our equations of motion still

come from the Euler-Lagrange equation:

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i}. \tag{4.24}$$

Chapter 5

The Bound Orbits in Special Relativity

5.1 The Action Principle in Special Relativity

Again, we will consider the system of a body in a gravitational well. Since our treatment of the classical two body problem put no limit on $\dot{r}$, the solution for the orbit must be wrong in the relativistic limit. We have a central gravitational potential of $U(r) = -\dfrac{GMm}{r}$, and so we have a Lagrangian of

$$L = -mc^2\sqrt{1 - \frac{v^2}{c^2}} + \frac{GMm}{r}.$$

Non-dimensionalizing as we did for the classical system, and using polar coordinates, we get:

$$L = -\sqrt{1 - \dot{r}^2 - r^2\dot{\theta}^2} + \frac{1}{r}. \tag{5.1}$$

We also find it useful to non-dimensionalize the energy, (given by the sum of the free-particle Hamiltonian and the potential energy). The total energy in special relativity has a term associated with the rest mass of the object, with the rest being the relativistic energy. We will split this term off, and call the remaining energy E_{rel}, and so:

$$E = E_{\text{rel}} + mc^2 \tag{5.2}$$

Thus when we non-dimensionalize using equation (3.6a) we get:

$$e + 1 = \frac{1}{\sqrt{1 - \dot{r}^2 - r^2\dot{\theta}^2}} - \frac{1}{r}, \tag{5.3}$$

where we modify (3.6a) so that $e = \dfrac{E_{\text{rel}}}{mc^2}$. We do this since e is the component of the energy which is related to the motion and gravitational potential, and so will the part we vary to describe different orbits. Note that when $v \ll c$, E_{rel} becomes the classical energy from earlier chapters.

We should use different symbols for the non-dimensionalized coordinates and quantities, but that becomes cumbersome. So we will use the same symbols for our convenience. It will be obvious from context whether or not an equation is non-dimensionalized, and can usually be determined

at a glance by the presence/absence of c terms.

From using the θ coordinate in the Euler-Lagrange equation (4.24), we get:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} = 0 \tag{5.4a}$$

This constant quantity $\dfrac{\partial L}{\partial \dot{\theta}}$ is the angular momentum, l, and is given by

$$l = \frac{r^2\dot{\theta}}{\sqrt{1 - \dot{r}^2 - r^2\dot{\theta}^2}} \tag{5.4b}$$

These quantities for e and l are similar to those we got in the classical treatment, and they reduce to the originals in the limit $v^2 = \dot{r}^2 + r^2\dot{\theta}^2 \ll 1$:

$$e = \frac{1}{2}(\dot{r}^2 + r^2\dot{\theta}^2)^2 + \frac{1}{r} + \mathcal{O}(v^3) \approx \frac{1}{2}(\dot{r}^2 + r^2\dot{\theta}^2)^2 + \frac{1}{r}, \tag{5.5a}$$

$$l = r^2\dot{\theta}\left(1 + \mathcal{O}(v^2)\right) \approx r^2\dot{\theta}, \tag{5.5b}$$

as we got before in equations (3.7a) and (3.7b).

5.2 The Pseudopotential

As before, we can find a useful quantity by setting $\dot{r} = 0$ and substituting the angular momentum (5.4b) into the total energy (5.3). In the special relativistic case this yields:

$$U_\psi = -1 - \frac{1 - \sqrt{l^2 + r^2}}{r} \tag{5.6}$$

However, unlike the classical case, this is *not* an effective potential. We cannot write the total energy as the sum of a portion corresponds to some motion, $\dot{r}$, and a component corresponding to the position, r. To distinguish the quantity we have just derived and the classical effective potential, we shall call U_ψ the *pseudopotential*.

Despite the pseudopotential not having the same physical interpretation as an effective potential, it is not useless. It still acts as a lower bound on e, since the energy with $\dot{r} \neq 0$ must be greater than the energy if $\dot{r} = 0$. It can also still find circular orbits since $\dot{r} = 0$ for any circular orbit, and so $e_{\text{circ}} = U_{\psi,\text{min}}$.

Taking the derivative of U_ψ from equation (5.6) and setting it equal to zero we find a minimum at:

$$r_c = l\sqrt{l^2 - 1}. \tag{5.7}$$

This allows us to find a minimum energy, which is:

$$e_{\text{min}} = U_\psi(r_c) = -1 + \frac{\sqrt{l^2 - 1}}{l}. \tag{5.8}$$

This is the minimum energy of a bound orbit can have.

Note that e_{min} ceases to have a real value for $l < 1$. We can also see from figure 5.1 that the pseudopotential looks qualitatively different for $l = 1$. It would appear that if l is too low then

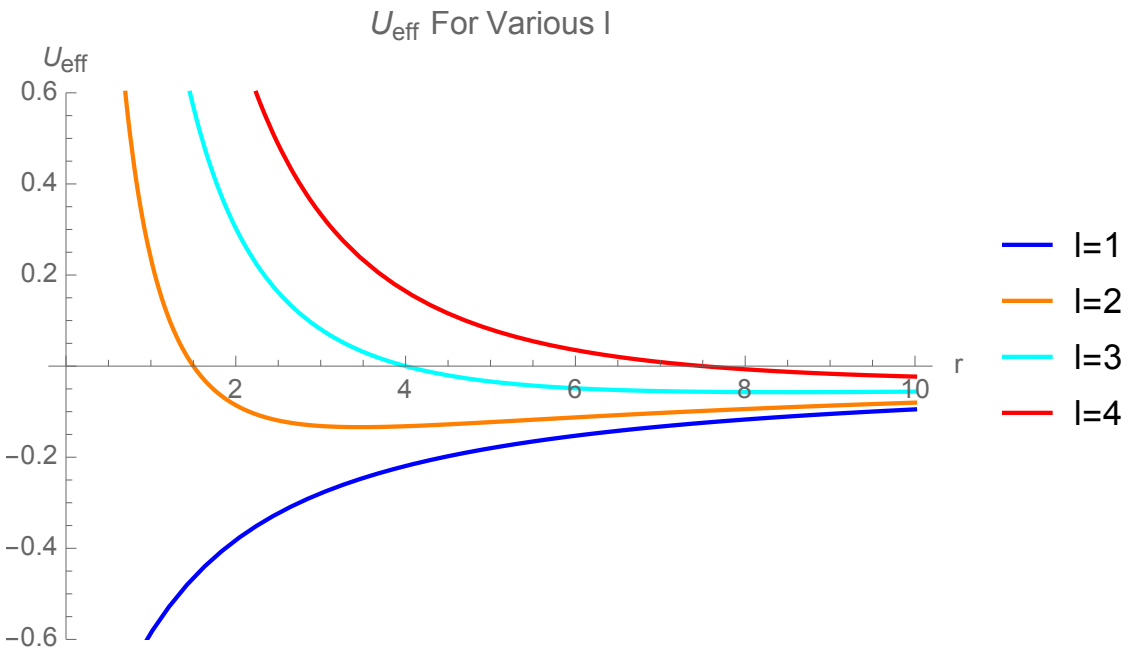


Figure 5.1: The pseudopotential for various values of l. When $l = 1$ the pseudopotential goes to -1 as r becomes small. For all other values of l there is a barrier before $r = 0$, and so we have bound orbits for $e < 0$. The minima of these plots correspond to circular orbits.

the object will fall in to the center. This was not a feature we saw of non-relativistic orbits, where we could not fall in so long as $l \neq 0$.

We can also see from the pseudopotential that we have an upper limit on what energies are allow for bound orbits. Since the limit of $U_\psi(r)$ as r approaches infinity is zero, the maximum energy for a bound orbity is:

$$e_{\max} = 0 \tag{5.9}$$

We will see that these observations will agree with the results from our solution.

5.3 Solving the Equation of Motion

The special relativistic system has an analytic solution much like the non-relativistic one did, and it will be found in a similar way.

As before, we want to solve for $\dot{\theta}$ so that we can rewrite our equations to remove the time dependence. We will use the same technique as we used in the classical problem and substitute $u = \dfrac{1}{r}$ and $\dot{r} = \dot{\theta} r' = -\dot{\theta}\dfrac{u'}{u^2}$. By manipulating equation (5.4b), we get:

$$\dot{r} = \frac{lu'}{\sqrt{1 + l^2u^2 + l^2u'^2}}. \tag{5.10}$$

By using equation (5.10) we can rewrite the total energy (5.3) as:

$$e + 1 = \sqrt{1 + l^2u^2 + l^2u'^2} - u \tag{5.11}$$

We can then solve for u' and get

$$\frac{du}{d\theta} = \pm\frac{\sqrt{(e+1+u)^2 - 1 - l^2u^2}}{l}. \tag{5.12}$$

By separating the differential we can integrate, and so:

$$\pm\int d\theta = \int \frac{l\,du}{\sqrt{(e+1+u)^2 - 1 - l^2u^2}}. \tag{5.13a}$$

As in the non-relativistic case, we have the choice of initial conditions by choosing the additive constants in the anti-derivatives. We will again choose them so that they are both zero, and so:

$$\pm\,\theta = \int \frac{l\,du}{\sqrt{(e+1+u)^2 - 1 - l^2u^2}}. \tag{5.13b}$$

It is not immediately obvious how to solve this integral. However, the term in the radical can be cleaned up by completing the square. Then with some clever substitutions we can get:

$$r(\theta) = \frac{r_0}{1 + \varepsilon\cos\eta\theta}, \tag{5.14}$$

where

$$\eta = \frac{\sqrt{l^2 - 1}}{l}, \tag{5.15a}$$

$$\varepsilon = \sqrt{\frac{e^2l^2 + 1 + 2el^2}{(e+1)^2}}, \tag{5.15b}$$

and

$$r_0 = \frac{l^2 - 1}{e+1}, \tag{5.15c}$$

5.4 Analysis

Equation (5.14) looks very similar to our original solution for r, but has an important distinguishing feature; The factor of η in the cosine. This factor causes the orbit to precess, and so relativistic orbits will not form the closed ellipses of Newtonian orbits. An example of such an orbit can be seen in figure 5.2.

This precession factor has another interesting implication. If $l < 1$, then η is imaginary, and so the position becomes $r(\theta) = \dfrac{r_0}{1 + \varepsilon\cosh|\eta|\theta}$. This causes the orbit to collapse in on the central body. This behavior can be seen in figure 5.3. This makes sense given what we saw from the effective potential in section 5.2.

We should also note that ε and r_0 are different here than they were in the classical case. The eccentricity is more complicated and the scale factor r_0 is smaller. However, we will see that these still limit to their original values in the non-relativistic limit. Also, if for $e = 0$, we get $\varepsilon = 1$, which indicates that, as we saw in the pseudopotential, the (non-rest-mass) energy of a bound orbit cannot exceed zero.

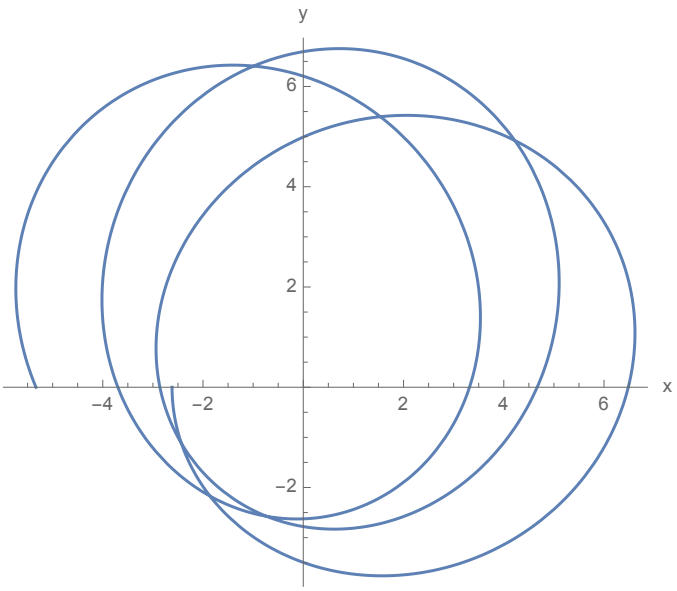


Figure 5.2: This is the orbit, $r(\theta) = \dfrac{r_0}{1 + \varepsilon \cos \eta\theta}$, for the energy and angular momentum values of -1.9 and 2.1 respectively. The orbit is near elliptical, but does not close back on itself. Instead it precesses about the origin.

5.4.1 Non-Relativistic Limits

In order to fully understand the non-relativistic limits, we would like to know the behavior of l and e when v is always very much less than c.

From equation (5.11) we can solve for l in terms of e, u and u'. This yields:

$$l = \sqrt{\frac{(e + 1 + u)^2 - 1}{u^2 + u'^2}}. \tag{5.16}$$

Since we want $v \ll c$ at all points on the orbit, let us consider when the object is at the minimum distance from the star, and so has the greatest velocity:

$$u_{\max} = \frac{1}{r_{\min}} = \frac{1 + \varepsilon}{r_0} = (e + 1)\frac{1 + \varepsilon}{l^2 - 1}, \tag{5.17}$$

$$u' = 0. \tag{5.18}$$

However, e is itself a function of u. From equation (5.3) we know that $e + 1 = \dfrac{1}{\sqrt{1 - v^2}} - u$. Plugging this into (5.17) and solving for u yields:

$$u_{\max} = \frac{1}{\sqrt{1 - v^2}} \frac{1 + \varepsilon}{l^2 + \varepsilon}. \tag{5.19}$$

Equations (5.16), (5.18), and (5.19) gives us:

$$l = v\frac{l^2 + \varepsilon}{1 + \varepsilon}, \tag{5.20}$$

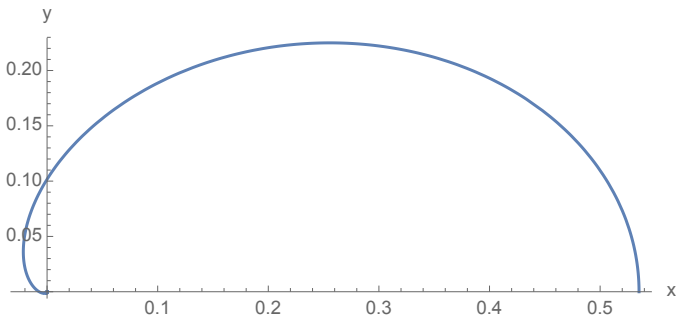


Figure 5.3: This is the "orbit", $r(\theta) = \dfrac{r_0}{1 + \varepsilon \cos \eta\theta}$, for the energy and angular momentum values of -1.5 and 0.5 respectively. This causes η to be imaginary, and so the object falls in to the center and cannot have a stable orbit.

which means that:

$$l = \frac{1+\varepsilon}{2v} + \sqrt{-\varepsilon + \frac{(1+\varepsilon)^2}{4v^2}}, \tag{5.21}$$

where we only concern ourselves with the positive solution since we assume l is positive. For a bound orbit the eccentricity is always between zero and one, so the limit of l as v becomes small is positive infinity. In other words, our non-relativistic limit on l is:

$$l \gg 1. \tag{5.22}$$

From here we know that the maximum energy of a bound orbit is zero, and the minimum is given by equation (5.8). However, in the non-relativistic limit, where l becomes large, the minimum becomes:

$$e_{\min} = -1 + \frac{\sqrt{l^2 - 1}}{l} \approx 0. \tag{5.23}$$

Thus in the non-relativistic limit, e is forced to be near zero, and approaches zero at the same rate that l approaches infinity. In other words:

$$|e| \ll 1. \tag{5.24}$$

By approaching our e and l limits our orbital parameters become:

$$\begin{aligned} \eta &\approx \frac{\sqrt{l^2}}{l} = 1, \\ \varepsilon &\approx \sqrt{\frac{0 + 1 + 2el^2}{1^2}} = \sqrt{1 + 2el^2}, \\ r_0 &\approx \frac{l^2}{1} = l^2. \end{aligned}$$

All of which line up with the values we got in the classical case.

Chapter 6

Relativistic Electrodynamics

One of the nice things about electromagnetism is that it already works in relativity with no need for modification. This is unsurprising since Maxwell's equations were a motivating factor in Einstein's theory. However, there are transformation rules: what is an electric field to one observer may be a magnetic field to another, because what was a stationary charge in the first frame is a current in the second.

The issue with electromagnetism is not how it changes under relativity, but it is how we represent it, and that starts with the electromagnetic 4-potential.

6.1 The 4-Potential

Maxwell's equations can be split into two categories: the homogeneous and inhomogeneous equations. The homogeneous equations are:

$$\nabla \times \mathbf{E} = -\frac{\mathbf{B}}{\partial t}, \tag{6.1a}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{6.1b}$$

These equations ensure that we can express the electric and magnetic fields in terms of an electric potential V and a magnetic vector potential $\mathbf{A}$:

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V \tag{6.2a}$$

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{6.2b}$$

For example, we can express the x component these of these fields as:

$$E_x = -\frac{\partial}{\partial t} A_x - \frac{\partial}{\partial x} V, \tag{6.3a}$$

$$B_x = \frac{\partial}{\partial y} A_z - \frac{\partial}{\partial z} A_y, \tag{6.3b}$$

and we can write the other components in a similar manner by cycling (x, y, z) in the equations.

Suppose we have a 4-vector, which we shall call the *4-potential*:

$$\mathbb{A} = (V, A_x, A_y, A_z), \tag{6.4}$$

with components A^μ. This should feel like a natural way to combine the two classical potentials into one 4-vector, though I will not go though argument to prove it transforms as a 4-vector.

Consider then the expression:

$$F^{\mu\nu} = \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu, \tag{6.5}$$

where we define the operator $\partial^\mu = \dfrac{\partial}{\partial x_\mu}$ and similarly $\partial_\mu = \dfrac{\partial}{\partial x^\mu}$. We should be careful to note that the derivatives in the above equation are with respect to the *covariant* coordinates, and that $x^0 = -x_0$, and so $\partial^0 = -\partial_0$. This ensures that $F^{\mu\nu}$ transforms covariantly, which is the tensor equivalent of transforming like a 4-vector. We can also see that $F^{\mu\nu} = -F^{\nu\mu}$ and that for $\mu = \nu$, $F^{\mu\nu} = 0$.

Note that for $\mu = 0$,

$$F^{0\nu} = \partial^0 A^\nu - \partial^\nu A^0 = -\frac{\partial}{\partial t} A^\nu - \frac{\partial}{\partial x_\nu} V$$

Which is equal to components of electric field for $\mu \neq 0$. Also, when both μ and ν are non-zero and unequal $F^{\mu\nu}$ looks like components of the magnetic field.

These $F^{\mu\nu}$ are components of an object called the *Electromagnetic tensor*:

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix}, \tag{6.6}$$

so $F^{\mu\nu}$ is the μth row and νth column of F.

The electromagnetic tensor in this form is covariant, as we can tell from the upper indices. We can construct a contravarient electromagnetic tensor by lowering the indices, which is accomplished by multiplying twice by $\eta_{\mu\nu}$:

$$F_{\mu\nu} = \eta_{\mu\alpha}\eta_{\beta\nu}F^{\alpha\beta} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix}. \tag{6.7}$$

6.2 The 4-Current

If we pick units such that $\varepsilon_0 = \mu_0 = 1$, (which is consistent with our previous unit choice of $c = 1$), we can write the inhomogeneous Maxwell equations in terms of the charge density ρ and current density $\mathbf{J}$ as:

$$\nabla \cdot \mathbf{E} = \rho, \tag{6.8a}$$

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} - \mathbf{J}. \tag{6.8b}$$

We can right away note that we can rewrite (6.8a) as:

$$\partial_\nu F^{0\nu} = \rho. \tag{6.9}$$

This should be immediately suggestive that we should create a 4-vector with ρ as the zeroth component. The natural partner of charge density is of course the current density, and so we shall define the *4-current* as:

$$\mathbb{J} = (\rho, J_x, J_y, J_z). \tag{6.10}$$

To show this is a useful expression consider:

$$\begin{aligned}
\partial_\nu F^{1\nu} &= \frac{\partial F^{10}}{\partial x^0} + \frac{\partial F^{11}}{\partial x^1} + \frac{\partial F^{12}}{\partial x^2} + \frac{\partial F^{13}}{\partial x^3} \\
&= -\frac{\partial E_x}{\partial t} + \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} \\
&= \left(-\frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} \right)_x \\
&= J^1 = J_x
\end{aligned}$$

The results are similar for $\mu = 2$ and 3, and the three together get equation (6.8b). This also allows us to write

$$\partial_\nu F^{\mu\nu} = J^\mu, \tag{6.11}$$

which gives the full information of the two inhomogeneous Maxwell equations.

6.3 Electromagnetic Lagrangian and Energy Density

We should note that the inner product of the electromagnetic tensor with itself gives us:

$$F_{\mu\nu} F^{\mu\nu} = 2\left(E^2 - B^2\right), \tag{6.12}$$

which is a quantity very much like the energy density:

$$u = \frac{1}{2}\left(E^2 + B^2\right), \tag{6.13}$$

but it is off by a factor of 4 and has the electric and magnetic components with opposite sign: for the energy density we want the sum of the squares, not the difference. Similarly, we can construct a term similar to the potential energy with:

$$J^\mu A_\mu = -\rho V + \mathbf{J} \cdot \mathbf{A} \tag{6.14}$$

Now, while (6.12) is not the energy density, it is related to another quantity with the same units: the Lagrangian density.

The Lagrangian density for electromagnetism is:

$$\begin{aligned}
\mathcal{L}_{\text{EM}} &= \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J^\mu A_\mu \\
&= \frac{1}{4} \left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) \eta^{\mu\alpha} \eta^{\beta\nu} \left(\partial_\alpha A_\beta - \partial_\beta A_\alpha\right) - J^\mu A_\mu
\end{aligned} \tag{6.15}$$

To see that this is the proper Lagrangian, we plug it into equation (2.14) with $j = \nu$ and

$\phi_i = A_\mu$. This yields:

$$-\frac{\partial \mathcal{L}_{\text{EM}}}{\partial A_\mu} = -\partial_\nu \frac{\partial \mathcal{L}_{\text{EM}}}{\partial \left(\partial_\nu A_\mu\right)},$$

and so:

$$J^\mu = -\partial_\nu \left(\partial^\nu A^\mu - \partial^\mu A^\nu\right) = -\partial_\nu F^{\nu\mu} = \partial_\nu F^{\mu\nu}, \tag{6.16}$$

which is identical to equation (6.11). So $\mathcal{L}_{\text{EM}}$ yields the proper equations, and is thus clearly the proper choice.

Now that we have the Lagrangian, we can find the Hamiltonian density:

$$\begin{aligned}\mathcal{H}_{\text{EM}} &= \left(\partial_0 A_\mu\right) \frac{\partial \mathcal{L}_{\text{EM}}}{\partial \left(\partial_0 A_\mu\right)} - \mathcal{L}_{\text{EM}}, \\ &= \frac{1}{2}\left(E^2 + B^2\right) - \rho V + \mathbf{J}\cdot\mathbf{A} + (\nabla V)\cdot\mathbf{E},\end{aligned} \tag{6.17}$$

which is almost what we expect for the energy density.

Now, when you integrate the last term over all space using integration by parts we get:

$$\iiint \nabla V \cdot \mathbf{E} dx^3 = VE\big|_{-\infty}^{\infty} - \iiint V\left(\nabla\cdot\mathbf{E}\right) dx^3, \tag{6.18}$$

where if we assume that the electric field vanishes at infinity and there are no source terms, (and so $\nabla \cdot \mathbf{E} = 0$), then equation (6.18) becomes zero.

Since (6.17) is the Hamiltonian *density*, we can ignore terms which integrate to zero, and so the electromagnetic Hamiltonian density in the absence of sources is:

$$\mathcal{H}_{\text{EM}}(\rho = 0, \mathbf{J} = \mathbf{0}) = \frac{1}{2}\left(E^2 + B^2\right), \tag{6.19}$$

which is exactly our expected energy density.

The terms which mattered to the source free Hamiltonian density were:

$$F_{0\mu}F^{0\mu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

The presence of zeros in the indices suggests that this is one term in a larger object. The fact that there is an upper zero an a lower zero means we should have a two index object, for if we were to put the same index in both those slots we would simply sum over that index, which does not seem right since with zeros there we have a physical term we are already familiar with. Also, the second term in this expression would appear to have no dependence on the zeroes, however, we can remedy this by putting an η^{00} there, which is just one and so does not effect the value. This would make our new tensor:

$$T^{\mu\nu} = \eta^{\mu\beta}F_{\beta\alpha}F^{\nu\alpha} - \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}. \tag{6.20}$$

We call the tensor $T^{\mu\nu}$ from equation (6.20) the *electromagnetic stress-energy tensor*. We have seen that T^{00} corresponds to the energy density of the electric and magnetic fields. It is natural

to ask what the other terms correspond to. Since $T^{\mu\nu}$ is symmetric with respect to μ and ν, we need only explore one half of the tensor, as the other will be the same.

Let us start with the rest of the $\mu = 0$ row:

$$\begin{aligned} T^{01} &= \eta^{0\beta} F_{\beta\alpha} F^{1\alpha} - \frac{1}{4} \eta^{01} F_{\alpha\beta} F^{\alpha\beta} \\ &= -S_x \end{aligned} \tag{6.21}$$

Where S_x is the x component of the Poynting vector. Similarly, we find that:

$$T^{02} = -S_y, \tag{6.22}$$

$$T^{03} = -S_z. \tag{6.23}$$

So we see that the energy-stress tensor is also related to the energy-flux density: the rate at which electromagnetic radiation delivers energy to a surface.

The other terms of the stress energy tensor $T^{\mu\nu}$ for $\mu, \nu \geq 1$ form the classical Maxwell stress tensor. This is why we call $T^{\mu\nu}$ the energy-stress tensor. It is the relativistic object which relates the energy-density and the electromagnetic stress.

Chapter 7

The Effects of Relativistic Electromagnetism on Bound Orbits

In chapter 5 we saw how the invariant of the speed of light, one of the most important aspects of special relativity, affected the shape of closed orbits. Here we will examine the effect of one of the other important results of relativity: the mass-energy equivalence.

We will do this by imagining that the body our object of interest is orbiting is charged. Classically this would change nothing about the orbits. The electric field only affects charged objects, so as long as the orbiting body is not charged we would expect the orbits to be unchanged. However, while the electric field will not exert an electric force on our satellite, it does contain energy. Under special relativity this energy *is* mass, and so the electric field will exert a *gravitational* force.

7.1 Charged Star

We have so far ignored that the object we are orbiting has a radius, however we it is now important since we need to find the total energy in the electric field between the star and the orbiting satellite. Consider the electric field from a star with radius R and charge q:

$$\mathbf{E}(r) = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}\hat{r} \tag{7.1}$$

We can use the shell theorem which states that for any inverse square law in a spherically symmetric situation we need only consider the volume within a sphere up to the radius of interest, as the effects from positions farther away will all cancel out. This means the energy stored in the electric field within the orbit is:

$$\begin{aligned} U_{\text{EM}} &= \frac{\varepsilon_0}{2}\iiint E(r')^2 dV' \\ &= \frac{q^2}{8\pi\varepsilon_0}\left(\frac{1}{R} - \frac{1}{r}\right) \end{aligned} \tag{7.2}$$

Thus, by equation (4.19), the "mass" of the electric field inside the orbit is:

$$M_{\text{EM}} = \frac{U_{\text{EM}}}{c^2} = \frac{q^2}{8\pi\varepsilon_0 c^2}\left(\frac{1}{R} - \frac{1}{r}\right). \tag{7.3}$$

We don't want to carry all these annoying constants around, so we will substitute in:

$$\kappa = \frac{q^2}{8\pi\varepsilon_0 M c^2}, \tag{7.4}$$

where M is the mass of the larger body. This allows us to rewrite equation (7.3) as:

$$M_{\text{EM}} = M\kappa\left(\frac{1}{R} - \frac{1}{r}\right). \tag{7.5}$$

This makes the gravitational force:

$$F(r) = -\frac{Gm(M + M_{\text{EM}})}{r^2} = -\frac{GmM}{r^2} - \frac{GmM\kappa}{Rr^2} + \frac{GmM\kappa}{r^3}, \tag{7.6}$$

where m is the mass of the orbiting satellite and G is the gravitational constant. This makes the gravitational potential:

$$U(r) = -\int_{\infty}^{r} F(r')dr' = -\frac{GmM}{r'} - \frac{GmM\kappa}{Rr'} + \frac{GmM\kappa}{2r'^2}\bigg|_{\infty}^{r} = -\frac{GmM}{r} - \frac{GmM\kappa}{Rr} + \frac{GmM\kappa}{2r^2} \tag{7.7}$$

Note that before when we did not care about the star's radius we also ignored the possibility of crashing into the star and only considered $r \to 0$ as falling in. Now, we must also assume that we are in the region *outside* the star, i.e., $r > R$. Thus we will replace the notion of $r \to 0$ with $r \to R$.

7.2 Finding the Equations of Motion

Unfortunately, unlike before, this system does not have an analytic solution even if we consider r a function of θ. To find the shapes of the orbits we must instead use numerical methods. To do this we will do as we did before and use angular momentum to remove the $\dot{\theta}$ terms and then solve for $\dot{r}$ as a function of r. This allows us to use Runge-Kutta methods to numerically solve the differential equation given initial conditions. But before any of this we must first set up our Lagrangian.

If we non-dimensionalize we can construct the Lagrangian as:

$$L = -\sqrt{1 - \dot{r}^2 - r^2\dot{\theta}^2} + \frac{1}{r} + \frac{\kappa}{Rr} - \frac{\kappa}{2r^2}, \tag{7.8}$$

where we omit writing our dimensionless quantities with tildes for notation convenience.

We should note that when $\kappa = 0$, we return the Lagrangian for the uncharged case in equation (5.1). This means that our solution in this section should reduce to the previous case when $\kappa = 0$. In particular, we can find values for the precession factor, the eccentricity, and scale factor, and they should approach the values from equations (5.15a) through (5.15c) when κ goes to zero.

Since we did not add or change any $\dot{\theta}$ terms in the Lagrangian, we can make the same derivations for angular momentum as we did in chapter 5. This makes the angular momentum:

$$l = \frac{r^2\dot{\theta}}{\sqrt{1 - \dot{r}^2 - r^2\dot{\theta}^2}}. \tag{7.9}$$

The total energy of this system is given by the Hamiltonian as defined in equation (2.7). However, since the canonical momenta of this system are the same as in chapter 5, the energy is the same as in equation (5.3) except with the modified potential of equation (7.7). It is thus:

$$e + 1 = \frac{1}{r}\sqrt{\frac{l^2 + r^2}{1 - \dot{r}^2}} - \frac{1}{r} - \frac{\kappa}{Rr} + \frac{\kappa}{2r^2}. \tag{7.10}$$

We can solve the above to get equations for $\dot{\theta}$ and $\dot{r}$, which yields:

$$\dot{r}^2 = 1 - \frac{l^2 + r^2}{r^2\left(e + 1 + \frac{1}{r} + \frac{\kappa}{Rr} - \frac{\kappa}{2r^2}\right)^2} \tag{7.11a}$$

$$\dot{\theta} = \frac{l}{r^2\left(e + 1 + \frac{1}{r} + \frac{\kappa}{Rr} - \frac{\kappa}{2r^2}\right)} \tag{7.11b}$$

To find the shape of the orbits we will rewrite our equations to instead think of r as a function of θ. We can do this as we did before by remembering that $\dot{r} = \dot{\theta}r'$, where r' is the derivative of r with respect to θ.

Thus equations (7.11a) and (7.11b) become:

$$r'^2 = \frac{r^4}{l^2}\left[\left(e + 1 + \frac{1}{r} + \frac{\kappa}{Rr} - \frac{\kappa}{2r^2}\right)^2 - 1\right] - r^2. \tag{7.12}$$

It is preferable when using numerical methods to use second order differential equations when we have a sign ambiguity as we do in the above equation. To this end we then differentiate with respect to θ. This gives us an equation for r'':

$$r'' = \frac{2r^3}{l^2}\left[\left(e + 1 + \frac{1}{r} + \frac{\kappa}{Rr} - \frac{\kappa}{2r^2}\right)^2 - 1\right] - \frac{r^4}{l^2}\left(e + 1 + \frac{1}{r} + \frac{\kappa}{Rr} - \frac{\kappa}{r^2}\right)\left(\frac{1}{2r^2} + \frac{\kappa}{Rr^2} - \frac{\kappa}{r^3}\right) - r \tag{7.13}$$

This cannot be solved analytically, but can be solved numerically. However, numerical methods require knowing our initial conditions. So we want a way of knowing which initial conditions give us bound orbits. To do this we again create a pseudopotential, which will also give us intuition into the physics of the system.

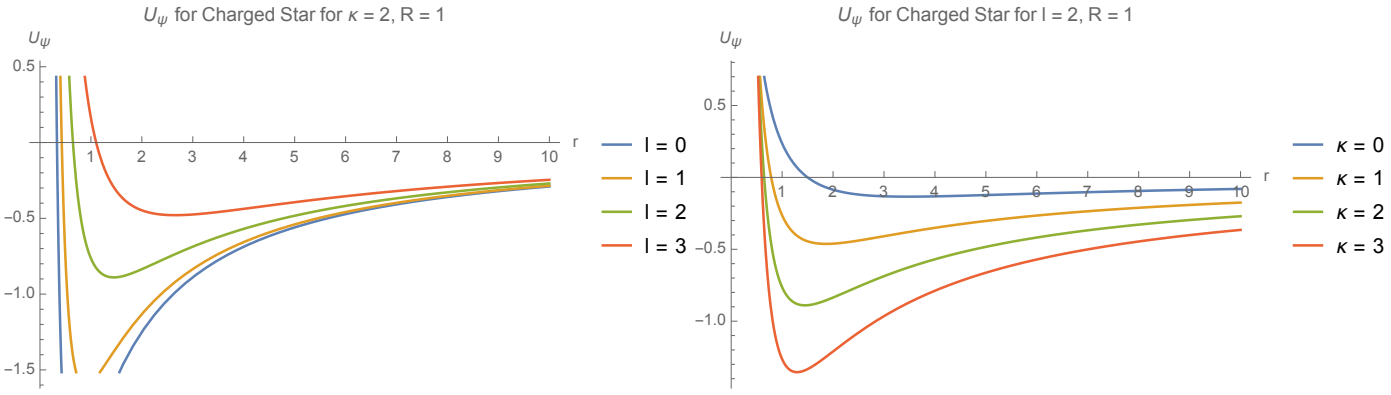
7.3 Pseudopotential With Charge

As before we can create a pseudopotential by evaluating the energy e at $\dot{r} = 0$, which becomes:

$$U_\psi = e\,(\dot{r} = 0) = -1 + \frac{1}{r}\sqrt{l^2 + r^2} - \frac{1}{r} - \frac{\kappa}{Rr} + \frac{\kappa}{2r^2}, \tag{7.14}$$

which, it should be noted, is the equal to the pseudopotential from equation (5.6) but with the modified potential term. A plot of the pseudopotential can be seen in figure 7.1.

Note that we can only get meaningful information when $r > R$, as this was assumed when we



(a) (b)

Figure 7.1: Figure (a) plots the pseudopotential for various values of l with κ fixed at 2. Figure (b) plots the pseudopotential for various values of κ with l fixed at 2. Both have $R = 1$. Note that the graphs for $l = 0$ and $l = 1$ with $\kappa = 2$ have local minima, seemingly implying that there can be bound orbits in these cases. This is fallacious as the minima occur at an r less than 1, and so would be inside the star. This is unphysical, and so they can be discounted.

set up our initial equations. So while it is possible to get turning points at radii less than R, these are unphysical and are not true orbits.

We will again find it useful to find the radius and energy of circular orbits. To do this we try to minimize U_ψ with respect to r. This results in the following equation for radius of circular orbit, r_c:

$$\left.\frac{\partial U_\psi}{\partial r}\right|_{r_c} = \frac{1}{r_c^2}\left(1 + \frac{\kappa}{R} - \frac{l^2}{\sqrt{l^2 + r_c^2}}\right) - \frac{\kappa}{r_c^3} = 0, \tag{7.15}$$

which can be rewritten as a quartic equation of r_c:

$$r_c^4(\kappa+R)^2 - 2\kappa r_c^3 R(\kappa+R) + r_c^2\left(-l^4R^2 + l^2(\kappa+R)^2 + \kappa^2R^2\right) - 2\kappa l^2 r_c R(\kappa+R) + \kappa^2 l^2 R^2 = 0. \tag{7.16}$$

Luckily, quartics can be solved explicitly. There exists an explicit formula for the roots of a quartic function, however, we will gain little by writing it out. Instead we plot the quartic and so can see the roots clearly. A plot of the solution for various l can be seen figure 7.2. Note that for $\kappa = 0$, equation (7.16) becomes:

$$R^2\left(r_c^4 + l^2 r_c^2\left(l^2 + 1\right)\right) = 0, \tag{7.17}$$

Which has the radius found in equation (5.8) as a solution. This confirms that our equation reproduces the behavior we found in the uncharged case.

It continues to be true that the energy of circular orbit is the minimum allowed energy for a bound orbit, so for our numerical solutions, we will want to only use energies greater than it.

7.4 Numerical Methods to Solve the Equations of Motion

Our numerical calculations were done using Mathematica and an adaptive step 4th-order Runge-Kutta method to calculate the shape of the orbit. The code I wrote to implement the method can

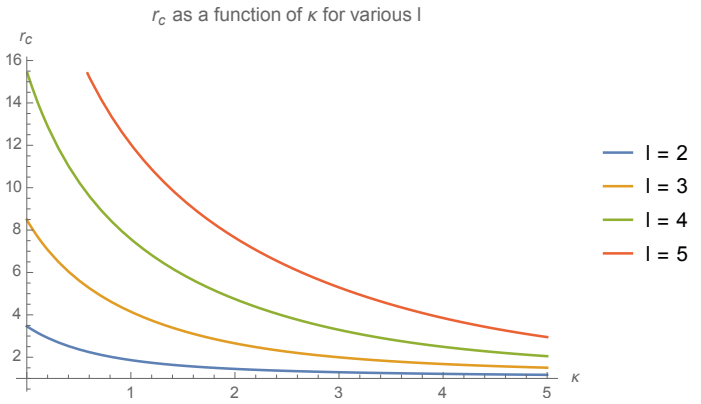


Figure 7.2: A plot of the circular radius for $R = 1$ as a function of κ for several values of l. We can see that the eventually cut off, at which point there is no circular orbit possible. Also, we must cut off the plot when $r_c = 1$, because below that the orbit is unphysical.

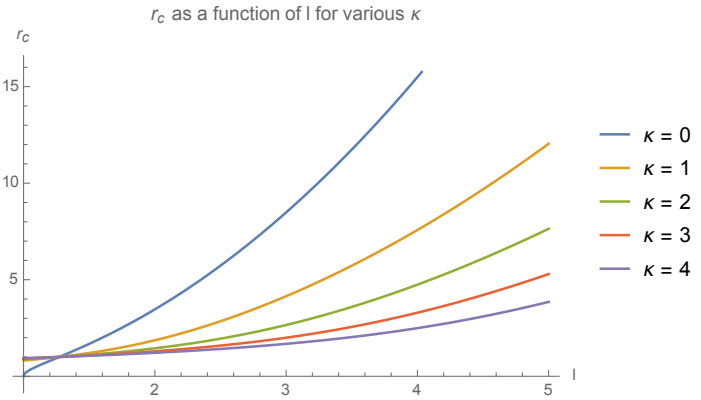


Figure 7.3: A plot of the circular radius for $R = 1$ as a function of l for several values of κ. We can see that the eventually cut off, at which point there is no circular orbit possible. Also, we must cut off the plot when $r_c = 1$, because below that the orbit is unphysical.

be seen in the appendix. Applying this method to our differential equations, we can generate a list of r and θ values, corresponding such that the nth r-value occurs at the nth angle in the θ list. This allows us to plot the orbits. An example of an orbit generated using this method can be seen in figure 7.4.

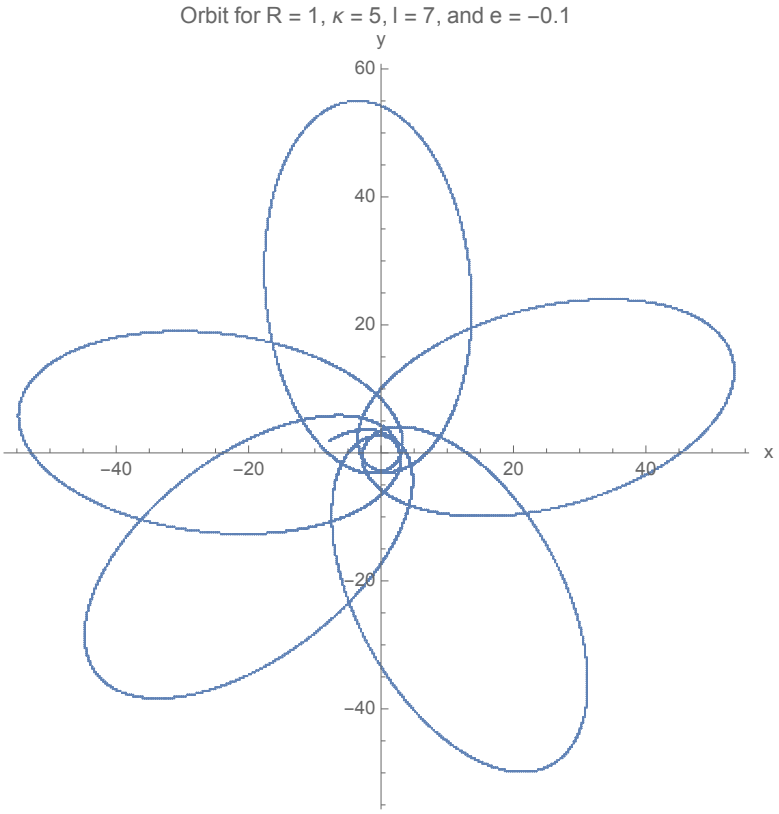


Figure 7.4: A plot of the numerical approximation the orbit around a charged star with $R = 1$, $\kappa = 5$, $l = 7$ and $e = -0.1$. At a glance it may look very similar to the solutions to the non-charged star, but this is not the case.

To ensure that all orbits we generate are bound, we can take our energies to be fraction (less than one) times the energy of a circular orbit because the allowed energies for a bound orbit are between the energy of a circular orbit and zero.

Once we generate a numeric orbit, we would like to extract several quantities: the characteristic radius, r_0; the eccentricity, ε; and the precession angle, η. To do we define quantities we believe to will become those found in equations (5.15a) to (5.15c) when $\kappa = 0$. Then we run the code in the appendix called Charged Orbital Analysis to obtain the quantities from the numerical data.

We get an approximation for ε by recalling equation (5.14), and noting that in the uncharged case:

$$r_{\min} = \frac{r_0}{1+\varepsilon}, \tag{7.18a}$$

$$r_{\max} = \frac{r_0}{1-\varepsilon}, \tag{7.18b}$$

which can be solved for ε, yielding:

$$\varepsilon = \frac{r_{\max} - r_{\min}}{r_{\max} + r_{\min}} \tag{7.19}$$

Since $r_{\min}$ and $r_{\max}$ are easy to obtain from a numeric solution, we shall use (7.19) as our approximation for the eccentricity. Equation (7.18a) can be trivially rearranged to find r_0 in terms of $r_{\min}$ and ε, both of which we can now get from our numerical solution. We will take our approximation

of r_0 as:

$$r_0 = r_{\min}(1 + \varepsilon). \tag{7.20}$$

To find η, we start by finding the angular differences between local maxima, and then average them over the whole orbit to minimize numerical defects, then divide 2π by the result. This works to find the precession factor since, in a precessing orbit, the angular difference between two local maxima is such that $\eta\theta = 2\pi$. Finding the local maxima is fairly simple, we simply compare a point to its nearest neighbors and see if it is greater. We then find the difference in the θ values corresponding to the maxima from one to the next. We can write a formula which tell us how to find η as:

$$\eta = \frac{2\pi}{\overline{\Delta\Theta}_{\max}}, \tag{7.21}$$

where $\Delta\Theta$ is the difference between the θ values of adjacent point in our numerical data

To see that these approximations of r_0, ε, and η are actually what we want, we can extract them from the orbit in figure 7.4 and then overlay a plot of (5.14) using our values of them. If they indeed give us the desired information the two plot should look similar. This is done in figure 7.5, and we can see the two orbits almost match, except that the fit is "thinner" than the numerical solution. This can be better seen in figure 7.6, where we plot the data except that we multiply the θ coordinate of each point by the η found from (7.21). Also plotted is the ellipse whose scale factor and eccentricity are those found by Charged Orbital Analysis. We can see that data is not elliptical, and produces and orbit whose bulbs are wider than would be expected if it were. However, this is only noticeable for small l, i.e., the most relativistic values of l. This can be seen in figure 7.7, where we repeat the above process for a large value of l.

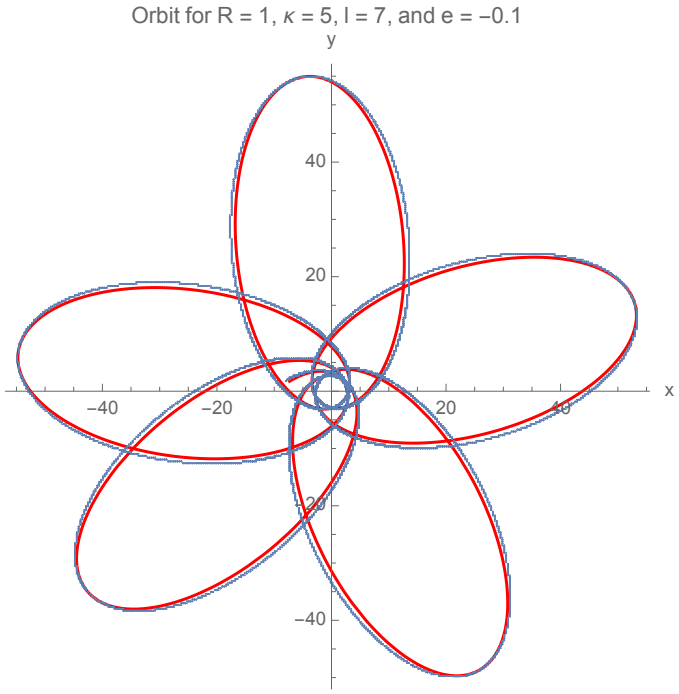


Figure 7.5: A plot of the numerical approximation the orbit around a charged star with $R = 1$, $\kappa = 5$, $l = 7$ and $e = -0.1$ in blue, along with an non-charged approximation orbit using the values of r_0, ε, and η, 5.14, 0.907, and 0.821 respectively, obtained from the numerical orbit plotted in red. The two look extremely similar, and so we can say with fair certainty that our approximations are good.

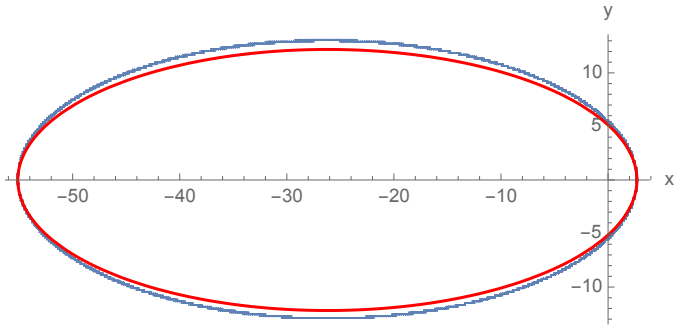
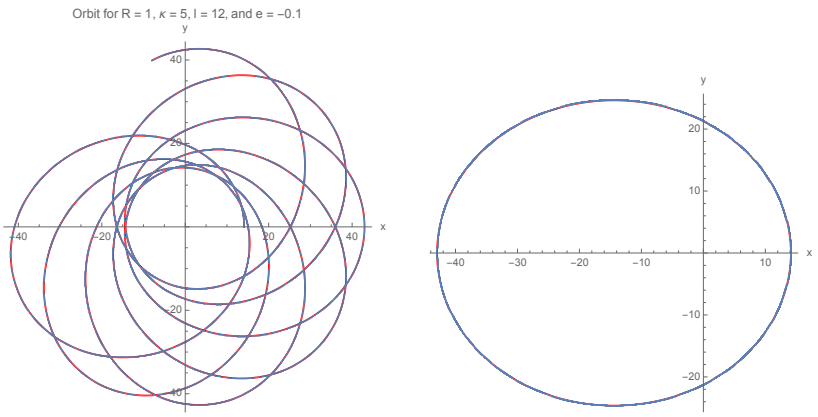


Figure 7.6: A plot of the numerical approximation the orbit around a charged star with $R = 1$, $\kappa = 5$, $l = 7$ and $e = -0.1$ in blue, along with an non-charged approximation orbit using the values of r_0, ε, and η, 5.13, 0.907, and 0.820 respectively, obtained from the numerical orbit plotted in red. The two look extremely similar, and so we can say with fair certainty that our approximations are good.

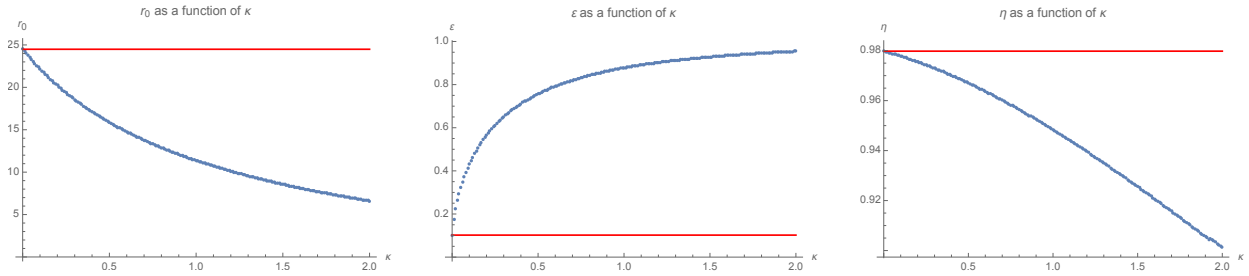
To verify that these numerical solutions match with what we found in the uncharged case, we simply pick values of e, l, and R, and then find the characteristics for various κ. We then compare the characteristics at $\kappa = 0$ with what we would expect from our earlier formulas. This is done in figure 7.8, and we see that they agree at the limit. It should be noted that the behaviors for all



(a) (b)

Figure 7.7: Analogous plots to those in figures 7.5 and 7.6, except using $l = 12$. These plots show that the orbits are much closer to the uncharged case for large l.

the parameters is monotonic, and so we do not expect there to be critical values of κ, (other than zero), which dramatically change to effects of the charge.



(a) (b) (c)

Figure 7.8: Plots of r_0, ε, and η as functions of κ. These were made keeping e and l constant at -0.02 and 5 respectively. The red bars indicate the values for each orbital characteristic we expect for the give e and l values as given in equations (5.15a),(5.15b), and (5.15c). The plots clearly converge on their expected $\kappa = 0$ values.

Chapter 8

Manifolds, Metrics, and Covariant Derivatives

Before we can move on to Einstein's general theory of relativity we need to establish the geometry we will be using, much as we constructed Minkowski space before we started working with special relativity. The key idea used in general relativity is that gravity is expressed as the *curvature* of spacetime.

However, what exactly curvature means in spacetime is non-obvious. The traditional metaphor is bowling balls on a sheet: The bowling ball will sink into the sheet, creating an indentation around it, and any object which rolls near it will bend in the well created by the bowling ball. This makes for a good insight, but does note extend nicely to the spacetime. The idea being that the metaphor uses extrinsic curvature, curvature which is seen by someone not on the surface. We cannot leave the confines of spacetime, and so need a notion of the intrinsic curvature.

What we want is a notion of curvature which is inherent to the space itself, and not reliant on it curving in an ambient direction. To develop such a notion, we must first develop a language to talk about these spaces in: the language of Riemannian manifolds.

8.1 Manifolds, Tangent Spaces, and Riemannian Metrics

In general a *manifold* is any space which is locally Euclidean. What this means is that in any neighborhood of any point on an n-dimensional manifold, or n-manifold, the space will look like $\mathbb{R}^n$. We will only consider manifolds which are themselves subsets of Euclidean space. Note that with our definition, the dimension of the manifold is not the dimension of the Euclidean space it is a subset of, but rather the dimension of the manifold itself. For example consider the sphere, which is a subset of $\mathbb{R}^3$: it is two dimensional, in that around each point you can put a two dimensional coordinate grid.

Now, we define a path in a manifold M as a smooth function $\gamma\colon \mathbb{R} \to M$. We say a path goes through a point p is there exists a $t \in \mathbb{R}$ such that $\gamma(t) = p$. We can take derivatives of the path at a point p. This allows us to define directions one can go from a point. In fact, we can use this to define vectors tangent to M at a point. If M is some n-dimensional manifold, then at every point p on the manifold, the set T_pM is the set of vectors tangent to M at point p, and is called the *tangent space* of M at p. To say this more formally, T_pM is the vector space composed of vectors $\gamma'(t)$, where γ is a path in M and $\gamma(t) = p$. The tangent space will have dimension equal to that of the manifold. For example, if our manifold is a sphere, a two dimensional manifold, the tangent

space at a point p will be the set of vectors which lie in the plane tangent to the sphere at p, and so will be isomorphic to $\mathbb{R}^2$.

What tangent spaces mean for us is that we can no longer think of vectors as free floating arrows, now they must be based at a specific point, and two vectors based at different points are *not* the same object. For example, consider two points, p_1 and p_2, of equal height on the upper hemisphere of a sphere, neither on the equator nor at the pole. Consider the two paths which connect each of the points to the pole, and consider the unit vectors generated by these paths. Both can be thought of as the "up" unit vector at these points, however, they are not the same vector, as both will have a slight tilt to them depending where on the hemisphere they are. Essentially, what tangent spaces do is allow us to consider direction locally, and not globally.

To finalize this geometry we need only define the metric. For a manifold, M, a *Riemannian metric* g is a function which maps vectors from the same tangent space, $v, w \in T_pM$, for all $p \in M$, to a real numbers with the following properties:

- Symmetric: $g(v, w) = g(w, v)$
- Positive definite: $g(v, v) \geq 0$ and $g(v, v) = 0$ iff $v = 0$
- Bilinear: $g(u + v, w) = g(u, w) + g(v, w)$ and $g(ku, w) = kg(u, w)$

We will call a manifold equipped with a Riemannian metric a *Riemannian manifold*.

The metric is analogous to the dot product, and in the same way we can use the dot product to define lengths and angles, we can use the metric g. We will say the length of a vector v is:

$$\|v\| = \sqrt{g(v, v)}$$

and we will define the angle between to non-zero vectors v, w in the same tangent space as:

$$\cos\theta_{v,w} = \frac{g(v, w)}{\|v\|\|w\|}$$

Also, the fact that g is linear means that we can represent it as a matrix. Suppose we are on a 2-dimensional manifold. Then all the tangent spaces are planes, and we can write the vectors v and w in terms of a basis. Let $\mathbf{e}_1$ and $\mathbf{e}_2$ form a basis for some tangent space T_pM, with $v = v_1\mathbf{e}_1 + v_2\mathbf{e}_2$ and $w = w_1\mathbf{e}_1 + w_2\mathbf{e}_2$ as vectors in T_pM. Thus:

$$\begin{aligned} g(v, w) &= g(v_1\mathbf{e}_1 + v_2\mathbf{e}_2, w_1\mathbf{e}_1 + w_2\mathbf{e}_2) \\ &= \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} g(\mathbf{e}_1, \mathbf{e}_1) & g(\mathbf{e}_1, \mathbf{e}_2) \\ g(\mathbf{e}_2, \mathbf{e}_1) & g(\mathbf{e}_2, \mathbf{e}_2) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \end{aligned}$$

This is done here in only two dimensions, but can be trivially extended to the n-dimensional case.

This means that to know how g acts on any vectors, we need only know how it acts on a basis. We will denote the matrix form of the metric as:

$$g = \begin{pmatrix} g(\mathbf{e}_1, \mathbf{e}_1) & g(\mathbf{e}_1, \mathbf{e}_2) \\ g(\mathbf{e}_2, \mathbf{e}_1) & g(\mathbf{e}_2, \mathbf{e}_2) \end{pmatrix}$$

And we will label the entries as $g_{ij} = g(\mathbf{e}_i, \mathbf{e}_j)$. Since the metric is invertible (do to positive definiteness), we write the entries of the inverse matrix as g^{ij}. We can use this notation with the summation notation we set up in section 4.2.1 to then say that $g_{ij}g^{jk} = \delta_i^k$, the Dirac delta

function. Also, note that the metric is symmetric, and so $g_{ij} = g_{ji}$. We shall similarly take the inverse metric to be symmetric.

Similarly, we can use index notation to rewrite our definition of length as:

$$\|v\| = \sqrt{v^i v^j g_{ij}} \tag{8.1}$$

This means that the arclength of a curve $\gamma(t) = (x^1(t), x^2(t), \ldots, x^n(t))$ from t_0 to t_1 is:

$$\tau = \int_{t_0}^{t_1} \sqrt{\frac{dx^i}{dt}\frac{dx^j}{dt} g_{ij}}\, dt \tag{8.2}$$

8.2 The Geodesic Equation

If we call the term in the integral L, then we can find the curve which extremizes length between $\gamma(t_0)$ and $\gamma(t_1)$ using the calculus of variations we used in chapter 2. We call such curves geodesics. Geodesics will become important to physics for the same reason as before: in general relativity, objects will follow paths which extremize this length. We will later make this connection more concrete.

In any case, the equations which define geodesics are the Euler-Lagrange equation:

$$\frac{d}{dt}\frac{\partial L}{\partial dx^l/dt} = \frac{\partial L}{\partial x^l}. \tag{8.3}$$

Also note, that we write L as the derivative of τ with respect to t. This allows us to reparametrize in terms of arclength should we need to.

The left side of the (8.3) is a simple derivative if we assume the coordinates x^i and the derivatives of the coordinates $\dfrac{dx^i}{dt}$ are independent:

$$\begin{aligned}\frac{\partial L}{\partial x^l} &= -\frac{1}{2L}\frac{dx^i}{dt}\frac{dx^j}{dt}\frac{\partial g_{ij}}{\partial x^l} \\ &= -\frac{L}{2}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}\frac{\partial g_{ij}}{\partial x^l},\end{aligned}$$

where in the last step we convert t derivatives to τ derivatives by the relation $\dfrac{d}{dt} = L\dfrac{d}{d\tau}$. The other side is:

$$\begin{aligned}\frac{\partial L}{\partial dx^l/dt} &= -\frac{1}{2L}\left(\frac{dx^j}{dt}g_{lj} + \frac{dx^i}{dt}g_{il}\right) \\ &= -\frac{dx^i}{d\tau}g_{il}.\end{aligned}$$

Next, we take the t derivative, which we can convert into an arclength derivative:

$$\begin{aligned}\frac{d}{dt}\frac{\partial L}{\partial dx^l/dt} &= L\frac{d}{d\tau}\left(-\frac{dx^i}{d\tau}g_{il}\right)\\ &= -L\left(\frac{d^2x^i}{d\tau^2}g_{il} + \frac{1}{2}\left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j}\right)\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}\right),\end{aligned}$$

where in the last step we duplicate the derivative of the metric and divide by half to compensate. Putting both sides of equation (8.3) together yields:

$$\frac{d^2x^i}{d\tau^2}g_{il} = -\frac{1}{2}\left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\right)\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}$$

In order to make our equations look nicer, we will multiply both sides of the above by g^{kl} in order to isolate the second derivative term. This gives us:

$$\frac{d^2x^k}{d\tau^2} + \frac{1}{2}g^{kl}\left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\right)\frac{dx^i}{d\tau}\frac{dx^j}{d\tau} = 0.$$

We will define the *Christoffel symbol* Γ^k_{ij} as:

$$\Gamma^k_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l}\right) = \frac{1}{2}g^{kl}\left(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}\right), \tag{8.4}$$

where in the second version we write the partial derivative with respect to x^i as ∂_i. This symbol allows us to neatly write the geodesic equation succinctly as:

$$\frac{d^2x^k}{d\tau^2} + \Gamma^k_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau} = 0, \tag{8.5}$$

which must hold for each k. This gives us one equation for each dimension.

8.2.1 Notation of the Metric

We earlier introduced a way of expressing the metric as components of a matrix and then used that notation in equation (8.1). If we consider a vector of an infinitesimal length, $d\tau$, we can rewrite equation (8.1) as:

$$d\tau^2 = g_{ij}dx^i dx^j \tag{8.6}$$

This is really just another way we can write the metric, but it gives us a nice physical interpretation. We can thus express the metric in a way using infinitesimal displacements, which is a natural way of thinking about it in physics. It also tells us how we can integrate over length in a Riemannian manifold, as we simply take the square root of $d\tau^2$.

We will use this notation in later sections.

8.3 Lorentzian manifold

However, this is not enough yet. This is a way of talking about space. We need a mathematical construct which will correspond to spacetime. To do this we must slightly modify how our metric

works. This will lead to what is called a Lorentzian manifold.

A *Lorentzian manifold* will have a (Lorentzian) metric which will be identical to that of Riemannian metric except we replace the requirement of positive definiteness with being non-degenerate. This means that if $g(v, w) = 0$ for all $w \in T_pM$, then $v = 0$. We also impose the restriction that each tangent space, T_pM, has a basis such that $g_{11} = 1$ and that $g_{\mu\mu} = -1$ for all $\mu \neq 1$. This allows us to have three "spatial" dimensions and one "temporal" dimension. We will use Greek subscripts to differentiate Lorentzian manifolds from Riemannian ones.

Now that we allow the metric to be negative, we can split all vectors into three categories, just as we did in special relativity. We will call vectors v for whom $g(v, v) > 0$ *time-like*, $g(v, v) < 0$ *space-like*, and $g(v, v) = 0$ either *null* or *light-like*. A path γ is said to be time-like if $g(\gamma'(t), \gamma'(t)) < 0$ for all t. These will be the paths we will consider to be physical. We must also modify our formula for distance; in a Lorentzian manifold the length of a vector will be:

$$\|v\| = \sqrt{|v^\mu v^\nu g_{\mu\nu}|} \tag{8.7}$$

We do this because our modified metric allows for negative values, which would otherwise cause imaginary lengths. We would like our notion of length to remain a positive real number, and so we take the absolute value of the metric to compensate for this. However, this modification does not effect the geodesic equation, and we could rederive it in the same exact way as before. To see this, note that:

$$\frac{d}{dx}L = \frac{d}{dx}\sqrt{\left|\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}g_{\mu\nu}\right|} = \frac{\frac{d}{dx}\left|\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}g_{\mu\nu}\right|}{2\sqrt{\left|\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}g_{\mu\nu}\right|}} = \frac{1}{2L}\text{sign}\left(\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}g_{\mu\nu}\right)\frac{d}{dx}\left(\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}g_{\mu\nu}\right),$$

where x is some arbitrary quantity with respect to which we can take a derivative. This equation shows that differs from what we would get without the absolute value by at worst at sign, and at best nothing at all. Since both sides of the Euler-Lagrange equation have derivatives of L, the factor of $\text{sign}\left(\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}g_{\mu\nu}\right)$ cancels out, and so we can derive the geodesic equation the same as we could in a Riemannian manifold.

The concepts of a Lorentzian manifold should all be sounding very familiar. This is because we have actually already dealt with a Lorentzian manifold. Minkowski space, discussed in section 4, is a simple example of a Lorentzian manifold. We did not need to talk about it as a manifold before though, since it is a *flat* manifold. The metric is just a modified identity matrix, and so there is little the material from this section adds to the picture to justify the complexity and generality. However, now that we wish to move to dealing with a curved spacetime, it is now necessary.

8.4 Covariant Derivatives and Curvature

An issue arises from having different vector spaces associated with every point on the manifold. How can one take a derivative if the vector space is changing? To do this we define the covariant derivative, which turn out to be essential for defining curvature.

The *covariant derivative* on a manifold M with metric g, denoted as either $\nabla(\mathbf{X}, \mathbf{Y})$ or $\nabla_{\mathbf{X}}\mathbf{Y}$, is a function which maps a pair of vector fields to a third vector field satisfying the following

properties:

1. The Leibniz Rule: $\nabla_{\mathbf{X}}(f\mathbf{Y}) = D_{\mathbf{X}}(f)\mathbf{Y} + f\nabla_{\mathbf{X}}\mathbf{Y}$, where f is any smooth function $f\colon M \to \mathbb{R}$ and $D_{\mathbf{X}}(f)$ is the directional derivative $g\left(\mathbf{X}, \mathrm{grad}(f)\right)$
2. $\nabla_{f\mathbf{X}}\mathbf{Y} = f\nabla_{\mathbf{X}}\mathbf{Y}$
3. bilinearity of ∇ : $\nabla_{\mathbf{X}+\mathbf{Y}}\mathbf{Z} = \nabla_{\mathbf{X}}\mathbf{Z} + \nabla_{\mathbf{Y}}\mathbf{Z}$, $\nabla_{\mathbf{X}}(\mathbf{Y}+\mathbf{Z}) = \nabla_{\mathbf{X}}\mathbf{Y} + \nabla_{\mathbf{X}}\mathbf{Z}$, and $\nabla_{\mathbf{X}}\left(c\mathbf{Y}\right) = c\nabla_{\mathbf{X}}\mathbf{Y}$ for all $c \in \mathbb{R}$
4. Metric Compatibility: $D_{\mathbf{X}}(g(\mathbf{Y}, \mathbf{Z})) = g(\nabla_{\mathbf{X}}\mathbf{Y}, \mathbf{Z}) + g(\mathbf{Y}, \nabla_{\mathbf{X}}\mathbf{Z})$
5. Torsion Free: $\nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} = [\mathbf{X}, \mathbf{Y}]$, where $[\cdot, \cdot]$ denotes the *Lie bracket*, $[\mathbf{X}, \mathbf{Y}] = D_{\mathbf{X}}(Y^{\nu})\mathbf{e}_{\nu} - D_{\mathbf{Y}}(X^{\mu})\mathbf{e}_{\mu}$

It is a fact that there is only one function which satisfies these properties for any manifold M and metric g.

The motivation for the covariant derivative is that it is a generalization of the directional derivative. It tells us how a vector field $\mathbf{Y}$ changes in the direction of $\mathbf{X}$ at every point on the manifold. It is called covariant since it transforms covariantly.

Since the covariant derivative is bilinear, we can determine how it acts by simply applying it to the basis vectors. We can do this be considering the basis vectors as constant vector fields. Applying property 5 of the definition of the covariant derivative we find:

$$\nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\nu} - \nabla_{\mathbf{e}_{\nu}}\mathbf{e}_{\nu} = [\mathbf{e}_{\nu}, \mathbf{e}_{\mu}] = 0,$$

where the Lie bracket must be zero because the basis fields are constant. This means that:

$$\nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\nu} = \nabla_{\mathbf{e}_{\nu}}\mathbf{e}_{\mu}, \tag{8.8}$$

or equivalently, the covariant derivative is symmetric on the basis vectors.

Now, all we need to do is determine $\nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\nu}$. By property 4 of the covariant derivative, we know the following three equations:

$$\begin{aligned} D_{\mathbf{e}_{\mu}}g(\mathbf{e}_{\nu}, \mathbf{e}_{\xi}) &= g(\nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\nu}, \mathbf{e}_{\xi}) + g(\mathbf{e}_{\nu}, \nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\xi}) && (A) \\ D_{\mathbf{e}_{\nu}}g(\mathbf{e}_{\xi}, \mathbf{e}_{\mu}) &= g(\nabla_{\mathbf{e}_{\nu}}\mathbf{e}_{\xi}, \mathbf{e}_{\mu}) + g(\mathbf{e}_{\xi}, \nabla_{\mathbf{e}_{\nu}}\mathbf{e}_{\mu}) && (B) \\ D_{\mathbf{e}_{\xi}}g(\mathbf{e}_{\mu}, \mathbf{e}_{\nu}) &= g(\nabla_{\mathbf{e}_{\xi}}\mathbf{e}_{\mu}, \mathbf{e}_{\nu}) + g(\mathbf{e}_{\mu}, \nabla_{\mathbf{e}_{\xi}}\mathbf{e}_{\nu}) && (C) \end{aligned}$$

By adding A and B and subtracting C and noting that $D_{\mathbf{e}_{\mu}} = \partial_{\mu}$ we get:

$$\partial_{\mu}g_{\nu\xi} + \partial_{\nu}g_{\mu\xi} - \partial_{\xi}g_{\mu\nu} = 2g(\nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\nu}, \mathbf{e}_{\xi}). \tag{8.9}$$

We know that the result of $\nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\nu}$ is some vector field, and hence can be broken down into its components. We will say that $\nabla_{\mathbf{e}_{\mu}}\mathbf{e}_{\nu} = X^{o}_{\mu\nu}\mathbf{e}_{o}$, and so if we solve for $X^{o}_{\mu\nu}$, we will know how the covariant derivative will act on the basis vectors. Thus equation (8.9) becomes:

$$\partial_{\mu}g_{\nu\xi} + \partial_{\nu}g_{\mu\xi} - \partial_{\xi}g_{\mu\nu} = 2g(X^{o}_{\mu\nu}\mathbf{e}_{o}, \mathbf{e}_{\xi}) = 2X^{o}_{\mu\nu}g_{o\xi}, \tag{8.10a}$$

$$X^{o}_{\mu\nu} = \frac{1}{2}g^{o\xi}\left(\partial_{\mu}g_{\nu\xi} + \partial_{\nu}g_{\mu\xi} - \partial_{\xi}g_{\mu\nu}\right) = \Gamma^{o}_{\mu\nu}, \tag{8.10b}$$

which is exactly the Christoffel symbol from equation (8.4). Thus the covariant derivative on the basis vectors becomes:

$$\nabla_{\mathbf{e}_\mu}\mathbf{e}_\nu = \Gamma^{\xi}_{\mu\nu}\mathbf{e}_\xi. \tag{8.11}$$

We can thus use properties 1 through 3 of the covariant derivative to find how it acts on any vector fields:

$$\nabla_{\mathbf{X}}\mathbf{Y} = X^\mu \nabla_{\mathbf{e}_\mu}\mathbf{Y} = X^\mu\left(\partial_\mu(Y^\nu)\mathbf{e}_\nu + Y^\nu\nabla_{\mathbf{e}_\mu}\mathbf{e}_\nu\right) = D_{\mathbf{X}}(Y^\nu)\mathbf{e}_\nu + X^\mu Y^\nu\Gamma^{\xi}_{\mu\nu}\mathbf{e}_\xi \tag{8.12}$$

We can now use the covariant derivative to define what it means for space to be curved. Note that in Euclidean space the Christoffel symbols are all zero since the metric is constant in the usual coordinates. This means that second term in equation (8.12) is zero, and so:

$$\nabla_{\mathbf{X}}\mathbf{Y} - \nabla_{\mathbf{Y}}\mathbf{X} = D_{\mathbf{X}}(Y^\nu)\mathbf{e}_\nu - D_{\mathbf{Y}}(X^\mu)\mathbf{e}_\mu = [\mathbf{X},\mathbf{Y}]. \tag{8.13}$$

Since this holds for Euclidean space, it holds for all flat spaces regardless of coordinates. In flat space, by applying equation (8.13), we can assert the following identity:

$$\nabla_{[\mathbf{X},\mathbf{Y}]}\mathbf{Z} = \nabla_{\nabla_{\mathbf{X}}\mathbf{Y}-\nabla_{\mathbf{Y}}\mathbf{X}}\mathbf{Z} = \nabla_{\nabla_{\mathbf{X}}\mathbf{Y}}\mathbf{Z} - \nabla_{\nabla_{\mathbf{Y}}\mathbf{X}}\mathbf{Z} = \nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}\mathbf{Z} - \nabla_{\mathbf{Y}}\nabla_{\mathbf{Z}}\mathbf{Z}$$

Thus, in flat space

$$\nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}\mathbf{Z} - \nabla_{\mathbf{Y}}\nabla_{\mathbf{Z}}\mathbf{Z} - \nabla_{[\mathbf{X},\mathbf{Y}]}\mathbf{Z} = \mathbf{0} \tag{8.14}$$

But this is zero as a vector field. We would like to use this to define the curvature, however, our intuition for curvature suggests we should want a scalar value. If we want a scalar, we can apply the metric. We will define the *Riemann Curvature Tensor* as:

$$R(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}) = g(\nabla_{\mathbf{X}}\nabla_{\mathbf{Y}}\mathbf{Z} - \nabla_{\mathbf{Y}}\nabla_{\mathbf{Z}},\mathbf{Z} - \nabla_{[\mathbf{X},\mathbf{Y}]}\mathbf{Z},\mathbf{W}) \tag{8.15}$$

Which we can fully understand from how it acts on the basis vectors:

$$R_{\mu\nu\xi o} = R(\mathbf{e}_\mu,\mathbf{e}_\nu,\mathbf{e}_\xi,\mathbf{e}_o) = \left(\partial_\mu\Gamma^{\lambda}_{\nu\xi}\right)g_{\lambda o} - \left(\partial_\nu\Gamma^{\lambda}_{\mu\xi}\right)g_{\lambda o} + \Gamma^{\lambda}_{\nu\xi}\Gamma^{\sigma}_{\mu\lambda}g_{\sigma o} - \Gamma^{\lambda}_{\mu\xi}\Gamma^{\sigma}_{\nu\lambda}g_{\sigma o} \tag{8.16}$$

In a sense, the Riemann curvature tensor measures the failure of parallel transport to preserve a vector. The technicalities are more complicated, but the idea is that we take a vector and move it along a parallelogram defined by two directions, then the when the vector returns to its initial position it may or may not point in the same direction. In Euclidean space it will, but on a sphere it will not. This behavior on the sphere can be seen in figure 8.1. This change is what forms our notion of curvature on a manifold. A more rigorous discussion of how curvature, parallel transport, and the Riemann curvature tensor are related is given by Pe'er (4).

The Riemann curvature tensor has several nice properties. Those being:

- $R(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}) = -R(\mathbf{Y},\mathbf{X},\mathbf{Z},\mathbf{W})$
- $R(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}) = -R(\mathbf{X},\mathbf{Y},\mathbf{W},\mathbf{Z})$
- $R(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}) = R(\mathbf{Z},\mathbf{W},\mathbf{X},\mathbf{Y})$
- $R(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathbf{W}) + R(\mathbf{Y},\mathbf{Z},\mathbf{X},\mathbf{W}) + R(\mathbf{Z},\mathbf{X},\mathbf{Y},\mathbf{W}) = 0$

These symmetries reduce the number of calculations needed to know the form of the Riemann curvature tensor.

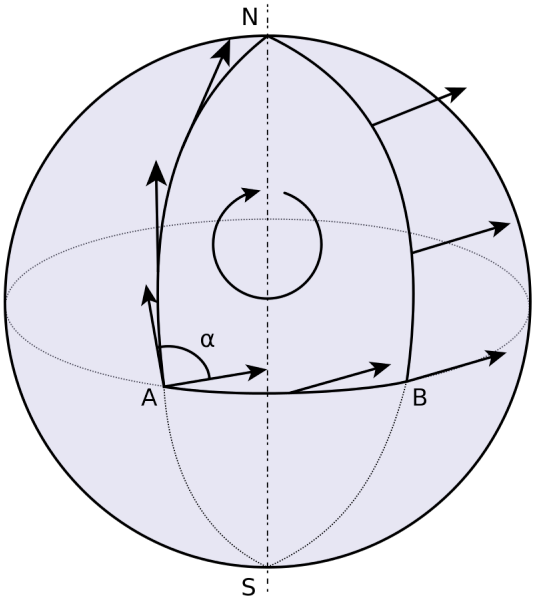


Figure 8.1: The parallel transport of a vector on a sphere from points A to N to B to A. The result of this process is a vector which forms an angle of α with respect to initial vector. Image by Fred the Oyster [CC BY-SA 4.0 (`http://creativecommons.org/licenses/by-sa/4.0`)], via Wikimedia Commons.

However, the Riemann curvature tensor provides more information than we need. We define the *Ricci curvature tensor*, Ric as:

$$\mathrm{Ric}_{\mu\nu} = -g^{oo}R_{\mu o\nu o} = -g^{oo}\left(\left(\partial_\mu \Gamma^\lambda_{o\nu}\right) g_{\lambda o} - \left(\partial_o \Gamma^\lambda_{\mu\nu}\right) g_{\lambda o} + \Gamma^\lambda_{o\nu}\Gamma^\sigma_{\mu\lambda} g_{\sigma o} - \Gamma^\lambda_{\mu\nu}\Gamma^\sigma_{o\lambda} g_{\sigma o}\right), \tag{8.17}$$

and can be considered a contraction of the Riemann curvature tensor. Due to the symetries of the Riemann curvature tensor, we also have:

$$\mathrm{Ric}_{\mu\nu} = \mathrm{Ric}_{\nu\mu} \tag{8.18}$$

We will prefer the Ricci curvature tensor to the Riemannian one due to Einstein's insight that the mass/energy of a physical system should relate to its curvature. As we saw in section 6, the stress energy tensor is of rank 2, and so we want a curvature tensor also of rank 2. However, the Ricci curvature alone is not enough. We can repeat the process done in equation (8.17) of summing over two of the indices on the Ricci curvature tensor, this yields the *scalar curvature* or the *Ricci scalar*:

$$R = g^{\mu\nu}\mathrm{Ric}_{\mu\nu} \tag{8.19}$$

We can now write out Einstein's field equation. If we use units such that the speed of light and Newton's gravitational constant are both 1, a Lorentzian manifold represents a physical system when:

$$\mathrm{Ric}_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{8.20}$$

where $T_{\mu\nu}$ is the stress energy tensor, representing the energy and mass of the system, and Λ is the *cosmological constant*. The cosmological constant is a theorized constant which has to do with the expansion of the universe, and we will assume it to be zero for the rest of this paper. It can

be shown with little effort that when the the stress-energy tensor is zero, then Einstein's field equation becomes:

$$\mathrm{Ric}_{\mu\nu} = 0, \tag{8.21}$$

which is called Einstein's vacuum equation[1].

Equation (8.20) is what governs general relativity. These equations allow us to determine how mass and energy will curve space, and then this curvature tells us what paths objects will follow.

Now that we have an equation which gives a condition for a manifold to represent a physical system, it is natural to ask what is the simplest space which satisfies (8.21). This is, of course, Minkowski space. In Minkowski space the metric is constant over all space, and hence the Christoffel symbols are all zero. This makes the Riemann curvature zero everywhere, which in turn means $\mathrm{Ric}_{\mu\nu} = 0$. This example is rather uninteresting, so we will now consider the simplest non-trivial example of a metric which satisfies Einstein's vacuum equation: the Schwarzchild metric.

[1]If we do not assume the cosmological constant is zero, the vacuum equation becomes $\mathrm{Ric}_{\mu\nu} = \Lambda g_{\mu\nu}$

Chapter 9

The Schwarzchild Solution

We would now like to compare the effects of general relativity on the orbits of satellites around stars to those we found classically and from special relativity. However, before we can do that we must first figure out what the curvature caused by a star is. To do this we find the Schwarzchild solution to Einstein's field equation.

9.1 The Simplest Non-Trivial Solution

We will consider a metric which is spherically symmetric and time independent metric. The idea being that a star, itself being spherically symmetric, should cause a spherically symmetric geometry and that gravitational field created by the star is not changing with time, and so neither should the metric. The metric will thus look like:

$$ds^2 = f(r)dt^2 - g(r)dr^2 - r^2 d\phi^2 - r^2 \sin^2 \phi d\theta^2 \tag{9.1}$$

Where f and g are functions of r which we must solve for. Note we will do this derivation in a fully dimensional setting and then later non-dimensionalize. This will give us physical insight into some of the constants of integration we choose.

To satisfy Einstein's vacuum equation, the Ricci curvature tensor must be zero. In particular, we must have:

$$\begin{aligned}
&\mathrm{Ric}_{00} = 0 \Rightarrow 4f(r)f'(r)g(r) - rf(r)f'(r)g'(r) - rf(r)^2 g(r) + 2rf(r)f''(r)g(r) = 0, && (A)\\
&\mathrm{Ric}_{11} = 0 \Rightarrow 4f(r)^2 g'(r) + rf(r)f'(r)g'(r) + rf(r)^2 g(r) - 2rf(r)f''(r)g(r) = 0, && (B)\\
&\mathrm{Ric}_{22} = 0 \Rightarrow rf(r)g'(r) - rf'(r)g(r) - 2f(r)g(r) + 2f(r)g(r)^2 = 0, && (C)
\end{aligned}$$

where we assume that $f(r)$ and $g(r)$ are both actually dependent on r, and so we ignore the case which returns the Minkowski metric in spherical coordinates.

Equations (A) and (B) above can be combined to yield:

$$f'(r)g(r) + f(r)g'(r) = 0$$

or

$$f(r)g(r) = K$$

for some constant K. By substituting the above into (C) we can find a differential equation for

$g(r)$:

$$rg'(r) - g(r)[1 - g(r)] = 0$$

Which has the solution:

$$g(r) = \left(1 - \frac{2GM}{c^2 r}\right)^{-1}, \tag{9.2}$$

where $\dfrac{2GM}{c^2}$ is a constant of integration chosen so that it will become physically meaningful and have the proper units. We know that f must be the reciprocal of g times the constant K. Taking K to be c^2, our proposed solution to Einstein's equation has the form:

$$ds^2 = \left(1 - \frac{2GM}{c^2 r}\right) c^2 dt^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 - r^2 d\phi^2 - r^2 \sin^2 \phi d\theta^2, \tag{9.3}$$

which we call the *Schwarzchild metric*. Here, c is the speed of light, G is Newton's gravitational constant, and M is the mass of the gravitational source. The factor $\dfrac{GM}{c^2}$ is called the *Schwarzchild radius*, r_s, and has an important physical meaning: If an object is within the Schwarzchild radius of a gravitational source, it is impossible for it to escape. It does not matter how much external force is used to try and pull the object out, it will always fall into what is called the *singularity*. This leads to objects called *black holes*, so called since not even light can escape from the sphere of radius r_s, called the *event horizon*, which surrounds then.

Note that the Schwarzchild radius is the unit we chose for our non-dimensionalization we did in section 3.2: this was part of the reason why. From here on we will revert to our non-dimensionalization.

9.2 Geodesics in the Schwarzchild Metric

We are interested in the geodesics of the Schwarzchild metric. This will be done, not by applying the geodesic equation (8.5), but rather by finding conserved quantities. These quantities will then simplify the situation and connect to our physical constants in the non-relativistic and special relativistic cases.

To start off, we will use the spherical symmetry of the Schwarzchild metric. The symmetry suggest that any geodesic should reside in a plane[1]. We can understand all planar geodesics, geodesics who are confined in some plane of motion, by studying the planar geodesics with $\phi = \dfrac{\pi}{2}$ and $\dot{\phi} = 0$, where dots indicate derivatives with respect to proper time. So an understanding of the $\phi = \dfrac{\pi}{2}$, $\dot{\phi} = 0$ case is sufficient for understanding the action of test particles in the Schwarzchild metric.

Under this assumption the metric becomes:

$$d\tau^2 = \left(1 - \frac{2}{r}\right) dt^2 - \left(1 - \frac{2}{r}\right)^{-1} dr^2 - r^2 d\theta^2, \tag{9.4}$$

where we can ignore the ϕ component entirely since we are confined to a plane. The Schwarzchild

[1]Using the word "plane" is somewhat an abuse of the word. Removing a degree of freedom in this case allows us to go from 3+1 space to 2+1 space. We use the word "plane" to refer to the fact that its motion is confined to two *spatial* dimensions, and not that the motion is fully two dimensional.

metric of equation (9.3) yields a path length of:

$$\tau = \int_{t_0}^{t_1} \sqrt{\left(1-\frac{2}{r}\right)\left(\frac{dt}{d\lambda}\right)^2 - \left(1-\frac{2}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 - r^2\left(\frac{d\theta}{d\lambda}\right)^2}\, d\lambda = \int_{t_0}^{t_1} L d\lambda, \tag{9.5}$$

where λ is an arbitrary parameter which describes the path, and L is defined as:

$$L = \sqrt{\left(1-\frac{2}{r}\right)\left(\frac{dt}{d\lambda}\right)^2 - \left(1-\frac{2}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 - r^2\left(\frac{d\theta}{d\lambda}\right)^2}, \tag{9.6}$$

for any parameter λ. Note that this holds for *all* parameters λ, and in particular holds for $\lambda = \tau$. Thus:

$$\tau = \int_{t_0}^{t_1} L d\tau, \tag{9.7}$$

which in differential form becomes:

$$d\tau = L d\tau. \tag{9.8}$$

Note that this means that not only is L non-dimensional, but that $L = 1$. We can substitute the $d\tau$ in equation (9.7) with the left-hand side of equation (9.8) to get:

$$\tau = \frac{1}{c^2}\int_{t_0}^{t_1} L^2 d\tau. \tag{9.9}$$

Thus the geodesic equations, which come from the Euler-Lagrange equations, are:

$$\frac{\partial L^2}{\partial x^\mu} = \frac{d}{dt}\frac{\partial L^2}{\partial \dot{x}^\mu}.$$

Since L^2, using τ as our parameter, is independent of of t and θ, we have two conserved quantities, e and l, defined by:

$$\begin{aligned}\frac{\partial L^2}{\partial \dot{t}} &= 2\left(1-\frac{2}{r}\right)\dot{t} = 2e \\ \frac{\partial L^2}{\partial \dot{\theta}} &= -2r^2\dot{\theta} = -2l,\end{aligned}$$

The quantities e and l are analogous to the energy and angular momentum from previous sections, but they are not quite the same. We will explore this more when we talk about the non-relativistic limit. Regardless of their relation to earlier constants, these conserved quantities allow us to simplify the equations of motion.

Since $L = 1$, and so $L^2 = 1$, we have that:

$$\left(1-\frac{2}{r}\right)\dot{t}^2 - \left(1-\frac{2}{r}\right)^{-1}\dot{r}^2 - r^2\dot{\theta}^2 = 1, \tag{9.10}$$

which we can clean up by substituting in the quantities e and l as defined above to get:

$$\dot{r}^2 = e^2 - 1 + \frac{2}{r} - \frac{l^2}{r^2} + \frac{2l^2}{r^3} \tag{9.11}$$

This gives us a differential equation for r as a function of τ. However, proper time is not the most useful parameter for orbits; we also want to know r as a function of θ so we can compare to our earlier work. We it will also simplify our equations to write them in terms of $u = \frac{1}{r}$. Using primes to denote derivatives with respect to θ, we can write $\dot{r}$ as $-\dot{\theta}\frac{u'}{u^2}$, and so the equation which describes the orbit is:

$$u'^2 = \frac{1}{l^2}\left(e^2 - 1\right) + \frac{2}{l^2}u - u^2 + 2u^3. \tag{9.12}$$

By taking another derivative with respect to θ we can turn this into a second order differential equation:

$$u'' = \frac{1}{l^2} - u + 3u^2. \tag{9.13}$$

This can be integrated to solve for the shape of an orbit using the numerical methods of section 7.4. However, before we preform the numerical integrations, we need to know the limits on bound orbits. This will be done by examining the effective potential.

9.3 General Relativistic Effective Potential

If we multiply equation (9.11) by $\frac{1}{2}m$ we get:

$$\frac{1}{2}\left(e^2 - 1\right) = \frac{1}{2}\dot{r}^2 - \frac{1}{r} + \frac{l^2}{2r} - \frac{l^2}{r^3} \tag{9.14}$$

Which looks like the sum of kinetic and potential energy terms equaling a constant, which can be loosely interpreted as the total energy. The "actual" total energy is the constant e, however, $\frac{1}{2}\left(e^2 - 1\right)$ is constant if and only if e is as well, so the above equation is still analogous to the total energy equation. Interpreting it as such gives us an effective potential of:

$$U_{\text{eff}}(r) = -\frac{1}{r} + \frac{l^2}{2r^2} - \frac{l^2}{r^3} \tag{9.15}$$

Note that this is a true effective potential since it is separable into terms involving r and a classical one-dimensional kinetic term. This makes it much nicer to deal with than the pseudopotentials of sections 5.2 and 7.3. It looks the same as the effective potential from section 3.3 with an additional quadrupole term. This quadrupole term causes some interesting behavior.

Of perneial interest are circular orbits. By taking the derivative of U_{eff} and setting it equal to zero we can find them. This yields:

$$r_c^2 - l^2 r_c + 3l^2 = 0, \tag{9.16}$$

where r_c is the radius of a circular orbit. Since this is a quadratic equation, it will in general have two solutions, those being:

$$r_c = \frac{l^2}{2}\left(1 \pm \sqrt{1 - \frac{12}{l^2}}\right). \tag{9.17}$$

We will call the inner solution r_{in} and the outer r_{out}. There are no circular orbit, and thus no bound orbits at all, when $l < 2\sqrt{3}$. We can also see that the two solutions agree when $l = 2\sqrt{3}$.

This gives us a lower bound on the angular momentum term similar to what we found in the special relativistic treatment.

We can see this behavior in the plots of U_{eff} in figure 9.1. Note that all the plots limit to negative infinity and zero for r approaching zero and positive infinity respectively. The plot for $l = \sqrt{3}$ has no critical points, the plot for $l = 2\sqrt{3}$ has one, and the rest have two. For the graphs which have two, we get a local maximum and a local minimum. This means that there are two possible circular orbits, a stable one and an unstable one. These orbits become the same at the critical value of $l = 2\sqrt{3}$.

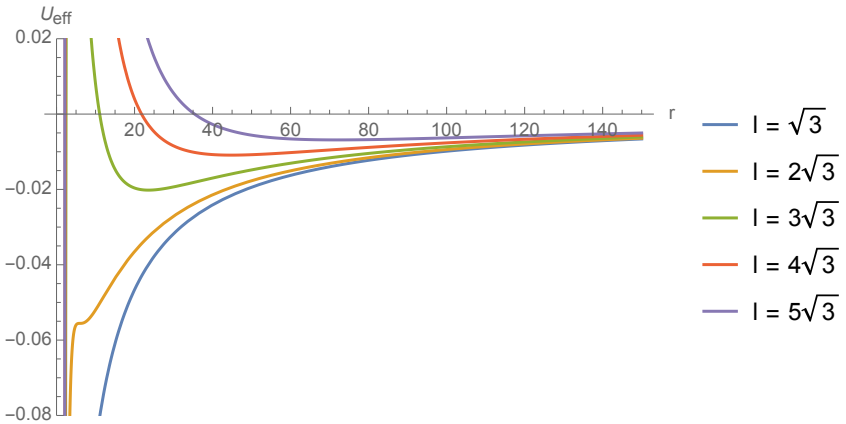


Figure 9.1: A plot of U_{eff} for various values of l. Note that for $l = \sqrt{3}$ there is no point with a slope of zero, and thus no circular orbits. For $l = 2\sqrt{3}$, we have one critical point which is a saddle point. All other curves have two critical points: one maximum and one minimum. All curves approach negative infinity as r approaches zero, and approach zero as r approaches infinity.

The radius which yields the minimum is the larger solution, r_{out}. Plugging this into equation yields the minimum value of e needed for a bound orbit. A plot of this can in figure 9.2

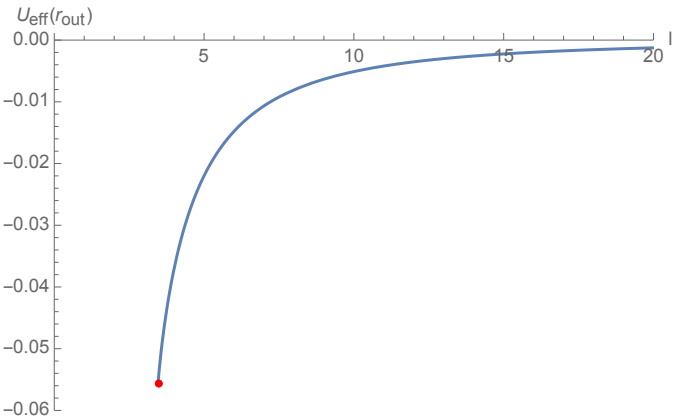


Figure 9.2: A plot of $U_{\text{eff}}(r_{\text{out}})$ as a function of l. The red dot denotes $l = 2\sqrt{3}$. The plot fails to be real valued for $l < 2\sqrt{3}$. This quantity corresponds to a minimum "total energy" term.

What we learn from the that $U_{\text{eff}} \to -\infty$ as $r \to 0$ is it is possible to fall into the star for any angular momentum provided you are at a distance less than r_{in}. This property is similar the way

in special relativity it was possible to fall in for certain non-zero angular momenta. This is in fact part of a stronger principle of general relativity: that it is impossible *not* to fall in if you are at a radius less than the *Schwarzchild radius*, $r_s = \dfrac{2GM}{c^2}$ or in our non-dimensionalization, $r_s = 2$. To see that this is consistent, note that the smallest r_{in} can be occurs at $l = 2\sqrt{3}$, and so $r_{\text{in}} > 3 > r_s$.

This means that no matter what, r_{out} is always greater that the Schwarzchild radius, and so we do not have to worry about our bound orbits being unphysical.

9.4 Bound Orbits in General Relativity

We can plot the closed orbits by using the Runge-Kutta method used for solving the charged special relativistic orbits. First, given a value of l, we find a valid value of e by picking one which is greater than $U_{\text{eff}}(r_{\text{out}})$ and less than both $U_{\text{eff}}(r_{\text{in}})$ and zero. Then we pick the initial position by finding the largest real root of:

$$e = U_{\text{eff}} = -\frac{1}{r} + \frac{l^2}{2r^2} - \frac{l^2}{r^3}, \tag{9.18}$$

which will correspond to the maximum radius of a closed orbit.

An example of such a closed orbit can be seen in figure 9.3 plotted using $l = 3\sqrt{3}$ and $e = -0.005$. It looks very much like the precessing ellipses from special relativity, however, the orbits are not quite elliptical.

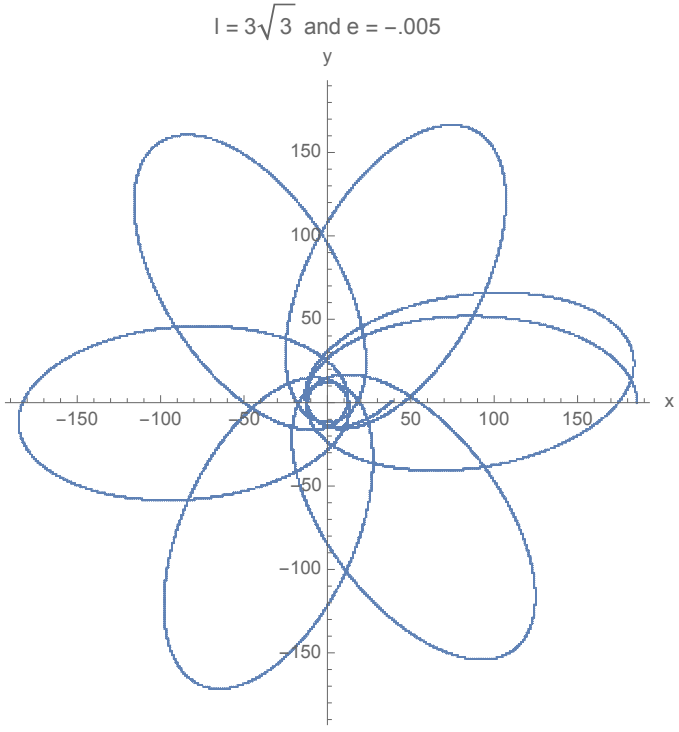


Figure 9.3: A bound orbit in general relativity with $l = 3\sqrt{3}$ and $e = -0.005$. At first glance the orbit appears elliptical, but this is not the case.

Using the same methods used in section 5.4 we can find approximations for the precession factor η, the eccentricity ε, and the scale factor r_0. For the example orbit in figure 9.3 we find

$\eta \approx 0.854$, $\varepsilon \approx 0.879$, and $r_0 \approx 22.467$. This yields a precessing ellipse which fits the orbit as can be seen in figure 9.4. The fit is very good, but it has wider lobes than the actual orbit does. We can see this behavior better in figure 9.5, which was generated by multiplying the θ coordinates of the orbit by η to collapse all the lopes into one near-ellipse. This makes clear the fact that GR orbits are non-elliptical. However, we can see from figure 9.6 we can see that as l gets larger, we quickly approach the precessing ellipses of special relativity.

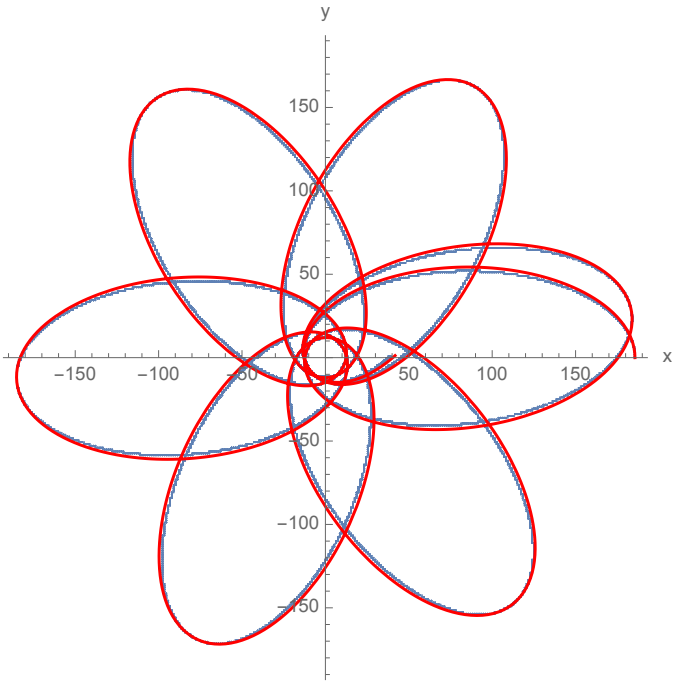


Figure 9.4: The bound orbit with $l = 3\sqrt{3}$ and $e = -0.005$ plotted in figure 9.3, but fitted with the precessing ellipse $r(\theta) = \dfrac{22.467}{1 - 0.879\cos(0.854\theta)}$. The fit works pretty well, but we can see that the lobes of the fit are slightly wider than the orbit.

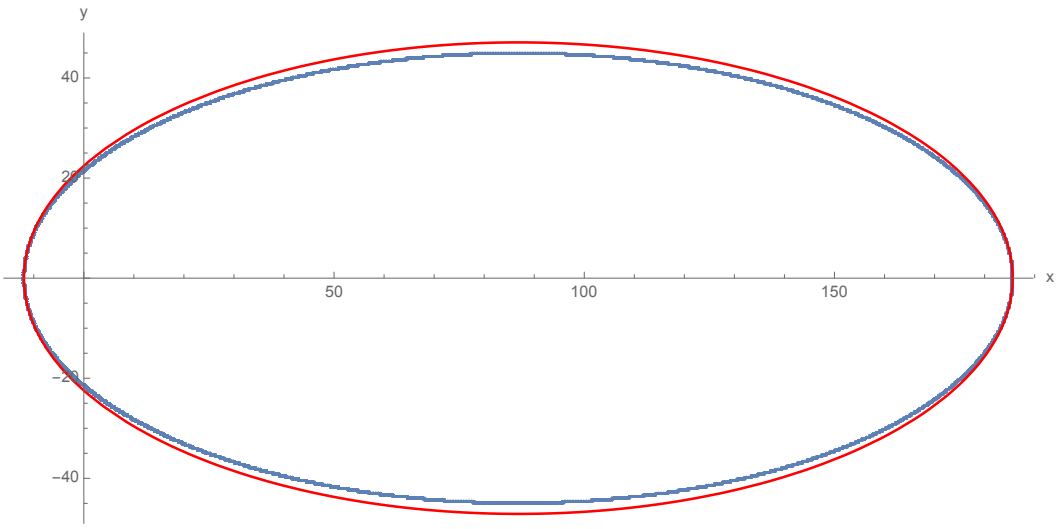


Figure 9.5: The orbit of figure 9.3 collapsed into one lobe and then compared to the elliptical approximation $r(\theta) = \dfrac{22.467}{1 - 0.879\cos(\theta)}$. The fix works pretty well, but we can see that the lobes of the fit are slightly wider than the orbit.

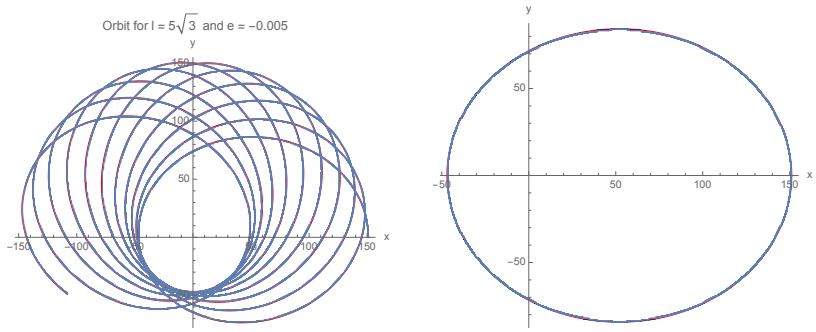
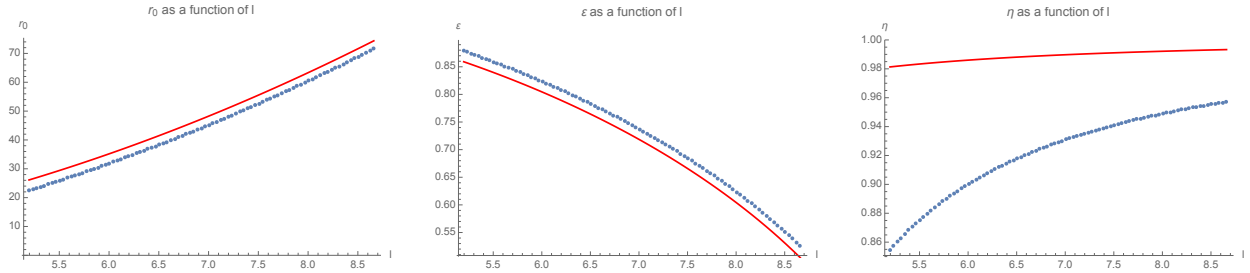


Figure 9.6: Analogous plots to those in figures 9.4 and 9.5, except using $l = 5\sqrt{3}$. These plots show that the orbits are much closer to the special relativistic case for larger l.

This suggests that SR is a good approximation of general relativity for large l. This, however, is not accurate. If we look at figure 9.7, looking at how r_0, ε, and η change with l we can see that they do not seem to approach their SR values. In fact, it is not possible to take a limit of the Schwarzchild solution and get anything which looks like special relativity without ignoring gravity entirely.

9.5 Limit Cases of the Schwarzchild Metric

We will first consider the *Newtonian limit* of the Schwarzchild metric: that the Schwarzchild radius is small, $r \gg 1$, and that the total speed of the object is much less than the speed of light, i.e.,



(a) (b) (c)

Figure 9.7: Plots of r_0, ε, and η as functions of l for $e = -0.005$. The red curve indicate the value for each orbital characteristic we expect for the given e value as given in equations (5.15a),(5.15b), and (5.15c). The plots do not seem to approach the curves for larger l values.

$v \ll 1$. The first assumption means that we are very far from the star and can ignore terms of order $\frac{1}{r^2}$. The second means that the speed of the orbiting object is much less than the speed of light, and so we can ignore v^2 terms.

Using (9.5) and time as our parameter λ, the Schwarzchild metric yields a proper time of:

$$\tau = \int \sqrt{\left(1 - \frac{2}{r}\right) - \left(1 - \frac{2}{r}\right)^{-1} \dot{r}^2 - r^2\dot{\theta}^2}\, dt$$

Where we here use dots to denote time derivatives. This is consistent with dots for derivatives with respect to proper time in the non-relativistic limit, as when $v \ll c$, the proper time approaches ordinary time. Expanding the integrand to first order in $\frac{1}{r}$ we get:

$$\begin{aligned} L &= \int \sqrt{1 - \frac{2}{r} - \left(1 - \frac{2}{r}\right)^{-1} \dot{r}^2 - r^2\dot{\theta}^2} \\ &\approx \sqrt{1 - \frac{2}{r} - \left(1 + \frac{2}{r}\right)\dot{r}^2 - r^2\dot{\theta}^2} \\ &\approx \sqrt{1 - \frac{2}{r} - v^2 - \frac{2}{r}\dot{r}^2} \end{aligned} \tag{9.19}$$

We note that $\frac{2}{r}\dot{r}^2$ is small compared to either term before it, and so we ignore it in the Newtonian limit. This means that::

$$\begin{aligned} L &\approx \sqrt{1 - \frac{2}{r} - v^2} \\ &\approx 1 - \frac{1}{r} - \frac{1}{2}v^2 \end{aligned}$$

which looks like the classical Lagrangian except with an additive constant 1 and a scale factor of -1. Since such things will not effect the extremization of the arclength, we follow a geodesic in

the Newtonian limit if and only if we extremize the quantity:

$$S = \int \left[\frac{1}{2}v^2 + \frac{1}{r}\right] dt, \tag{9.20}$$

which is precisely the classical action we would expect from a star and a satellite. This justifies our choice of constants in equation (9.2), since it allowed our NR limit to have the correct form.

This demonstrates how general relativity does return our classical equations in the Newtonian limit. However, GR does not limit to special relativity as treated before.

Starting from (9.19), we proceed without imposing any limit on v, but still requiring $r \gg 1$. We can factor L to get:

$$L \approx \sqrt{1-v^2}\sqrt{1 - \frac{2}{r}\left(1-\dot{r}^2\right)}$$

We then approximate for $r \gg 1$ and get:

$$L \approx \sqrt{1-\frac{v^2}{c^2}}\left(1 - \frac{1}{2}\frac{r_s}{r}\left(1-\dot{r}^2\right)\right) = \sqrt{1-v^2} - \frac{1}{r} + \sqrt{1-v^2}\frac{\dot{r}^2}{r}$$

If it were not for the third term this would be a scalar multiple of the Lagrangian for special relativity, however, this term is non-negligible in the SR limit since it is non-constant and of linear order in $\frac{1}{r}$. This is a demonstration of the incompatibility of special relativity with gravity: when one accounts for gravitational forces in a relativistic setting, it is necessary to do a full general relativistic treatment.

Chapter 10

Conclusions

We have seen how relativity be thought of as an extension of the principle of least action. We first started using special relativity to geometrize the kinetic portion of the Lagrangian. However, this result did not allow us to account for gravity and the action principle was not fully geometric. General relativity fixed this by saying mass and energy distort spacetime, and that objects follow paths which extremize length in this curved space.

To demonstrate this we examined the effects of relativity on the bound orbits of satellites orbiting stars. Classically, orbits are closed ellipses and these orbits existed for all possible angular momenta. Under special relativity, our first attempt to geometrize physics, we found that bound orbits remained elliptical,but precessed about the star and that there was a lower limit on angular momentum for stable orbits. General relativity, which fully geometrizes gravity, also has precessing orbits, but they are not elliptical and actually have no analytic formula. General relativity also has unstable orbits for all possible angular momenta in that if an object is within the Schwarzchild radius of a black hole, it will be unable to escape regardless of angular momentum.

Additionally, we discussed how applying a charge to a star affects orbits in special relativity. The charge creates an electric field whose energy effects the gravitational force. This causes the orbits to no longer be elliptical, and moreover have no analytic solution. We also talked about how one can work with electric and magnetic fields in general relativity, but lack the time to fully explore them. Given time, I would have looked at how the charge effected the orbits of satellites, especially the orbital parameters r_0, ε, and η, and seen if it could limit to the SR case.

Chapter 11

Appendix

11.1 Runge-Kutta Method

RK4Step[func_, xPoint_, tPoint_, dt_]:=Module[{k1, k2, k3, k4},
k1 = dtfunc[xPoint, tPoint];
k2 = dtfunc $\left[\frac{\text{k1}}{2.} + \text{xPoint}, \frac{\text{dt}}{2} + \text{tPoint}\right]$;
k3 = dtfunc $\left[\frac{\text{k2}}{2.} + \text{xPoint}, \frac{\text{dt}}{2.} + \text{tPoint}\right]$;
k4 = dtfunc[k3 + xPoint, dt + tPoint];
Return $\left[\frac{\text{k1}+2.\text{k2}+2.\text{k3}+\text{k4}}{6.} + \text{xPoint}\right]$;
]
CKStep[func_, xPoint_, tPoint_, dt_]:=Module[{k1, k2, k3, k4, k5, k6},
k1 = dtfunc[xPoint, tPoint];
k2 = dtfunc $\left[\frac{\text{k1}}{5.} + \text{xPoint}, \frac{\text{dt}}{5.} + \text{tPoint}\right]$;
k3 = dtfunc $\left[\frac{3.\text{k1}}{40.} + \frac{9.\text{k2}}{40.} + \text{xPoint}, \frac{3.\text{dt}}{10.} + \text{tPoint}\right]$;
k4 = dtfunc $\left[\frac{3.\text{k1}}{10.} - \frac{9.\text{k2}}{10.} + \frac{6.\text{k3}}{5.} + \text{xPoint}, \frac{3.\text{dt}}{5.} + \text{tPoint}\right]$;
k5 = dtfunc $\left[-\frac{11.\text{k1}}{54.} + \frac{5.\text{k2}}{2.} - \frac{70.\text{k3}}{27.} + \frac{35.\text{k4}}{27.} + \text{xPoint}, \text{dt} + \text{tPoint}\right]$;
k6 = dtfunc $\left[\frac{1631.\text{k1}}{55296.} + \frac{175.\text{k2}}{512.} + \frac{575.\text{k3}}{13834.} + \frac{44275.\text{k4}}{110592.} + \frac{253.\text{k5}}{4096.} + \text{xPoint}, \frac{7.\text{dt}}{8.} + \text{tPoint}\right]$;
Return $\left[\left\{\frac{37.\text{k1}}{378.} + \frac{250.\text{k3}}{621.} + \frac{125.\text{k4}}{594.} + \frac{512.\text{k6}}{1771} + \text{xPoint}, \frac{2825.\text{k1}}{27648.} + \frac{18575.\text{k3}}{48384.} + \frac{13525.\text{k4}}{55296.} + \frac{277.\text{k5}}{14336.} + \frac{\text{k6}}{4.} + \text{xPoint}\right\}\right]$;
]
stepAdapt[func_, xPoint_, tPoint_, dt_, error_]:=Module[{xPointRK4, xPointRK5, δt},
{xPointRK5, xPointRK4} = CKStep[func, xPoint, tPoint, dt];
If[$\|$xPointRK4 − xPointRK5$\|$ = 0,
$\delta\text{t} = \text{dt}\sqrt[5]{\frac{\text{error}}{\|\text{xPointRK5}\|}}$,
$\delta\text{t} = \text{dt}\sqrt[5]{\frac{\text{error}}{\|\text{xPointRK4}-\text{xPointRK5}\|}}$;
];
Return[δt];
]
AdaptiveStepRK4[func_, x0_, dt_, steps_, error_]:=Module[{points, xPoint, tPoint, j, k1, k2, k3, k4, δt, xPoint5},
xPoint = x0;
tPoint = 0;
points = Table[{dt(i − 1.), x0}, {i, 1, steps + 1}];
For[$j = 0, j \leq \text{steps} - 1, j{+}{+}$,
δt = stepAdapt[func, xPoint, tPoint, dt, error];
tPoint = δt + tPoint;
{xPoint5, xPoint} = CKStep[func, xPoint, tPoint, δt];
points[[j + 2]] = {tPoint, xPoint};
];
Return[points];
]

11.2 Charged Orbital Analysis

$$
\begin{aligned}
&\text{OrbitalAnalysisCharged[l\_, e\_, }\kappa\text{\_, R\_, dt\_, steps\_, error\_]:=}\\
&\quad \text{Module}[\{\text{x0, ff, list, rlist, }\theta\text{list, rmin, rmax, }p,\epsilon\text{, r0, index, }\eta\},\\
&\quad \text{x0} = x/.\,\text{NSolve}\left[e = \tfrac{\sqrt{l^2+x^2}}{x} - \tfrac{\kappa}{Rx} + \tfrac{\kappa}{x^2} - \tfrac{1}{x} - 1, x\right][[2]];\\
&\quad \text{ff(r\_, t\_)}:= \left\{ r[[2]],\ \frac{\left(2r[[1]]^3\right)\left(\left(e - \frac{\kappa}{2r[[1]]^2} + \frac{\kappa}{Rr[[1]]} + \frac{1}{r[[1]]} + 1\right)^2 - 1\right)}{l^2} - \right.\\
&\qquad\qquad \left. \frac{r[[1]]^4\left(-\frac{\kappa}{r[[1]]^3} + \frac{\kappa}{Rr[[1]]^2} + \frac{1}{r[[1]]^2}\right)\left(e - \frac{\kappa}{2r[[1]]^2} + \frac{\kappa}{Rr[[1]]} + \frac{1}{r[[1]]} + 1\right)}{l^2} - r[[1]] \right\};\\
&\quad \text{list} = \text{AdaptiveStepRK4}\left(\text{ff}, \{x0, 0\}, \text{dt, steps, error}\right);\\
&\quad \text{rlist} = \text{Table}[\text{list}[[i, 2, 1]], \{i, 1, \text{Length}[\text{list}]\}];\\
&\quad \text{If}[\neg \text{rlist} \in \mathbb{R}, \text{Return}[\{\text{Null, Null, Null}\}]];\\
&\quad \theta\text{list} = \text{Table}[\text{list}[[i, 1]], \{i, 1, \text{Length}[\text{list}]\}];\\
&\quad \text{rmin} = \min[\text{rlist}];\\
&\quad \text{rmax} = \max[\text{rlist}];\\
&\quad p = \tfrac{\text{rmin}}{\text{rmax}};\\
&\quad \epsilon = \tfrac{1-p}{p+1};\\
&\quad \text{r0} = \text{rmin}(\epsilon + 1);\\
&\quad \text{index} = \{\};\\
&\quad \text{For}[i = 3, i < \text{Length}[\text{rlist}] - 2, i{+}{+},\\
&\qquad \text{If}[\text{rlist}[[i]] > \text{rlist}[[i+2]] \wedge \text{rlist}[[i]] > \text{rlist}[[i+1]] \wedge \text{rlist}[[i]] > \text{rlist}[[i-1]] \wedge \text{rlist}[[i]] > \text{rlist}[[i-2]],\\
&\qquad\quad \text{AppendTo}[\text{index}, i]\\
&\qquad];\\
&\quad];\\
&\quad \eta = \frac{2\pi}{\text{Mean}[\text{Differences}[\text{Table}[\theta\text{list}[[\text{index}[[i]]]], \{i, 1, \text{Length}[\text{index}]\}]]]};\\
&\quad \text{Return}[\{\text{r0}, \epsilon, \eta\}];\\
&]
\end{aligned}
$$

11.3 GR Orbital Analysis

$$
\begin{aligned}
&\text{OrbitalAnalysisGR}[\text{l\_}, \text{e\_}, \text{dt\_}, \text{steps\_}, \text{error\_}]:= \\
&\quad \text{Module}[\{\text{x0}, \text{ff}, \text{list}, \text{rlist}, \theta\text{list}, \text{rmin}, \text{rmax}, p, \epsilon, \text{r0}, \text{index}, \eta\}, \\
&\quad \text{x0} = \max\left[r /. \text{NSolve}\left[e = -\tfrac{l^2}{r^3} + \tfrac{l^2}{2r^2} - \tfrac{1}{r}, r\right]\right]; \\
&\quad \text{ff}(\text{r\_}, \text{t\_}) := \left\{r[[2]], \tfrac{1}{l^2} + 3r[[1]]^2 - r[[1]]\right\}; \\
&\quad \text{list} = \text{AdaptiveStepRK4}\left(\text{ff}, \left\{\tfrac{1}{\text{x0}}, 0\right\}, \text{dt}, \text{steps}, \text{error}\right); \\
&\quad \text{rlist} = \text{Table}\left[\tfrac{1}{\text{list}[[i,2,1]]}, \{i, 1, \text{Length}[\text{list}]\}\right]; \\
&\quad \text{If}[\neg \text{rlist} \in \mathbb{R}, \text{Return}[\{\text{Null}, \text{Null}, \text{Null}\}]]; \\
&\quad \theta\text{list} = \text{Table}[\text{list}[[i, 1]], \{i, 1, \text{Length}[\text{list}]\}]; \\
&\quad \text{rmin} = \min[\text{rlist}]; \\
&\quad \text{rmax} = \max[\text{rlist}]; \\
&\quad p = \tfrac{\text{rmin}}{\text{rmax}}; \\
&\quad \epsilon = \tfrac{1-p}{p+1}; \\
&\quad \text{r0} = \text{rmin}(\epsilon + 1); \\
&\quad \text{index} = \{\}; \\
&\quad \text{For}[i = 3, i < \text{Length}[\text{rlist}] - 2, i{+}{+}, \\
&\qquad \text{If}[\text{rlist}[[i]] > \text{rlist}[[i+2]] \wedge \text{rlist}[[i]] > \text{rlist}[[i+1]] \wedge \text{rlist}[[i]] > \text{rlist}[[i-1]] \wedge \text{rlist}[[i]] > \text{rlist}[[i-2]], \\
&\qquad\quad \text{AppendTo}[\text{index}, i] \\
&\qquad]; \\
&\quad]; \\
&\quad \eta = \tfrac{2\pi}{\text{Mean}[\text{Differences}[\text{Table}[\theta\text{list}[[\text{index}[[i]]]], \{i, 1, \text{Length}[\text{index}]\}]]]}; \\
&\quad \text{Return}[\{\text{r0}, \epsilon, \eta\}]; \\
&]
\end{aligned}
$$

References

[1] Callahan, J. *The Geometry Of Spacetime*, (Springer, New York, 2000).

[2] Griffiths, D. *Introduction To Electrodynamics*, (Pearson, 2013).

[3] Herman, R. L. *Derivation Of The Geodesic Equation And Defining The Christoffel Symbols*, (University of North Carolina Wilmington, 2008).

[4] Pe'er, A. *Curvature*, 2014, `http://www.physics.ucc.ie/apeer/PY4112/Curvature.pdf`