Physics C161; Problem Set #1

prof. kasen

due Friday, 1/26, at midnight

Problem 1: Why Three Dimensions?

As far as we can tell, we live in a 4-dimensional spacetime with 3 spatial dimensions and 1 time dimension. But why 3+1? Wy not 1 spatial dimension, or 5, or 10^{100}?

While this is an ongoing philosophical question, in this problem you will give a compelling argument why 3D space is special: planetary orbits are not stable in more than $N = 3$ spatial dimensions. It therefore seems impossible for long-lived solar systems – and hence intelligent life – to develop in such a spacetime. Perhaps universes with $N > 3$ exist, but if they do we could never find ourselves within one[1].

This problem reviews some important concepts of orbital dynamics that we will return later in the course when we discuss black holes, gravitational lensing, galactic dark matter, etc...

Consider a planet of mass, m, orbiting a much larger central mass M. The orbit is confined to a plane and so we can describe the planet location with polar coordinates r and ϕ. The total energy of the planet is the sum of the kinetic and potential energies

$$E = \frac{1}{2}mv_r^2 + \frac{1}{2}mv_\theta^2 - \frac{GMm}{r} \tag{1}$$

where $v_r = \partial r/\partial t$ is the velocity in the radial direction and $v_\phi = r\partial\phi/\partial t$ the velocity in the angular direction. The last term is the Newtonian gravitational potential[2].

A important technique for solving dynamical problems is to identify *conserved quantities*. In addition to the energy, the angular momentum $L = mv_\phi r$ is a conserved quantity for a mass moving in a central potential. We can then remove the v_ϕ variable and write the energy equation as

$$E = \frac{1}{2}mv_r^2 + \frac{L^2}{2mr^2} - \frac{GMm}{r} \tag{2}$$

Notice that this equation looks like an energy equation for 1-dimensional motion if we write it as

$$E = \frac{1}{2}mv_r^2 + V_{\rm eff} \qquad \text{where } V_{\rm eff}(r) = \frac{L^2}{2mr^2} - \frac{GMm}{r} \tag{3}$$

where $V_{\rm eff}$ is called the *effective potential*.

[1] This is one expression of the *anthropic principle*. The anthropic principle comes in many forms, but the over-arching idea is that our universe must have properties that allow intelligent life to develop, otherwise we would never be around to ask questions. One popular realization of this is the *multiverse*, which posits that many different universes exist, each perhaps with different numbers of spacetime dimensions and physical laws . Of course we find ourselves in the small subset of those universes where intelligent life can develop.

[2] We'll use Newtonian physics in this problem, but later in the class we will do similar calculations and see how orbits behave differently in general relativity.

The effective potential is an incredibly useful construct for gaining intuition into orbital dynamics. The radial motion of the planet can be visualized as a ball rolling on the effective potential "hill" (see Figure 1). The gravitational potential term ($-GMm/r$) tends to pull the planet towards the center while the $L^2/2mr^2$ terms is a "centrifugal force" that tends to throw the planet away from the center. We see in the Figure that there is a stable equilibrium (i.e., a minimum in V_{eff}) where the planet can remain at a fixed radius (i.e., a circular orbit). At this location the inward and outward radial forces balance.

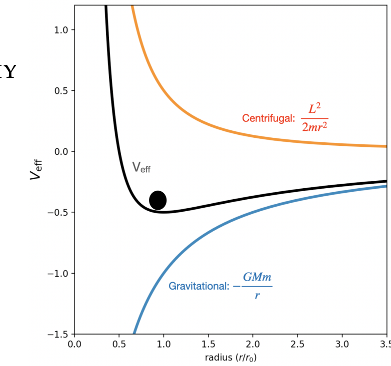

Figure 1: Plot of the effective potential. In a circular orbit the planet remains can be imagined as sitting at the minimum of V_{eff} at a fixed r coordinate. For an elliptical orbit, the planet can be imagined as rolling back and forth in the minimum, changing its distance over the orbit.

1a) Find the radius, r_0 at which the effective potential has an extremum, i.e., $\partial V_{\text{eff}}/\partial r = 0$. This gives the location of the minimum of V_{eff} seen in Figure 1 and will be the radius of a circular orbit for a planet with angular momentum L.

To show that the equilibrium point r_0 is *stable*, we must prove that the point is a minimum (i.e., a bowl-shape) and not a maximum (i.e., a peak). The mathematical condition for this is that the second derivative (i.e., curvature) of V_{eff} evaluated at the point r_0 must be positive

$$\left.\frac{\partial^2 V_{\text{eff}}}{\partial r^2}\right|_{r_0} > 0 \tag{4}$$

1b) Show that r_0 is a stable equilibrium.

Now let's see what would happen in a universe with *more* than 3 spatial dimensions. The key difference is that the gravitational field will have a different dependence on r. In the usual 3D case, the gravitational field lines from a point mass spread out over the surface of a sphere (with area = $4\pi r^2$). As a result, the gravitational force falls off with the familiar $1/r^2$ force law, and the gravitational potential (which is the integral of the force) falls off like $1/r$. In N spatial dimensions, the field lines spread out over a N-dimensional *hypersphere* (or N-sphere), which has area $\propto r^{N-1}$. The gravitational force will then fall of like r^{N-1} and the gravitational potential energy of a planet of mass m will be

$$V_N(r) = -\frac{G_N M m}{r^{N-2}} \tag{5}$$

where G_N is a constant which will depend on N (see the margin note if you want more details). [3]

For concreteness, let's consider $N = 5$ spatial dimensions (and you can optionally do the general case after). We will continue to take the orbit to be in a single plane. The effective potential is then

$$V_{\text{eff}} = \frac{L^2}{2mr^2} - \frac{G_N M m}{r^3} \tag{6}$$

[3] The gravitational field in higher dimensions follows from Gauss' Law, which states that the flux of a field through some surface is proportional to the mass, M, enclosed

$$\oint \vec{g} \cdot d\vec{A} = -kM$$

where $\vec{g}$ is the gravitational field and k is a constant. Choosing the surface to be an N-sphere (i.e., the generalization of a sphere to higher dimensions) the field is radial and constant over the surface, so the above integral becomes

$$g A_N = -kM$$

where A_N is the surface area of an N-sphere, given by $A_N = C_N r^{N-1}$, where the constant C_N can be calculated by doing the angular integrals. Thus

$$g = -\frac{k}{C_N}\frac{M}{r^{N-1}}$$

The force on a test mass m will be $F = gm$. To get the gravitational potential energy of the test mass we integrate the force over distance

$$V_N(r) = -\int_r^\infty F dr = -\frac{k}{(N-2)C_N}\frac{Mm}{r^{N-2}}$$

which holds for $N > 2$.

Looking up derivations of the area of hypersphere to get C_N we find

$$V_N(r) = -\frac{\Gamma(\frac{N}{2})}{2\pi^{N/2}(N-2)}\frac{kMm}{r^{N-2}}$$

where Γ is the gamma function. Thus we have shown that

$$V_N(r) = -\frac{G_N M m}{r^{N-2}}$$

where

$$G_N = \frac{k\Gamma(\frac{N}{2})}{2\pi^{N/2}(N-2)}$$

For $N = 3$, for example, $\Gamma(3/2) = \sqrt{\pi}/2$ and we find $G_3 = k/4\pi$, so the constant k is just 4π times Newton's gravitational constant G.

1c) Find the radius, r_0, of a circular orbit for a planet with mass m and angular momentum L moving through 5-dimensional space.

1d) Determine whether a circular orbit in 5D space is stable or not.

1e) To visualize what is happening, sketch or plot up the effective potential V_{eff} as a function of radius in a space of with 5 dimensions. To do so, first write V_{eff} as as a function of the *dimensionless*[4] radius $\tilde{r} = r/r_0$ (which expresses lengths in units of r_0) in which case you can show

$$V_{\text{eff}} = \frac{G_N M m}{r_0^3}\left[\frac{3}{2}\frac{1}{\tilde{r}^2} - \frac{1}{\tilde{r}^3}\right] \tag{7}$$

comment: You have shown that there are no stable circular orbits in $N = 5$ spatial dimensions, and can understand why. In 3D, the gravitational potential goes like $-1/r$ while the centrifugal term goes like $1/r^2$. At large r the gravitational term dominates, and pulls the planet towards the center, while at small r, the centrifugal term dominates and "pushes" the planet away from the center. The result is a radius of stable equilibrium.

In 5D, however, the gravitational potential has a stronger dependence on r (going like $-1/r^3$) and the opposite thing happens. At small r, gravity dominates and pulls the planet to the center, while at large r the centrifugal dominates and pushes the planet off to infinity. There is still an equilibrium point at $r = r_0$ but it is an unstable one (a peak in V_{eff}) and any little perturbation will make the planet fly off in one direction or another.

If you want to, you can repeat this calculation for any general value N and show that (for the same reason) there are no stable orbits for any higher dimensions $N > 3$.

aside: The fact that there are no stable planetary orbits for $N > 3$ does not necessarily preclude the emergence of life – perhaps life could somehow still form in some system that, although unstable, lasted long enough for evolution to take place. However, an analogous calculation can be carried out for the orbital of an electron in an atom, where you can show from quantum mechanics[5] that there are no stable bound states of the hydrogen atom if $N > 3$. It is even harder to imagine how life could emerge if bound atoms don't exist...

What about lower dimensional universes, with $N < 3$? The question is considered in a paper by Max Tegmark who suggests that a universe with $N = 1$ or 2 is just too simple for intelligent life to originate. For example, it is hard to imagine life without complex molecules, helical structures like DNA, brain nerves crossing...

We considered spatial dimensions, but why is there only 1 dimension of time? Tegmark points out that with more than 1 time

[4] The technique of writing equations in dimensionless form is one of the most powerful ways of gaining intuition into an equation. The idea is that instead of measuring r in units of, say, meters, we can choose to measure it in units of r_0, which actually has physical meaning in the problem (as it identifies the minimum point in V_{eff}). When we rewrite the effective potential in terms of a scaled radial coordinate $\tilde{r} = r/r_0$. we get Eq. 7 which (without all those constants) is easier to visualize and work with. If you go back to problems 1c) and 1d) they will be much simpler to solve using V_{eff} in this form.

[5] In quantum mechanics, instead of solving for the dynamics of a "ball rolling around" in the V_{eff} curve, you would be solving Schrodinger's equation for the radial wavefunctions (electron "clouds") that sit within a "potential well". As you can see in your plot, for $N > 3$, the effective potential V_{eff} has no minimum, and so no there is no potential well within which a bound state wavefunction can be enclosed.

dimension the laws of physics (assuming they continue to be described by differential equations) would no longer obey a uniqueness theorem. That is, given a set of initial conditions, there are more than one mathematical solutions for how the system evolves in time. It is hard to imagine how intelligent life could emerge in a universe with no predictability. I'm not sure we can even conceptualize what such a "multi-time" universe would be like.

These arguments are summarized in Tegmark's figure in the margin. While the question of the dimensionality of our universe it not a closed one, it does seem like 3+1 is a sweet spot for us.

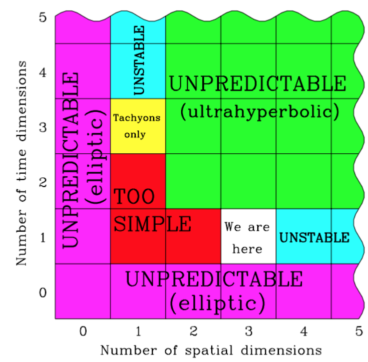

Figure 2: Figure from Tegmark illustrating the specialness of having 3 spatial and 1 time dimension.

aside: You may have heard that some physics theories, notably string theory, *do* posit the existence of extra spatial dimensions. However, these dimensions are wrapped up on some length scale, d, much smaller than commonly encountered lengths. On scales $r \gg d$, the extra dimensions don't affect the gravitational potential term, and so planetary orbits remain stable. It is only if we probe gravity on length scales $< d$ that we would notice that gravitational field lines are "spreading out" in the extra dimensions and so the gravitational force deviates from the $1/r^2$ law. This is in fact how experimentalists are trying to look for extra dimensions (nothing has been found yet).

Problem 2: The Metric of Curved Space

By 1908 or so, Einstein had come up with most of the conceptual foundations for general relativity. Yet it took him years more of labor (and mistakes) to complete the theory. In his autobiography, he explained: "Why were another seven years required for the construction of the general theory of relativity? The main reason lies in the fact that it is not so easy to free oneself from the idea that coordinates must have an immediate metric meaning".

This is also a major hurdle we face in studying black holes and cosmology. We need to realize that coordinates are just labels – like little stickers or flags that we plant throughout spacetime to uniquely identify each point. The labels themselves don't tell us the full geometry the space – e.g., the distance between points or angles between lines. To talk about the geometry we need to supply additional mathematical structure – the *metric* – which takes in coordinate labels and gives back physical distances.

In 2D Cartesian coordinates for example, the metric[6] of ordinary *Euclidean space* is essentially the Pythagorean theorem

$$dl^2 = dx^2 + dy^2 \tag{8}$$

which gives the physical length dl between any two points separated by (infinitesimally small) coordinate spacings of dx and dy.

[6] Strictly speaking, what I am writing down is called the *line element*, i.e., the equation for the length of a small line segment dl. The metric is actually a tensor that can be used to derive this equation for the line element. However, the line element contains the information of the metric that we are interested in, so I'll often just call it the "metric".

This flat Euclidean space is homogenous and isotropic, but it is not the only such space in 2D. The surface of sphere is a curved 2D space that is homogenous (i.e., the curvature is the same at all points). A hyperboloid (saddle shape) is also a homogenous 2D surface, with a curvature in the opposite sense (negative) of the sphere.

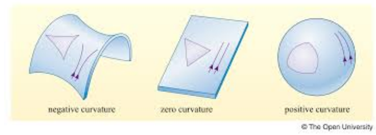

Figure 3: The 3 types of homogenous-isotropic 2D spaces. A hyperboloid has negative curvature and a sphere positive curvature, with the same degree of curvature at all points in the space.

Let's derive the metric that describes the 2D surface curved like a sphere, and then show that it leads to a fundamentally different geometry than the Euclidean one you have been taught. In particular, in curved space the circumference of a circle $C \neq 2\pi R$.

To do this, we will use a trick of embedding a curved 2D space within a flat 3D Euclidean space. In Cartesian coordinates, the 3D Euclidean metric is the generalized Pythagorean theorem

$$dl^2 = dx^2 + dy^2 + dz^2 \tag{9}$$

The surface of a sphere is a subset of the 3D space, consisting of all points that are fixed distance from the origin, that is the points that follow the constraint

$$x^2 + y^2 + z^2 = a^2 \tag{10}$$

where a is the radius of the sphere. We can use Eq. 10 to eliminate dz from the metric Eq. 9 and so derive the metric of a 2D spherically curved surface.

Cartesian coordinates are not all that useful for this purpose. Instead of x and y, it is more convenient to use polar coordinates r and ϕ given by

$$x = r\cos\ \phi \qquad y = r\sin\phi$$

2a) Show that the 3D Euclidean metric in cylindrical coordinates is[7]

$$dl^2 = dr^2 + r^2 d\phi^2 + dz^2 \tag{11}$$

[7] In this class we will take a refreshingly unrigorous approach to dealing with differentials. To relate a differential dx (i.e., a small change in the x coordinate) to the differentials in another coordinate system, dr and $d\phi$, we can just use the chain rule as if we were taking a derivative. So from $x = r\cos\phi$ we can write the differential

$$dx = dr\cos\phi - r\sin\phi d\phi$$

and similarly for dy.

comment: If we want to extract the metric of a flat 2D plane embedded in this 3D space, we could take only the subset of points of constant z, say $z = 0$. Then since z is constant on this flat surface, $dz = 0$ and the metric becomes

$$dl^2 = dr^2 + r^2 d\phi^2 \quad \text{(2D Euclidean space, polar coordinates)}$$

This is the same metric of flat Euclidean space as Eq. 8, just written in polar coordinates instead of Cartesian coordinates.

2b) To get the metric of a spherical 2D surface, we can use the equation for the subset of points on a sphere (Eq. 10) to eliminate dz and z from the 3D Euclidean metric Eq. 11. Show that the metric of a 2D spherical surface is

$$dl^2 = \frac{dr^2}{1 - r^2/a^2} + r^2 d\phi^2 \tag{12}$$

comment: We see that the metric for a spherical surface[8] differs from that of a flat plane by the factor of $1/(1 - r^2/a^2)$ in the first term. In the limit $a \to \infty$ (a very big sphere), this term becomes close to 1 and the spherical surface becomes more and more similar to a flat plane (in the same way that the earth seems locally flat).

Later in the class, we will see that the metric of the entire universe is essentially a 3D analogue of what you derived here (but with an additional factor to account for the expansion of space). The metric of a black hole will also have some similarities (but won't have a constant curvature everywhere).

[8] The surface of a hyperboloid is given by the subset of points

$$dx^2 + dy^2 - dz^2 = -a^2$$

From which one finds a metric

$$dl^2 = \frac{dr^2}{1 + r^2/a^2} + r^2 d\phi^2$$

which differs only by a sign. This is the case of negative constant curvature.

2c) Instead of r, we could choose to use a coordinate θ defined by $r = a \sin\theta$. Replacing r with θ show that the metric becomes

$$dl^2 = a^2 d\theta^2 + a^2 \sin^2\theta d\phi^2 \tag{13}$$

This is a convenient coordinate system for a spherical surface. The angles θ and ϕ are essentially like latitude and longitude that specify any point on a sphere.

comment: In this problem we derived the metric of a 2D curved space by embedding it within 3D space, but note that the final result only refers to two coordinates (e.g., θ and ϕ). So while the 3rd dimension was useful for visualizing the curvature, in the end it was nothing more than a mathematical artifice. A 2D metric like Eq. 13 has *intrinsic* curvature (i.e., a non-Euclidean geometry) without ever referring to a 3rd spatial coordinate. Similarly, the 3D space of our universe can have an intrinsic curvature without us having to invoke some 4th spatial dimension within which it curves.

Problem 3: Measurements in Curved Space

A METRIC DEFINES THE geometry of a space. We saw above that the metric of 2D Euclidan (i.e., flat, uncurved) space in Cartesian coordinates is $dl^2 = dx^2 + dy^2$, and can also be written in polar coordinates, (r, ϕ) as

$$dl^2 = dr^2 + r^2 d\phi^2 \tag{14}$$

Given this metric we can derive the standard results of Euclidean geometry, including that $C = 2\pi R$. Imagine drawing a circle at a radial coordinate $r = r_0$. We calculate the circumference by integrating the metric dl from $\phi = 0$ to $\phi = 2\pi$. Since we hold $r = r_0$ constant as we integrate around the circle, we have $dr = 0$. The distance around the circumference is then

$$C = \int dl = \int_0^{2\pi} r_0 d\phi = 2\pi r_0 \tag{15}$$

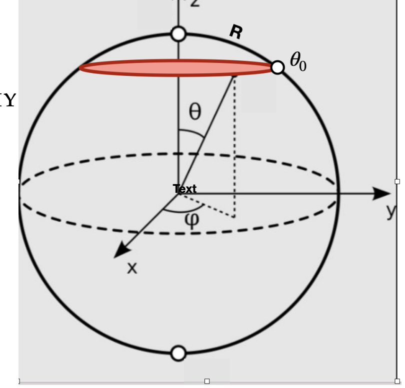

Figure 4: Drawing a circle on a curved spherical surface. The circle (red solid line) goes around in ϕ at a fixed $\theta = \theta_0$. The measured radius of the circle, R, is the distance from the north pole ($\theta = 0$) to the location $\theta = \theta_0$.

Similarly, the radius of the circle is gotten by integrating the metric in the radial direction (setting $d\phi = 0$ since ϕ is constant along this path) from the center $r = 0$ to $r = r_0$

$$R = \int dl = \int_0^{r_0} dr = r_0 \tag{16}$$

Putting Eq. 15 and Eq 16 together we have the expected result $C = 2\pi R$.

To give an example of how geometry is different curved space, let's measure the circumference of a circle drawn on a 2D spherical surface. Labeling positions with the standard spherical angle coordinates, the metric you derived above for a radius of curvature a is

$$dl^2 = a^2 d\theta^2 + a^2 \sin^2\theta d\phi^2 \tag{17}$$

Although it is not clear just from looking at it, the metric Eq. 17 represents an intrinsically different geometry than that of Eq. 14. We can show this by calculating the circumference of a circle in this space.

3a) Consider a circle that goes along the sphere at fixed latitude (i.e., at a fixed $\theta = \theta_0$). Integrate the metric Eq. 17 over ϕ to calculate the distance (i.e., circumference) of this circle.

3b) The radius of the circle is the distance from the circle center (at the north pole, $\theta = 0$) to the coordinate $\theta = \theta_0$ heading due south (i.e., at fixed ϕ). Use the metric Eq. 17 to calculate the distance R of this path

3c) Combine your results above to get an expression for circumference C in terms of radius R.

3d) Note that the circumference is not given by $C = 2\pi R$. To highlight the differences, consider the limit of a very small circle, $R \ll a$ and Taylor expand[9] your expression.

[9] The expansion of $\sin(x)$ is

$$\sin(x) \approx x - \frac{x^3}{3!} + ...$$

for $x \ll 1$, where the ... indicates higher order terms. You need keep only the first two terms in the series for this problem.

3e) Someone living on the surface of a sphere could determine that space was curved by drawing a circle and measuring R and C. Show that the radius of curvature of the space a would be given by

$$\frac{1}{a^2} = \lim_{R\to 0}\left[\frac{3}{\pi R^3}(2\pi R - C)\right] \tag{18}$$

where we wrote the limit as $R \to 0$ since we made the Taylor expansion approximation $R \ll a$.

comment: The quantity $1/a^2$ provides a nice measure of the degree of curvature of the sphere. As a becomes small, the curvature $1/a^2$ becomes large, while as $a \to \infty$ the surface begins to look more and more locally flat and the curvature goes to zero.

For an arbitrary 2D surface (not just a sphere) we can define a quantity called the *Gaussian curvature*, K_G using the same deviation of the circumference from $2\pi R$

$$K_G = \lim_{R\to 0}\left[\frac{3}{\pi R^3}(2\pi R - C)\right]$$

At some point on the surface, K_G describes the curvature by quantifying how it is locally equivalent to that of a sphere of radius a. For a sphere, the Gaussian curvature has the same value $K_G = 1/a^2$ at all points, but for a more complicated 2D surface, K_G could vary from point to point. On the surface of a sphere, we found $C < 2\pi R$ so the Gaussian curvature is positive. However, in some curved spaces we will find that $C > 2\pi R$ and the curvature will be negative. This is the case of a saddle like surface.

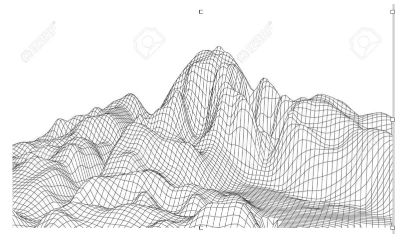

Figure 5: An arbitrarily curved 2D surface. The Gaussian curvature varies from place to place. At any single point on the surface, one can calculate K_G by drawing a small circle and comparing C to $2\pi R$. The Gaussian curvature essentially tells you $1/a^2$ where a is the effective radius of curvature at that point.

In general, it will not be immediately clear just by looking at a metric like Eq. 17 whether or not we are dealing with a curved space or not. We would need to carry out a mathematical procedure like the above to determine if the curvature is non-zero anywhere. For 2D surfaces, the K_G (as a function of position) completely describes the intrinsic geometry of the space. For 3D spaces, however, one number is not sufficient to fully characterize the geometry, and we instead need to calculate a multi-component tensor (called the Reimann tensor) to determine the curvature. We won't go much into such tensors in this class, but see a reference in general relativity if you want a more detailed picture.